
Modeling and analysis of non-Gaussian long-range correlated data

Extreme value theory and effective sample sizes
with applications to precipitation data

DISSERTATION

zur Erlangung des akademischen Grades

Doctor rerum naturalium
(Dr. rer. nat.)

vorgelegt dem

Bereich Mathematik und Naturwissenschaften
der Technischen Universität Dresden

von

Dipl. Math. KATJA POLOTZEK
geboren am 22.09.1987 in Erfurt, Deutschland.

Tag der Einreichung: 18.11.2021

Preface

The present thesis is the result of my inspiring time as a PhD student at the Max Planck Institute for the Physics of Complex Systems in Dresden, Germany. I would first like to thank my supervisor, Prof. Holger Kantz, who has always been open to my research interests and supported my work with his invaluable expertise. I address special thanks to my supervisor, Prof. Axel Voigt, and to the two of them for their never declining support and patience especially during the final phase of my PhD studies.

I am highly grateful for the opportunity of meeting a number of highly interesting people on my way. Many thanks to Carl, Benedict and Alex for recurring fruitful and enjoyable both scientific and personal discussions. There is no way of an overvaluation of getting to know people that simply push your personal development to higher levels, particular thanks to Stephan, Jens and my beloved Daniel.

Never would I have been able to complete my thesis without the dedicated support of my family and close friends. I send love and kisses to my mother and grandmother, my brother and to Steffi und Helen, who all patiently waited for my return to less occupied times. I hope that somehow and somewhere also my father and grandfather will perceive that I actually finished PhD studies.

Katja Polotzek
Dresden, November 2021

Extended abstract

The present thesis deals with different areas concerning both the mathematical theory and application of non-Gaussian stationary stochastic processes with special focus on risk assessment and statistical uncertainty due to the presence of long-range correlations (LRC) by an infinite sum of the autocorrelation function (acf). The topics under consideration are

- (I) Data models for non-Gaussian long-range correlated processes
- (II) The theory of effective sample sizes for stationary processes
- (III) Extreme value theory for stationary and especially meta-Gaussian processes
- (IV) Application of the concepts (I) to (III) to empirical precipitation records

The major results listed in Chapter 10 are twofold. First, data models obtained from nonlinear and memoryless transformations of Gaussian processes are capable of reproducing key statistical quantities of empirical non-Gaussian LRC precipitation records without deploying the physics of the underlying dynamics concretely. Second, incorporating such a model into existing stochastic theory allows for specifying and improving the impact of long-range correlations on statistical inference about empirical data, in particular, about their sample mean and return levels of extreme events.

The introduction in Chapter 1 describes the need for a robust and parametric modeling of non-Gaussian LRC data in general and, in particular, in applications with precipitation time series. Such models form the basis of reliable conclusions about key statistics listed in topic IV below and serve as an ingredient for simulations of systems of higher complexity, for example for weather generators or global climate models.

I Modeling of non-Gaussian long-range correlated processes

Chapter 2 summarizes existing knowledge about Gaussian and non-Gaussian stochastic processes regarding stationarity, marginal distributions and autocorrelations. In particular, autoregressive-fractionally-integrated-moving-average (ARFIMA) processes provide discrete-time Gaussian LRC data models deploying a single parameter for describing LRC and an arbitrary number of autoregressive (AR) and moving-average parameters for short-range correlations (SRC). ARFIMA processes obey a power-law decaying acf that sums up to an infinite correlation time by an exponent between negative unity and zero.

Pointwise nonlinear transformations $Y_t = g(X_t)$ of Gaussian processes X_t have been applied to the modeling and generation of this kind of data before. Chapter 3 recaps previous knowledge about the mathematical theory of such meta-Gaussian processes. Chapter 6 establishes a truncated-Gaussian-power (tGp) transformation to an ARFIMA process for the modeling of empirical non-Gaussian LRC daily precipitation amounts.

The main challenges of meta-Gaussian ARFIMA modeling are the choice of an appropriate transformation g along with its parametrization and a proper adjustment of the SRC and LRC parameters such that the transformed process Y_t exhibits desired statistical properties. The Hermite expansion of square-integrable transformations g yields an analytic formula for the acf of transformed Gaussian processes. Section 3.3 repeats this meta-Gaussian procedure for the squared and lognormal process X_t^2 and e^{X_t} , respectively, and adds a closed-form description for the acf of the absolute value $|X_t|$ process with a Gaussian processes X_t each.

Section 3.2 elaborates in detail that the acf of the transformed process is a convex function of the original acf independent of specific time lags. Instantaneous and memory-less transformations $g(X_t)$ thus decrease or at most keep the strength of autocorrelations. The Hermite approach further yields an asymptotic power-law behavior of the acf of a transformed Gaussian LRC process given the power-law exponent of the original acf. Only even transformations are capable of changing the power-law asymptotics of the acf, which is not satisfied in typical applications.

For an estimation of the LRC parameter from empirical data Chapter 4 describes three of the existing methods for the numerical detection of LRC, namely of the Hurst effect. Rescaled-range statistics, detrended fluctuation analysis (DFA) and the wavelet transform obey different capabilities of detrending. Such estimates are subject to possible bias or misinterpretation due to nonstationarities or strong SRC. Section 4.3 adds a detailed description of the previously known sensitivity of the estimators to seasonalities, polynomial trends or intrinsic nonstationarities.

II The theory of effective sample sizes for meta-Gaussian processes

The presence of correlations increases the uncertainty of statistical quantities by slowing down the convergence of statistical estimators to their stochastic limit. The accuracy of estimates for correlated data thus effectively compares to the one of a lower number of independent and identically distributed (iid) samples. A comparison of the variance σ_N^2 of the sample mean $\frac{1}{N} \sum_{i=1}^N X_i$ of $N \in \mathbb{N}$ identically distributed and correlated random variates X_1, \dots, X_N to the variance $\tilde{\sigma}_N^2$ of the sample mean of N iid samples of the same distribution has been used for the definition of an asymptotic effective sample size by $N_{\text{eff}} := \lim_{N \rightarrow \infty} \frac{\tilde{\sigma}_N^2}{\sigma_N^2} N$ before.

Chapter 5 determines a closed form representation of the variance of the sample mean for LRC processes with a power-law decaying acf. Introducing a finite-size decorrelation time establishes the notion of a finite-size effective sample size even for processes without a characteristic time scale. The obtained closed form representation of the effective sample size allows for a calculation of effective sample sizes not only asymptotically but also as a finite-sample property. Section 5.2 specifies these results for meta-Gaussian processes by combining the concepts of the Hermite expansion from Chapter 3 and finite-size decorrelation times.

The finite-size effective sample sizes allow for different interpretations of statistical inference. Precise knowledge about the variance of the sample mean provides confidence intervals of this estimator for the mean of LRC Gaussian and meta-Gaussian processes. A comparison between the given and effective sample sizes N and N_{eff} , respectively, gives a quantitative impression about the impact of correlations. As an example, the sample mean of 1000 standard Gaussian AR samples with exponentially decaying acf $e^{k \ln(0.3)}$ still exhibits the same variance as only 539 standard Gaussian iid samples.

III Extreme value theory for stationary and meta-Gaussian processes

Under mild conditions on the strength of the dependence among individual samples universally the three-parameter generalized extreme value (GEV) distribution well describes the asymptotic distribution of the maximum $M_n := \max\{X_1, \dots, X_n\}$ of $n \in \mathbb{N}$ samples X_1, \dots, X_n of a stationary process as the block size $n \rightarrow \infty$. By the extremal types theorem for appropriate norming constants $a_n \in \mathbb{R}_{>0}$ and $b_n \in \mathbb{R}$, the normed maximum $a_n(M_n - b_n)$ converges in distribution to one out of three extreme value distributions (EVD) as $n \rightarrow \infty$. Chapter 7 summarizes classical EVT for iid processes as well as EVT for stationary processes with dependencies both with particular focus on rates of convergence in the extremal types theorem and corresponding norming constants.

The maximum M_n of Gaussian samples is a prominent example for particularly slow convergence to an asymptotic Type-I or Gumbel EVD. Chapter 8 specifies EVT for meta-Gaussian processes $g(X_t)$ and derives conditions on the transformation g , so that a Taylor expansion yields appropriate norming constants. These norming constants prove that the maximum of tGp samples applied to precipitation amounts in Chapter 6 logarithmically approaches an asymptotic Gumbel shape analogously to Gaussian samples.

A maximum likelihood (ML) estimate of the location, scale and shape parameters for adjusting a GEV distribution to empirical block maxima is the usual approach to estimating return levels for prescribed return periods. Nonstationarity, a slow convergence or clustering of extremes are capable of confounding such estimates. In case of slow convergence in the Gumbel class, full GEV estimates excel Gumbel estimates with fixed zero shape parameter but accompanied by higher statistical uncertainty. Chapter 9 improves the validity and statistical uncertainty of return level estimates by incorporating outside knowledge about the distribution and dependency structure of the individual samples if existent. Section 9.2 elaborates a procedure for comparing estimates by EVT to estimates by modeling the marginal distribution directly. For stationary processes with only weak SRC and LRC such alternative estimates provide return levels with smaller confidence intervals.

For certain marginal distributions the Gumbel estimate provides valid return levels despite a slow convergence due to the similarity of the tail of the maximum distribution to the exponential tail of the Gumbel distribution such as for the Gamma or tGp distribution with exponents close to two. Section 9.3 deploys this observation for further improving the statistical uncertainty of return level estimates by comprising knowledge

about the marginal and the limit Gumbel distribution to a modified Gumbel estimate. There exist norming constants a_n and b_n based on the marginal distribution which force the tail of the distribution of $a_n(M_n - b_n)$ to coincide with the limit Gumbel tail in a prescribed return level. Choosing norming constants giving coincidence in a return level in the neighborhood of the return level with a desired return period instead of ML-estimated norming constants yields the desired return level with less statistical uncertainty compared to a GEV or direct estimate by the marginal distribution.

IV Application of the concepts I to III to empirical precipitation records

Chapter 6 establishes and validates a parsimonious meta-Gaussian model for mid-latitude daily precipitation amounts based on a tGp-transformed ARFIMA process with an alternative methodology for the generation of prescribed correlations. The methodology I provides this robust approach to not only simulations of precipitation time series but also to statistical inferences about the daily mean, annual totals and extreme events applying the developed concepts II and III, respectively.

Five parameters suffice to capture the non-Gaussianity of the data and their SRC and LRC. Three parameters adjust the tGp model to the empirical marginal distribution of daily amounts. A least square estimate of the model to the empirical survival function in logarithmic scaling achieves particular focus on extreme events by discriminating small and emphasizing large amounts. The estimation of one AR parameter introduces the conditional probability of the occurrence of two consecutive days with a noticeable amount of 4 mm. This method for non-Gaussian processes is an alternative to previous ML estimates of AR parameters for Gaussian models. The last parameter describes the weak but proved significant numerically estimated LRC in the precipitation data. The tGp transformation retains the asymptotics of the acf, so that the estimated Hurst exponent directly transfers to the LRC parameter of the underlying ARFIMA model.

Section 6.3 deploys sample means and variances, p-p and q-q plots, waiting times and return level estimates for a detailed assessment and comparison of the statistics of the model and the data. For a random selection of 20 European data sets the stationary meta-Gaussian ARFIMA model well reproduces the statistics of daily and annual total amounts, SRC and LRC, wet and dry spells and annual maximum amounts. Assessing the impact of LRC by the effective sample size, the tGp-transformed ARFIMA model adjusted to precipitation amounts recorded at the “Fichtelberg” station in Germany reveals a statistical uncertainty of the sample mean of the given 103 years of daily measurements of effectively only about 14 years of independent daily measurements.

Section 9.2.2 and Section 9.3.4 obtain 100-year return levels of extreme precipitation events by the comparison of return level estimates and the modified Gumbel approach. Assured weak SRC and LRC among daily amounts yield estimates with decreased statistical uncertainty using either directly the return level or the norming constants obtained from tGp model for the marginal distribution.

Contents

Preface	iii
Extended abstract	v
Contents	ix
Table of contents	ix
List of figures	xiii
List of tables	xv
1 Introduction	1
1.1 Modeling and simulation of non-Gaussian long-range correlated processes	2
1.2 Statistical uncertainty due to prominent correlations	3
1.3 Extreme value theory for stationary processes	5
1.4 Contemporary precipitation models	6
Structure of the thesis	9
Notation	9
2 Gaussian and non-Gaussian stochastic processes	11
2.1 Gaussian processes	11
2.2 Stationary processes	12
2.3 Classification of the marginal distribution	13
2.4 Long-range correlated processes	15
2.5 Autoregressive-fractionally-integrated-moving-average (ARFIMA) processes	17
2.6 Numerical generation of ARFIMA time series	21
3 Meta-Gaussian stochastic processes	23
3.1 Hermite polynomial approach to correlations	24
3.2 Implications for correlation properties	27
3.3 Example transformations	31
3.3.1 Generated marginal distributions	31
3.3.2 Generated autocorrelation functions	33
3.4 Closing remarks	36
3.4.1 Inverse relation between generated and original correlations	36
3.4.2 Generating the same marginal distribution but different correlations	40
3.4.3 Alternative exponential model with long-range correlations	40

4	Numerical estimation of long-range correlations	45
4.1	Estimation based on the empirical autocorrelation function	45
4.2	Estimation based on fluctuations on different temporal scales	47
4.2.1	Rescaled-range analysis	49
4.2.2	Detrended fluctuation analysis	51
4.2.3	Wavelet analysis	53
4.3	Sensitivity of the estimators	61
4.4	Closing remarks	65
4.4.1	Anomalous scaling of stochastic processes	65
4.4.2	Probabilistic properties of the estimators	68
5	Effective sample size of stationary processes	69
5.1	Asymptotics and uncertainty of the sample mean	69
5.1.1	Limit distribution of the sample mean	70
5.1.2	Sample mean approach to an effective sample size	73
5.1.3	Sample mean approach to a characteristic time scale	77
5.1.4	Asymptotics of the effective sample size	79
5.2	Effective sample size for meta-Gaussian processes	81
5.2.1	Finite-sample interpretation	81
5.2.2	Large sample behavior	84
5.3	Closing remarks and outlook	92
5.3.1	Interpretation of the effective sample size	92
5.3.2	Anomalous scaling and the effective sample size	92
5.3.3	Link to large deviation theory	93
6	Data model for local daily precipitation amounts	95
6.1	Model design	95
6.1.1	Stationarity and annual cycle	95
6.1.2	Long-range correlations	98
6.1.3	Truncated-Gaussian-power model	99
6.2	Model estimation	100
6.2.1	Tail-oriented marginal distribution	101
6.2.2	Long-range correlations by detrended fluctuation analysis	103
6.2.3	Short-range correlations by conditional probabilities	104
6.2.4	Step-by-step modeling procedure	105
6.3	Model validation	106
6.3.1	Daily amounts	107
6.3.2	Long-range correlations	112
6.3.3	Short-range correlations	113
6.3.4	Wet and dry spells	115

6.3.5	Effective sample size and daily mean	117
6.3.6	Annual totals	118
6.3.7	Annual maxima	120
6.4	Closing remarks and outlook	123
6.4.1	Statistical representation of empirical data	123
6.4.2	The tail of the marginal distribution	124
6.4.3	The skewness of precipitation amounts	125
6.4.4	Limitations of the model and outlook	126
7	Extreme value theory	127
7.1	Classical extreme value theory	127
	Norming constants	129
7.1.1	Extreme value distributions	130
7.1.2	Domains of attraction	134
7.2	Extreme value theory for stationary processes	138
7.2.1	The conditions $D(u_n)$ and $D'(u_n)$	138
	Extremal index	140
7.2.2	The conditions $D(u_n)$ and $D'(u_n)$ for Gaussian processes	144
7.3	Rates of convergence	144
7.3.1	Rate of convergence in the classical setting	145
7.3.2	Rate of convergence for stationary processes	153
7.3.3	Ultimate and penultimate approximations	155
7.4	Statistical inference by extreme value theory	158
7.4.1	Return levels and return periods of extremes	159
7.4.2	Block maxima approach	161
7.4.3	Peaks over threshold approach	165
	Generalized Pareto distribution	165
7.4.4	Statistical inference for stationary processes	172
8	Extreme value theory for meta-Gaussian processes	175
8.1	Norming constants	176
8.2	Rate of convergence	179
8.2.1	Rate of convergence in the classical setting	179
8.2.2	Rate of convergence for stationary meta-Gaussian processes	181
8.3	Closing remarks	185
8.3.1	Effective block size of stationary processes	185
8.3.2	Large deviations in extreme value theory	191

9	Improved estimation of return levels by incorporating outside knowledge	193
9.1	Definitions of annual return levels	195
9.2	Comparison of return level estimates	196
9.2.1	Application to synthetic data	198
9.2.2	Application to daily precipitation amounts	203
9.3	Estimation of return levels by norming constants in the Gumbel domain	206
9.3.1	Error reduction for return levels by norming constants	208
9.3.2	Statistical inference by norming constants	212
9.3.3	Application to synthetic data	212
9.3.4	Application to daily precipitation amounts	213
9.4	Closing remarks and outlook	216
10	Conclusion	219
10.1	List of central contributions	220
10.2	Open questions and impulses for future research	224
	Appendix	227
A	Mehler's formula	227
B	Specific distributions	228
C	Convergence of distribution functions	229
D	Weak dependence structures of stochastic processes	231
E	Estimated model parameters	232
F	Influence of tails in extreme value theory	235
G	Mean residual life plot	237
	Abbreviations	238
	Symbols	239
	Bibliography	241
	Confirmation	259

List of figures

1.1	Distribution of daily precipitation amounts at the city of Potsdam, Germany	7
2.1	Theoretical and empirical acfs of different Gaussian processes	20
2.2	Sample time series of different Gaussian processes	22
3.1	Monotonic relation between positive acfs	30
3.2	Example densities of transformed Gaussians	32
3.3	Ansatz for inverse mapping of acfs	37
3.4	Comparison of approximations of acfs	38
3.5	Density of the pairwise sum of squared Gaussian processes	42
3.6	Acfs of a non-memoryless transformation of AR(1) and ARFIMA processes	43
4.1	Comparison of biased and unbiased estimation of the acf	46
4.2	acfs of different Gaussian processes in semi- and double-logarithmic scaling	47
4.3	Numerical estimation of LRC: stationary processes	48
4.4	Conceptions of rescaled-range analysis	50
4.5	Visualization of detrended fluctuation analysis	52
4.6	Generalized Haar wavelets	57
4.7	Numerical estimation of LRC: nonstationary processes	61
4.8	Numerical estimation of LRC: additive polynomial trends	62
4.9	Numerical estimation of LRC: polynomial trends and fluctuations	63
4.10	Numerical estimation of LRC: SRC, frequencies and crossovers	64
4.11	Sample trajectories with anomalous diffusion	67
5.1	Non-Gaussian sample mean distributions for meta-Gaussian processes	72
5.2	Gaussian sample mean distribution for stationary Gaussian processes	76
5.3	Effective sample sizes for meta-Gaussian example processes	83
5.4	Effective sample sizes of transformed AR(1) processes	87
5.5	Effective sample sizes of transformed ARFIMA(0, d , 0) processes	89
5.6	Range of the prefactor in the asymptotics of the effective sample size	91
6.1	Strength of seasonality of precipitation and temperature data	96
6.2	Trends in mean and variance of annual precipitation and temperature	97
6.3	Significance of LRC in precipitation data	98
6.4	Model validation: survival function and density	108
6.5	Model validation: cumulative distribution function	109
6.6	Model validation: daily mean and standard deviation	110
6.7	Model validation: probability plots	111
6.8	Model validation: long-range correlations	112

List of figures

6.9	Model validation: short-range correlations	114
6.10	Model validation: waiting times	116
6.11	Model validation: daily mean and standard deviation	118
6.12	Model validation: annual total amounts	119
6.13	Model validation: Mahalanobis distance	121
6.14	Model validation: annual maximal amounts	122
6.15	Model validation: estimated shape parameter	124
6.16	Model validation: skewness of daily amounts and annual totals	126
7.1	Unnormed cdfs of maxima for two example distributions	128
7.2	Unnormed pdfs of maxima for two example distributions	129
7.3	Types of extreme value distributions	132
7.4	Normed pdfs and cdfs of maxima of iid Gaussian samples	137
7.5	Example stationary process with extremal index $\neq 1$	142
7.6	Rate of convergence for maxima of iid Gaussian samples	152
7.7	Rate of convergence for maxima of ARFIMA samples	156
7.8	Penultimate GEV estimate of the shape parameter for iid Gaussian data	158
7.9	Concepts of the block maxima and peaks over threshold approaches	159
7.10	GEV pdf ML adjusted to block maxima of iid Gaussian samples	162
7.11	Exemplified estimation of 100-year return level by block maxima	164
7.12	GPD pdf ML estimated for threshold exceedences of iid Gaussian samples	168
7.13	Mean residual life plot for two example distributions	169
7.14	Exemplified estimation of 100-year return level by peaks over threshold	171
7.15	Peaks over threshold for a Pareto distribution	172
7.16	GEV and Gumbel estimates of 100-year return level of ARFIMA(0, d , 0)	174
8.1	Rate of convergence for maxima of iid tGp samples	182
8.2	Rate of convergence for maxima of LRC tGp samples	183
8.3	Rate of convergence for maxima of transformed ARFIMA samples	184
8.4	Maxima distribution for extremal index unequal 1	189
9.1	Estimation of return levels: tGp-transformed ARFIMA(0, d , 0) processes	200
9.2	Estimation of return levels: tGp-transformed AR(1) processes	201
9.3	Estimation of return levels: extremal index	202
9.4	Comparison of return levels: 20 precipitation data sets	204
9.5	Comparison of return levels: confidence intervals for example data sets	205
9.6	Return levels by norming constants: tail similarity	207
9.7	Alternatively normed pdfs of maxima of iid Gaussian and tGp samples	211
9.8	Modified Gumbel estimates for synthetic data	214
9.9	Modified Gumbel estimates for precipitation data sets	215

List of tables

2.1	Examples of light-, heavy- and fat-tailed distributions	14
3.1	Fitted parameter values of inverse estimation of acfs	39
4.1	Generalized Haar wavelets	59
5.1	Asymptotics of N_{eff} of transformed AR(1) processes	87
5.2	Calculation of the prefactor of the asymptotics of N_{eff}	88
5.3	Asymptotics of N_{eff} of transformed ARFIMA(0, d , 0) processes	89
6.1	Example stations for model validation	106
6.2	Parameter values of adjusted tGp model	107
6.3	Model validation: central statistical quantities	109
6.4	Model validation: effective sample sizes	117
7.1	Extreme value distributions	131
7.2	Examples of the three classes of extreme value distributions	135
7.3	Norming constants for the three types of extreme value distributions	136
7.4	Selection of norming constants for maxima of iid Gaussian samples	147
B.1	Collection of frequently used distributions	228
E.1	Model validation: estimated parameters for all data sets	232
E.2	Model validation: central statistical quantities for all data sets	233
E.3	Model validation: Hurst exponents for all data sets	234
F.1	Characterization of domains of attraction of extreme value distributions	236

1. Introduction

The phenomenon of persistence of fluctuations in geophysical measurements was first described by Hurst in the 1950's [82, 83] and is commonly known as the Hurst effect. Above- or below-average values then have the tendency to occur as clusters [29]. It has been observed in data from various fields of research ever since with precipitation records among them as one of the key indicators for construction planning in flood protection or for adequate measures in water supply. Precipitation time series exhibit the Hurst effect by long periods of drought or heavy rain and directly affect human life regarding climate change (IPCC 2021 [121]).

The presence of long-range dependencies by slowly decreasing autocorrelations crucially affects key statistical quantities such as sample means or return levels of extreme events in general. Municipal regulations typically demand protection against extreme precipitation or flooding events that occur once in a century while empirical records mostly reach back about 60 to 120 years with at most one observation of such a rare event. Reliable statistical inference and realistic simulations raise the need of mathematical modeling as the basis of adequate political or economic decision making by for example persons federally responsible and insurance or financial service providers, respectively. Models for precipitation serve as an ingredient of weather generators or global climate models (GCM) and thus allow for forecasts and simulations in broader context such as future directions of agriculture or the atmosphere itself.

Fractional Gaussian noise and autoregressive-fractionally-integrated-moving-average (ARFIMA) processes are well known data models for Gaussian long-range correlated (LRC) observations. Such data driven models do not describe underlying physical dynamics precisely but are capable of reproducing key statistics of empirical time series and thus of reproducing realistic time series. The however non-Gaussian and LRC nature of many observables of interest require additional model conceptions. A previously known approach to non-Gaussian models is an appropriate nonlinear transformation of a Gaussian process. Aiming at a parametric non-Gaussian LRC model we apply nonlinear transformations to ARFIMA processes and elaborate how to adjust such a model to a marginal distribution and correlation structure (Sect. 1.1) in general.

Both qualitative and quantitative assessments of statistical properties improve statistical inference about LRC data. With special focus on statistical uncertainty we establish effective sample sizes (Sect. 1.2) and extreme value theory (Sect. 1.3) for general stationary and especially transformed Gaussian LRC processes. Applying these concepts to the modeling of non-Gaussian LRC precipitation records, we aim at realistic simulations of occurrences of precipitation and improving the estimation of return levels of rare events by incorporating our model for daily amounts into extreme value theory (Sect. 1.4).

1. Introduction

1.1. Modeling and simulation of non-Gaussian long-range correlated processes

Hurst established long-range dependence in time series in his seminal work on water-runoffs when he was concerned with the planning of the storage capacity of the Great Dam of the river Nile [82, 83]. He studied the accumulation of the difference between annual river inflow and a hypothetical constant water outflow and repeated the remarkable observation of persistence also for other natural phenomena like gauge heights by the Roda (Cairo) Nilometer dated back to the year 640, thickness of tree-rings with records of up to 900 years and varves for 4,000 years, along with rainfall, temperature and pressure measurements, sunspot numbers and even wheat prices [83, Fig. 2]. Ever since, persistent behavior has been found in data from various fields of research, not only in geo- and atmospheric physics for, e.g., global mean temperature anomalies [119], wind speeds [91] and air pollution [86] but also in biology and chemistry, e.g., for DNA sequences [147], neural oscillations [71] and molecular orientation [170], and even in computer science [103, 166], economics [11] and finance [51, 162]. LRC have been observed for precipitation amounts accumulated over time windows of different lengths, such as minutes [122, 149], months [132] and years [30, 70, 146]. This kind of dependency in the data is stronger and more prominent for smaller periods of accumulation and loses intensity for larger ones. Due to the abundance of data on daily precipitation amounts, we concentrate on time series of 24h accumulated amounts [20].

The synthesis of non-Gaussian LRC samples typically employs elaborate numerics with a high number of hidden parameters for obtaining both a prescribed marginal distribution and autocorrelations. Serinaldi and Lombardo generate surrogate data by Davies and Harte's algorithm based on spectral properties [169]. Leland models ethernet traffic by an aggregation of non-Gaussian variates to obtain LRC [103]. Scherrer obtains a gamma distribution for internet traffic by the sum of the squares of independent Gaussian LRC processes [166]. For models of precipitation amounts introducing LRC into synthetic time series of large sample size is possible by specifying correlations on the a larger (annual) time scale and then disaggregating the samples to a smaller (daily) time scale [144] or by a copula-based method [145]. Hosseini et al. give an approach to explicitly accounting for temporal dependencies on an annual time scale between different daily rainfall amounts by considering a high number of previous amounts for conditional probabilities [80]. They obtain gamma distributed precipitation amounts by inserting these temporal correlations directly. The model process, however, essentially represents a Markov process of high order. All the above methods for the numerical generation of non-Gaussian LRC samples raise major algorithmic effort by sophisticated algorithms along with non-interpretable model parameters. We elaborate an alternative approach by applying a nonlinear transformation to an LRC ARFIMA model.

1.2. Statistical uncertainty due to prominent correlations

Non-Gaussian stochastic processes $Y_t = g(X_t)$ obtained from pointwise nonlinear transformations of Gaussian processes $(X_t)_{t \in \mathbb{N}}$ are occasionally referred to as “meta-Gaussian” processes amongst other terms (cp. Rem. 3.1). Commonly monotonic transformations $g : \mathbb{R} \rightarrow \mathbb{R}$ yield non-Gaussian models since by inverse sampling the monotonic transformation $g := F^{-1} \circ \Phi$ generates a desired marginal distribution with cumulative distribution function (cdf) F by concatenation of the corresponding quantile function F^{-1} and the standard Gaussian cdf Φ . For the extreme value theory of meta-Gaussian processes the monotonicity is a crucial property since it retains the order of maximal values among the underlying Gaussian and the transformed samples. In general, an Hermite expansion of the transformation g provides a closed form formula of the autocorrelation function (acf) of the transformed process given the acf of the original Gaussian process for all square-integrable transformations $g \in \mathcal{L}^2(\mathbb{R})$ [22, 161]. As an example, the squared process X_t^2 relates to financial and economic science and its acf is the square of the original one [44, 137].

Janicki applies non-parametric inverse sampling by using the empirical quantile function of correlated non-Gaussian small area income and poverty data to estimate Hermite coefficients [85]. Aiming at a non-Gaussian LRC data model for precipitation amounts based on the Hermite approach, Papalexiou approximates LRC by the Yule-Walker equations or a finite sum of a high number of first-order autoregressive Gaussian processes [138] and estimates the acf of the original Gaussian process directly by the empirical acf and Hermite formula describing the acf of the transformed process. Such an estimate is, however, subject to fluctuations around zero for large time lags. Aiming at a parsimonious, numerically convenient and robust approach to data models for non-Gaussian LRC data with application to precipitation records, we estimate LRC based on the Hurst exponent and incorporate them by transforming an ARFIMA process.

1.2. Statistical uncertainty due to prominent correlations

A typical approach to statistical inference about non-Gaussian data is the Box-Cox method of power transformations to obtain Gaussianity and apply usual statistical methods for Gaussian random variates [160]. The square root of non-Gaussian wind speed measurements, for example, is close to Gaussianity and was modeled by ARFIMA processes [72]. Our inverse method of transforming a Gaussian process to a process with a non-Gaussian marginal distribution requires an accurate estimation of model parameters regarding the transformation and the correlations of the underlying ARFIMA process and then allows for the assessment of statistical quantities based on this model.

1. Introduction

Correlations fall into two essential categories. Short-range correlations (SRC) typically decay exponentially with effects on short-time scales only. LRC, however, asymptotically vanish such slowly that the sum of the acf over all time lags becomes infinite as for example for power-law decaying autocorrelations with an exponent between negative unity and zero. For this power-law setting several methods for the estimation of the Hurst exponent have been proposed [176], among them Hurst's original rescaled-range or R/S -statistics [82], the detrended fluctuation analysis (DFA) [147], and wavelet transforms [4]. These methods can estimate LRC more robustly than a direct estimation of the power-law decay in a double-logarithmic plot of the empirical acf. Fluctuations of the acf around zero, in particular, logarithms of negative values, impede reliable inferences about the rate of the decay of the acf. Note that in real-world data the temporal horizon is always finite, such that it is impossible to decide about the origin of persistent empirical correlations. It may either lie in strong SRC that do not transfer beyond automatically or in LRC that will survive [128].

The presence of correlations affects the rate of convergence of statistical quantities and thus increases statistical uncertainty. By the central limit theorem the sample mean of iid random variates converges in distribution to their mean at rate $\tilde{\sigma}_N \propto N^{-1/2}$ as the sample size N increases. For SRC processes with an acf decaying faster than any power-law an asymptotic effective sample size of $N_{\text{eff}} \sim \tilde{\alpha}N$ as $N \rightarrow \infty$ has been known before with the factor $\tilde{\alpha} \in \mathbb{R}_{>0}$ given by the finite correlation time of the process. Applying the mean square displacement, previous studies have also shown that the standard deviation of the sample mean for Gaussian LRC processes X_t with a power-law decaying acf $\varrho_X(k) \propto k^{-\gamma}$, $\gamma \in (0, 1)$, as time lags $k \rightarrow \infty$ decays at the slower rate $N^{-\gamma/2}$ as $N \rightarrow \infty$. A corresponding asymptotic effective sample size is, however, only known up to $N_{\text{eff}} \sim \alpha N^{-\gamma}$ as $N \rightarrow \infty$ with an unknown prefactor $\alpha \in \mathbb{R}_{>0}$. We aim at a quantitative description of the variance of the sample mean as a finite-sample property of stationary Gaussian and non-Gaussian processes. The concepts of LRC and meta-Gaussian processes serves for the modeling of fluctuations in a system and inference about sampling errors.

1.3. Extreme value theory for stationary processes

Besides expected mean values in Section 1.2, the occurrence and magnitude of extreme events are further statistical quantities crucial for risk assessment. A value $q \in \mathbb{R}$ satisfying $F(q) = 1 - \frac{1}{m}$ is called a return level with return period $m \in \mathbb{N}$ of a random variable with cumulative distribution function F . Low frequency return levels are commonly considered extreme if the exceedance probability $\frac{1}{m} \in (0, 1)$ is less than 5% to 1%. Extreme value theory (EVT) is a tool for statistical inference about such extremes. The extremal types theorem states the generalized extreme value (GEV) distribution as the universal limit distribution of the maximum of a sequence of length $n \in \mathbb{N}$ of independent and identically distributed (iid) samples as the sample size n increases to infinity. The tail of the marginal distribution determines the shape parameter of the asymptotic GEV distribution among three different types of extreme value distributions (EVD) with zero, negative or positive shape, called the Gumbel, Fréchet or Weibull distribution, respectively. The key difference between the three types of EVDs lies in the asymptotics of the tails of their probability density functions (pdf), that capture the probability of the occurrence of rare extremes. The Gumbel class comprises light- to heavy-tailed distributions with tails decaying exponentially or at least faster than any power-law, while the Fréchet class captures fat-tailed distributions with power-law tails. In either case, extreme events of arbitrary magnitude are possible, whereas Weibull distributions exhibit a finite upper endpoint. The validity of these universal results remains valid also in the presence of SRC or power-law decaying LRC [101].

Adjusting an EVD distribution to empirical block maxima yields extreme events as return levels of an estimated GEV distribution. As an advantage this method is always applicable even if only for example annual maximum measurements of gauge heights or precipitation amounts are available. The presence of additional information about the marginal distribution allows for a more detailed assessment of return level estimates obtained from EVT. The tail of the marginal distribution provides knowledge about which of the three types of EVDs are valid for given empirical data. At the same time, for certain distributions in the Gumbel class, namely the Gaussian and certain meta-Gaussian distributions [67, 68], a particular slow convergence of the maximum distribution to a limit Gumbel shape is known such that for finite block size estimates of for example annual return levels applying a Gumbel estimate crucially confounds the resulting return levels. We aim at improving statistical uncertainty of return level estimates for stationary processes in two ways. First, knowledge about the marginal distribution and the strength of both SRC and LRC among individual samples allows for assessing the validity of return level estimates obtained from EVT. Second, we apply an appropriate model for the individual measurements of empirical time series to obtain return levels directly from the marginal distribution if correlations are weak and we incorporate such a model into EVT by defining GEV parameters based on the marginal distribution, which are usually only estimated parameters.

1.4. Contemporary precipitation models

The occurrence and intensity of precipitation is affected by a multitude of atmospheric processes, which evolve on many different temporal and spatial scales. Modeling approaches are either physical, purely data driven or hybrids of the two. Data models are a convenient choice to capture the statistical properties of the outcome of the highly complex physical dynamics underlying precipitation. As an application of the meta-Gaussian approach to the modeling of non-Gaussian LRC data based on the transformation of a Gaussian ARFIMA process we use a truncated-Gaussian-Power transformation (tGp) to describe the marginal distribution of daily precipitation records.

Some earlier studies [88, 159] found LRC in daily precipitation and point out that in general this indicator of persistence is rather weak but still significant [88, 89, 159]. Mid-latitude precipitation further shows prominent intermittency by periods of several days of zero precipitation. These zero measurements fundamentally influence correlations.

A classical approach to modeling daily precipitation statistics are two-part models, in which the occurrence or absence of precipitation and its positive amounts are generated independently [104, 106, 184]. Correlations between different occurrences are commonly introduced by a Markov chain of first or second order. Recent studies explicitly address correlations between different precipitation amounts by modified Markov chain approaches [34, 136].

There is ongoing discussion on the most appropriate choice of a model distribution for daily precipitation amounts. In particular, their tail behavior is crucial for the estimation of large precipitation events. Most global studies with focus on large amounts find tails heavier than exponential [135, 143, 168]. By arguments from atmospheric physics, Wilson and Toumi [185] deduced a stretched exponential tail with a universal shape parameter as an approximation for the extreme regime. The geographic location and the climatic zone have strong influence on which distribution is most realistic. Case specific suggestions range from the light-tailed exponential, mixed-exponential or gamma distribution [104, 154] and the heavy-tailed generalized gamma [142] or lognormal [106] distribution to fat-tailed Burr-type distributions [140] and q -exponentials [186]. As a remark, since none of the aforementioned distributions is stable under convolution with itself, it is also evident that the distribution will change if the period of accumulation is changed, for example, hourly data will follow a different distribution than daily data. Figure 1.1 shows the highly non-Gaussian empirical density of daily precipitation recorded at the city of Potsdam in Germany from the year 1893 until 2018 [49].

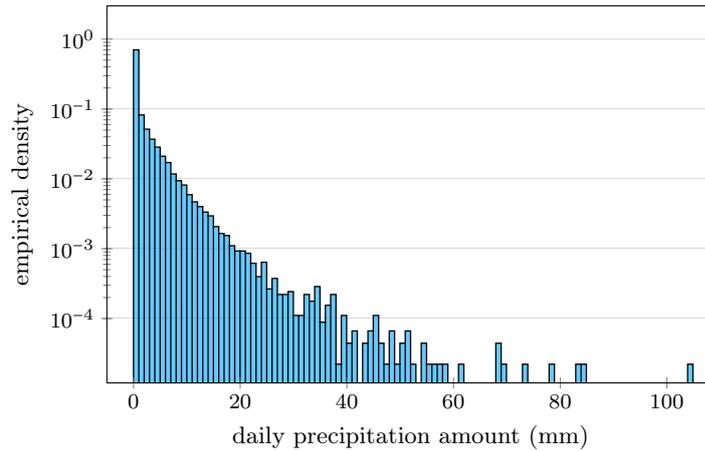


Fig. 1.1 Distribution of daily precipitation amounts at the city of Potsdam, Germany.

To include dependencies between precipitation amounts also transformed Gaussian processes have been applied before [14, 15]. As mentioned in Section 1.1 meta-Gaussian models can generate a prescribed distribution by inverse sampling. The intermittency that precipitation time series naturally exhibit is automatically incorporated into such a model when applying truncated, also called mixed-type, distributions, that generate a point mass at zero. Correlations can directly be defined by the underlying Gaussian process, which is then transformed adequately to obtain a certain distribution. Recent studies include also physical knowledge in the sense that the underlying (spatio-temporally correlated) Gaussian process describes atmospheric dynamics, which are then transformed appropriately. On that account, tGp transformations of Gaussian SRC processes have been used to model the distribution of precipitation amounts and their dynamics [6, 164, 171]. Without explicitly pointing out the property of LRC, Baxevani and Lennartsson use an underlying Gaussian process with a temporally hyperbolically (and spatially exponentially) decaying spatio-temporal autocorrelation function [18]. Transforming a process, however, does not preserve its temporal correlations, so that additional adjustments of the correlations are necessary to attain prescribed correlations.

By Section 1.1 an approach to directly estimating the acf of the underlying Gaussian process is expanding the transformation in Hermite polynomials. A historical note on Hermite series in precipitation modeling is given in [145]. Guillot applies this method to the spatial behavior of rainfall events with an exponentially decaying acf and a truncated gamma distribution for the rainfall amounts [64]. Alternatively, Papalexiou fits a function that maps the autocorrelations of the transformed to the autocorrelations of the underlying Gaussian process [138].

1. Introduction

We provide a five-parameter meta-Gaussian model for mid-latitude daily precipitation amounts. Three parameters describe the location, scale and shape of the tGp distribution and one parameter each describes the SRC and LRC of the underlying ARFIMA process. Section 1.1 introduces the robust estimation of LRC by the asymptotic strength of fluctuations on the empirical data. Box and Jenkins established a method for the identification and estimation of Gaussian SRC ARIMA models based on the acf and partial autocorrelations in their seminal work on time series analysis [27]. Our precipitation data and the corresponding model, however, have a non-Gaussian, strongly asymmetric marginal distribution, so that we formulate a different approach to the estimation of the autoregressive parameter based on conditional probabilities of consecutive pairs of days with at least a prescribed amount.

Typically precipitation amounts are measured by reading off the height of collected water in a rain gauge with a precision of 0.1 mm. Hence, historical records in the range of one millimeter and below have to be treated with care, due to measurement errors. Mainly evaporation strongly affects measurements of this magnitude. Modern measurement instruments increase the accuracy by applying laser detection. Noticeable drizzling rain ranges from about 0.1 to 0.5 mm/h, weak and moderate showers from about 0.6 mm/h to 12 mm/h, while heavy rain and storm reach 30 mm/h and more than 50 mm/h, respectively. After periods of some days of such precipitation intensities the danger of flooding arises. Note that $1 \text{ mm} = 1 \text{ l/m}^2$. For details on the classification and measurement of precipitation refer to the World Meteorological Organization (WMO) or any national weather service like the German (DWD) or British (MetOffice) Meteorological Office, respectively. Most studies estimate the marginal distribution by maximum likelihood or method of moments approaches. As the tail is naturally represented only poorly in empirical data, this may lead to an underestimation of extremal events [20]. Such an effect was addressed by for example entropy based parameter estimation [140]. We estimate the marginal distribution in a tail-oriented way discriminating small and emphasizing large amounts. We validate this approach by a detailed statistical analysis of key quantities such as daily and annual mean and variance, duration of droughts and extreme precipitation events and derive a robust and easily applicable data model for daily precipitation. Applying the effective sample introduced in Section 1.2, we quantify the statistical uncertainty of daily mean and annual total amounts. Incorporating the tGp LRC model into EVT we derive and validate return levels of extreme precipitation with improved uncertainty.

Structure of the thesis

The present thesis derives theoretical results on effective sample sizes for stationary processes and especially meta-Gaussian processes in Chapter 5 and on extreme value theory for meta-Gaussian processes in Chapter 8. Chapter 6 contains results on the design, estimation and validation of a tGp LRC model for daily precipitation amounts together with an application of the concept of effective sample sizes to empirical precipitation data. Chapter 9 establishes a procedure for the assessment of return level estimates obtained from EVT along with a method for incorporating meta-Gaussian models into EVT for improving statistical uncertainty of return level estimates and applies these theoretical results to empirical precipitation data.

The Chapters 2 to 4 and 7 provide existing knowledge on Gaussian and meta-Gaussian processes, the numerical detection of LRC and EVT for both iid and stationary processes, respectively, in form of mathematical theorems mainly, without proof for later application or generalization. If the proof of a theorem is later used for deriving further results, then these methods are introduced at least heuristically for later referencing.

Every chapter obeys a final section with closing remarks or an outlook. These sections link the either previously known or newly derived contents of the respective chapter to other fields of research or opens question for further research. Chapter 10 concludes the main findings and lists the central contributions of the present thesis

Notation

Throughout the present thesis the notation X_t contextually refers to either a stochastic processes $(X_t)_{t \in \mathbb{N}_{\geq 0}}$ or a random variable representing an individual component. Statistical quantities belonging to a process X_t with an own symbol carry the process as an index, for example ϱ_X denotes the acf of a process X_t . The symbols Φ and ϕ denote the standard Gaussian cdf and pdf in all chapters. In the end of the thesis a list of abbreviations and general or specific symbols is given.

2. Gaussian and non-Gaussian stochastic processes

Gaussian stochastic processes exhibit an outstanding importance in probability theory and modeling applications. Not only the exceptional role of the normal distribution by the central limit theorem but also the particular convenience of the definition of Gaussian processes just by their first and second moments reasons their prominent position. Nonlinear transformations of Gaussian processes transfer these advantages to the modeling of non-Gaussian stochastic processes.

Let $(X_t)_{t \in \mathbb{T}}$ be a stochastic process. The individual variates X_t are indexed by a family \mathbb{T} of indices. Typical choices of the set \mathbb{T} are subsets of integers or reals and provide a temporal interpretation of the order of the variates. Brownian motion or the Ornstein-Uhlenbeck process are prominent examples of *continuous-time* stochastic processes indexed by $\mathbb{T} \subseteq \mathbb{R}$. Such processes allow for the formulation of stochastic integrals and stochastic differential equations. *Discrete-time* processes are considered as an approximation of continuous-time processes observed at equally-spaced points $t \in \mathbb{T} \subseteq \mathbb{Z}$ in time [24]. Aiming at modeling daily precipitation amounts in Chapter 6, indexing by natural numbers $\mathbb{T} = \mathbb{N}$ properly captures the discrete temporal resolution of the empirical data and is assumed in what follows. Our focus lies on stationary non-Gaussian stochastic processes with long-range correlations that are the result of a pointwise transformation of a Gaussian process.

Based on the textbook [137] the Sections 2.2 and 2.4 recap the general concepts of stationarity and long-range correlations for Gaussian (Sect. 2.1) and general non-Gaussian processes. Section 2.3 introduces a classification of distributions by means of their tail behavior. ARFIMA models (Sect. 2.5) are a well-established approach to the generation of discrete-time Gaussian long-range correlated data (Sect. 2.6). Nonlinear transformations of ARFIMA processes provide models for non-Gaussian long-range correlated processes (cp. Chap. 3) such as daily precipitation amounts (cp. Chap. 6).

2.1. Gaussian processes

The property of Gaussianity describes a Gaussian character among arbitrary parts of a stochastic process. Not only the distribution of the process is assumed Gaussian at any single point in time but also the joint distribution of any finite ensemble of members of the process.

Definition (Gaussian process): A stochastic process $(X_t)_{t \in \mathbb{N}}$ is called a *Gaussian process* if the joint distribution of any finite number $s \in \mathbb{N}$ of variates X_{t_1}, \dots, X_{t_s} at points $t_i \in \mathbb{N}$ in time with $i = 1, \dots, s$ is a multivariate Gaussian distribution.

2. Gaussian and non-Gaussian stochastic processes

By the definition of a multi-dimensional normal distribution Gaussian processes $(X_t)_{t \in \mathbb{N}}$ are uniquely determined yet by their first and second moments $\mathbb{E}[X_t]$ and $\text{Cov}(X_s, X_t)$ for indices $s, t \in \mathbb{N}$, respectively. The multivariate normal distribution of subsets of a Gaussian processes ensures that sums over such collections follow a normal distribution as well. Model formulation is therefore particularly straightforward, for example, for linear Gaussian processes such as the ARFIMA model introduced in Section 2.5.

2.2. Stationary processes

Stationarity describes the property of an unchanged stochastic behavior among different parts of a stochastic process $(X_t)_{t \in \mathbb{N}}$. Denote by $F_{(X_{t_1}, \dots, X_{t_s})} : \mathbb{R}^s \rightarrow [0, 1]$ the joint cdf of a finite number $s \in \mathbb{N}$ of variates X_{t_1}, \dots, X_{t_s} of the process at points $t_1, \dots, t_s \in \mathbb{N}$ in time, if such a cdf exists for all parts of the process.

Definition (Strict stationarity): A stochastic process $(X_t)_{t \in \mathbb{N}}$ is called *strictly* or *strict-sense stationary* if for all finite numbers $s \in \mathbb{N}$ of points $t_1, \dots, t_s \in \mathbb{N}$ in time and time lags $k \in \mathbb{N}_0$ the joint distributions of the variates X_{t_1}, \dots, X_{t_s} and the variates $X_{t_1+k}, \dots, X_{t_s+k}$ with shifted temporal indices satisfy

$$\mathbb{P}(X_{t_1+k} \leq x_1, \dots, X_{t_s+k} \leq x_s) = \mathbb{P}(X_{t_1} \leq x_1, \dots, X_{t_s} \leq x_s) \quad (2.1)$$

for all points $(x_1, \dots, x_s) \in \mathbb{R}^s$. In other words, if existent, the joint cdfs of such ensembles satisfy

$$F_{(X_{t_1+k}, \dots, X_{t_s+k})}(x_1, \dots, x_s) = F_{(X_{t_1}, \dots, X_{t_s})}(x_1, \dots, x_s). \quad (2.2)$$

Strict-sense stationarity considers the entire joint distribution of arbitrary parts of a stochastic process. An alternative to this broad view involves only the first and second moments of the process and is called *weak*, *second-order* or *wide-sense* stationarity.

Definition (Second-order stationarity): A stochastic process $(X_t)_{t \in \mathbb{N}}$ with time-independent mean $\mathbb{E}[X_t] =: \mu \in \mathbb{R}$ and variance $\text{Var}(X_t) =: \sigma^2 \in \mathbb{R}_{>0}$ is called *second-order* or *wide-sense stationary* if there exists an *autocorrelation function* (acf) ϱ_X such that

$$\varrho_X(k) = \frac{\text{Cov}(X_t, X_{t+k})}{\sigma^2} \quad (2.3)$$

for all points $t \in \mathbb{N}$ in time and time lags $k \in \mathbb{N}_0$.

2.3. Classification of the marginal distribution

Note that in view of equation (2.3) the autocovariance function $\text{Cov}(X_s, X_t)$ of a second-order stationary process depends on the time lag $k = |s - t|$ only and not on the particular points $s, t \in \mathbb{N}$ in time. In what follows we refer to second-order stationary processes whenever using the term “stationary”.

The simplest examples of (strictly and second-order) stationary processes are *independent and identically distributed* (iid) processes. Such processes are commonly referred to as *white noise processes*. A stationary Gaussian process $(X_t)_{t \in \mathbb{N}}$ is entirely determined by a Gaussian marginal distribution $N(\mu, \sigma^2)$ with constant mean $\mu \in \mathbb{R}$ and variance $\sigma^2 \in \mathbb{R}_{>0}$ and a time-independent acf ϱ_X in virtue of equation (2.3). Therefore, the stationarity and the fluctuations of Gaussian processes are directly assessable, which provides a particularly simple approach to model formulation.

Remark 2.1: Strictly stationary processes are not necessarily weakly or second-order stationary, since the latter property assumes finite first and second moments. For Gaussian processes the two properties are equivalent, since given first and second moments fully characterize all of their marginal distributions.

2.3. Classification of the marginal distribution

Joint distributions of arbitrary subsets of a stochastic process are called *marginal distributions*. If the number $s \in \mathbb{N}$ of variates in such a subset is finite like in the definitions of Gaussian (Sect. 2.1) or stationary (Sect. 2.2) processes, the related joint distribution is commonly referred to as an *s-dimensional marginal distribution*. All marginal distributions of strictly stationary processes equate to each other. For modeling purposes the one-dimensional marginal distribution is of particular importance because it directly describes an observed quantity, which is for example a daily precipitation amount in our study in Chapter 6. In what follows we refer to this distribution whenever using the term “marginal distribution”.

The tail behaviour of stochastic distributions is of special interest because it allows for limit theorems concerning the sums of random variables (cp. Sect. 5.1) or the occurrence of their extreme values (cp. Sect. 7.1.2). We classify distributions based on their tails according to [48].

Definition (Heavy-tailed distributions): The distribution of a random variable X is called *heavy tailed* if for all constants $t \in \mathbb{R}_{>0}$ the *moment generating function* M satisfies

$$M(t) := \mathbb{E}[e^{tX}] = \infty. \quad (2.4)$$

2. Gaussian and non-Gaussian stochastic processes

If the distribution of the random variable X has a pdf f_X , then equation (2.4) reads

$$\int_{\mathbb{R}} e^{tx} f_X(x) dx = \infty. \quad (2.5)$$

If there exists a constant $t \in \mathbb{R}_{>0}$ such that the expectation (2.4) is finite, then the distribution is commonly called *light tailed*.

Distributions with an at least exponentially or faster decaying pdf are light tailed. A distribution whose pdf decays slower than every exponential function is heavy tailed. In the literature of applied sciences occasionally any distribution with a tail decaying slower than the one of the Gaussian distribution is called heavy tailed. In the physics literature power-law decaying pdfs are commonly referred to as heavy tailed. We classify more specifically by denoting such distributions as *fat tailed*.

Definition (Fat-tailed distributions): A distribution with pdf f is called *fat tailed* if there exists an exponent $\alpha \in \mathbb{R}_{>1}$ such that

$$f(x) \propto x^{-\alpha} \quad (x \rightarrow \infty). \quad (2.6)$$

The constraint $\alpha > 1$ on the power-law exponent in proportionality (2.6) is necessary to ensure integrability since pdfs integrate to unity on the entire probability space. Including the concepts of regularly or slowly varying functions a generalization of the perception of fat tails is possible. Further classes of heavy-tailed distribution are subexponential, long-tailed or dominatedly varying distributions [48, Sect. 1.4.1]. In Chapter 6 we apply a power of a truncated Gaussian distribution to the modeling of daily precipitation amounts. The resulting truncated-Gaussian-power (tGp) distribution is heavy but not fat-tailed (Sect. 6.1.3). Table 2.1 states some examples of light-, heavy- and fat-tailed distributions. Note that any fat-tailed distribution is also heavy tailed.

Light-tailed distributions	Heavy-tailed distributions	Fat-tailed distributions
Gaussian	Lognormal	Pareto
Exponential	truncated-Gaussian power	Cauchy
Gamma		Beta
		α -stable, $\alpha \in (0, 2)$

Table 2.1 Examples of light-, heavy- and fat-tailed distributions

Note that the notion of α -stable distributions refers to densities with tail behavior $\propto x^{-\alpha-1}$ as $x \rightarrow \infty$ in contrast to definition (2.6). For parameters $\alpha \in [1, 2]$ these distributions exhibit an infinite variance with infinite mean for $\alpha \in (0, 1]$. Special cases are the normal distribution with $\alpha = 2$ and the Cauchy distribution with $\alpha = 1$.

2.4. Long-range correlated processes

The concept of long-range correlations (LRC) in time series was established by Hurst in 1951 in his fundamental work on water storage capacities of the Great Dam of the river Nile [82, 83]. He formulated the *rescaled-range analysis* or *R/S-statistics* (Sect. 4.2.1) for the numerical detection of LRC in empirical observations. Ever since, LRC have been found in data sets from various fields of science (cp. Sect. 1.2). Hurst observed that fluctuations of empirical gauge heights deviated from what uncorrelated Gaussian in- and outflow would have generated. He identified that the increase of the range R of the water levels normalized by the square root S of their sample variance on average follows a power law

$$\frac{R}{S}(s) \propto s^H \quad (s \rightarrow \infty) \quad (2.7)$$

as time $s \in \mathbb{N}$ increases (cp. relation (4.8)). The scaling or *R/S-exponent* H in relation (2.7) is known as the *Hurst exponent* and a matter of current research [33, 107]. See Chapter 4 and relation (4.3) for a more general perception of this exponent. While Gaussian white noise input gives the exponent $H = \frac{1}{2}$ in the power law (2.7) [22], Hurst detected growth exponents $\frac{1}{2} < H < 1$ in the hydrological data. Suchlike deviations from uncorrelated behavior in the *R/S-exponent* is known as the *Hurst effect* ever since [61]. Mandelbrot [116] showed that for stationary processes the exponent $H = \frac{1}{2}$ is robust against non-Gaussian marginal distributions even with infinite variance. More precisely, infinite second moments are neither necessary nor sufficient for exponents $H \neq \frac{1}{2}$ in relation (2.7) [115]. In fact, long-range correlations [114] and nonstationarities [33] are possible origins of the Hurst effect as well (see Rems. 2.2 and 4.7).

Remark 2.2: Processes with infinite variance exhibit the so-called *Noah effect* [112] of the occurrence of events from the far tail of the marginal distribution. The naming originates from the bible's story about the Noachian flood. For marginal distributions with infinite first or second moments extreme events of such high extents feature a higher probability of becoming real. Stochastic processes with long-range correlations show a tendency of long periods of higher or lower values. Mandelbrot called such a behavior the *Joseph effect* according to the biblical story of Joseph that describes seven years of high precipitation followed by seven years of drought [113]. See Remark 4.7 for recent generalizations of the Hurst exponent such as the *Moses effect*.

2. Gaussian and non-Gaussian stochastic processes

The acf (2.3) provides access to the correlation structure of stationary stochastic processes and the definition of long- and short-range correlations.

Definition: A stationary stochastic process $(X_t)_{t \in \mathbb{N}}$ is said to exhibit (*temporal*) *long-range correlations* (LRC) if its acf ϱ_X (2.3) is not absolutely summable, in other words, if

$$\tau := \sum_{k=0}^{\infty} |\varrho_X(k)| = \infty. \quad (2.8)$$

If the sum in (2.8) is finite, then the process is said to have *short-range correlations* (SRC).

Remark 2.3: An alternative denomination of long-range and short-range correlations are the terms *long-* and *short-memory*, respectively [56, 137, 163]. To obviate misleading overlap with the notion of memory in the field of (stochastic) differential delay equations [130] we employ the terms LRC and SRC. Correlations are a statistical effect essentially, while memory typically suggests a physical causality.

The sum (2.8) is the discrete-time analogon of the *correlation time* of a time-continuous stochastic process. An in continuous time $t \in \mathbb{R}_{\geq 0}$ exponentially decaying quantity $N(t) = N(0)e^{-\frac{t}{T}}$, in general, features a typical mean life time $T \in \mathbb{R}_{> 0}$ that mimics a temporal scale in the system by $\frac{N(0)}{N(T)} = e$. As $T = \int_0^{\infty} e^{-\frac{t}{T}} dt$ in view of purely exponential decay, the more general equation

$$T := \int_0^{\infty} \varrho_X(h) dh \quad (2.9)$$

serves as a definition of the correlation time T of a continuous-time and stationary stochastic process $(X_t)_{t \in \mathbb{R}}$ with an acf ϱ_X . As for LRC processes a finite mean correlation time, and thus a typical temporal scaling, does not exist by equation (2.8), such processes are called *scale free* or *scale invariant*. In Chapter 5 we incorporate correlation times in the concept of effective samples sizes for meta-Gaussian LRC processes.

The presence of a power-law decaying acf occasionally serves as a definition of LRC processes [46, 88]. An asymptotic shape of the acf ϱ_X like

$$\varrho_X(k) \propto k^{-\gamma} \quad (k \rightarrow \infty) \quad (2.10)$$

for time lags $k \in \mathbb{N}$ and exponents $0 < \gamma \leq 1$ is a sufficient condition for LRC. An acf that decays to zero more rapidly than a power law (e.g., exponentially) or is constantly zero (uncorrelated behavior), so that the correlation time (2.8) converges, yields an SRC process.

2.5. Autoregressive-fractionally-integrated-moving-average (ARFIMA) processes

As a remark, a stochastic process with a power-law decaying acf and an exponent $\gamma > 1$ in relation (2.10) still exhibits a finite correlation time. Such correlation behavior shall be referred to as *intermediate* or *intermediate memory* [137]. For more details on categorizing correlations in time series including an also commonly used description by the power spectral density [84, Sect. I.6] or by a linear decomposition of the process see [28] or [151].

2.5. Autoregressive-fractionally-integrated-moving-average (ARFIMA) processes

For the representation of the different types of correlations described in Section 2.4 a variety of Gaussian stochastic models is available. Graves et al. [61] give a historical overview of the most prevalent approaches suchlike. The ARFIMA, sometimes also called FARIMA, process introduced below is one of the discrete-time LRC models among them. Our modeling of non-Gaussian daily precipitation amounts in Chapter 6 bases on a nonlinear transformation of an ARFIMA process.

Autoregressive-integrated-moving-average models, abbreviated as ARIMA(p, d, q) with parameters $p, q, d \in \mathbb{N}_{\geq 0}$ allow for the modeling of certain discrete-time Gaussian SRC processes [27]. The values p and q determine the numbers of involved *autoregressive* (AR) and *moving-average* (MA) components, respectively (cp. Eq. (2.15)). The integration parameter d describes a degree of differencing of previous values of the process. For all values $d \in \mathbb{N}_{>0}$ ARIMA models are nonstationary. Special cases, however, are the stationary *autoregressive* AR(p) ($d = 0 = q$) and *autoregressive-moving-average* ARMA(p, q) ($d = 0$) models. As a common feature all these models are SRC processes as there exists an asymptotically exponentially decaying envelope of their acfs. The acf ϱ_X of an AR(1) process $(X_t)_{t \in \mathbb{N}}$ with

$$X_t = \varphi X_{t-1} + \varepsilon_t \quad (2.11)$$

and $X_0 = 0$, an AR parameter $\varphi \in \mathbb{R}$ such that $|\varphi| < 1$ and a zero-mean Gaussian white noise process $(\varepsilon_t)_{t \in \mathbb{N}}$ with variance σ_ε^2 for time lags $k \in \mathbb{N}_0$ reads

$$\varrho_X(k) = \varphi^k \quad (2.12)$$

with variance

$$\text{Var}(X_t) = \frac{\sigma_\varepsilon^2}{1 - \varphi^2}. \quad (2.13)$$

2. Gaussian and non-Gaussian stochastic processes

Mandelbrot established the concept of *fractional Brownian motion* B_H (fBm) with a parameter $0 < H < 1$ for the modeling of Gaussian LRC processes [112]. The fBm is a continuous-time nonstationary Gaussian process defined by a fractional integral of Gaussian white noise. It has a fractional nature by statistical self-similarity, in the sense that for all scaling and temporal parameters $a, t \in \mathbb{R}_{>0}$ the variates $B_H(at)$ and $a^H B_H(t)$ share the same distribution. The increment process $B_H(t+1) - B_H(t)$ is called *fractional Gaussian noise* (fGn) and is stationary with a power-law decaying acf ϱ_{fGn} that satisfies [176]

$$\varrho_{\text{fGn}}(h) \sim H(2H-1)h^{2H-2} \quad (h \rightarrow \infty) \quad (2.14)$$

for self-similarity parameters $H \neq \frac{1}{2}$ and time lags $h \in \mathbb{R}_{\geq 0}$. The symbol “ \sim ” in relation (2.14) denotes *asymptotic equivalence*. For parameters $\frac{1}{2} < H < 1$ a fGn process obeys positive LRC, hence, is an LRC process with $0 < \gamma = 2 - 2H < 1$ in accordance with definition (2.8). For further details on negative LRC consult [112].

Hosking [78] and Granger and Joyeux [60] extended the toolbox of Gaussian models by *autoregressive-fractionally-integrated-moving-average processes* ARFIMA(p, d, q) with parameters $p, q \in \mathbb{N}_{\geq 0}$ and $d \in (0, \frac{1}{2})$ to get hands on discrete-time stationary Gaussian LRC processes. As a generalization of the aforementioned ARIMA models ARFIMA processes still allow for the modeling of SRC effects by including p autoregressive and q moving-average components. Such SRC are not explicitly inherent to fGn [78]. Using the notion of a backshift operator $BX_t := X_{t-1}$ for time indices $t \in \mathbb{N}$, an ARFIMA(p, d, q) process is defined as the solution $(X_t)_{t \in \mathbb{N}}$ to the fractional difference equation

$$\Phi(B)X_t = \Theta(B)(1-B)^{-d}\varepsilon_t, \quad (2.15)$$

where $(\varepsilon_t)_{t \in \mathbb{Z}}$ is a zero-mean Gaussian white noise process with variance $\sigma_\varepsilon^2 \in \mathbb{R}_{>0}$. The functions $\Phi(B) := 1 - \varphi_1 B - \dots - \varphi_p B^p$ and $\Theta(B) := 1 - \theta_1 B - \dots - \theta_q B^q$ in equation (2.15) denote polynomial autoregressive and moving-average operators, respectively. The coefficients $\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q \in \mathbb{R}$ are such that Φ and Θ do not have common roots [137].

Remark 2.4 (Rem. 3.2 in [137]): Under appropriate assumptions the solution to equation (2.15) is unique. This is not necessarily the case if the process is defined by the equation

$$\Phi(B)(1-B)^d X_t = \Theta(B)\varepsilon_t. \quad (2.16)$$

If a process X_t solves equation (2.16) with $d > 0$, then for all random variables V with finite variance the process $X_t + V$ is a solution, too.

2.5. Autoregressive-fractionally-integrated-moving-average (ARFIMA) processes

By a binomial expansion the fractional difference operator in definition (2.15) reads

$$(1 - B)^{-d} := \sum_{j=0}^{\infty} \psi_j B^j \quad \text{with} \quad \psi_j := \frac{\Gamma(d+j)}{\Gamma(d)\Gamma(1+j)}. \quad (2.17)$$

An ARFIMA(0, d , 0) process $(X_t)_{t \in \mathbb{N}_0}$ has the infinite moving-average representation

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}. \quad (2.18)$$

One can show that the coefficients ψ_j , $j \in \mathbb{N}_0$, satisfy $\psi_j \sim \frac{j^{d-1}}{\Gamma(d)}$ as $j \rightarrow \infty$ [78]. Hence, the sum (2.18) converges in the space \mathcal{L}_2 with $\text{Var}(X_t) = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_j^2 < \infty$ and

$$\text{Var}(X_t) = \frac{\Gamma(1-2d)}{\Gamma(1-d)^2}. \quad (2.19)$$

The moving-average representation (2.18) serves as a direct approach to the numerical generation of synthetic ARFIMA time series recapped in Section 2.6. We apply this method in Section 6.2.4 and Section 6.3 for the validation of our model of daily precipitation amounts and the generation of confidence intervals, respectively.

Irrespective of the values of the AR and MA parameters p and q , the acf ϱ_X of an ARFIMA(p, d, q) process asymptotically behaves like

$$\varrho_X(k) \propto k^{2d-1} \quad (k \rightarrow \infty). \quad (2.20)$$

For every $d \in (0, \frac{1}{2})$ under mild conditions (i.e., all roots of the operators Φ and Θ in equation (2.15) lie outside the closed unit disc in the complex plane) the ARFIMA(p, d, q) process is stationary, causal and invertible [137, Thm. 3.4] and obeys positive LRC. With $0 < \gamma = 1 - 2d < 1$ it is an LRC process in the sense of definition (2.8). The parameters p and q account for SRC effects that decay exponentially while the parameter d describes the power-law decay (2.20) of the acf ϱ_X [78].

In Chapter 6 we apply an ARFIMA(1, d , 0) process with LRC parameters $0 < d < \frac{1}{2}$. For details on the existence of solutions of equation (2.15), more general conditions on the parameter d for negative LRC and on the polynomials Φ and Θ , and for further properties of the ARFIMA(p, d, q) process see [27], [78] and [137]. By equation (2.15) an ARFIMA(1, d , 0) process given by $(1 - \varphi B)X_t = (1 - B)^{-d}\varepsilon_t$ can be represented by

$$X_t = \varphi X_{t-1} + \tilde{X}_t \quad (2.21)$$

with an ARFIMA(0, d , 0) process $(\tilde{X}_t)_{t \in \mathbb{N}_0}$. The latter process satisfies $\tilde{X}_t = (1 - B)^{-d}\varepsilon_t$ and equation (2.18). The AR and LRC parameters in equation (2.21) are chosen such that $\varphi \in \mathbb{R}$ with $|\varphi| < 1$ and $0 < d < \frac{1}{2}$, respectively, to ensure stationarity and invertibility. Therefore, an ARFIMA(1, d , 0) process is a first-order autoregression with ARFIMA(0, d , 0) perturbations [78].

2. Gaussian and non-Gaussian stochastic processes

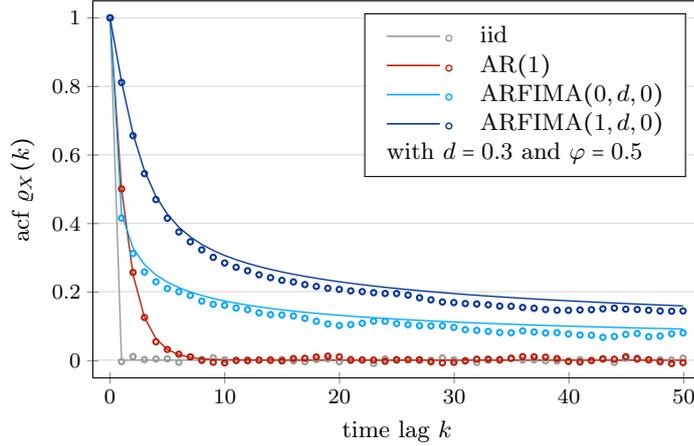


Fig. 2.1 Theoretical (solid lines) and empirical (circles) acfs of a sample time series X_1, \dots, X_N of length $N = 36,500$ of an iid, an AR(1), an ARFIMA(0, d , 0) and an ARFIMA(1, d , 0) process with standard Gaussian marginal distribution. The solid lines are linear interpolations of the analytically known values of the acfs at time lags $k \in \mathbb{N}_0$.

The acf $\varrho_{\tilde{X}}$ of an ARFIMA(0, d , 0) process \tilde{X}_t and ϱ_X of an ARFIMA(1, d , 0) process X_t are analytically known [78] and read

$$\varrho_{\tilde{X}}(k) = \frac{\Gamma(1-d)}{\Gamma(d)} \cdot \frac{\Gamma(k+d)}{\Gamma(k-d+1)} \quad \text{and} \quad (2.22)$$

$$\varrho_X(k) = \varrho_{\tilde{X}}(k) \cdot \frac{{}_2F_1(1, d+k, 1-d+k; \varphi) + {}_2F_1(1, d-k, 1-d-k; \varphi) - 1}{(1-\varphi) {}_2F_1(1, 1+d, 1-d; \varphi)} \quad (2.23)$$

for time lags $k \in \mathbb{N}_0$ with

$$\varrho_X(1) = \frac{(1+\varphi^2) {}_2F_1(1, d, 1-d; \varphi) - 1}{\varphi ({}_2F_1(1, d, 1-d; \varphi) - 1)} \quad (2.24)$$

for unity time lag $k = 1$. Therein, the function ${}_2F_1$ is the hypergeometric function. In Section 6.2.3 we use the lag-1 acf (2.24) for the fit of the SRC parameter of our model for daily precipitation amounts. Asymptotically, the acf (2.22) and (2.23) satisfy

$$\varrho_{\tilde{X}}(k) \sim \frac{\Gamma(1-d)}{\Gamma(d)} k^{2d-1} \quad (k \rightarrow \infty) \quad \text{and} \quad (2.25)$$

$$\varrho_X(k) \sim \frac{1+\varphi}{(1-\varphi)^2} \cdot \frac{1}{{}_2F_1(1, 1+d, 1-d; \varphi)} \cdot \varrho_{\tilde{X}}(k) \quad (k \rightarrow \infty). \quad (2.26)$$

Figure 2.1 shows the analytical and empirical acfs in linear scaling for four of the prototypes of discrete-time stationary Gaussian processes, namely of an iid, an AR(1), an ARFIMA(0, d , 0) and an ARFIMA(1, d , 0) process with AR and LRC parameter $\varphi = 0.5$ and $d = 0.3$, respectively. The sample time series are generated by the algorithm described in Section 2.6 and the acfs are estimated by the *sample acf* (4.1) recapped in Section 4.1. For semi- and double-logarithmic visualizations see Figure 4.2. Chapter 4 provides comments on the numerical estimation of the strength correlations in stochastic processes.

2.6. Numerical generation of ARFIMA time series

When numerically generating synthetic time series obtaining prescribed SRC and LRC accurately is a particular challenge. Several methods for this purpose have been proposed [180, Sect. 3.3], such as an approximative moving-average method, the method by McLeod and Hipel, the Durbin-Levinson algorithm or methods based on the Yule-Walker equations [139] or the FFT power spectrum [39]. In early years of computational studies smaller computing power required a direct generation of prescribed correlations to ensure that already short time series exhibit the correct correlations. Contemporary computing devices allow for an as accurate representation of given correlations also by approximate methods.

For synthesizing ARFIMA(1, d , 0) model time series for specific values of the AR and LRC parameter φ and d , respectively, we follow an algorithm formulated by Hosking [79] based on equation (2.21). An ARFIMA(1, d , 0) time series $(X_t)_{t=0}^{N-1}$ of length $N \in \mathbb{N}$ is directly obtained by equation (2.21) as follows.

1. Generation of an ARFIMA(0, d , 0) time series $(\tilde{X}_t)_{t=-L}^{N-1}$ by truncating the sum in its moving-average representation (2.18), starting with $\tilde{X}_{-L} = 0$ for a number $L \in \mathbb{N}$
2. Execution of $\tilde{N} := N + L$ auto-regressions starting with X_{-L} to obtain a time series $(X_t)_{t=-L}^{N-1}$ of length \tilde{N} by using the fractionally differenced noise \tilde{X}_t as an input for equation (2.21)
3. Ensuring accomplishment also of the desired SRC by omission of the L transients X_{-L}, \dots, X_{-1}

Aiming at an ARFIMA(1, d , 0) process with AR parameter φ , Hosking proposes the choice $|\varphi|^L \leq \varepsilon = 0.01$ for the number $L \in \mathbb{N}$ of transients. We employ a smaller tolerance of $\varepsilon = 0.001$ to further increase accuracy and make use of available computing power. For generating a member \tilde{X}_t of the ARFIMA(0, d , 0) time series we include $2\tilde{N}$ values of a white noise process $(\varepsilon_t)_{t=-2\tilde{N}-1}^{\tilde{N}-1}$. Hence, every value \tilde{X}_t and so X_t carries information about at least N noises ε_t in the past and reflects the given LRC. Applying a pseudo-random number generator¹, we obtain the Gaussian white noise by the Mersenne Twister. Reducing the above procedure for the generation of ARFIMA(1, d , 0) time series to the ARFIMA(0, d , 0) or AR part, respectively, provides an algorithm for synthesizing suchlike processes as well (cp. Fig. 2.1).

Figure 2.2 visualizes a sample time series for each of the four aforementioned processes in Figure 2.1 with the same AR and LRC parameters $\varphi = 0.5$ and $d = 0.3$. In dependence of the correlation structure the different characteristics of these processes become visible. The iid process exhibits a purely random behavior without any apparent relation between the values of the process. By relation (2.12) the autocorrelations of the AR(1) process decay exponentially. Consecutive values of such processes show a tendency of being close to each other. Pretty soon, though, the white noise perturbations take over and yield a

¹from the software environment MATLAB R2018b

2. Gaussian and non-Gaussian stochastic processes

steady alternation between positive and negative values. The LRC in the ARFIMA(0, d , 0) time series imply noticeable dependencies among its values. Longer periods of values below or above the average zero value are observable. This behavior is even more prominent for the ARFIMA(1, d , 0) time series. Due to the positive AR parameter in this example, the process does not only stay positive or negative, respectively, for long periods but also bears a tendency of keeping its directions along the way.

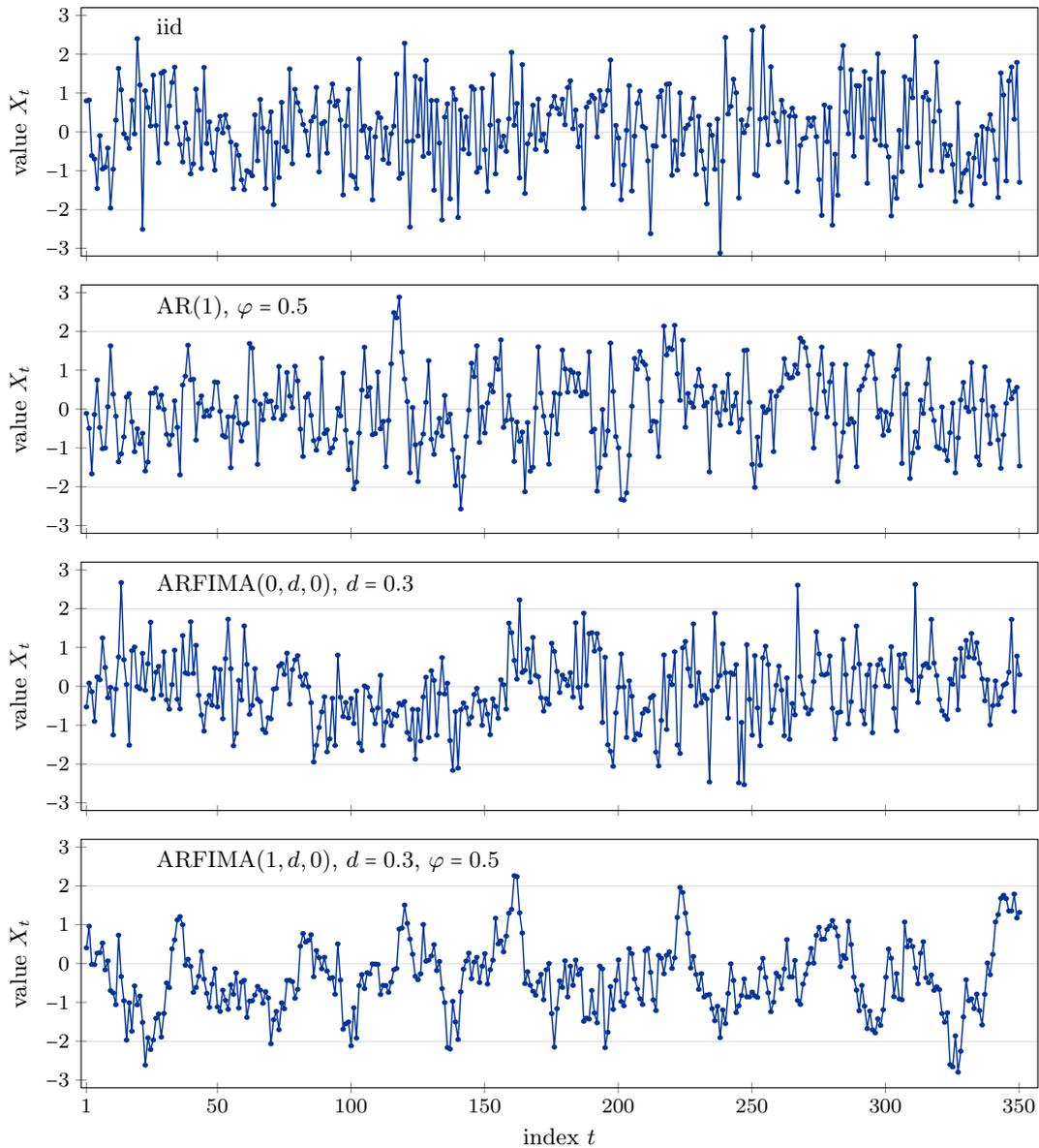


Fig. 2.2 Sample time series X_1, \dots, X_N of length $N = 350$ of an iid, an AR(1), an ARFIMA(0, d , 0) and an ARFIMA(1, d , 0) process (from top to bottom) with standard Gaussian marginal distribution each.

3. Meta-Gaussian stochastic processes

In the context of bivariate distributions Kelly and Krzysztofowicz [92] coined the phrase “meta-Gaussian” for distributions of monotonic transformations of (multivariately) normally distributed random variables. Meta-Gaussian models allow for discontinuous marginal distributions, which have been widely used in hydrometeorological modeling [63]. The truncated-Gaussian-power distribution that we use for modeling daily precipitation amounts in Chapter 6 is such a discontinuous distribution due to a point mass at the zero value. Section 1.4 and Section 6.1.3 provide details on precipitation modeling in general and our application of this specific distribution, respectively.

Definition (Meta-Gaussian process): Let $(X_t)_{t \in \mathbb{N}}$ be a Gaussian process. A stochastic process $(Y_t)_{t \in \mathbb{N}}$ is called a *meta-Gaussian process* if there exists a nondecreasing monotonic function $g: \mathbb{R} \rightarrow \mathbb{R}$ such that

$$Y_t = g(X_t) \tag{3.1}$$

pointwise for all indices $t \in \mathbb{N}$.

A variety of marginal distributions involving any kind of tail behaviour (cp. Sect. 2.3) results from nonlinear transformations of Gaussian variates. Any desired distribution can be generated by *inverse sampling* based on the *probability integral transform* as follows. For every random variable X with cdf F_X the random variable $F(X)$ follows a standard uniform distribution. Then the random variable $Y := F_Y^{-1}(F_X(X))$ obeys any prescribed cdf F_Y with quantile function F_Y^{-1} . Note that by the monotonicity of the two functions the concatenation $F_Y^{-1} \circ F_X$ of any quantile function F_Y^{-1} and any cdf F_X yields a nondecreasing transformation in equation (3.1). Models based on the inverse sampling idea are popular in hydrological and geophysical modeling (see Sect. 1.4) because of the convenience of Gaussian processes (see Sect. 2.1) and the straightforward access to models with non-Gaussian distributions by pointwise transformations.

Remark 3.1: The usage of the phrase “meta-Gaussian” and its definition is not consistent across scientific disciplines. In the geophysical literature we also find the term “parent-Gaussian” [141] without demanding monotonic increase in definition (3.1) explicitly. In physics, mathematics or computer science we also find phrases like “MNL” (memoryless nonlinear transformations) [178], special cases of “Gaussian subordinate” processes [1, 22, 56] or usage of the concept without a specific denotation [39]. See Section 1.1 for a clarification of the term “meta-Gaussian”.

3. Meta-Gaussian stochastic processes

Some authors refer to pointwise transformations of stochastic processes as *memoryless* [39, 73, 178] since processes defined by equation (3.1) do not gain additional dependencies among the different values of the process. In contrast, Section 3.4.3 states a transformation of a Gaussian process which is not considered memoryless. The correlations of a stochastic process nevertheless cannot be expected invariant under nonlinear pointwise transformation. Depending on the particular transformation the correlations of the transformed process are a function of the original correlations. For transformations of Gaussian processes, in particular, for meta-Gaussian processes the relation between these two acfs is analytically known as recapped in Section 3.1. Based on these insights Section 3.2 draws inference about the properties of the correlations of the transformed process. In Section 3.3 we study the resulting marginal distributions and acfs for some example transformations. The closing remarks in Section 3.4 broaden the scope of correlations of transformed Gaussian processes to more general settings.

3.1. Hermite polynomial approach to correlations

A nonlinear transformation (3.1) presumably changes the mean and the variance of the marginal distribution of a stationary stochastic process. Also the acf is thus expected to change even under a memoryless transformation of the process as both quantities affect correlations. Note that affine linear transformations yield affine linear changing of the mean and standard deviation of the process' marginal distribution, so that by its very definition the acf remains unchanged then. How the acf of an original Gaussian process changes precisely under a nonlinear transformation can be determined by an Hermite polynomial approach [22, 161]. This method is valid for square-integrable transformations, which extends its applicability beyond meta-Gaussian to processes generated by also non-monotonic transformations.

Let $(X_t)_{t \in \mathbb{N}}$ be a stationary and centered (i.e., zero-mean) Gaussian process with standard deviation $\sigma \in \mathbb{R}_{>0}$, acf ϱ_X and marginal pdf

$$f_X(x) := \frac{1}{\sigma} \phi\left(\frac{x}{\sigma}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \quad (3.2)$$

for values $x \in \mathbb{R}$. Unless specified otherwise, in what follows, the symbol f_X accompanied by the symbol F_X denote the pdf (3.2) and cdf, respectively, of a zero-mean Gaussian $N(0, \sigma^2)$ variate with arbitrary standard deviation (cp. Tab. B.1).

Denote by $\mathcal{L}^2 := \mathcal{L}^2(\mathbb{R}, \mu_{f_X})$ the Hilbert space of all with respect to the measure $\mu_{f_X}(A) := \int_A f_X(x) dx$ on Borel sets $A \subseteq \mathbb{R}$ square-integrable functions on the real space \mathbb{R} . The space \mathcal{L}^2 features the standard scalar product

$$\langle g, h \rangle := \int_{\mathbb{R}} g(x)h(x)f_X(x) dx \quad (3.3)$$

3.1. Hermite polynomial approach to correlations

for functions $g, h \in \mathcal{L}^2$. Consider a pointwise nonlinear \mathcal{L}^2 -transformation g (not necessarily nondecreasing) and the transformed process $(Y_t)_{t \in \mathbb{N}}$ with $Y_t := g(X_t)$. Such a transformation keeps the first $\mathbb{E}[Y_t] = \langle g, 1 \rangle$ and second $\mathbb{E}[Y_t^2] = \langle g, g \rangle$ moment of the marginal distribution finite, so that the resulting process Y_t is again stationary. Furthermore, every \mathcal{L}^2 -function exhibits an *Hermite polynomial* series expansion.

Definition (Hermite polynomials): For unit variance $\sigma^2 = 1$ and indices $j \in \mathbb{N}_0$ the *Hermite polynomials* $H_j : \mathbb{R} \rightarrow \mathbb{R}$ are defined as

$$H_j(x) := (-1)^j \frac{d^j}{dx^j} \left(e^{-\frac{x^2}{2}} \right) e^{\frac{x^2}{2}} \quad (3.4)$$

for values $x \in \mathbb{R}$. For arbitrary standard deviations $\sigma \in \mathbb{R}_{>0}$ by definition (3.4) the *generalized Hermite polynomials* are given by

$$H_j^{\sigma^2}(x) := \sigma^j H_j \left(\frac{x}{\sigma} \right). \quad (3.5)$$

The first four generalized Hermite polynomials read

$$H_0^{\sigma^2}(x) = 1, \quad H_1^{\sigma^2}(x) = x, \quad H_2^{\sigma^2}(x) = x^2 - \sigma^2 \quad \text{and} \quad H_3^{\sigma^2}(x) = x^3 - 3\sigma^2 x. \quad (3.6)$$

In the \mathcal{L}^2 -Hilbert space equipped with the scalar product (3.3) the (generalized) Hermite polynomials form an orthogonal basis [161, Prop. 6.3.2] since by equation (3.5), we find

$$\langle H_i^{\sigma^2}, H_j^{\sigma^2} \rangle = \sigma^{2j} j! \delta_{ij} \quad (3.7)$$

for indices $i, j \in \mathbb{N}_0$, where δ_{ij} denotes the Kronecker delta. Normalizing the scalar product (3.7) to unity, every function $g \in \mathcal{L}^2$ can be represented uniquely by

$$g = \sum_{j=0}^{\infty} \frac{\alpha_j}{\sigma^{2j} j!} H_j^{\sigma^2} \quad \text{with Hermite coefficients} \quad \alpha_j := \langle g, H_j^{\sigma^2} \rangle. \quad (3.8)$$

The series representation (3.8) of the function g implies how the acf ϱ_Y of the transformed process Y_t depends on the acf ϱ_X of the original Gaussian process X_t . Theorem 3.2 provides this result for centered Gaussian processes with arbitrary marginal standard deviations $\sigma \in \mathbb{R}_{>0}$ other than unity also, which is typically not elaborated in the literature but required explicitly for our fit procedure in Chapter 6. The necessary calculation using generalized Hermite polynomials and *Mehler's formula* [123] is given in Appendix A.

3. Meta-Gaussian stochastic processes

Theorem 3.2 (cp. Thm. 3.2 in [161]): *Let X_t be a stationary Gaussian process with nonnegative acf ϱ_X and marginal distribution $N(0, \sigma^2)$ with mean zero and standard deviation $\sigma \in \mathbb{R}_{>0}$. Then at time lags $k \in \mathbb{N}_0$ the acf ϱ_Y of the process $Y_t := g(X_t)$ obtained by a pointwise \mathcal{L}^2 -transformation g reads*

$$\varrho_Y(k) = \frac{\sum_{j=1}^{\infty} \frac{\alpha_j^2}{\sigma^{2j} j!} \varrho_X(k)^j}{\sum_{j=1}^{\infty} \frac{\alpha_j^2}{\sigma^{2j} j!}}. \quad (3.9)$$

Proof. Using Mehler's formula (see Appx. A) and equation (A.5), it can be shown [44] that for time lags $k \in \mathbb{N}_0$ we have

$$\mathbb{E}[Y_t Y_{t+k}] = \sum_{j=0}^{\infty} \frac{\alpha_j^2}{\sigma^{2j} j!} \varrho_X(k)^j. \quad (3.10)$$

Plugging in equation (3.10) into $\varrho_Y(k) = \frac{1}{\text{Var}(Y_t)} (\mathbb{E}[Y_t Y_{t+k}] - \mathbb{E}[Y_t]^2)$ implies equation (3.9). Mind that $\alpha_0 = \langle g, H_0^{\sigma^2} \rangle = \mathbb{E}[Y_t]$, so that the first summand in equation (3.10) cancels out, and note that the denominator in equation (3.9) is a series representation of the variance of the marginal distribution of the process Y_t . \square

Remark 3.3: The present thesis focusses on stochastic processes with nonnegative acfs and thus recites and proves Theorem 3.2 under this assumption only. The proof of the general theorem including negative correlations can be found in [137, Sect. 7.1]. Example 3.10 comments on effects on negative correlations. Our scope is the modeling of precipitation data with positive LRC in Chapter 6.

Without loss of generality in what follows we consider centered stationary Gaussian processes X_t with zero-mean marginal distribution $N(0, \sigma^2)$. By the orthogonality of the Hermite polynomials additive constants affect the first Hermite coefficient $\alpha_0 = \mathbb{E}[X_t]$ only. In consistency with the definition (2.3) of the acf this coefficient cancels out in the Hermite polynomial representation (3.9) just as affine linear transformations do not change correlations. Variances σ^2 other than unity, however, cannot be neglected for nonlinear transformations, as the acf ϱ_Y might depend nonlinearly on this factor, which does not cancel out generally (cp. the acfs ϱ_Z in eq. (3.12) of Ex. 3.4 and ϱ_{eX} in eq. (3.39)).

Example 3.4: Let X_t be a Gaussian process with marginal distribution $N(0, 1)$ and acf $\varrho_X > 0$. Define the polynomial transformations

$$g := H_0 + H_2 \quad \text{and} \quad h := H_2 + H_3 \quad (3.11)$$

as the sum of two Hermite polynomials (3.6) each and consider the processes

$$Y_t := g(X_t) \quad \text{and} \quad Z_t := h(X_t).$$

Note that the function g gives the pointwise square of the process X_t . The Hermite expansion (3.8) yields the Hermite coefficients $\alpha_0 = \mathbb{E}[X_t^2] = 1$ and $\alpha_2 = 2$ for the transformation g and $\alpha_2 = 2$, $\alpha_3 = 6$ for the transformation h . In either case, we have $\alpha_j = 0$ otherwise, including $\alpha_0 = \mathbb{E}[Z_t] = 0$ for the function h . A short calculation gives $\text{Var}(Y_t) = 2$ and $\text{Var}(Z_t) = 8$. Inserting these values into equation (3.9), we obtain $\varrho_Y = \varrho_X^2$ and $\varrho_Z = \frac{1}{4}\varrho_X^2 + \frac{3}{4}\varrho_X^3$. In the general setting of a process X_t with an $N(0, \sigma^2)$ marginal distribution with standard deviation $\sigma \in \mathbb{R}_{>0}$ applying generalized Hermite polynomials analogously yields

$$\varrho_Y = \varrho_X^2 \quad \text{and} \quad \varrho_Z = \frac{1}{1 + 3\sigma^2} (\varrho_X^2 + 3\sigma^2 \varrho_X^3). \quad (3.12)$$

The acf ϱ_Z depends nonlinearly on the variance σ^2 , though, the acf ϱ_Y does not. Figure 3.1 shows the polynomial relation between the original acf ϱ_X of the process X_t and the acfs ϱ_Y and ϱ_Z of the transformed processes Y_t and Z_t , respectively. As a remark, we find $\varrho_Z \rightarrow \varrho_X^3$ as $\sigma \rightarrow \infty$, so that the nonlinear dependence of the acf ϱ_Z on the variance σ^2 is crucial for small values only and diminishes as the variance increases.

3.2. Implications for correlation properties

Theorem 3.2 allows for direct access of the acf $\varrho_Y(k)$ of the transformed process Y_t at time lags $k \in \mathbb{N}_0$. The implications of equation (3.9) for the properties of the acf ϱ_Y are twofold. As a consequence of the instantaneous impact of a memoryless transformation, the acf of the transformed process is a function of the original acf *independent of the time lag* k . Equation (3.9) also bears the *asymptotics* of the acf of the transformed process. In Sections 6.1.2 and 6.2.2 we use this relation to adjust the LRC parameter of the ARFIMA process in our meta-Gaussian model for daily precipitation amounts.

3. Meta-Gaussian stochastic processes

Definition (Hermite rank): The smallest nonzero index $J \in \mathbb{N}_{>0}$, for which the Hermite coefficient $\alpha_J \neq 0$ in the Hermite polynomial expansion (3.8) is non-vanishing, is called the *Hermite rank* of the transformation g .

Considering positive correlations, the Hermite rank J of a function $g \in \mathcal{L}^2$ determines the asymptotic behavior of the acf ϱ_Y of the transformed process Y_t . Since the acf ϱ_X satisfies $\varrho_X(k) \rightarrow 0$ as the time lag $k \rightarrow \infty$, the degree of the (infinite) polynomial in equation (3.9) determines the asymptotic rate of decay of the acf ϱ_Y . The Hermite rank J quantifies the lowest order of this series expansion and yields the asymptotic behavior

$$\varrho_Y(k) \propto \varrho_X(k)^J \rightarrow 0 \quad (k \rightarrow \infty). \quad (3.13)$$

If the Hermite rank J is larger than unity, then the transformation increases the speed of the decay of the autocorrelations.

Remark 3.5: The coefficient for the index $j = 1$ in the Hermite expansion (3.8) reads

$$\alpha_1 = \int_{\mathbb{R}} g(x) x f_X(x) dx. \quad (3.14)$$

Every transformation g that is not even hence obeys the Hermite rank $J = 1$. Therefore, without further symmetry assumptions on the function g the transformation does not change the asymptotic behavior (2.10) of a Gaussian LRC process by relation (3.15).

The proportionality (3.13) is valid for any rate of decay of the original acf ϱ_X . The acf of the transformed process $Y_t = g(X_t)$ of a Gaussian SRC process with exponentially decaying acf $\varrho_X(k) \propto e^{k \ln |a|}$ ($k \rightarrow \infty$) with a basis $a \in \mathbb{R}$ satisfying $|a| < 1$ decays more rapidly at exponential rate $e^{Jk \ln |a|}$ as time lags $k \rightarrow \infty$ (cp. the acf (2.12) of an AR(1) process). If the process X_t has LRC in the sense of definition (2.10), then the acf ϱ_Y follows a power law with leading order

$$\varrho_Y(k) \propto k^{-\gamma^J} \quad (k \rightarrow \infty). \quad (3.15)$$

If the exponent γ of the underlying LRC process X_t satisfies $\gamma \in (0, \frac{1}{J}]$, then the transformed process Y_t obeys LRC as well. Otherwise, if $\gamma \in (\frac{1}{J}, 1]$, we obtain intermediate that means SRC. In the language of ARFIMA, processes with LRC parameter $d \in [\frac{1}{2} - \frac{1}{2J}, \frac{1}{2})$ maintain LRC but map to SRC (i.e., intermediate) for $d \in (0, \frac{1}{2} - \frac{1}{2J})$. The higher the Hermite rank of a transformation is, the larger is the parameter range of LRC processes that become SRC processes.

3.2. Implications for correlation properties

Beyond asymptotics, equation (3.9) allows for further insight into the specific shape of the relation between the two acfs ϱ_X and ϱ_Y .

Lemma 3.6: *Let $\varrho_X, \varrho_Y : [0, 1] \rightarrow [0, 1]$ be nonnegative acfs of a stationary Gaussian process $(X_t)_{t \in \mathbb{N}}$ and a transformed process $Y_t = g(X_t)$ with a function $g \in \mathcal{L}^2$. Then the relation $\mathcal{T} : \varrho_X \mapsto \varrho_Y$*

(i) *reduces the strength of correlations at fixed time lags $k \in \mathbb{N}$ by*

$$\varrho_Y(k) < \varrho_X(k), \quad (3.16)$$

(ii) *and thus reduces the correlation time by*

$$\tau_Y := \sum_{k=0}^{\infty} \varrho_Y(k) < \sum_{k=0}^{\infty} \varrho_X(k) =: \tau_X \quad (3.17)$$

(iii) *and is strictly monotonically increasing by*

$$\varrho_Y(k_1) < \varrho_Y(k_2) \quad \text{if} \quad \varrho_X(k_1) < \varrho_X(k_2) \quad (3.18)$$

for time lags $k_1, k_2 \in \mathbb{N}$, and

(iv) *is convex in the sense that*

$$\mathcal{T}\left(t\varrho_X(k_1) + (1-t)\varrho_X(k_2)\right) \leq t\varrho_Y(k_1) + (1-t)\varrho_Y(k_2) \quad (3.19)$$

for all values $t \in [0, 1]$ and time lags $k_1, k_2 \in \mathbb{N}$ with $\varrho_X(k_1) < \varrho_X(k_2)$.

Proof. (i) For positive correlations $\varrho_X(k) \in (0, 1)$ we find

$$\sum_{j=1}^{\infty} \frac{\alpha_j^2}{\sigma^{2j} j!} \varrho_X(k)^j < \varrho_X(k) \sum_{j=1}^{\infty} \frac{\alpha_j^2}{\sigma^{2j} j!} = \varrho_X(k) \text{Var}(Y_t), \quad (3.20)$$

so that equation (3.9) implies inequality (3.16) for all time lags $k \in \mathbb{N}$.

(ii) Inequality (3.17) follows from applying the estimate (3.16) to the definition (2.8) of the correlation time.

(iii) Analogously to estimate (3.20), we obtain the monotonicity (3.18).

(iv) Convexity (3.19) follows from equation (3.9) by the convexity of power functions as

$$\left(t\varrho_X(k_1) + (1-t)\varrho_X(k_2)\right)^j \leq t\varrho_X(k_1)^j + (1-t)\varrho_X(k_2)^j \quad (3.21)$$

for all exponents $j \in \mathbb{N}$. □

3. Meta-Gaussian stochastic processes

Note that for the two limit cases $\varrho_X(k) \rightarrow 1$ and $\varrho_X(k) \rightarrow 0$ as $k \rightarrow 0$ and $k \rightarrow \infty$, respectively, by equation (3.9) we have $\varrho_Y(k) \rightarrow 1$ and $\varrho_Y(k) \rightarrow 0$, as well. The term “memoryless” stems from the effect (3.16) since a pointwise transformation cannot add dependence to a process, in other words, increase correlations at fixed time lags. Section 3.4.3 gives a transformation of a Gaussian process that increases the strength of given correlations and is thus not memoryless.

Figure 3.1 visualizes the observations of Lemma 3.6 for the two polynomial transformations given in Example 3.4. For memoryless transformations the convex and monotonic image ϱ_Y of the acf ϱ_X lies below the diagonal by inequality (3.18). Section 3.4 relates equation (3.9) to a recent study [139] of a functional fit to the inverse relation $\varrho_Y \mapsto \varrho_X$ with visualizations for some example transformations in Figure 3.4.

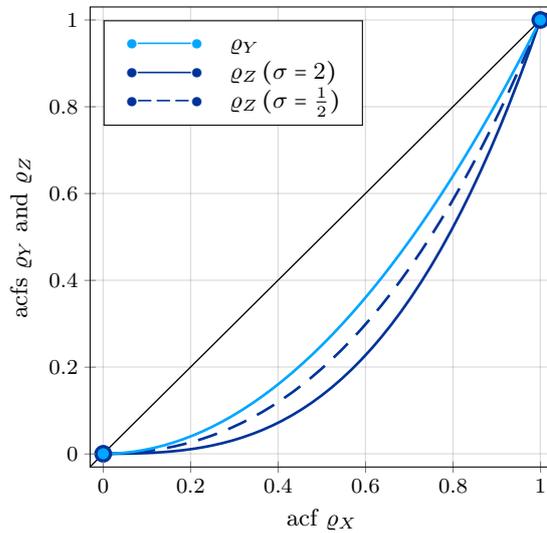


Fig. 3.1 Relation between the original acf ϱ_X of a Gaussian process X_t with marginal distribution $N(0, \sigma^2)$ and the acfs ϱ_Y and ϱ_Z (3.12) generated by the polynomial example transformations $g(x) = x^2$ and $h(x) = x^3 + x^2 - 3\sigma^2x - \sigma^2$ generalized from the equations (3.11) for arbitrary marginal standard deviations $\sigma \in \mathbb{R}_{>0}$.

See [10] for further reading on the sensitivity of the Hermite rank and on a central limit theorem representation of pointwise transformed LRC processes based on the Hermite polynomial approach. As a remark, one could also use Appell polynomials instead to generalize to transformations of *non-Gaussian* processes [85]. Due to definition (3.4), Hermite polynomials are a particularly convenient choice when transforming *Gaussian* processes. Palma [137] deals with the autocorrelations of the square of various non-Gaussian processes.

3.3. Example transformations

The pointwise square of stochastic processes is one of the transformations studied in detail earliest in the literature for linear processes with SRC or LRC filters [137, Sect. 7.2]. Further examples obtained by inverse sampling are given in [39] amongst others such as meta-Gaussian processes with uniform, exponential or Pareto marginal distribution. Having in mind the modeling of precipitation amounts, we consider transformations that yield distributions of nonnegative values only, namely the square, lognormal and absolute-value process. Section 3.3.1 deals with the marginal distributions we obtain by these nonlinear memoryless transformations of Gaussian processes. The related acfs are the subject of Section 3.3.2 along with their asymptotic behavior. For each of the three example maps we have a closed form of the acf of the transformed process. In general, however, by equation (3.9) the Hermite approach of Section 3.1 only yields a series representation of the acf typically with Hermite coefficients calculated numerically.

Beyond the Hermite polynomial approach Section 3.4.3 visualizes a non-instantaneous transformation of a Gaussian process providing a non-Gaussian LRC model with approximately exponential marginal distribution.

3.3.1. Generated marginal distributions

Before focusing on transformations of zero-mean Gaussian processes, we gain an impression of the variety of non-Gaussian models transformations of arbitrary Gaussian processes allow for. Let X be an $N(\mu, \sigma^2)$ -distributed Gaussian variate with mean $\mu \in \mathbb{R}$ and standard deviation $\sigma \in \mathbb{R}_{>0}$. Denote by F_X and f_X the cdf and pdf of the random variable X , respectively (cp. Tab. B.1). We consider the transformed variates $Y = X^2$, $Y = e^X$ and $Y = |X|$. The standard change-of-variables technique together with the perception $F_Y(y) = P(Y \leq y)$ of the cdfs F_Y of the random variables Y imply

$$F_{X^2}(y) = F_X(\sqrt{y}) - F_X(-\sqrt{y}), \quad f_{X^2}(y) = \frac{1}{2\sqrt{y}}(f_X(\sqrt{y}) + f_X(-\sqrt{y})), \quad (3.22)$$

$$F_{e^X}(y) = F_X(\ln y), \quad f_{e^X}(y) = \frac{1}{y}f_X(\ln y) \quad \text{and} \quad (3.23)$$

$$F_{|X|}(y) = F_X(y) - F_X(-y), \quad f_{|X|}(y) = f_X(y) + f_X(-y) \quad (3.24)$$

for all values $y \in \mathbb{R}_{>0}$ ($y \in \mathbb{R}_{\geq 0}$ for eqs. (3.24)). The positive real line is the support of each of the generated distributions (including additionally zero for the absolute-value transformation). Figure 3.2 shows the densities f_Y derived in equations (3.22), (3.23) and (3.24) for prescribed example parameters of the original Gaussian variable X .

3. Meta-Gaussian stochastic processes

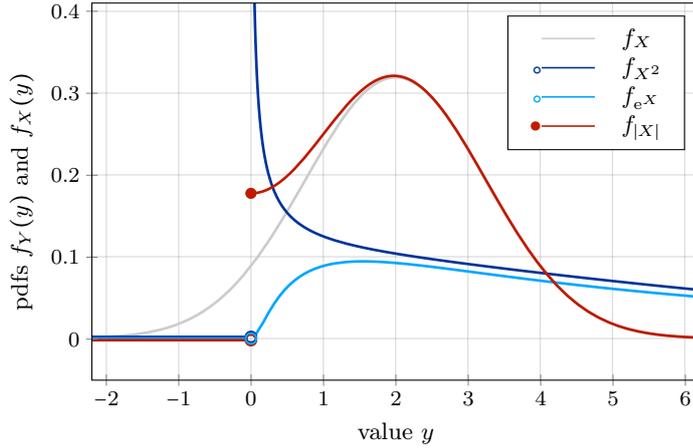


Fig. 3.2 Densities of the three transformed variates $Y = X^2$, $Y = e^X$ and $Y = |X|$ of a Gaussian variate $X \sim N(2, 1.25)$. Note that by equation (3.24) we find $f_{|X|}(0) = 2f_X(0)$.

Formulae, in particular, for the first two moments of a Gaussian and a lognormal distribution are well known as equation (3.25) and (3.26), respectively. Note that the square of a standard Gaussian variate follows a $\chi^2(1)$ -distribution (see Appx. B). Involving equation (3.24), a change of variables implies the mean $\mathbb{E}[|X|] = \int_0^\infty y f_{|X|}(y) dy$ for the absolute-value transformation. Equation (3.27) states the result of this integral obtained by straight forward but lengthy calculations².

$$\mathbb{E}[X^2] = \mu^2 + \sigma^2 \quad \text{Var}(X^2) = 4\mu^2\sigma^2 + 2\sigma^4 \quad (3.25)$$

$$\mathbb{E}[e^X] = e^{\mu + \frac{\sigma^2}{2}} \quad \text{Var}(e^X) = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1) \quad (3.26)$$

$$\mathbb{E}[|X|] = \frac{2\sigma}{\sqrt{2\pi}} e^{-\frac{\mu^2}{2\sigma^2}} + \mu \left(1 - 2\Phi\left(-\frac{\mu}{\sigma}\right) \right) \quad \text{Var}(|X|) = \mathbb{E}[X^2] - \mathbb{E}[|X|]^2 \quad (3.27)$$

In virtue of Section 2.3 we analyze the tails of the three example distributions and, for the sake of clarity, consider transformations of *centered* Gaussian variates for that purpose. Let $X \sim N(0, \sigma^2)$ be a zero-mean Gaussian random variable with standard deviation $\sigma \in \mathbb{R}_{>0}$, pdf f_X defined by equation (3.2) and corresponding cdf F_X . By the equations (3.22), (3.23) and (3.24) and due to the equality $F_X(-x) = 1 - F_X(x)$, the cdfs and as their derivatives the pdfs of the transformed variates read

$$F_{X^2}(y) = 2F_X(\sqrt{y}) - 1 \quad f_{X^2}(y) = \frac{1}{\sqrt{y}} f_X(\sqrt{y}) \quad (3.28)$$

$$F_{e^X}(y) = F_X(\ln y) \quad f_{e^X}(y) = \frac{1}{y} f_X(\ln(y)) \quad (3.29)$$

$$F_{|X|}(y) = 2F_X(y) - 1 \quad f_{|X|}(y) = 2f_X(y) \quad (3.30)$$

²verified by software WOLFRAM MATHEMATICA 11.3.0.0 for rearrangement of symbolic formulae

for values $y \in \mathbb{R}_{>0}$ ($y \in \mathbb{R}_{\geq 0}$ for the absolute-value transformation). The asymptotics of the densities (3.28), (3.29) and (3.30) reveal the categories these distributions belong to.

$$f_{X^2}(y) \propto y^{-\frac{1}{2}} e^{-\frac{y}{2\sigma^2}} = \mathcal{O}\left(e^{-\frac{y}{2\sigma^2}}\right) \quad (y \rightarrow \infty) \quad (3.31)$$

$$f_{e^X}(y) \propto y^{-1} e^{-\frac{(\ln y)^2}{2\sigma^2}} = y^{-1-\frac{1}{2\sigma^2} \ln y} \quad (y \rightarrow \infty) \quad (3.32)$$

$$f_{|X|}(y) \propto e^{-\frac{y^2}{2\sigma^2}} = \mathcal{O}\left(e^{-\frac{y^2}{2\sigma^2}}\right) \quad (y \rightarrow \infty) \quad (3.33)$$

In the sense of Section 2.3 the pdfs (3.31) and (3.33) of the variates generated by the square and absolute-value transformation of a Gaussian are light tailed, as they decay faster than an exponential function. The pdf (3.32) of a lognormal distribution satisfies $e^{-ty} = \mathcal{O}(f_{e^X}(y))$ as $y \rightarrow \infty$ for all constants $t \in \mathbb{R}_{>0}$ and is thus heavy tailed by definition (2.5). The tail of the density of a lognormal distribution decays slower than any exponential function. In Section 6.1.3 we analogously analyze the properties of the density and asymptotics of the truncated-Gaussian-power distribution.

3.3.2. Generated autocorrelation functions

Let $(X_t)_{t \in \mathbb{N}}$ be a zero-mean Gaussian process with marginal distribution $N(0, \sigma^2)$ with standard deviation $\sigma \in \mathbb{R}_{>0}$ and acf ϱ_X . To study the influence of the three transformations on the acf ϱ_X we apply the Hermite polynomial approach (Sect. 3.1). All equations below involving acfs are to be understood pointwise at time lags $k \in \mathbb{N}_0$ based on the instantaneous effect of the transformation (cp. Sect. 3.2).

Example 3.4 deals with the acf of the square of a centered Gaussian process with arbitrary variance. In case of the square and absolute-value transformation, we shall assume unit variance $\sigma^2 = 1$ because their acfs satisfy $\varrho_{|aX|} = \varrho_{|X|}$ and $\varrho_{(aX)^2} = \varrho_{X^2}$ for all factors $a \in \mathbb{R}$. For the exponential transformation we need to account for variances $\sigma^2 \neq 1$ explicitly in the calculation of the acf ϱ_{e^X} as multiples of the variates X_t do not cancel out.

By equation (3.12) we have $\varrho_{X^2} = \varrho_X^2$ for the squared process. For the exponential transformation we assume the standard deviation $\sigma \in \mathbb{R}_{>0}$ arbitrary and thus involve the generalized Hermite polynomials (3.5). By induction and basic integral arguments the scalar product (3.3) yields the Hermite coefficients $\alpha_j = \sigma^{2j} e^{\frac{\sigma^2}{2}}$ for all indices $j \in \mathbb{N}_{\geq 0}$. Note that $\alpha_j = \sqrt{e}$ is constant for all indices j in case of $\sigma^2 = 1$, which is a popular example [39, 44] of the Hermite polynomial approach. Involving the mean and variance (3.26), we obtain the result (3.39) by the Hermite series representation (3.9) of the acf ϱ_{e^X} and the series expansion of the exponential function.

3. Meta-Gaussian stochastic processes

For the absolute-value process we find $\alpha_0 = \mathbb{E}[|X_t|] = \frac{2}{\sqrt{2\pi}}$ by equation (3.27) and $\alpha_j = 0$ for all uneven indices $j \in \mathbb{N}$ because of the symmetry of the transformation. Using partial integration, we have the auxiliary calculation

$$\begin{aligned} \int_a^b x \frac{d^j}{dx^j} \left(e^{-\frac{x^2}{2}} \right) dx &= \left[x \frac{d^{j-1}}{dx^{j-1}} \left(e^{-\frac{x^2}{2}} \right) - \frac{d^{j-2}}{dx^{j-2}} \left(e^{-\frac{x^2}{2}} \right) \right]_a^b \\ &= (-1)^{j-1} \left[e^{-\frac{x^2}{2}} (xH_{j-1}(x) + H_{j-2}(x)) \right] \end{aligned} \quad (3.34)$$

for integral limits $a, b \in \mathbb{R} \cup \{-\infty, \infty\}$. For even indices $j \in \mathbb{N}_{\geq 2}$ definition (3.4) of the Hermite polynomials and the scalar product (3.7) imply

$$\begin{aligned} \langle |\cdot|, H_j \rangle &= \frac{(-1)^j}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |x| \frac{d^j}{dx^j} \left(e^{-\frac{x^2}{2}} \right) dx \\ &= \frac{(-1)^j}{\sqrt{2\pi}} \left(- \int_{-\infty}^0 x \frac{d^j}{dx^j} \left(e^{-\frac{x^2}{2}} \right) + \int_0^{\infty} x \frac{d^j}{dx^j} \left(e^{-\frac{x^2}{2}} \right) \right) \\ &= \frac{(-1)^j}{\sqrt{2\pi}} \left((-1)^j \left[e^{-\frac{x^2}{2}} (xH_{j-1}(x) + H_{j-2}(x)) \right]_{-\infty}^0 \right. \\ &\quad \left. + (-1)^{j-1} \left[e^{-\frac{x^2}{2}} (xH_{j-1}(x) + H_{j-2}(x)) \right]_0^{\infty} \right) \quad (\text{by (3.34)}) \\ &= \frac{(-1)^j}{\sqrt{2\pi}} \left((-1)^j (H_{j-2}(0) - 0) + (-1)^{j-1} (0 - H_{j-2}(0)) \right) \\ &= \frac{(-1)^j}{\sqrt{2\pi}} \cdot 2(-1)^j H_{j-2}(0) = \frac{2}{\sqrt{2\pi}} H_{j-2}(0). \end{aligned} \quad (3.35)$$

The values of the Hermite polynomials at the point zero satisfy $H_j(0) = (-1)^{\frac{j}{2}}(j-1)!!$ for even indices $j \in \mathbb{N}_{\geq 2}$ [2, equs. (22.4.8),(22.5.18)] and are known as the *Hermite numbers*. The doubled exclamation mark denotes the *double factorial*, which is defined by $(2m-1)!! := \frac{(2m)!}{2^m m!}$ for natural numbers $m \in \mathbb{N}_{\geq 1}$ and $(-1)!! := 1$. By the result (3.35) the Hermite coefficients (3.8) of the absolute-value transformation read

$$\alpha_0 = \frac{2}{\sqrt{2\pi}}, \quad \alpha_j = 0, \quad j \in \mathbb{N}_{>0} \text{ uneven}, \quad \alpha_j = \frac{2}{\sqrt{2\pi}} (-1)^{\frac{j-2}{2}} (j-3)!!, \quad j \in \mathbb{N}_{>0} \text{ even}. \quad (3.36)$$

Note that $\alpha_0 = \mathbb{E}[|X_t|]$ meets equation (3.27) just as $\text{Var}(|X_t|) = 1 - \frac{2}{\pi} = \frac{\pi-2}{\pi}$ for mean $\mu = 0$ and standard deviation $\sigma = 1$. Plugging in these two moments and the Hermite coefficients (3.36) into the Hermite series representation (3.9) of the acf $\varrho_{|X|}$ gives

$$\varrho_{|X|} = \frac{2}{\pi-2} \sum_{j=2, j \text{ even}}^{\infty} \frac{((j-3)!!)^2}{j!} \varrho_X^j. \quad (3.37)$$

Symbolic numerics³ give the closed-form result (3.40).

Remark 3.7: For AR(1) processes one can derive the result (3.40) for the acf $\varrho_{|X|}$ directly by plugging in the notion (2.11) of the original process into the definition (2.3) of the acf of the transformed process and considering the bivariate distribution of $(|X_t|, |X_{t+k}|)$ based on the bivariate Gaussian distribution of (X_t, X_{t+k}) for time lags $k \in \mathbb{N}$. Such an elementary but extensive calculation also yields equation (3.40) but is not easily transferable to, for example, ARFIMA processes. The Hermite approach, however, is more general and applicable to any stationary Gaussian process.

We conclude that both transformations the square and the absolute value obey Hermite rank $J = 2$. The exponential function, however, has Hermite rank $J = 1$. The closed-form solutions and asymptotics of the acfs of the transformed processes read

$$\varrho_{X^2} = \varrho_X^2 \tag{3.38}$$

$$\varrho_{e^X} = \frac{e^{\sigma^2 \varrho_X} - 1}{e^{\sigma^2} - 1} \sim \frac{\sigma^2}{e^{\sigma^2} - 1} \varrho_X \tag{3.39}$$

$$\varrho_{|X|} = \frac{2}{\pi - 2} \left(\sqrt{1 - \varrho_X^2} + \varrho_X \arcsin(\varrho_X) - 1 \right) \sim \frac{1}{\pi - 2} \varrho_X^2 \tag{3.40}$$

as $\varrho_X \rightarrow 0$. By equation (3.39) and (3.40) the leading order of the asymptotics of the acfs (i.e., the Hermite rank J) becomes equivalently apparent by a series expansion of the exact formulae in the neighborhood of the argument $\varrho_X = 0$.

In agreement with the power-law relation (3.15) the asymptotic decay of the acf of a Gaussian LRC process in the sense of definition (2.8) with an acf $\varrho_X(k) \propto k^{-\gamma}$ as time lags $k \rightarrow \infty$ remains unchanged under the exponential function (erroneously disproven in [44], cp. Rem. 3.8). If the process is squared, though, or if its absolute value is taken, then LRC processes with exponents $\frac{1}{2} < \gamma < 1$ become SRC processes, due to intermediate correlations, while for exponents $0 < \gamma < \frac{1}{2}$ the LRC asymptotics are preserved with a change in the exponent of the decrease (3.15) of the acf to $\varrho_X(k) \propto k^{-\gamma J}$ as $k \rightarrow \infty$.

As a remark, if the process X_t is a long-range anticorrelated ARFIMA(0, d , 0) process with LRC parameter $d \in (-1, 0)$, then the pointwise-squared process X_t^2 is an SRC process [163].

Remark 3.8: Other than asserted in [44] the asymptotic behavior of a given hyperbolically decaying acf ϱ_X is preserved instead of accelerated (cp. relation (3.39)) by the *exponential transformation*. This example does not serve as a counterexample for the invalidity of Theorem 3.2 with an infinite Hermite series (3.9) as erroneously argued therein. The Hermite polynomial approach is valid also for transformations with infinite Hermite series expansion [161, Thm. 6.3.4].

³by the software WOLFRAM MATHEMATICA 11.3.0.0 for rearrangement of symbolic formulae

3.4. Closing remarks

The modeling of non-Gaussian stochastic processes by transformations of Gaussian processes is an active field of current research. As an alternative to the Hermite polynomial approach, a recent study [139] deals with an *inverse approach* to fitting the acf of the underlying Gaussian processes appropriately. In Section 3.4.1 we apply this method to the processes in Example 3.4 and Sections 3.3 and draw a relation between the two approaches.

Beyond the setting of Chapter 3 more general conditions than *positive* correlations and *memoryless* transformations allow for stochastic models with broader effects and properties. Section 3.4.2 gives an impression of the acf of transformations of Gaussian processes with *negative* correlations. Example 3.10 therein is a side product of a study on the synthesis of multivariately non-Gaussian random vectors by transforming a multivariate Gaussian [73]. In Section 3.4.3 we formulate a model for a correlated process with an approximately exponentially distributed marginal distribution by a *non-memoryless* transformation.

3.4.1. Inverse relation between generated and original correlations

The Hermite polynomial approach of Section 3.1 describes how the acf ϱ_X of a stationary Gaussian process X_t maps to the acf ϱ_Y of pointwise transformations $Y_t = g(X_t)$ of the original process. In a recent study Papalexiou provides a functional form of the inverse relation between the generated and the original acf [139]. Empirically observed autocorrelations of a non-Gaussian process thus directly determine the acf of the Gaussian process in a meta-Gaussian model for geophysical data.

The approximation of the acf ϱ_X given the acf ϱ_Y bases on the properties of the acf of the transformed process listed in Lemma 3.6 as follows. By the convexity (3.19) of the relation $\varrho_X \mapsto \varrho_Y$ the inverse relation is concave. The two monotonicities (3.16) and (3.18) give rise to concave and strictly monotonically increasing functions $\mathcal{T}_{1,2} : [0, 1] \rightarrow [0, 1] : \varrho_Y \mapsto \mathcal{T}_{1,2}(\varrho_Y) := \varrho_X$ that satisfy the two limits $\mathcal{T}_{1,2}(\varrho_Y) \rightarrow 0$ as $\varrho_Y \rightarrow 0$ and $\mathcal{T}_{1,2}(\varrho_Y) \rightarrow 1$ as $\varrho_Y \rightarrow 1$. An appropriate ansatz for these requirements are the functions

$$\mathcal{T}_1(\varrho_Y) = \frac{(1 + b_1 \varrho_Y)^{c_1} - 1}{(1 + b_1)^{c_1} - 1} \quad (3.41)$$

$$\mathcal{T}_2(\varrho_Y) = 1 - (1 - \varrho_Y^{b_2})^{c_2} \quad (3.42)$$

with parameters $b_1 \in (-1, 0) \subset \mathbb{R}$ and $c_1 \in \mathbb{R}_{\geq 1}$ or $b_1 \in \mathbb{R}_{>0}$ and $c_1 \in \mathbb{R}_{\leq 1}, c_1 \neq 0$, for the function (3.41) and $b_2 \in (0, 1] \in \mathbb{R}$ and $c_2 \in \mathbb{R}_{>1}$ for the function (3.42). The limits $c_1 = 1$ and $b_2 = c_2 = 1$ yield the identity $\varrho_X = \varrho_Y$. Figure 3.3 and Remark 3.9 comment on the parameter choice for these functions.

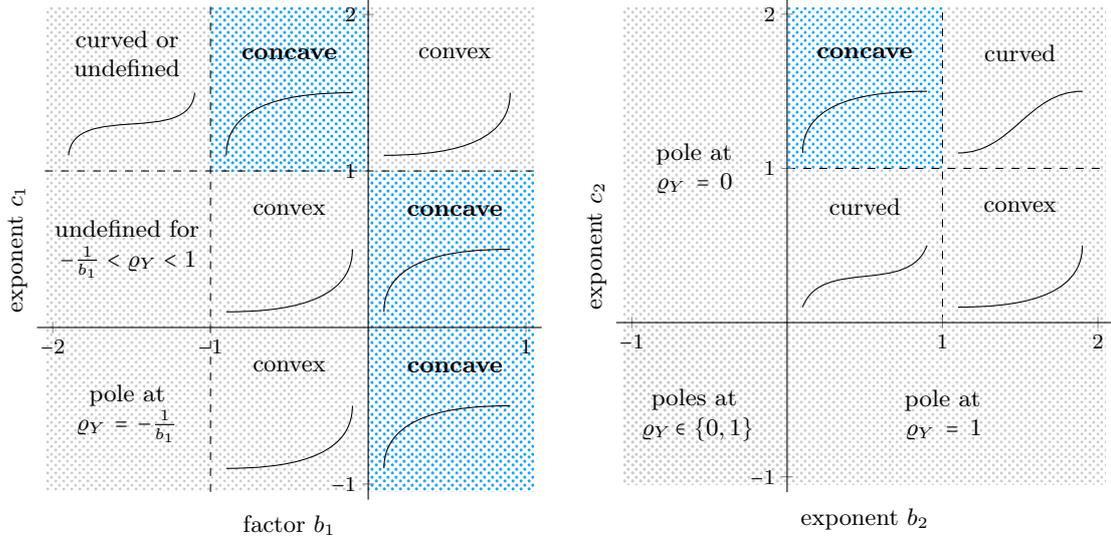


Fig. 3.3 Shape of the functions \mathcal{T}_1 (3.41) and \mathcal{T}_2 (3.42) on their parameter spaces.

Remark 3.9: The definitions (3.41) and (3.42) along with their first and second derivatives provide insights into the properties of these two functions in dependence of the parameter choice. A positive first derivative together with a negative second derivative imply strict increase and concavity.

For all parameters $c_1 \in \mathbb{R}_{\neq 0}$ the function \mathcal{T}_1 (3.41) is undefined if $b_1 = 0$ due to denominator zero. All values $c_1 \in \mathbb{R} \setminus \mathbb{N}_0$ require $b_1 \geq -1$ as $(1 + b_1 \varrho_Y)^{c_1}$ with $\varrho_Y \in [0, 1]$ has a pole at $-\frac{1}{b_1}$ or is undefined for $\varrho_Y \in (-\frac{1}{b_1}, 1]$ otherwise. As a remark, if $c_1 \in \mathbb{N}_{>0}$ and even, then the function \mathcal{T}_1 is defined for $b_1 \in \mathbb{R} \setminus \{0, -2\}$. In this case, for all values $b_1 \in \mathbb{R}_{<-1}$, however, the function \mathcal{T}_1 does not map onto the unit interval, since $\mathcal{T}_1(x) < 0$ for all $\varrho_Y \in (0, -\frac{2}{b_1})$ if $b_1 \in (-\infty, -2)$ and $\mathcal{T}_1(\varrho_Y) > 1$ for all $\varrho_Y \in (-(1 + \frac{2}{b_1}), 1)$ if $b_1 \in (-2, -1)$. For uneven values $c_1 \in \mathbb{N}$ the function \mathcal{T}_1 is strictly monotonically increasing from zero to unity but changes from concavity to convexity at $\varrho_Y = -\frac{1}{b_1}$. If $c_1 \in \mathbb{R}_{\neq 0}$ and $b_1 \in \mathbb{R}_{>-1}$, then the function \mathcal{T}_1 also increases strictly monotonically. Figure 3.3 (left panel) depicts the parameter areas with concavity.

The function \mathcal{T}_2 has the pole zero if $b_2 \in \mathbb{R}_{<0}$, unity if $c_2 \in \mathbb{R}_{<0}$ and both if both parameters are negative. It is strictly increasing for pairs of positive parameters. By the second derivative the curvature changes at $\varrho_Y = (\frac{b_2 - 1}{b_2 c_2 - 1})^{\frac{1}{b_2}}$ from concavity to convexity if $b_2, c_2 \in (0, 1)$ and conversely if $b_2, c_2 \in (1, \infty)$. Figure 3.3 (right panel) visualizes the parameter area valid for concavity.

3. Meta-Gaussian stochastic processes

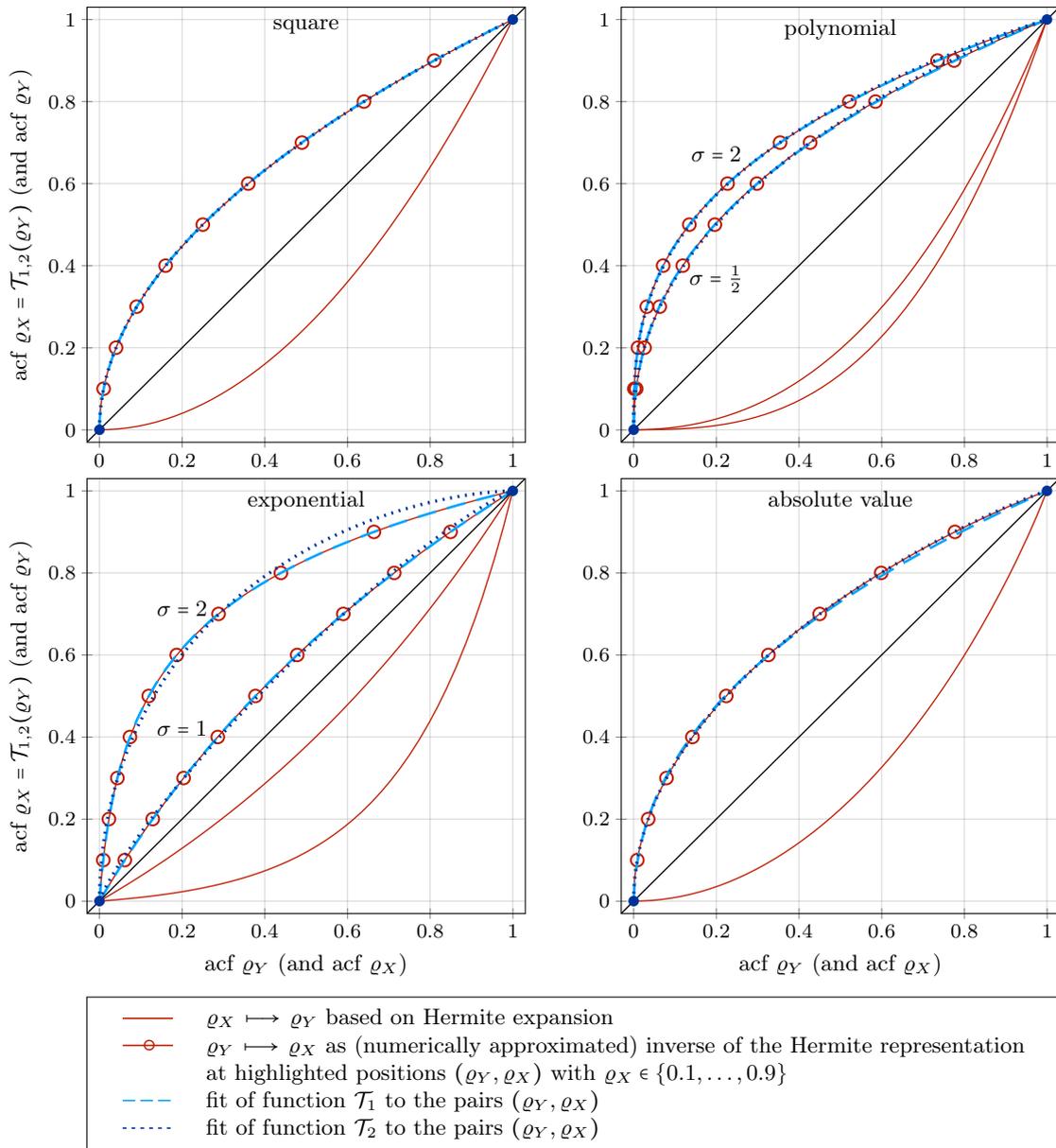


Fig. 3.4 Comparison of approximations of acfs. Given the acf ϱ_X of a Gaussian process X_t with marginal distribution $N(0, \sigma^2)$ and standard deviation $\sigma \in \mathbb{R}_{>0}$ the Hermite polynomial approach yields the acf ϱ_Y of the transformed processes Y_t . From top left to bottom right the four transformations are the examples (3.31), (3.12), (3.32) and (3.33). Given in turn the acf ϱ_Y of the transformed process Y_t a fit of the functions \mathcal{T}_1 or \mathcal{T}_2 provides an approximation of the required acf ϱ_X of the original Gaussian process X_t .

The results in Figure 3.4 depict an application of the inverse idea to the transformed processes in Example 3.4 and Section 3.3. Choosing analytically known pairs (ϱ_Y, ϱ_X) for a least-squares fit implies good agreement between the inverse of the Hermite representation of the acf ϱ_Y and the fitted functionals (3.41) and (3.42) with the parameter values shown in Table 3.1.

The fitted parameter values $\bar{b}_1, \bar{c}_1, \bar{b}_2$ and \bar{c}_2 in Table 3.1 relate to the Hermite rank J of the transformation g with $Y_t = g(X_t)$. As the parameter $b_1 \rightarrow \infty$ we find $\mathcal{T}_1(\varrho_Y) \rightarrow \varrho_Y^{c_1}$, so that $c_1 = \frac{1}{J}$ if $\varrho_Y = \varrho_X^J$ such as for the squared process with $J = 2$, $\bar{c}_1 \approx \frac{1}{2}$ and $\bar{b}_1 \gg 1$ depending on the chosen accuracy of the fit. Analogously, we obtain $\mathcal{T}_2(\varrho_Y) = \varrho_Y^{b_2}$ for $c_2 = 1$, so that $b_1 = \frac{1}{J}$ with $\bar{b}_1 \approx \frac{1}{2}$ and $\bar{c}_2 \approx 1$ for the square accordingly.

		\mathcal{T}_1		\mathcal{T}_2	
		\bar{b}_1	\bar{c}_1	\bar{b}_2	\bar{c}_2
square		163,841.0	0.498	0.499	0.999
polynomial	$\sigma = \frac{1}{2}$	3,619.1	0.404	0.453	1.056
	$\sigma = 2$	114,480.6	0.336	0.357	1.029
exponential	$\sigma = 1$	1.718	0.000	0.843	1.173
	$\sigma = 2$	53.598	0.000	0.519	1.616
absolute value		1,304.6	0.569	0.502	1.086

Table 3.1 The parameter values $\bar{b}_1, \bar{c}_1, \bar{b}_2$ and \bar{c}_2 result from least-squares fitting the two functions \mathcal{T}_1 and \mathcal{T}_2 to the analytically known pairs (ϱ_Y, ϱ_X) described in Figure 3.4.

In case of more complex relations between the acfs ϱ_Y and ϱ_X , the dominant summands in the Hermite expansion (3.9) determine the parameters c_1 and b_2 , respectively. If $J > 1$ then these parameters represent an *effective Hermite rank* \hat{J} in the neighborhood of the Hermite rank J as the leading order. Due to $\mathcal{T}_1(\varrho_Y) \propto \varrho_Y^{c_1}$ for $b_1 \gg 1$, we observe $\hat{J} := \frac{1}{\bar{c}_1}$ close to the Hermite rank J for the absolute-value example. Since the Hermite coefficient $\alpha_2 = 0$ vanishes for this transformation, the Hermite rank $J = 2$ dominates the effective Hermite rank $\hat{J} = 1.8$. The effective Hermite rank of the polynomial $H_2 + H_3$ depends on both the two leading Hermite coefficients $\alpha_2 \neq 0$ and $\alpha_3 \neq 0$ and the variance σ^2 of the underlying Gaussian process resulting in $\hat{J} = 2.5$ and $\hat{J} = 3$, respectively, compared to its Hermite rank $J = 2$.

For the function \mathcal{T}_2 the derivative $\mathcal{T}_2'(\varrho_Y) = b_2 c_2 (1 - \varrho_Y^{b_2})^{c_2 - 1} \varrho_Y^{b_2 - 1}$ induces $\hat{J} := \frac{1}{b_2}$ by matching the leading order $J - 1$ of the derivative of the Hermite expansion (3.9) in the neighborhood of $\varrho_Y = 0$. Again the effective Hermite rank $\hat{J} = 2$ fits the Hermite rank of the square and absolute-value transformation, and measures $\hat{J} = 2.2$ and $\hat{J} = 2.8$ for the polynomial with different variances. In case of Hermite rank $J = 1$ such as for the lognormal transformation the interpretation of the fitted parameters as effective Hermite ranks is no longer valid.

3. Meta-Gaussian stochastic processes

Figure 3.4 emphasizes that in general the acf ϱ_Y depends on the variance σ^2 of the original Gaussian process (see the polynomial and lognormal example). Difficulties arise when deciding about which of the two functional forms (3.41) and (3.42) is an appropriate choice for given autocorrelations of non-Gaussian data [139]. The deviations of the fitted function \mathcal{T}_2 for correlations larger than $\frac{1}{2}$ in the lognormal example depict this issue. Moreover, the quality of estimated acfs based on empirical data is not as high as for the analytically known pairs (ϱ_Y, ϱ_X) used for the fits in Figure 3.4.

3.4.2. Generating the same marginal distribution but different correlations

Applying different nonlinear transformations to Gaussian stationary processes might generate the same marginal distribution, though, differing acfs. Example 3.10 illustrates such a phenomenon for the χ^2 -distribution.

Example 3.10 (Ex. 2.5.1 in [73]): Consider a stationary Gaussian process X_t with marginal distribution $N(0, 1)$. Let Q be the quantile function of the $\chi^2(1)$ -distribution with one degree of freedom (see Appx. B). By its very definition a marginal $\chi^2(1)$ -distribution is generated straightforwardly by the transformation X_t^2 . Applying inverse sampling, the process $g(X_t)$ with $g := Q \circ \Phi$ exhibits the same marginal distribution.

While the marginal distributions of these two transformed processes coincide, their acfs do not. The transformation g is monotonically increasing contrary to the square, which forms a symmetric parabola. The Hermite coefficients of the two transformations hence differ. The square obeys Hermite rank $J = 2$, whereas the function $Q \circ \Phi$ features Hermite rank $J = 1$. The long-term behavior of the acfs of the two transformed processes hence differ despite exhibiting the same marginal distribution.

The different acfs become particularly apparent if the original process obeys negative autocorrelations. Then the process $g(X_t)$ has negative autocorrelations as well, while the process X_t^2 does not.

3.4.3. Alternative exponential model with long-range correlations

Exponentially distributed and correlated random numbers can for example be generated by inverse sampling applied to Gaussian processes [139] or spectral techniques using the Fourier transform [90, 169]. An alternative approach bases on the observation that the sum of the squares of two standard Gaussian iid variates follows a $\chi^2(2)$ -distribution with two degrees of freedom, which is the exponential distribution $\text{Exp}(\frac{1}{2})$. The pairwise sum of the square of a Gaussian process $(X_t)_{t \in \mathbb{N}}$ yields a correlated process with an exponential marginal distribution in leading order as follows.

For studying the marginal distribution let $(X, Y) \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma\right)$ be a bivariate Gaussian vector with (two-dimensional) zero mean, marginal standard deviation $\sigma \in \mathbb{R}_{>0}$ each and autocovariance matrix $\Sigma := \sigma \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ with correlation $\rho = \text{Corr}(X, Y) \in (0, 1)$. Using the equations (3.25) and (3.38), the mean and variance of the sum $X^2 + Y^2$ read

$$\begin{aligned} \mathbb{E}[X^2 + Y^2] &= \mathbb{E}[X^2] + \mathbb{E}[Y^2] = 2\sigma^2 \\ \text{Var}(X^2 + Y^2) &= \text{Var}(X^2) + \text{Var}(Y^2) + 2 \text{Cov}(X^2, Y^2) \\ &= 2\sigma^4 + 2\sigma^4 + 2 \cdot 2\sigma^4 \cdot \rho^2 = 4\sigma^4(1 + \rho^2). \end{aligned} \quad (3.43)$$

We obtain the pdf of the variate $X^2 + Y^2$ as the derivative of its cdf by considering the polar coordinates of the elliptic paraboloid

$$A_z := \{(x, y) \in \mathbb{R}^2 \mid x = r \sin \omega, y = r \cos \omega \text{ with } r \in [0, \sqrt{z}], \omega \in [0, 2\pi]\} \quad (3.44)$$

of height $z \in \mathbb{R}_{\geq 0}$. Define $\alpha := \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}}$ and $\beta := \frac{1}{2\sigma^2(1-\rho^2)}$. Then the joint pdf of the variates X and Y reads $f_{(X,Y)}(x, y) = \alpha e^{-\beta(x^2+y^2-2\rho xy)}$ (cp. Appx. A) and integration by substitution gives

$$\begin{aligned} \mathbb{P}(X^2 + Y^2 \leq z) &= \mathbb{P}((X, Y) \in A_z) \\ &= \int_0^{\sqrt{z}} \int_0^{2\pi} \alpha e^{-\beta(r^2 \sin^2 \omega + r^2 \cos^2 \omega - 2\rho r^2 \sin \omega \cos \omega)} r \, d\omega dr \\ &= \alpha \int_0^{\sqrt{z}} r e^{-\beta r^2} \int_0^{2\pi} e^{\beta \rho r^2 \sin(2\omega)} \, d\omega dr \\ &= 2\pi\alpha \int_0^{\sqrt{z}} r e^{-\beta r^2} I_0(\beta \rho r^2) \, dr = \int_0^z \alpha \pi e^{-\beta \tilde{z}} I_0(\beta \rho \tilde{z}) \, d\tilde{z}, \end{aligned} \quad (3.45)$$

where the symbol I_0 in equation (3.45) denotes the modified Bessel function of the first kind. The pdf $f_{X^2+Y^2}$ of the sum $X^2 + Y^2$ hence reads

$$f_{X^2+Y^2}(z) = \begin{cases} \alpha \pi e^{-\beta z} I_0(\beta \rho z), & z \geq 0 \\ 0, & z < 0. \end{cases} \quad (3.46)$$

The two limit cases of $\rho \rightarrow 0$ of vanishing and $\rho \rightarrow 1$ of full correlation or $d \rightarrow 0$ and $d \rightarrow \frac{1}{2}$ in the language of ARFIMA processes give the iid case of an $\text{Exp}(\frac{1}{2\sigma^2})$ -distribution and the scaled $\chi^2(1)$ -distribution of $2X^2$ using the pdf (3.28), respectively.

3. Meta-Gaussian stochastic processes

In the limit of $z \rightarrow \infty$ the Bessel function I_0 in equation (3.46) satisfies the asymptotic equivalence $I_0(z) \sim \frac{e^z}{\sqrt{2\pi z}}$ [2, equ. (9.7.1)]. The tail of the pdf thus decays at rate $e^{-\beta z(1-\varrho)} z^{-\frac{1}{2}}$ dominated by the exponential decrease. In a semi-logarithmic scaling the shape of the tail is a straight line with slope $-\frac{1}{2\sigma^2(1+\varrho)}$ (Fig. 3.5).

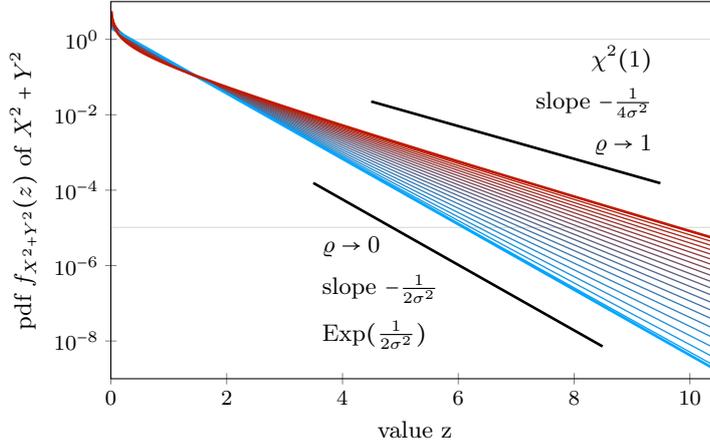


Fig. 3.5 Densities of the sum $X^2 + Y^2$ of bivariate Gaussian distributed variates X and Y with marginal standard deviation $\sigma = \frac{1}{2}$ and correlation $\varrho \in \{0.05, 0.1, \dots, 0.95\}$. The pdfs of the two limit distributions $\chi^2(2) = \text{Exp}(\frac{1}{2\sigma^2})$ with appropriately scaled for $\varrho \rightarrow 0$ and $\chi^2(1)$ for $\varrho \rightarrow 1$, respectively, delimit the range of possible distributions.

The correlation structure of the summed squares of a Gaussian process depend on how the single members of the process are added together. Let $(X_t)_{t \in \mathbb{N}}$ be a Gaussian process with standard Gaussian marginal distribution and acf ϱ_X . Given the Gaussian process X_t define a stochastic process $(Y_t)_{t \in \mathbb{N}}$ by the transformation

$$Y_t := X_t^2 + X_{t+1}^2. \quad (3.47)$$

This transformation is not pointwise like the examples in Section 3.3 but comprises two values of the original process X_t to one value of the process Y_t . By the equations (3.43) and (3.31) the variance and covariance of the process Y_t are $\text{Var}(Y_t) = 4(1 + \varrho_X(1)^2)$ and

$$\begin{aligned} & \text{Cov}(Y_t, Y_{t+k}) \\ &= \text{Cov}(X_t^2, X_{t+k}^2) + \text{Cov}(X_t, X_{t+k+1}^2) + \text{Cov}(X_{t+1}^2, X_{t+k}^2) + \text{Cov}(X_{t+1}, X_{t+k+1}^2) \\ &= 2\varrho_X(k)^2 + 2\varrho_X(k+1)^2 + 2\varrho_X(k-1)^2 + 2\varrho_X(k)^2 \\ &= 2(\varrho_X(k-1)^2 + 2\varrho_X(k)^2 + \varrho_X(k+1)^2) \end{aligned}$$

for time lags $k \in \mathbb{N}$, so that the acf ϱ_Y of the transformed process Y_t reads

$$\varrho_Y(k) = \frac{\varrho_X(k-1)^2 + 2\varrho_X(k)^2 + \varrho_X(k+1)^2}{2(1 + \varrho_X(1)^2)}. \quad (3.48)$$

Consider an AR(1) and an ARFIMA(0, d , 0) process X_t with an SRC parameter $\varphi \in (0, 1)$ and an LRC parameter $d \in (0, \frac{1}{2})$, respectively. Figure 3.6 shows the acfs of the original process X_t , the squared process X_t^2 and the transformed process Y_t (3.47). Plugging in the analytically known acfs (2.12) and (2.22) of an AR(1) and an ARFIMA(0, d , 0) process into equation (3.48), we obtain the asymptotic behavior

$$\varrho_Y(k) = \frac{1 + \varphi^2}{2\varphi^2} \varphi^{2k} \quad \text{and} \quad \varrho_Y(k) \sim \frac{2c}{1 + \varrho_X(1)^2} k^{4d-2} \quad (k \rightarrow \infty), \quad (3.49)$$

respectively, of the acfs ϱ_Y for time lags $k \in \mathbb{N}_{>0}$, where $c := \frac{\Gamma(1-d)}{\Gamma(d)}$. Note that in either case, the process Y_t bears stronger correlations than the process X_t^2 since additional dependencies enter the time series by combining different values of the process. In contrast, the correlations of the squared process fall below the original acf ϱ_X in agreement with inequality (3.16) with doubled exponent in the exponential and power-law decay, respectively, due to Hermite rank $J = 2$ in relation (3.31). Comparing the acfs ϱ_X and ϱ_Y at time lags $k = 1$ and $k = 2$ exemplifies that nonlinear transformations of a stochastic process might both increase or decrease the correlations with respect to specific time lags once the transformation is not memoryless as defined by the pointwise equation (3.1).

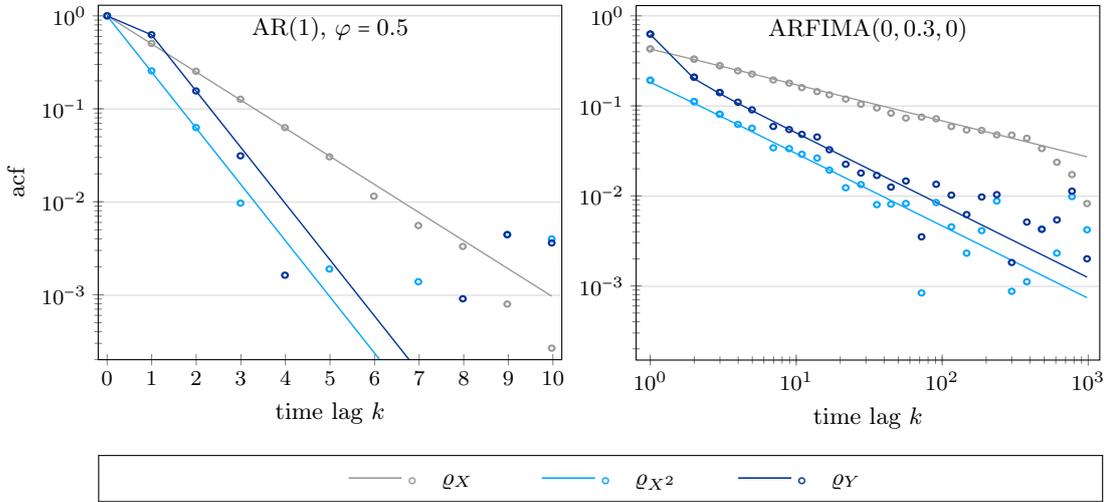


Fig. 3.6 Theoretical (solid lines) and estimated (circles) acfs of one sample time series of length $N = 100,000$ each of an AR(1) (left) and an ARFIMA(0, d , 0) (right) process X_t with AR parameter $\varphi = 0.5$ and LRC parameter $d = 0.3$, their squared processes X_t^2 and the transformed processes $Y_t = X_t^2 + X_{t+1}^2$. The solid lines are linear interpolations of the analytically known values of the acfs (2.12), (2.22), (3.31) and (3.48) at time lags $k \in \mathbb{N}_0$.

3. Meta-Gaussian stochastic processes

Note that by equation (3.43) the acf of the original process X_t affects the variance of the transformed process Y_t (3.47) whereas the opposite does not apply. Thus, an application of such a model requires a fit of the correlations of empirical exponentially distributed data by equation (3.48) prior to adjusting the marginal distribution of the model to the data by the variance σ^2 of the Gaussian process X_t .

As a remark, the process $\tilde{Y}_t := X_{2t}^2 + X_{2t+1}^2$ with distinct pairwise sums of members of an AR(1) or ARFIMA process X_t instead of overlapping pairs of summands as above obeys the acfs

$$\varrho_{\tilde{Y}}(k) = \varrho_Y(k) \cdot \varphi^{2k} \quad \text{and} \quad \varrho_{\tilde{Y}}(k) = \varrho_Y(k) \cdot 2^{4d-2}, \quad (3.50)$$

respectively, using the result (3.48). In other words, we find $\varrho_{\tilde{Y}}(k) = \varrho_Y(2k)$ for time lags $k \in \mathbb{N}$ in either case, which mimics the increased difference between the summands of the squared original process when generating the transformed process \tilde{Y} compared to the process Y_t above.

4. Numerical estimation of long-range correlations

Several approaches to the detection of the Hurst effect along with an estimation of the power-law exponent γ of the rate of the decay of the acf in relation (2.10) are available. Taqqu et al. [176] give a collection of such methods including the aggregated variance and differencing the variance methods, periodogram methods, the Whittle estimator and both the R/s -method (Sect. 4.2.1) introduced in Section 2.4 and the residuals of regression (Sect. 4.2.2). Beran's textbook [22, Chap. 5] provides a broad overview of further methods including maximum likelihood, least square and Bayesian [21] approaches. Residuals of regression is a more recent alternative name for the well-known *detrended fluctuation analysis* (DFA). The methods above can estimate LRC more robustly than a direct estimation of the power-law decay (2.10) in a double-logarithmic plot of the empirical acf. Section 4.1 is dedicated to the weaknesses of this approach. In Section 6.2 we apply R/s -analysis and DFA along with a wavelet approach (Sect. 4.2.3) for the estimation of LRC in precipitation time series.

4.1. Estimation based on the empirical autocorrelation function

A possible estimator $\bar{\varrho}_X$ of the acf ϱ_X of a stationary process at time lags $k \in \mathbb{N}_0$ based on $N \in \mathbb{N}$ temporally ordered members X_1, \dots, X_N of the process reads

$$\bar{\varrho}_X(k) := \frac{1}{(N-k)\sigma^2} \sum_{t=1}^{N-k} (X_t - \mu)(X_{t+k} - \mu) \quad (4.1)$$

if $\mu \in \mathbb{R}$ and $\sigma^2 \in \mathbb{R}_{>0}$ are the known mean and standard deviation of the process. The estimation of a power-law exponent γ in a relation $\bar{\varrho}_X(k) \propto k^{-\gamma}$ ($k \rightarrow \infty$) by equation (4.1) and a double-logarithmic plot bears the following subtleties. First, the estimator (4.1) is only unbiased if the mean μ and variance σ^2 of the process are known [148]. It is biased, however, if the two quantities are replaced by their empirical counterparts

$$\bar{\mu} := \frac{1}{N} \sum_{t=1}^N X_t \quad \text{and} \quad \bar{\sigma}^2 := \frac{1}{N-1} \sum_{t=1}^N (X_t - \bar{\mu})^2. \quad (4.2)$$

4. Numerical estimation of long-range correlations

Figure 4.1 visualizes the bias of the estimator (4.1) involving sample mean and variance (4.2) for an ARFIMA(0, 0.3, 0) process. This estimate falls below the theoretically known acf systematically resulting in an estimate of the exponent γ in the power-law decay (2.10) larger than actually present. The underestimation is recognizable in Figure 2.1 also. Therein the theoretical and empirical acfs slightly differ for the two ARFIMA processes for small time lags already.

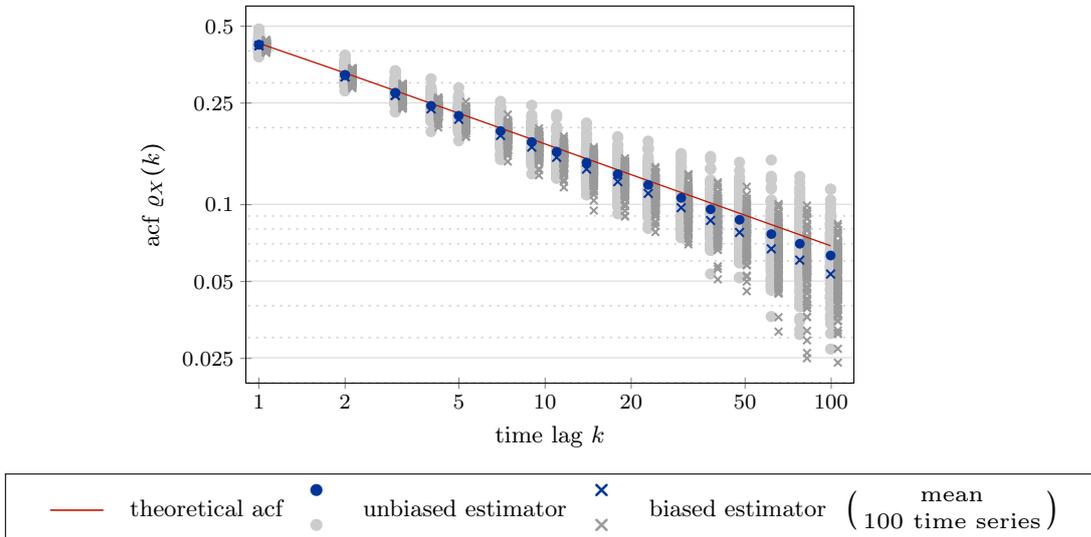


Fig. 4.1 Estimation of the theoretically known acf ϱ_X of an ARFIMA(0, 0.3, 0) process with standard Gaussian marginal distribution. For the same 100 sample time series of length $N = 36,500$ the distribution and mean of the estimator (4.1) is depicted using either the known mean and variance (unbiased) or their estimations (4.2) (biased). For the sake of clarity the biased estimates are slightly shifted to the right in the graph.

Second, in case of absent or exponentially decaying correlations or exponents γ close to unity (i.e., LRC parameters d close to zero in relation (2.20)), fluctuations of the empirical acf around zero and, in particular, logarithms of negative values, impede reliable inferences about the rate of the decay of the acf. Figure 4.2 shows estimates of the acfs of the example processes of the Figures 2.1 and 2.2 including the empirical acf of an ARFIMA(0, 0.1, 0) process. The left panel illustrates the fluctuations of the estimator (4.1) around zero for SRC data with vanishing results for time lags with a negative estimate. The right panel shows the difficulty of identifying SRC effects, in particular, for weak LRC because of strong fluctuations. Moreover, comparing the estimate to the theoretical acf reveals the bias when including the sample mean and variance (4.2). Chapter 5 is dedicated to a precise calculation of the increased variance of the sample mean for meta-Gaussian processes in the presence of correlations.

4.2. Estimation based on fluctuations on different temporal scales

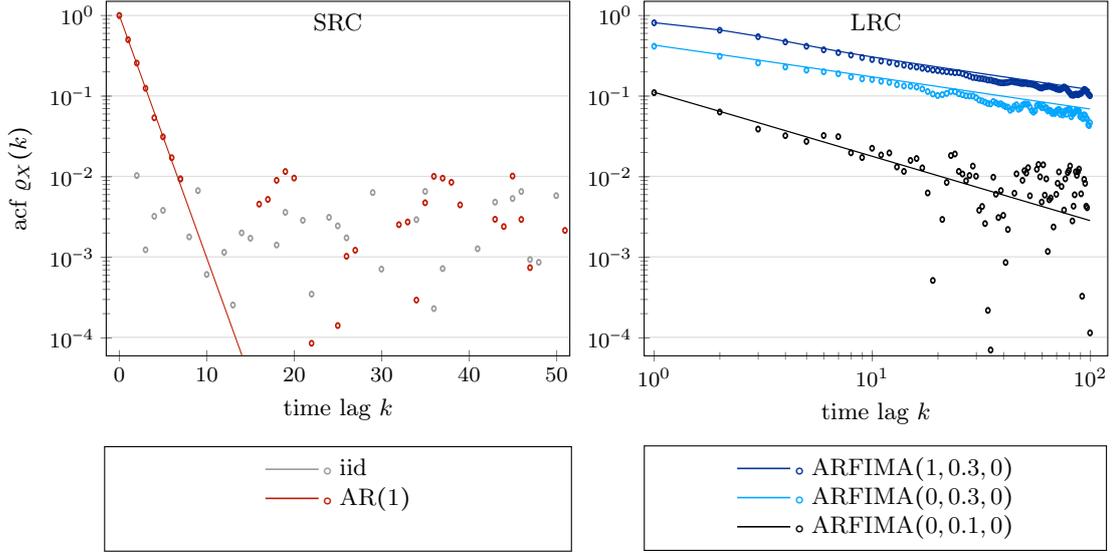


Fig. 4.2 Theoretical (solid line interpolation) and estimated (empty circles) acfs of Gaussian SRC (left) and LRC (right) processes. Using the AR parameter $\varphi = 0.5$ in either case, the estimates base on sample time series X_1, \dots, X_N of length $N = 36,500$ each of an iid and an AR(1) and three different ARFIMA processes (cp. Figs. 2.1 and 2.2).

4.2. Estimation based on fluctuations on different temporal scales

Given a time series of length $N \in \mathbb{N}$, the three methods we apply in Section 6.2 select time scales $s \in \mathbb{N}$ with $s < N$ and perform an estimate of the strength $\mathcal{F}(s)$ of fluctuations on this temporal scale. The respective quantities, namely the rescaled ranges (R/s -statistics), the fluctuation function (DFA) and the wavelet coefficients (wavelet analysis) are time averages over all disjoint time intervals or *windows* of length s contained in the data set.

The methods differ in their way of measuring the strength of fluctuations. Typically the maximal considered window size is one-third of the overall length N of the time series to ensure a reasonable number of windows for a later averaging [87]. The minimal window size depends on the particular method and certain properties of the data as discussed below. Visualizing the strength $\mathcal{F}(s)$ of the fluctuations in a time series versus the time scale s in a double-logarithmic plot, the asymptotic scaling of

$$\mathcal{F}(s) \propto s^\alpha \quad (s \rightarrow \infty) \tag{4.3}$$

identifies the correlation structure of the data by a growth exponent $\alpha \in \mathbb{R}_{>0}$. This power-law exponent α is commonly referred to as the *Hurst exponent* (cp. Equ. (2.7), Rem. 2.2 and Rem. 4.7).

4. Numerical estimation of long-range correlations

If a stochastic process has a finite correlation time in definition (2.8), then we find $\alpha = \frac{1}{2}$ [50], while $\alpha = 1 - \frac{\gamma}{2}$ [87] for LRC processes with a power-law decaying acf with exponents $0 < \gamma \leq 1$ in relation (2.10). For stationary processes this is true independently of their marginal distribution and, in particular, even if this distribution has fat tails [176].

Figure 4.3 exemplifies estimates of LRC for synthetic data with known Hurst exponents using the methods R/s -analysis, DFA and wavelet analysis. The three methods give similar estimates of the Hurst exponent for the depicted stationary stochastic processes. In case of the nonstationary fractional Brownian motion, R/s -statistics differ from DFA and the wavelet analysis. Sections 4.2.1, 4.2.2 and 4.2.3 give details on the implementation and interpretation of these methods and Section 4.4 provides a summary and remarks.

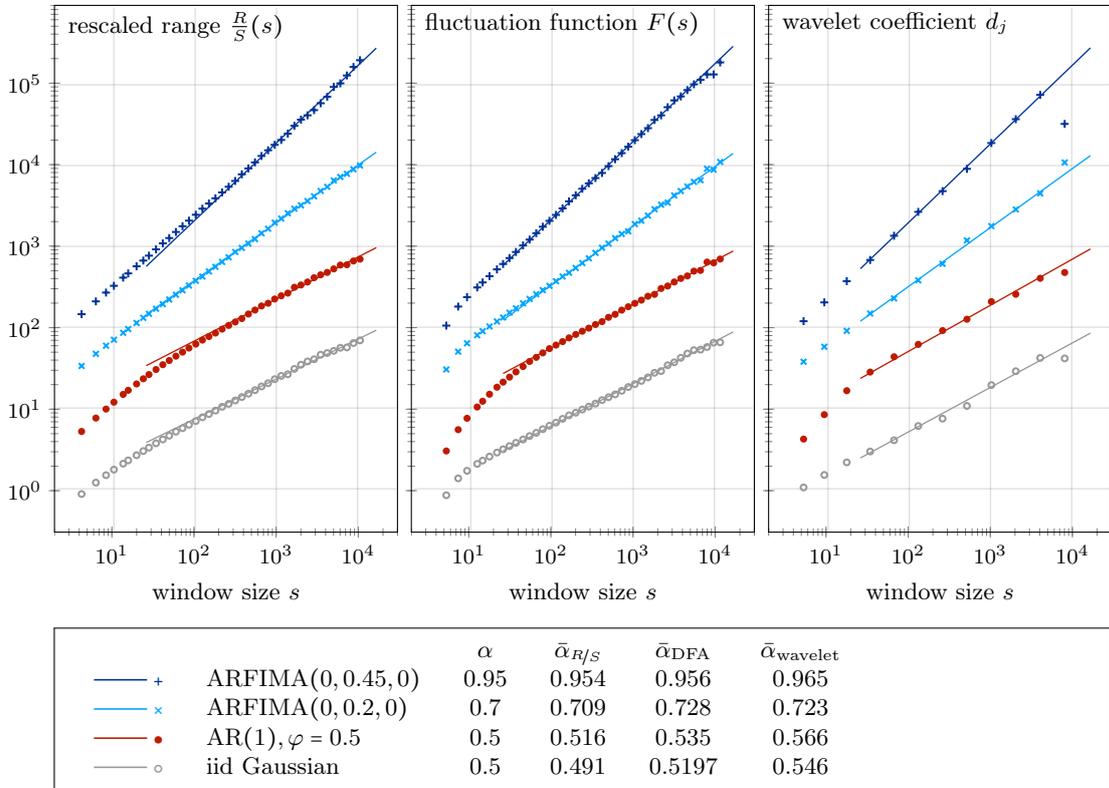


Fig. 4.3 Estimations $\bar{\alpha}$ for different stationary stochastic processes with prescribed Hurst exponent α by the three methods R/s -analysis, DFA(3) and wavelet analysis. The results are shifted vertically for better perceptibility.

4.2.1. Rescaled-range analysis

Hurst studied water levels of the river Nile that result from the accumulation of river in- and outflow [82] (cp. Sect. 2.4). As a natural quantity, the profile (4.4) of the water flow time series reflects the fluctuations of the gauge heights. To Hurst's surprise, the influx did not act as white noise, causing the water volume inside the reservoir to behave like a Brownian path, but he found some different scaling behavior by the following procedure. For R/s -analysis the input time series is divided into non-overlapping windows of equal size $s \in \mathbb{N}$ (cp. Sect. 4.2). In every window (X_1, \dots, X_s) of the time series we determine the *profile*

$$Y_t := \sum_{i=1}^t X_i \quad (4.4)$$

with indices $t = 1, \dots, s$, and its *range*

$$R(s) := \max_{1 \leq t \leq s} \left(Y_t - \frac{t}{s} Y_s \right) - \min_{1 \leq t \leq s} \left(Y_t - \frac{t}{s} Y_s \right) \quad \text{and} \quad (4.5)$$

$$S(s) := \frac{1}{s-1} \sum_{i=1}^s \left(X_i - \frac{Y_s}{s} \right)^2, \quad (4.6)$$

where the *scale* (4.6) denotes the window's sample variance S . The smallest possible window size is $s = 2$, so that both the range (4.5) and the sample variance (4.6) are well defined. For low window sizes fluctuations strongly affect the sample variance. This effect diminishes as the window size increases. The range (4.5) captures the highest and lowest positions of the profile compared to the straight line of uniform growth (with respect to the window mean $\mu_w := \frac{Y_s}{s}$) [161]. Note that this definition of the range R is equivalent to first centering the points in every window by subtraction of the window's sample mean before calculating the range (4.5) of the profile within the respective window by

$$\begin{aligned} \tilde{Y}_t &:= \sum_{i=1}^t (X_i - \mu_w) = Y_t - t\mu_w \quad \text{and} \\ \tilde{R}(s) &:= \max_{1 \leq t \leq s} \tilde{Y}_t - \min_{1 \leq t \leq s} \tilde{Y}_t = R(s). \end{aligned} \quad (4.7)$$

Figure 4.4 visualizes the two conceptions (4.5) and (4.7) of the R/s -approach.

4. Numerical estimation of long-range correlations

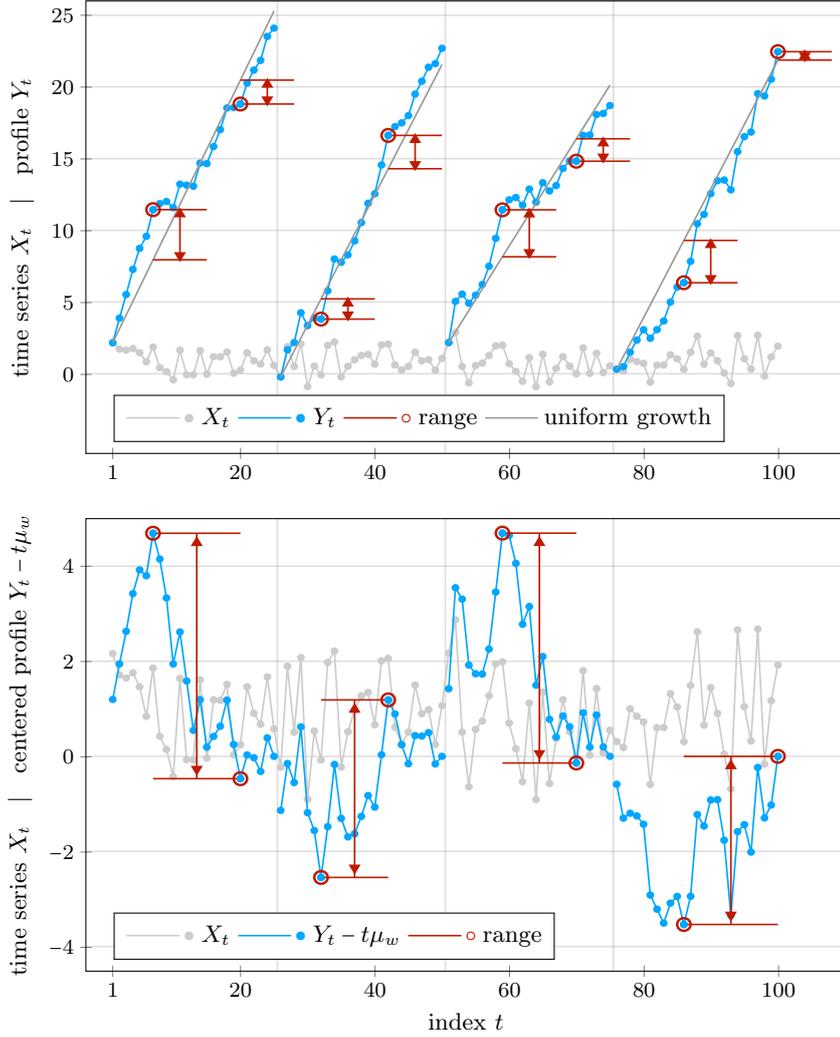


Fig. 4.4 Two perceptions of R/s -analysis with the profile Y_t (4.4) (top) and the profile \tilde{Y}_t (4.7) centered by the window's mean μ_w (bottom) for windows of size $s = 25$ and an iid time series X_t of length $N = 100$ with Gaussian marginal distribution $N(1, 1)$. The grid on the abscissa separates the four windows each. In each window an arrow marks the range between the maximal and minimal value of the (centered) profile of the (centered) window data. The sum of the two parts of the range per window in the top gives the range in bottom graph.

By averaging the rescaled range $\frac{R(s)}{S(s)}$ over all windows of size s under certain conditions [22] we obtain a scaling behavior

$$\frac{R}{S}(s) := \left\langle \frac{R(s)}{S(s)} \right\rangle \propto s^H \quad (s \rightarrow \infty). \quad (4.8)$$

4.2. Estimation based on fluctuations on different temporal scales

The angle brackets in definition (4.8) denote the sample mean of all rescaled ranges $\frac{R(s)}{S(s)}$ in windows of size s , so that the number of values entering the average differs depending on how many windows of size s the empirical time series contains. Numerically the slope of a linear regression of the points $(s, \frac{R}{S}(s))$ in a double-logarithmic scaling yields the R/s - or Hurst exponent H as shown for some synthetic time series in Figure 4.3.

The deliberate choice of the notation H for the parameter of fBm by Mandelbrot reflects that for fGn the R/s -analysis gives the growth exponent H in relation (4.8) [116]. In case of an iid process, the rescaled ranges (4.8) grow with an exponent $H = \frac{1}{2}$. This exponent is not only robust for marginal distributions with heavier tails than Gaussian but also in the presence of weak correlations like SRC [161].

The R/s -exponent is bounded from below by zero and from above by unity [111]. Nonstationarities in the data lead to Hurst exponents close to unity and difficulties may arise in distinguishing them from strong LRC effects [110]. Figure 4.3 shows the result of a R/s -analysis for a stationary ARFIMA process with strong LRC ($d = 0.45$) and a nonstationary process defined by $X_t + t$ with iid standard Gaussian components X_t and a linear trend $t \in \mathbb{N}$. The estimated Hurst exponents are close to unity in either case, so that a reliable inference about the origin of such a behavior of the rescaled ranges is impossible.

4.2.2. Detrended fluctuation analysis

The R/s -analysis lacks the ability of appropriately distinguishing intrinsically nonstationary processes like fBm or processes with additive or local trends from processes with strong LRC (cp. Fig. 4.7 and Sect. 4.2.1). Aiming at separating LRC from such nonstationarities, Peng et al. [147] formulated the detrended fluctuation analysis.

Similar to R/s -statistics, for DFA we consider the profile Y_t (4.4) of an input time series $(X_t)_1^N$ of length $N \in \mathbb{N}$. The profile series is divided into non-overlapping windows of equal size $s \in \mathbb{N}$ again. For DFA(q) in every window a polynomial of degree $q \in \mathbb{N}_{>0}$ is fitted to and then subtracted from the profile. In doing so, local trends are intended to be removed. Let p_i be the values of the fitted polynomial at times $i = 1, \dots, s$ in such a window (Y_1, \dots, Y_s) . In every window the fluctuations of the detrended profile are captured by

$$\tilde{F}^2(s) := \frac{1}{s} \sum_{i=1}^s (Y_i - p_i)^2. \quad (4.9)$$

The minimal adequate window size s is $q + 2$ as $\tilde{F}^2(s) = 0$ for all $1 \leq s \leq q + 1$ due to exact interpolation. Then the *fluctuation function* F is defined as

$$F(s) := \sqrt{\langle \tilde{F}^2(s) \rangle} \propto s^\alpha \quad (s \rightarrow \infty) \quad (4.10)$$

4. Numerical estimation of long-range correlations

by averaging over all windows of size s . The fluctuation function (4.10) reflects the average asymptotic behavior of fluctuations of the detrended profile as described below. Figure 4.5 visualizes the method DFA(3) for an example process with additive nonstationarities.

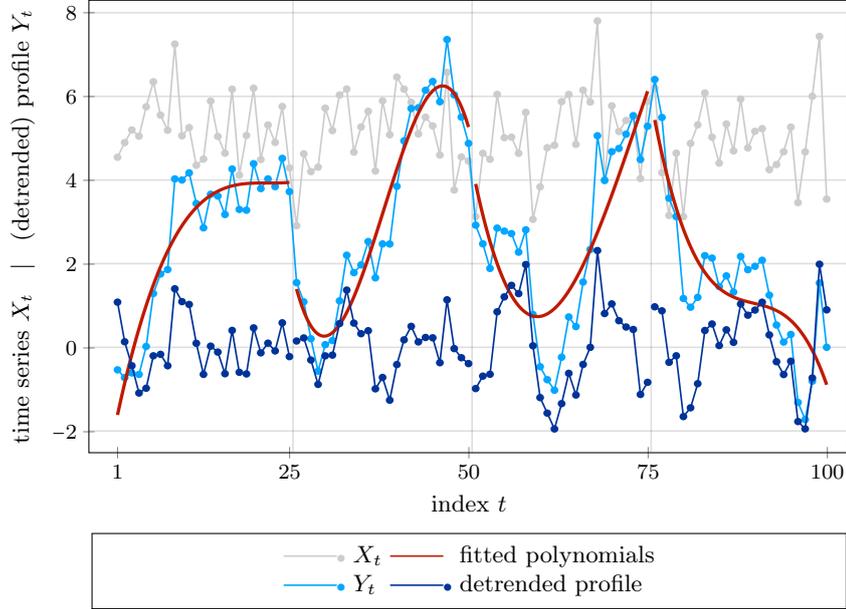


Fig. 4.5 Visualization of DFA(3) for a sample time series of length $N = 100$ of the example process $X_t = \varepsilon_t + 5 + \frac{1}{2} \sin\left(\frac{2\pi}{3}t\right)$ with iid standard Gaussian components ε_t , profile Y_t (4.4) and the detrended profile for window size $s = 25$, $t = 1, \dots, N$. The grid on the abscissa separates the four windows. In each window a polynomial of degree three is fitted to and subtracted from the profile.

Like for the R/s -exponent H (4.8), the DFA exponent α is numerically estimated as the slope of a linear regression of the points $(s, F(s))$ in a double-logarithmic scaling with values s up to one-third of the length of the input time series X_t .

If a stochastic process is stationary and has a power-law decaying acf with exponent γ in relation (2.10), then asymptotically as the window size increases the fluctuation function grows like the power law (4.10) with

$$\alpha = 1 - \frac{\gamma}{2}. \quad (4.11)$$

4.2. Estimation based on fluctuations on different temporal scales

Heuristics for this relation and for the growth rate (4.10) can be found in [87]. For stationary processes with finite variance, recently, it was derived that the fluctuation function (4.10) of DFA is indeed an exact transformation of the acf [76, 107] (cp. Sect. 4.4.1), so that relation (4.11) is satisfied. For the Gaussian toy models fGn and ARFIMA(p, d, q) and generalizations of them to stationary LRC processes with finite variance marginal distributions the R/s - and DFA exponents H and α , respectively, coincide [129, 176], that is $\alpha = H$ for $0 < H < 1$ and

$$\alpha = d + \frac{1}{2} \quad (4.12)$$

for ARFIMA processes as $\gamma = 1 - 2d$ in this case (cp. Sect. 2.5). In Section 6.2.2 we generalize relation (4.12) to meta-Gaussian processes for fitting a transformed ARFIMA model to non-Gaussian daily precipitation amounts.

As a remark, by DFA Hurst exponents $\alpha > 1$ have been found in empirical data (which is impossible with R/s -statistics). These findings motivated generalizations [33, 107] of the term (cp. Sect. 4.4.1). For fBm B_H with parameter $0 < H < 1$ DFA gives the scaling exponent $\alpha = H + 1$ [133] in relation (4.10), while its R/s -exponent (4.8) equals unity due to its nonstationarity. Figure 4.7 visualizes this difference in the results of these two methods in case of nonstationary processes.

4.2.3. Wavelet analysis

Wavelet analysis is a filtering technique for the detection of (temporal) scale invariance (cp. Sect. 2.4) in time series. This method has been widely used in signal and image processing [4] before it became a common tool in the field of long-range dependence [9, 89]. For a didactic introduction to wavelet transforms see [3]. An application-oriented summary is given in [137, Sects. 1.1.12, 4.5.5].

The description below embeds the concept of wavelet analysis in the field of LRC. Lemma 4.3 provides a generalization of the well-known *Haar wavelet* for distinguishing higher order polynomial trends in data from LRC along with visualizing the *generalized Haar wavelets* in Example 4.5. The numerical implementation of the wavelet analysis based on these generalized Haar wavelets is particularly straight forward and established in Lemma 4.6.

4. Numerical estimation of long-range correlations

Definition: (Wavelet) A *wavelet* is an integrable function $\psi : \mathbb{R} \rightarrow \mathbb{R}$ such that

$$\int_{\mathbb{R}} \psi(t) dt = 0. \quad (4.13)$$

Typically, wavelets are chosen as $\mathcal{L}^1(\mathbb{R})$ or $\mathcal{L}^2(\mathbb{R})$ functions (see below). A wavelet ψ is said to have $m \in \mathbb{N}_{>0}$ *zero moments* or *vanishing moments* if

$$\int_{\mathbb{R}} t^n \psi(t) dt = 0 \quad (4.14)$$

for all exponents $n = 0, \dots, m - 1$.

Given a wavelet ψ_0 , define a family $\{\psi_{j,k} : \mathbb{R} \rightarrow \mathbb{R} \mid j, k \in \mathbb{Z}\}$ of wavelets by

$$\psi_{j,k}(t) := 2^{-\frac{j}{2}} \psi_0(2^{-j}t - k). \quad (4.15)$$

The wavelet ψ_0 is commonly referred to as the *mother wavelet*, of which the functions (4.15) are *dilations* and *translations*. The factors 2^j with *octaves* $j \in \mathbb{Z}$ are called *scales*, the constants $k \in \mathbb{Z}$ denote integer shifts. The normalization by the factor $2^{\frac{j}{2}}$ ensures preserving the \mathcal{L}^2 -norm since $\int_{\mathbb{R}} \psi_{j,k}(t)^2 dt = \int_{\mathbb{R}} \psi_0(t)^2 dt$ for all values $j, k \in \mathbb{Z}$ [3]. Frequently used wavelets are the *Haar*, the *Daubechies*, the *poor man's* ($\psi(t) := \delta(t) - \delta(t - \tau)$, $t, \tau \in \mathbb{R}$) and the *Mexican hat* ($\psi(t) := \frac{d^2t}{dt^2} \phi(t)$, $t \in \mathbb{R}$) wavelet. The *original Haar wavelet* $\psi_0 : \mathbb{R} \rightarrow \mathbb{R}$ is defined by

$$\psi_0(t) := \begin{cases} 1, & t \in [0, 1/2) \\ -1, & t \in [1/2, 1) \\ 0, & \text{otherwise.} \end{cases} \quad (4.16)$$

Integration yields $\int_{\mathbb{R}} t^n \psi(t) dt = \frac{1}{n+1} (2^{-n} - 1)$ for all exponents $n \in \mathbb{N}_0$, so that the original Haar wavelet obeys $m = 1$ zero moment [137, Ex. 1.9]. As a natural generalization, step functions with a higher number of different constant pieces than the wavelet (4.16) allow for more vanishing moments. Lemma 4.3 provides a generalization of the original Haar wavelet with $m = 1$ zero moments to higher numbers m of vanishing moments. Example 4.5 and Figure 4.6 visualize these wavelets for numbers $m \in \{1, 3, 5, 7\}$.

4.2. Estimation based on fluctuations on different temporal scales

Given a mother wavelet ψ_0 , the functions (4.15) form a set of filters and base the method of the discrete wavelet transform of a stochastic process.

Definition (Discrete wavelet transform): For a stochastic process X_t the set of *wavelet coefficients*, also called *details*,

$$d_{j,k} := \int_{\mathbb{R}} X_t \psi_{j,k}(t) dt \quad (4.17)$$

with octaves and shifts $j, k \in \mathbb{Z}$ is called the *discrete wavelet transform* (DWT) of the process X_t .

Lemma 4.1 collects essential properties of the wavelet coefficients (4.17).

Lemma 4.1 (cp. [3]): *Let X_t be a stochastic process. Then the wavelet coefficients $d_{j,k}$ defined by equation (4.17) with octaves and shifts $j, k \in \mathbb{Z}$ feature the following properties.*

- (i) *If the process X_t has stationary increments, or is stationary itself, then for all octaves j the wavelet coefficients $d_{j,k}$ are stationary with respect to the shifts k .*
- (ii) *If the process X_t is self-similar with self-similarity parameter $0 < H < 1$ (cp. Sect. 2.5), then for arbitrary but fixed octave j the process $(d_{j,k})_{k \in \mathbb{Z}}$ satisfies*

$$d_{j,k} \stackrel{d}{=} 2^{j(H+1/2)} d_{0,k}. \quad (4.18)$$

- (iii) *If the process X_t is self-similar with stationary increments and self-similarity parameter $0 < H < 1$, zero mean and finite variance, then*

$$\mathbb{E}[d_{j,k}] = 0 \quad \text{and} \quad \mathbb{E}[d_{j,k}^2] = \mathbb{E}[d_{0,0}^2] \cdot 2^{j(2H+1)}. \quad (4.19)$$

Note that the expected value of the squared wavelet coefficients depends on the scale 2^j only and does not depend on the specific shift k .

Heuristics. For heuristical calculations of the properties in Lemma 4.1 see [3]. The stationarity (i) is due to the definition (4.17) of the wavelet coefficients $d_{j,k}$ by plugging in the stationarity of the process X_t .

The self-similarity of the process X_t induces $a^{-H} X_{at} \stackrel{d}{=} X_t$ for all factors $a \in \mathbb{R}_{>0}$. Applying a change of variables to definition (4.17) of the coefficients $d_{j,k}$ implies equation (4.18) in statement (ii).

4. Numerical estimation of long-range correlations

Interchanging expectation and deterministic integration in definition (4.17) yields the zero mean (4.19) of the wavelet coefficients $d_{j,k}$ in statement (iii). Combining the stationarity (i) of the coefficients $d_{j,k}$ with respect to the shifts k and the scaling (4.18) in statement (ii) reasons the scaling of their expected value in relation (4.19). \square

The scaling (4.19) allows for estimating the presence and strength of LRC in time series by the application of DWTs to a centered stochastic process X_t . For fBm with self-similarity parameter $0 < H < 1$ in this relation we find the exponent $j(2H + 1)$, and $j(2H - 1)$ for the corresponding fGn [167]. Assessing the scaling behavior of the root expectation of the squared wavelet coefficients $d_{j,k}^2$ thus allows for direct access of the Hurst exponent by DWT. Given a mother wavelet ψ_0 , for an octave $j \in \mathbb{Z}$ the root average

$$d_j := \sqrt{\langle d_{j,k}^2 \rangle} \quad (4.20)$$

is a non-parametric estimator of the root expectation (4.19) and independent of the shift $k \in \mathbb{Z}$ by the stationarity (i) in Lemma 4.1.

For an octave j the sample average $\langle d_{j,k}^2 \rangle$ in definition (4.20) denotes the sample mean of the squared DWTs (4.17) of all non-overlapping windows of the process X_t generated by the wavelets $\psi_{j,k}$ (4.15) by using different shifts $k \in \mathbb{Z}$. In a double-logarithmic scaling the slope of a linear regression of the pairs $(2^j, d_j)$ of scales and estimated wavelet coefficients depicts the scaling behavior (4.19).

Remark 4.2: Mind that the Hurst exponent reads H for fGn (cp. Sect. 4.2.1) and $H + 1$ for fBm (cp. Sect. 4.2.2). Neglecting the \mathcal{L}^2 -normalization by $2^{-\frac{j}{2}}$ in definition (4.15) yields the exponents $2jH$ and $2j(H + 1)$ in relation (4.19), respectively. Then in a double-logarithmic scaling the slope of a linear regression of the pairs $(2^j, d_j)$ coincides with these Hurst exponents without further computation.

The sample mean in definition (4.20) is unbiased if the number m of zero moments of the mother wavelet ψ_0 is larger than $H - 1$ as, in this case, the wavelet coefficients $d_{j,k}$ do not exhibit practically significant LRC [4]. In general, however, the sample mean of LRC data obeys strong uncertainty (cp. Chap. 5). The fluctuations of the estimates d_j over scales 2^j are a second order description of the process X_t and constitute a kind of ‘wavelet spectrum’ [3].

For the numerical implementation of wavelet analyses Lemma 4.3 provides generalized Haar wavelets with an arbitrary prescribed number of zero moments. Rybsky et al. [159] formulate an alternative construction of wavelets with a desired number of vanishing moments by higher-order differencing of a time series’ profile (4.4).

4.2. Estimation based on fluctuations on different temporal scales

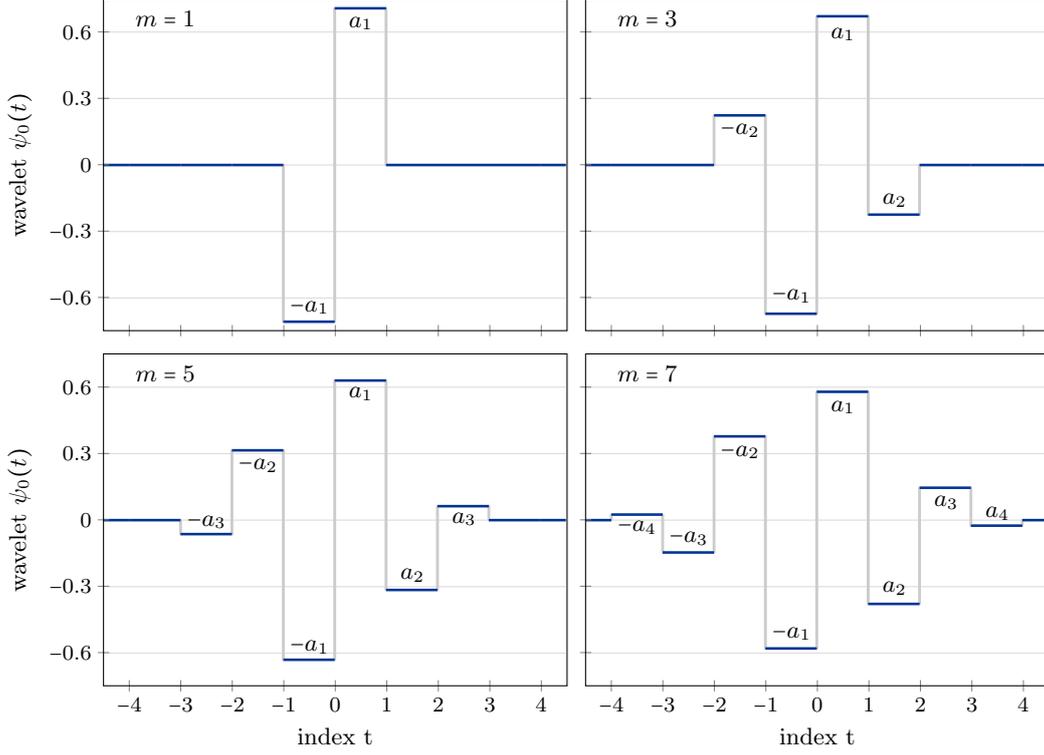


Fig. 4.6 Generalized Haar wavelets by Lemma 4.3 visualized for $m \in \{1, 3, 5, 7\}$ vanishing moments with step heights calculated in Example 4.5 and collected in Table 4.1.

Lemma 4.3 (Generalized Haar wavelets): *Let $m \in \mathbb{N}$ be a desired uneven number of zero moments and define the two intervals $I_i := [i - 1, i)$ and $J_i := [-i, -(i - 1))$. Then there exist unique constants $a_i \in \mathbb{R}$, $i = 1, \dots, \lceil \frac{m}{2} \rceil$, with $a_1 > 0$, such that the step function*

$$\psi_0(t) := \sum_{i=1}^{\lceil \frac{m}{2} \rceil} a_i \left(\chi_{I_i}(t) - \chi_{J_i}(t) \right) \quad (4.21)$$

is a wavelet with m zero moments and unit \mathcal{L}^2 -norm. (The symbol χ_A denotes the indicator function of a set A . Note that by the symmetry of the function (4.21) the integral (4.14) vanishes for all even exponents $n \in \mathbb{N}$.)

Proof. The step function (4.21) is antisymmetric with respect to the origin due to the symmetric choice of the intervals I_i and J_i , $i = 1, \dots, \lceil \frac{m}{2} \rceil$, and thus a wavelet by its very definition. For values $n \in \mathbb{N}$ elementary calculations yield

$$0 \stackrel{!}{=} \int_{\mathbb{R}} t^n \psi_0(t) dt = \int_{-\lceil \frac{m}{2} \rceil}^{\lceil \frac{m}{2} \rceil} t^n \psi_0(t) dt = \frac{(-1)^{n+1} + 1}{n + 1} \sum_{i=1}^{\lceil \frac{m}{2} \rceil} a_i (i^{n+1} - (i - 1)^{n+1}), \quad (4.22)$$

4. Numerical estimation of long-range correlations

which hence vanishes for all even numbers n . Setting equation (4.22) equal to zero for all uneven $n \in \{1, \dots, m-2\}$ gives a set of $\lfloor \frac{m}{2} \rfloor$ linear equations determining the $\lfloor \frac{m}{2} \rfloor$ parameters $a_1, \dots, a_{\lfloor \frac{m}{2} \rfloor}$. This under-determined set of constants becomes unique when additionally normalizing the wavelet ψ_0 to unit \mathcal{L}^2 -norm involving

$$1 \stackrel{!}{=} \int_{\mathbb{R}} \psi_0(t)^2 dt = \int_{\mathbb{R}} \sum_{i,j=1}^{\lfloor \frac{m}{2} \rfloor} a_i a_j (\chi_{I_i}(t) - \chi_{J_i}(t)) (\chi_{I_j}(t) - \chi_{J_j}(t)) dt \quad (4.23)$$

$$= \sum_{i=1}^{\lfloor \frac{m}{2} \rfloor} a_i^2 \int_{\mathbb{R}} \chi_{I_i}(t) + \chi_{J_i}(t) dt = 2 \sum_{i=1}^{\lfloor \frac{m}{2} \rfloor} a_i^2. \quad (4.24)$$

Note that the $\lfloor \frac{m}{2} \rfloor$ vectors $(i^{n+1} - (i-1)^{n+1})_{i=1}^{\lfloor \frac{m}{2} \rfloor}$, $n \in \{1, \dots, m-2\}$ uneven, used in equation (4.22) are linearly independent, so that together with demanding unit \mathcal{L}^2 -norm (4.24) and $a_1 > 0$ the heights $a_1, \dots, a_{\lfloor \frac{m}{2} \rfloor}$ of the steps of the wavelet (4.21) are unique. \square

Remark 4.4: If $m \in \mathbb{N}$ is even, then the wavelet ψ_0 obtained by the procedure established in Lemma 4.3 has $m+1$ zero moments since the integral (4.14) vanishes for all exponents $n = 0, \dots, m$. For uneven numbers m the integral (4.14) does not vanish for the exponent $n = m$ as the vector $(i^{m+1} - (i-1)^{m+1})_{i=1}^{\lfloor \frac{m}{2} \rfloor}$ is linearly independent from the $\lfloor \frac{m}{2} \rfloor$ vectors $(i^{n+1} - (i-1)^{n+1})_{i=1}^{\lfloor \frac{m}{2} \rfloor}$, $n \in \{1, \dots, m-2\}$ uneven. The obtained parameters $a_1, \dots, a_{\lfloor \frac{m}{2} \rfloor}$ thus cannot eliminate the integral (4.22) for the exponent $n = m$.

Example 4.5 applies Lemma 4.3 for calculating the generalized Haar wavelets (4.21) depicted in Figure 4.6.

Example 4.5: The sum in equation (4.22) shall be represented as the scalar product of the vector of parameters $(a_1, \dots, a_{\lfloor \frac{m}{2} \rfloor})$ and the vector of coefficients $(i^{n+1} - (i-1)^{n+1})_{i=1}^{\lfloor \frac{m}{2} \rfloor}$ for each uneven value $n \in \{1, \dots, m-2\}$. For the numbers $m \in \{3, 5, 7\}$ of vanishing moments Table 4.1 (top) exemplifies the vectors of coefficients for deriving the linear equations for the parameters $a_1, \dots, a_{\lfloor \frac{m}{2} \rfloor}$. Note that the number $\lfloor \frac{m}{2} \rfloor$ of different step heights depends on the desired number m of vanishing moments. The parameters are unique when assuming $a_1 > 0$ and involving a unit \mathcal{L}^2 -norm by equating the integral (4.24) to unity. Table 4.1 (bottom) and Figure 4.6 visualize the generalized Haar wavelets based on Lemma 4.3 for $m \in \{1, 3, 5, 7\}$ zero moments.

4.2. Estimation based on fluctuations on different temporal scales

	$n = 1$	$n = 3$	$n = 5$
$m = 3$	(1, 3)	–	–
$m = 5$	(1, 3, 5)	(1, 15, 65)	–
$m = 7$	(1, 3, 5, 7)	(1, 15, 65, 175)	(1, 31, 211, 781)

$m = 1$	$m = 3$	$m = 5$	$m = 7$
$a_1 = \frac{\sqrt{2}}{2} \approx 0.71$	$a_1 = \frac{3\sqrt{5}}{10} \approx 0.67$	$a_1 = \frac{5\sqrt{7}}{21} \approx 0.63$	$a_1 \approx 0.579$
–	$a_2 = -\frac{\sqrt{5}}{10} \approx -0.22$	$a_2 = -\frac{5\sqrt{7}}{42} \approx -0.32$	$a_2 \approx -0.378$
–	–	$a_3 = \frac{\sqrt{7}}{42} \approx 0.06$	$a_3 \approx 0.146$
–	–	–	$a_4 \approx -0.025$

Table 4.1 Coefficients (top table) of the linear equations (see Ex. 4.5) describing the parameters $a_1, \dots, a_{\lfloor \frac{m}{2} \rfloor}$ for the heights of the different steps of the generalized Haar wavelets (4.21) by regarding the integral (4.22) as a scalar product and equating it to zero for uneven exponents $n \in \{1, \dots, m - 2\}$ and deduced parameter values exemplified for $m \in \{1, 3, 5, 7\}$ vanishing moments (bottom table). The $\lfloor \frac{m}{2} \rfloor$ linear equations (4.22) together with the normalization (4.24) yield exact solutions for the coefficients with $a_1 > 0$.

The stepped shape of the generalized Haar wavelets (4.21) induce a particularly simple computation of the filter (4.17). Without loss of generality, consider the scaled but unshifted wavelets $\psi_{j,0}$ of the mother wavelet (4.21) centered at the argument $t = 0$ as depicted in Figure 4.6. For discrete-time processes only dilations ($j \in \mathbb{N}_0$) of this mother wavelet are valid. The reasoning of Lemma 4.6 is analogous for the shifted wavelets filtering different windows of the input data.

Given an octave $j \in \mathbb{N}_0$ and a desired number $m \in \mathbb{N}$ zero moments, by construction the infimum and supremum of the arguments $t \in \mathbb{Z}$, for which the scaled wavelets $\psi_{j,0}(t)$ are nonzero, are $t = \mp 2^j \lfloor \frac{m}{2} \rfloor$, respectively. The number of integer indices covered by the scaled wavelet $\psi_{j,0}$, in other words, of points included in the filtering, thus reads $2^{j+1} \lfloor \frac{m}{2} \rfloor + 1$. The $2 \lfloor \frac{m}{2} \rfloor + 1$ positions of the jumps of these step functions are the two bounds of the scaled support of the wavelet ψ_0 and all points within on an equidistant grid with accuracy 2^j . Let

$$(X_{-\lfloor \frac{m}{2} \rfloor}, X_{-(\lfloor \frac{m}{2} \rfloor - 1)}, \dots, X_{-1}, X_0, X_1, \dots, X_{\lfloor \frac{m}{2} \rfloor - 1}, X_{\lfloor \frac{m}{2} \rfloor})^T \quad (4.25)$$

enumerate these members of a stochastic process X_t within the considered window. Lemma 4.6 describes a simple numerical implementation of the filter (4.17) just using a scalar product of two real vectors.

4. Numerical estimation of long-range correlations

Lemma 4.6: *Let X_t be a discrete-time stochastic process with profile Y_t (4.4). Let $j \in \mathbb{N}_0$ be an octave and define*

$$\mathbf{a} := (a_{\lceil \frac{m}{2} \rceil}, a_{\lceil \frac{m}{2} \rceil - 1} - a_{\lceil \frac{m}{2} \rceil}, \dots, a_1 - a_2, -2a_1, a_1 - a_2, \dots, a_{\lceil \frac{m}{2} \rceil - 1} - a_{\lceil \frac{m}{2} \rceil}, a_{\lceil \frac{m}{2} \rceil})^T, \quad (4.26)$$

where $a_1, \dots, a_{\lceil \frac{m}{2} \rceil}$ are the heights of the steps of the generalized Haar wavelet (4.21) with $m \in \mathbb{N}$ vanishing moments given by Lemma 4.3. Analogous to notation (4.25) denote by

$$\mathbf{Y} := (Y_{-\lceil \frac{m}{2} \rceil}, Y_{-(\lceil \frac{m}{2} \rceil - 1)}, \dots, Y_{-1}, Y_0, Y_1, \dots, Y_{\lceil \frac{m}{2} \rceil - 1}, Y_{\lceil \frac{m}{2} \rceil})^T \quad (4.27)$$

the values of the profile Y_t corresponding to the location of the jumps of the scaled generalized Haar wavelet $\psi_{j,0}$. Then the filter (4.17) reduces to the scalar product

$$d_{j,k} = \mathbf{a}^T \mathbf{Y}. \quad (4.28)$$

Proof. For the sake of clarity define $s := \lceil \frac{m}{2} \rceil$. Then equation (4.28) follows from plugging in definition (4.21) into the filter (4.17) and using notation (4.27) and the property $\int_{c_1}^{c_2} X_t dt = Y_{c_2} - Y_{c_1}$, $c_1, c_2 \in \mathbb{R}_{>0}$, $c_2 \geq c_1$, since

$$\begin{aligned} \int_{\mathbb{R}} X_t \psi_{j,0}(t) dt &= \int_{\mathbb{R}} X_t \sum_{i=1}^s \chi_{I_i}(t) - \chi_{J_i}(t) dt \\ &= \sum_{i=1}^s a_i (Y_i - Y_{i-1} - (Y_{-(i-1)} - Y_{-i})) \\ &= \sum_{i=1}^s a_i Y_i - \sum_{i=0}^{s-1} a_{i+1} Y_i - \sum_{i=0}^{s-1} a_{i+1} Y_{-i} + \sum_{i=1}^s a_i Y_{-i} \\ &= a_s Y_{-s} + \sum_{i=1}^{s-1} Y_{-i} (a_i - a_{i+1}) - 2a_1 Y_0 + \sum_{i=1}^{s-1} Y_i (a_i - a_{i+1}) + a_s Y_s. \end{aligned} \quad (4.29)$$

The scalar product (4.28) captures the sum (4.29). \square

In the Figures 4.3 and 4.7 and in Chapter 6 we apply the generalized Haar wavelet with $m = 3$ vanishing moments for wavelet analyses of synthetic data and empirical precipitation amounts, respectively.

4.3. Sensitivity of the estimators

Figure 4.3 shows a similarly good performance of the three methods R/s -analysis, DFA and wavelet analysis when estimating Hurst exponents $\frac{1}{2} \leq \alpha < 1$ of stationary stochastic processes. The R/s -analysis, however, cannot distinguish between additive and intrinsic nonstationarities. This method estimates both the Hurst exponent $\alpha = 1.8$ of a fBm with self-similarity parameter $H = 0.8$ and $\alpha = \frac{1}{2}$ of an iid process with an additive linear trend close to unity (Fig. 4.7).

The methods of DFA and wavelet analysis were designed particularly for detecting local trends. Peng [147] originally formulated DFA applying a linear detrending. Both methods succeed in distinguishing local polynomial trends from LRC as the estimated Hurst exponents for the iid process with linear trend are close to $\frac{1}{2}$ either. Figure 4.7 also shows that they properly estimate the Hurst exponent of approximately 1.8 for the fBm above. Contrary, by construction R/s -statistics cannot result in Hurst exponents larger than unity [111].

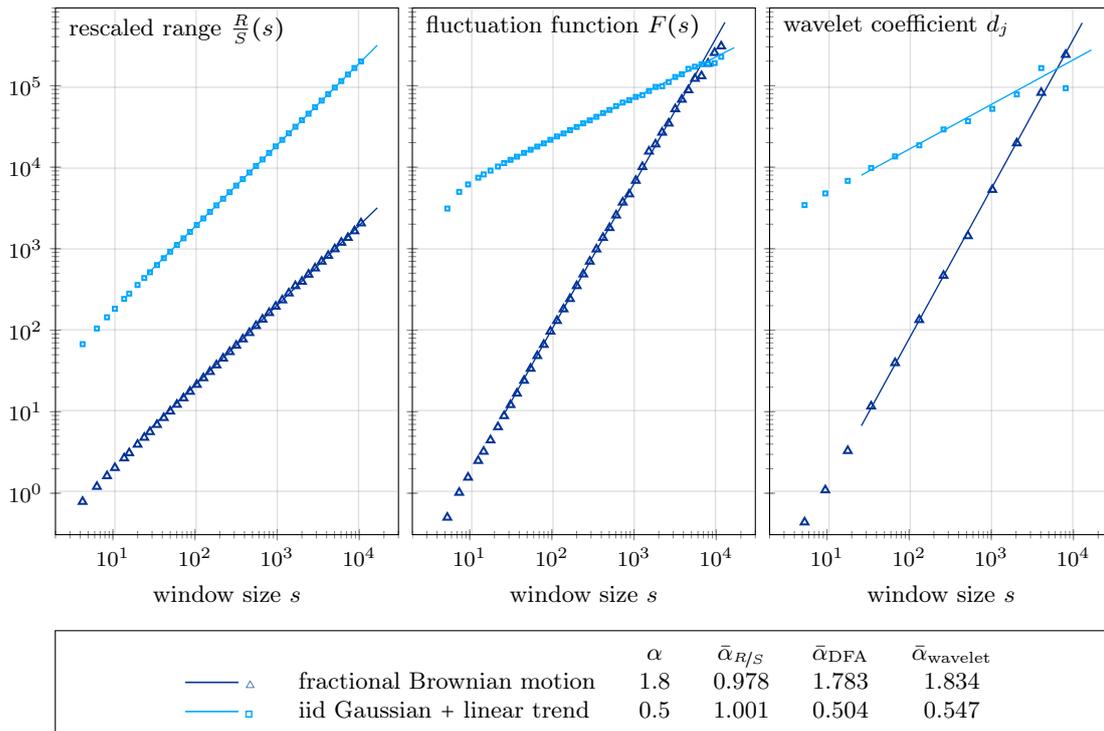


Fig. 4.7 Estimated Hurst exponents $\bar{\alpha}$ for an intrinsically nonstationary fBm with self-similarity parameter $H = 0.8$ and Hurst exponent $\alpha = H + 1 = 1.8$ and for a process with an additive linear trend by the three methods R/s -analysis, DFA(3) and wavelet analysis with $m = 3$ zero moments. The process $X_t + t$ is the sum of iid Gaussian process X_t and the linear trend $t \in \mathbb{N}$. The results are shifted vertically for better perceptibility.

4. Numerical estimation of long-range correlations

Numerical studies [81, 87] showed that higher order DFA(q) can remove additive polynomial trends of order $q \in \mathbb{N}$ in the profile and $q - 1$ in the original time series. Wavelets with $m \in \mathbb{N}$ zero moments are blind to polynomials of degrees up to $m - 1$, in other words, the wavelet coefficients of a stationary process X_t and processes $X_t + p(t)$ with such an additive polynomial p coincide [3]. In this case, DFA(q) corresponds to wavelet analysis with wavelets of order $m = q$. Note that by construction for even numbers $m \in \mathbb{N}$ the generalized Haar wavelets (4.21) obey $m + 1$ zero moments, whereas m zero moments for odd numbers m (cp. Rem. 4.4). The Figures 4.8 and 4.9 visualize the effect of detrending on the estimation of Hurst exponents and the relation between DFA and wavelet analyses using the generalized Haar wavelets.

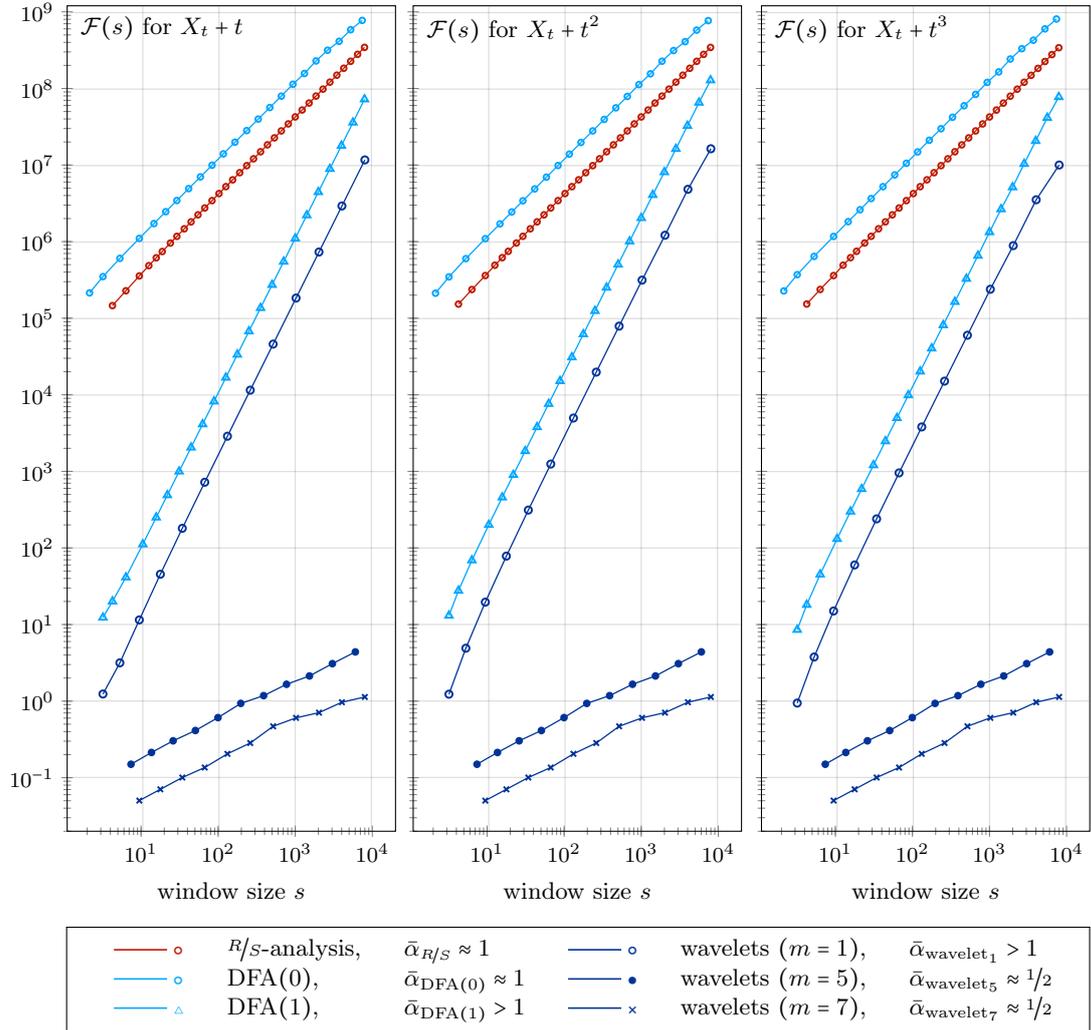


Fig. 4.8 Measured strengths \mathcal{F} (4.3) of fluctuations and estimated Hurst exponents $\bar{\alpha}$ of an iid standard Gaussian process X_t with different additive polynomial trends by the methods R/s -analysis, DFA(0), DFA(1) and wavelet analyses using the generalized Haar wavelets (4.21) with $m \in \{1, 5, 7\}$ zero moments. For these methods the results for the three examples coincide⁴. The latter two wavelet analyses detect the correct Hurst exponent of $\frac{1}{2}$ in either case. The results are shifted vertically for better perceptibility.

⁴See Figure 4.9 and below for comments on slopes > 1 .

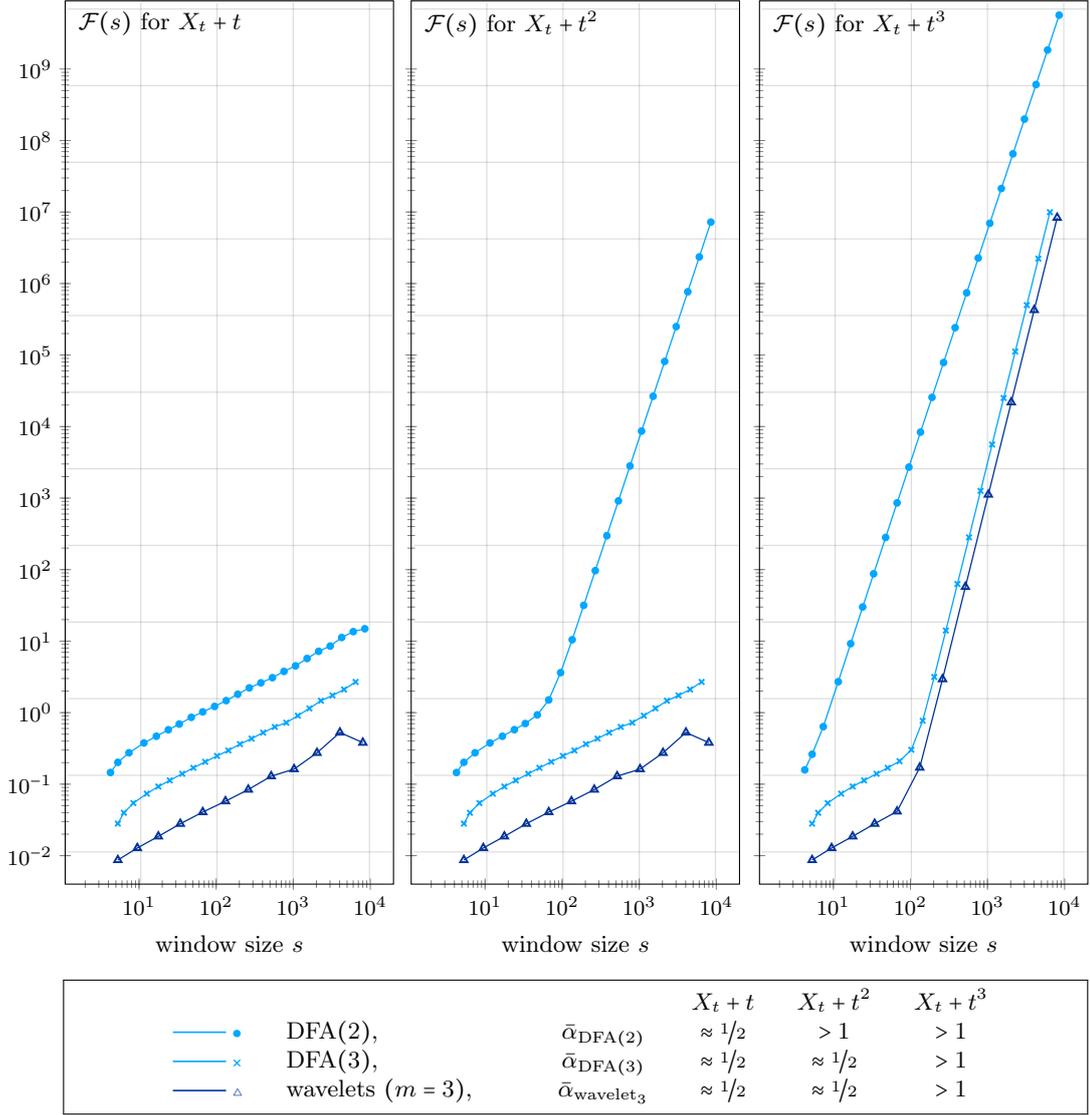


Fig. 4.9 Measured strengths \mathcal{F} (4.3) of fluctuations and estimated Hurst exponents $\bar{\alpha}$ of an iid standard Gaussian process X_t with different additive polynomial trends by the methods DFA(2), DFA(3) and wavelet analyses using the generalized Haar wavelets (4.21) with $m = 3$ zero moments. All these methods remove the linear trend consistently with the correct estimated Hurst exponent of $\frac{1}{2}$. DFA(2) cannot remove higher order polynomial trends just as both DFA(3) and the generalized Haar wavelets for orders higher than two. The results are shifted vertically for better perceptibility.

Figure 4.9 visualizes that the shape of the estimated strength \mathcal{F} (4.3) of fluctuations carries information beyond the asymptotic behavior described by the Hurst exponent. For smaller window sizes the impact of the added polynomial trends in the mid and right panel dominate the estimators only from a crossover point on. Until that the slope $\frac{1}{2}$ of the iid Gaussian process determines the measured fluctuations.

As a remark, the exact slopes > 1 for DFA and wavelet analysis remain an open question of further research. The slope of R/s -analysis is bounded by 1 [111]. The right panel of Figure 4.9 indicates that similarly an upper limit slope for the two other methods might depend on their order of detrending.

4. Numerical estimation of long-range correlations

The different orders of DFA and wavelet analysis with different detrending capabilities can yield insight into both LRC and types of local trends [159] and can help examine the trend structure in the data and distinguish it from LRC [17]. The ability of DFA and wavelet analysis of recognizing polynomial trends is nonetheless limited in the presence of strong SRC or nonstationarities on temporal scales beyond the empirical data. The origin of numerically detected LRC might always lie in strong SRC or the local trend in a periodic signal with low frequency. Figure 4.10 shows fluctuation functions by DFA(3) for an AR(1) process with SRC, an ARFIMA(1, 0.3, 0) process with LRC and a periodic signal with additive Gaussian white noise. Recent studies apply the location of crossover points of the fluctuation function (4.10) to the inference of characteristic time scales in time series [126–128].

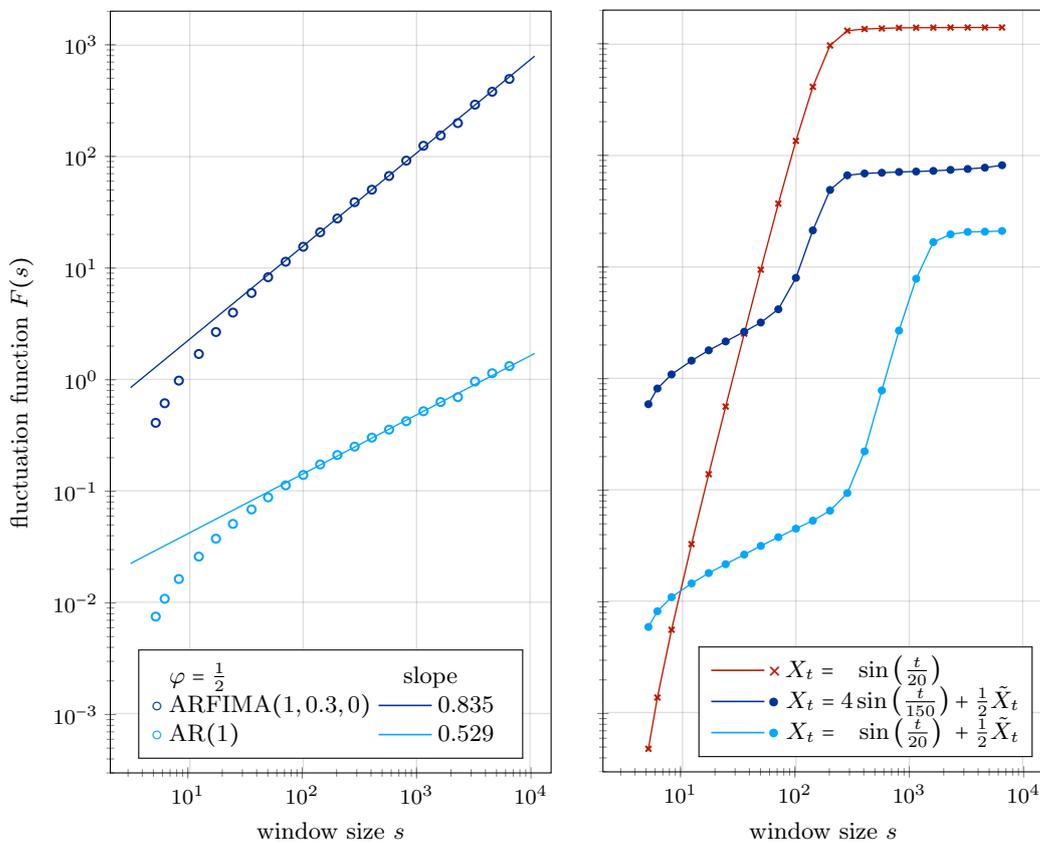


Fig. 4.10 Fluctuation function (4.10) by DFA(3) of an ARFIMA(1, 0.3, 0) and an AR(1) process X_t with AR parameter $\varphi = 0.3$ either (left) and deterministic sine time series X_t of different frequency and without and with additive standard Gaussian white noise \tilde{X}_t (right). Crossovers in the fluctuation functions indicate characteristic time scales of dominating effects such as SRC or local trends in periodic signals. A zero slope indicates constant strength of fluctuations. The results are shifted vertically for better perceptibility.

The estimates of LRC by DFA or wavelet analysis in practice can not only be affected by nonstationarity [13, 77] but also non-Gaussianity [16]. On this account, a double-checking of detected LRC by different methods is indispensable. In the Sections 6.1.1 and 6.1.2 we apply R/s -analysis, DFA and wavelet analysis to daily precipitation records to estimate the stationarity and correlation structure of this data.

4.4. Closing remarks

The detection and estimation of the origins of the scaling behavior of stochastic processes is an active field of research. Recent studies focus on relations between anomalous diffusion and LRC (Sect. 4.4.1). Also the quantification of the uncertainty of methods like DFA is a current matter of interest. Section 4.4.2 summarizes explicit formulae assigning confidence intervals to the estimators introduced in the Sections 4.2.1 to 4.2.3.

4.4.1. Anomalous scaling of stochastic processes

The temporal scaling of a dynamical process typically refers to its spreading behaviour determined by the (finite-sample) *mean squared displacement* (MSD)

$$\langle |x(t) - x(0)|^2 \rangle := \frac{1}{K} \sum_{k=1}^K |x^{(k)}(t) - x(0)|^2. \quad (4.30)$$

The MSD serves a natural measurement of the diffusion of particles $x(t) \in \mathbb{R}^d$, $d \in \mathbb{N}$, with initial value $x(0)$ at time $t = 0$ as time $t \in \mathbb{R}_{>0}$ evolves. The MSD is an ensemble average and the sum in definition (4.30) includes the temporal evolution of $K \in \mathbb{N}$ particles $x^{(k)}$ enumerated $k = 1, \dots, K$. Note that by *Birkhoff's Ergodic Theorem* the two averages over time and space coincide in ergodic systems [42] (cp. Sect. 5.3.2). The average spreading of particles that follow a standard Brownian motion $B(t)$ in $d \in \mathbb{N}$ dimensions yields an MSD that scales linearly with time as $\langle |B(t) - B(0)|^2 \rangle \sim 2dDt$ as $t \rightarrow \infty$ with diffusion coefficient $D \in \mathbb{R}_{>0}$. The asymptotic scaling behavior

$$\langle |x(t) - x(0)|^2 \rangle \propto t^\beta \quad (t \rightarrow \infty) \quad (4.31)$$

provides a classification of the spreading of diffusive processes by the exponent $\beta \in \mathbb{R}_{>0}$ into *normal diffusion* ($\beta = 1$) and *anomalous diffusion* ($\beta \neq 1$) or, more precisely, *subdiffusion* ($0 < \beta < 1$) and *superdiffusion* ($1 < \beta < 2$). An exponent $\beta = 2$ describes *ballistic motion* on constant velocity.

4. Numerical estimation of long-range correlations

For each of the above types of diffusion there exist mathematical toy models with a prescribed scaling of their MSD [152]. Deterministic models, such as iterations of piecewise linear functions [62] or Pomeau-Manneville-type maps [117] are capable of generating normal and both sub- and superdiffusion [55], respectively. Random walks or Brownian motions are examples of normally diffusive stochastic processes. Randomly chosen piecewise linear maps, however, can generate subdiffusion of different strength depending on the probability of choosing either map per iteration [165]. The aforementioned models deploy intermittency as a central mechanism yielding anomalous scaling behaviour [125, 189]. For continuous time random walk models it is known [94] that a waiting time distribution with density $\rho(n) \sim n^{-(1+\beta)}$ as time $n \rightarrow \infty$ implies an MSD exponent of β in relation (4.30). Figure 4.11 shows sample trajectories for some of these models. Compared to the normal diffusive sample with purely random behavior, the super- and subdiffusive ones remain longer in the status of moving to one direction or persisting at a current position, respectively.

Contemporary studies [5, 33, 107] focus on differentiating and detecting the origins of anomalous scaling. Estimators like DFA are only capable of detecting anomalous scaling of stochastic processes, which traces back to their autocorrelations by Hurst exponents other than the value $\frac{1}{2}$ in relation (4.3). The fluctuation function (4.10) is a transformation of the acf [76, 107] (see Sect. 4.2.2). Recent results generalize the Hurst effect by splitting up the Hurst exponent into a sum $H = L + J + M - 1$ referring to the Noah (L), Joseph (J) and a newly introduced *Moses (M) effect* [33], respectively (cp. Rem. 2.2). The single summands represent the different sources of *anomalous scaling* with growth exponents $H \neq \frac{1}{2}$, namely a fat-tailed (see Sect. 2.3) increment distribution, LRC and nonstationarities.

Remark 4.7: Following the notion of Remark 2.2, the *Moses effect* refers to the biblical story of Moses, who “led the Israelites after their Exodus from Egypt as they wandered through the wilderness having no stationary settlements” [33].

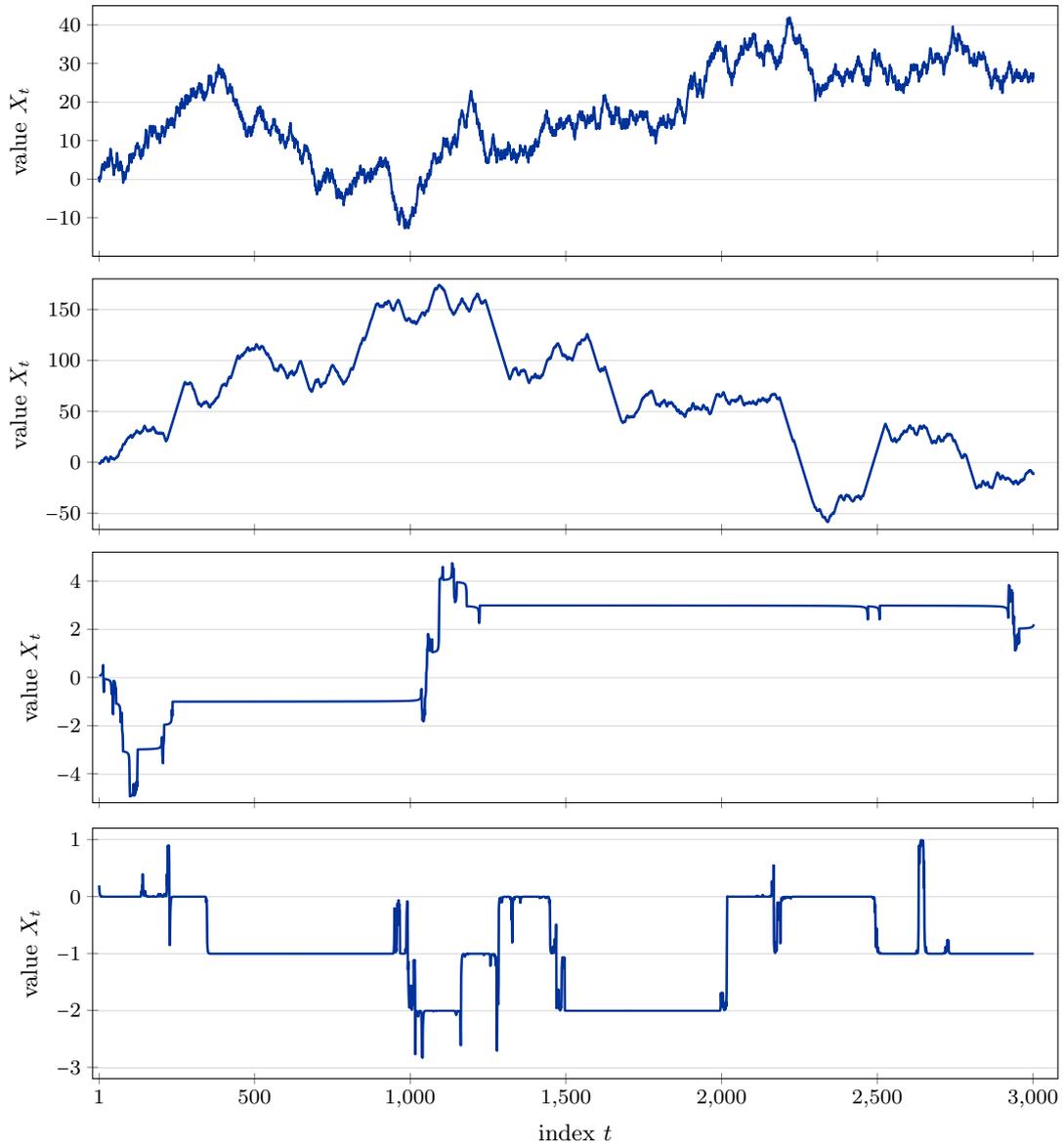


Fig. 4.11 Sample trajectories X_t , $1 \leq t \leq 3,000$, of a piecewise linear map, a super- and subdiffusive Pomeau-Manneville-type map and a random piecewise linear map (from top to bottom).

4. Numerical estimation of long-range correlations

4.4.2. Probabilistic properties of the estimators

For Gaussian processes there exist explicit formulae describing probabilistic properties of the rescaled range (2.7) (Sect. 4.2.1) and the fluctuation function (4.10) (Sect. 4.2.2). The distribution of wavelet coefficients (4.20) (Sect. 4.2.3) in a double logarithmic scaling is known also for non-Gaussian settings.

Let X_t be a fGn process with self-similarity parameter $0 < H < 1$. Then asymptotically the rescaled range satisfies [22, Sect. 5.4.2]

$$s^{-H} \frac{R}{S}(s) \xrightarrow{d} \sup_{0 \leq u \leq 1} \tilde{B}_H(u) - \inf_{0 \leq u \leq 1} \tilde{B}_H(u), \quad (4.32)$$

where $\tilde{B}_H(u) := B_H(u) - uB_H(1)$ denotes the *fractional Brownian bridge*. A generalization of relation (4.32) to linear LRC processes with finite variance is given in [161, Sect. 5.1]. The distribution of the rescaled ranges and thus confidence intervals for the estimated Hurst exponents are typically determined by Monte Carlo simulations [97, 183]. This approach helps quantify the uncertainty of the R/s -analysis in case of non-Gaussian data, as well [98].

Abry et al. [3, Sect. 4.2] provide an explicit formula for the variance of the estimated Hurst exponent by wavelet analysis even in the non-Gaussian case. For a large number $n_j \in \mathbb{N}$ of non-overlapping windows in the sample average (4.20) the squared wavelet coefficients d_j^2 are χ^2 random variables with n_j degrees of freedom. Then one can deduce an explicit description of the mean and variance of the slope of the wavelet coefficients in a double logarithmic scaling [137, Sect. 4.5.5].

Similarly, a recent study [172] derives probabilistic properties of the squared fluctuation function F^2 (4.10). It includes explicit formulae for the mean and variance of the LRC parameter obtained by DFA applied to white and fractional Gaussian noise and standard fractional Brownian motion.

5. Effective sample size of stationary processes

The presence of correlations affects the estimation of statistical quantities such as the sample mean and variance (4.2) of empirical data. The central limit theorem (CLT) ensures that the sample mean of iid finite-variance samples is asymptotically Gaussian. Deviations from this setting might, however, not only change the convergence rate of such observables but also their asymptotic probabilistic properties. Thus, usage of the classical confidence intervals for the sample mean might no longer be justified [56, Chap. 1].

Section 5.1 is dedicated to the asymptotics and the finite-sample variance of the sample mean for stationary processes. For certain meta-Gaussian LRC processes the limit distribution of the sample mean might even deviate from Gaussianity as summarized in Section 5.1.1 according to [22, Chap. 4.2]. In Section 5.1.2 the concept of an *effective sample size* provides a quantitative comparison of the uncertainty of the sample mean under the influence of correlations to the iid case by an effectively decreased number of samples. Correlated and thus in a way redundant samples carry the same information like a corresponding lower number of iid samples. Section 5.1.3 reflects the asymptotic behavior of the effective sample size and draws a relation to characteristic time scales in empirical data based on the uncertainty of the sample mean. Section 5.2 gives an exact calculation of the finite-sample and asymptotic variance of the sample mean and the effective sample size for meta-Gaussian processes. Some closing remarks in Section 5.3 draw relations of the effective sample size to anomalous scaling and large deviation theory.

5.1. Asymptotics and uncertainty of the sample mean

Let Y_1, \dots, Y_N be $N \in \mathbb{N}$ samples of a stationary stochastic process with mean $\mu_Y := \mathbb{E}[Y_i]$ and finite variance $\sigma_Y^2 \in \mathbb{R}_{>0}$ and denote by

$$\bar{\mu}_N := \frac{1}{N} \sum_{i=1}^N Y_i \quad \text{and} \quad \sigma_{\bar{\mu}_N}^2 := \text{Var}(\bar{\mu}_N) \quad (5.1)$$

the sample mean and the variance of the sample mean of N samples, respectively. If the variates Y_1, \dots, Y_N are iid random variables, then by the CLT the normalized sample mean $\sqrt{N} \left(\frac{\bar{\mu}_N - \mu_Y}{\sigma_Y} \right)$ converges in distribution to a standard Gaussian variate as the number N of samples approaches infinity. If further $\mathbb{E}[Y_i^3] < \infty$, then by the Berry-Esseen theorem the rate of this convergence with respect to the Kolmogorov-Smirnov distance reads $\frac{1}{\sqrt{N}}$ as $N \rightarrow \infty$ (i.e., there exists a constant $C \in \mathbb{R}_{>0}$ such that the cdf of the sample mean $\bar{\mu}_N$ satisfies $|\mathbb{P}(\bar{\mu}_N \leq x) - \Phi(x)| \leq \frac{C}{\sqrt{N}}$ for all values $x \in \mathbb{R}$).

5. Effective sample size of stationary processes

Different generalizations of the classical CLT exist by relaxing either of its conditions. The textbook [22] by Beran et al. provides a comprehensive collection of stochastic limit theorems for sums of random variables both in terms of convergence in distribution or of weak convergence of the sum's cdf as *functional central limit theorems*.

An example of a limit theorem for normalized sums of infinite-variance variates is given by α -stability. The sum of $N \in \mathbb{N}$ iid symmetric and fat-tailed random variables (cp. Sect. 2.3) with asymptotics $\propto |x|^{-\alpha-1}$ of their density as $|x| \rightarrow \infty$ tends to an α -stable variate with the same tail parameter $\alpha \in (0, 2]$ when normalized by $N^{-1/\alpha}$ [22, Sect. 4.3.2.3]. Hence, the rate of the convergence of the sample mean $\bar{\mu}_N$ reduces to $\frac{1}{N^{1-1/\alpha}}$ in this case compared to $\frac{1}{\sqrt{N}}$ in the Gaussian limit case of $\alpha = 2$ or the classical CLT, respectively.

In the presence of SRC along with finite variance, the Gaussian limit distribution and rate of convergence of the sample mean remain valid just as in the classical CLT. Finite-variance processes that are *mixing* or exhibit *m-dependence* or the Markov property (cp. Appx. D) satisfy this condition [56, Cap. 3]. If the decay of the acf is, however, too slow, then non-Gaussian limit distributions are possible even if the variates obey finite variance. Such a setting is occasionally referred to as “noncentral” limit theorems [22, Sect. 4.2.5.3] and is valid for example for certain meta-Gaussian LRC processes as reflected in Sect. 5.1.1.

5.1.1. Limit distribution of the sample mean

Violating the conditions of the CLT potentially changes the asymptotic properties of the sample mean (5.1). For stationary Gaussian LRC processes the limit distribution of the sample mean remains Gaussian. The speed of the convergence to the limit distribution is however slower than in the iid setting. Theorem 5.1 captures these two observations. Section 5.1.2 deals with the implications for the uncertainty of the sample mean due to a lower speed of its convergence in the presence of LRC.

Theorem 5.1 (cp. Thm. 4.2 in [22]): *Let X_t be a stationary standard Gaussian process with acf $\varrho_X(k) \propto k^{2d-1}$ as time lags $k \rightarrow \infty$ and with LRC parameter $d \in (0, \frac{1}{2})$. For $N \in \mathbb{N}$ samples of the process X_t define the sum $S_N(u) := \sum_{t=1}^{\lfloor Nu \rfloor} X_t$. Then*

$$N^{-(d+1/2)} L(N) S_N(u) \Longrightarrow B_H(u) \quad (u \in (0, 1)), \quad (5.2)$$

where $L(N)$ is a slowly varying function as $N \rightarrow \infty$ and B_H is a standard fractional Brownian motion with Hurst parameter $H = d + \frac{1}{2}$ (“ \Longrightarrow ” denotes weak convergence).

5.1. Asymptotics and uncertainty of the sample mean

Relation (5.2) yields a slowed convergence at the lower rate $\frac{1}{N^{1/2-d}}$ as $N \rightarrow \infty$ of the normalized sample mean $N^{1/2-d}\bar{\mu}_N$ for LRC parameters $d \in (0, \frac{1}{2})$ compared to $\frac{1}{N^{1/2}}$ for iid samples. Figure 5.2 visualizes the increased uncertainty of the sample mean for different ARFIMA time series.

Remark 5.2: Theorem 5.1 remains valid for parameters $d \in (-\frac{1}{2}, \frac{1}{2})$ if the acf ϱ_X satisfies $\sum_{k=-\infty}^{\infty} \varrho_X(k) = 0$ in the antipersistent case of $d \in (-\frac{1}{2}, 0)$ [22]. The setting of $d = 0$ represents the presence of at most SRC by a finite correlation time (2.8) (cp. Sect. 5.1).

For meta-Gaussian processes the limit distribution of the sample mean is analytically known and depends on the Hermite rank (cp. Sect. 3.2) of the transformation that defines the process. In the generic case of Hermite rank $J = 1$ (see Rem. 3.5), the sample mean (5.1) remains asymptotically Gaussian. Higher Hermite ranks result in non-Gaussian limits as Theorem 5.3 specifies.

Theorem 5.3 (cp. Thm. 4.4 in [22]): *Let X_t be a stationary standard Gaussian process with acf $\varrho_X(k) \propto k^{2d-1}$ as time lags $k \rightarrow \infty$ and with LRC parameter $d \in (0, \frac{1}{2})$. For $N \in \mathbb{N}$ samples of a meta-Gaussian process $Y_t = g(X_t)$ define the sum $S_N(u) := \sum_{t=1}^{\lfloor Nu \rfloor} Y_t$. If the Hermite rank J of the transformation $g \in \mathcal{L}^2$ satisfies $J(1-2d) < 1$, then*

$$N^{-(1-J(1/2-d))} L(N) S_N(u) \Longrightarrow \frac{\alpha_J^2}{J!} Z_{J,H}(u) \quad (u \in (0, 1)), \quad (5.3)$$

where $Z_{J,H}$ is an Hermite process with parameter $H = d + \frac{1}{2}$ and α_J is the J -th Hermite coefficient (3.8) (“ \Longrightarrow ” denotes weak convergence).

The condition $J(1-2d) < 1$ in Theorem 5.3 ensures LRC for the transformed process. If the Hermite rank J satisfied the opposite, then the correlation time (2.8) would be finite despite a power-law decaying acf. Such an SRC setting is introduced as “intermediate” in Section 2.4 and implies a Gaussian limit distribution (cp. Sect. 5.1).

Analogously to the implications of Theorem 5.1, relation (5.3) yields the rate $\frac{1}{N^{J(1/2-d)}}$ for the convergence of the normalized sample mean $N^{J(1/2-d)}\bar{\mu}_N$ as $N \rightarrow \infty$. The condition $J(1-2d) < 1$ implies $J(\frac{1}{2} - d) < \frac{1}{2}$, so that this convergence rate is slower than the rate $\frac{1}{N^{1/2}}$ of the iid setting.

5. Effective sample size of stationary processes

Figure 5.1 shows the distribution of the sample mean $\bar{\mu}_N$ for 1,000 time series each of length $N = 5,000$ for a Gaussian LRC ARFIMA(0, 0.3, 0) process X_t and the three meta-Gaussian example processes of Section 3.3, namely the lognormal e^{X_t} , the square X_t^2 and the absolute-value process $|X_t|$. Applying Theorem 5.1 and Theorem 5.3 yields an asymptotically Gaussian distribution for the sample mean of the original Gaussian and the lognormal process as both examples obey Hermite rank $J = 1$. Note that the Hermite process $Z_{1,H}$ in convergence (5.3) is the fBm and thus Gaussian. Both the square and the absolute value exhibit Hermite rank $J = 2$ resulting in the *Rosenblatt process* $Z_{2,H}$ in convergence (5.3). The sample mean for the squared process is a sum of dependent χ^2 -variates and follows the non-Gaussian Rosenblatt distribution [175]. The pdf of the Rosenblatt distribution lacks a closed-form description but can be approximated numerically [181].

The Gaussian ML-fitted densities in Figure 5.1 well represent the sample mean distributions of the processes X_t and e^{X_t} . Such Gaussians, however, do not well describe the Rosenblatt sample mean distribution of the processes X_t^2 and $|X_t|$ as visualized by the very skewness of the empirical distributions.

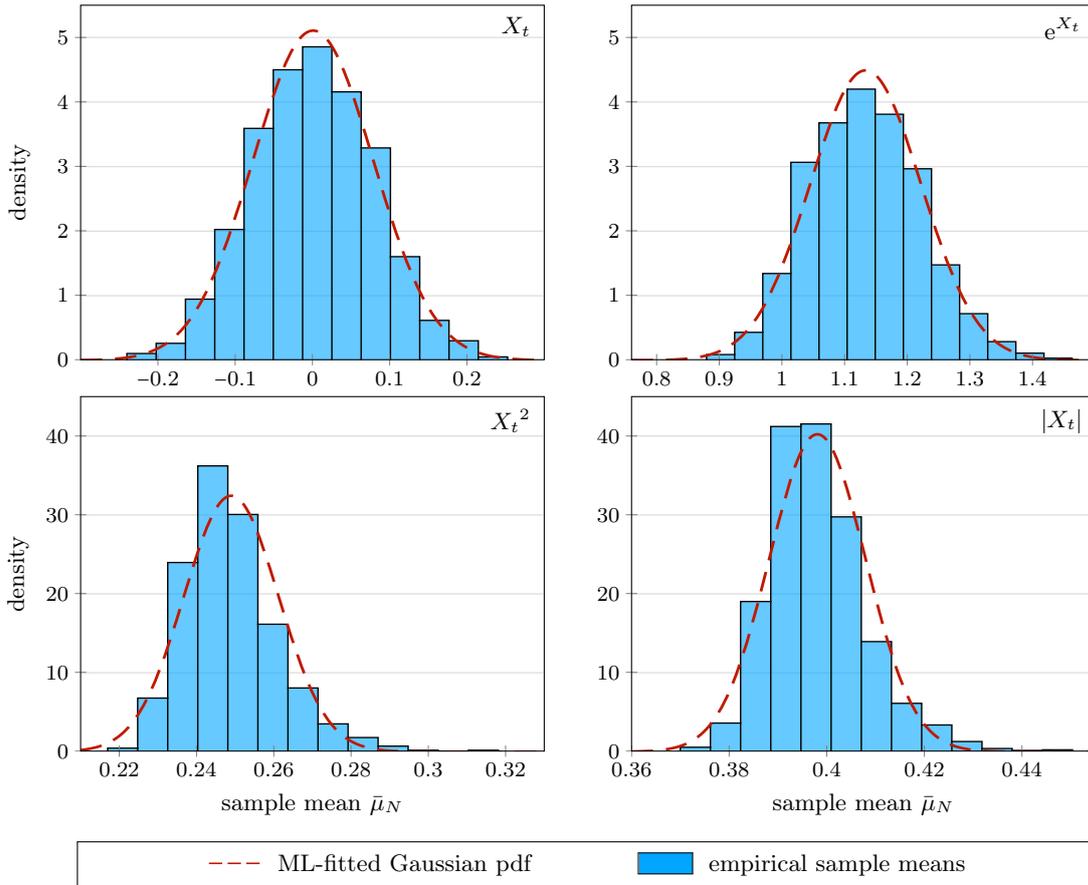


Fig. 5.1 Histograms of the sample means $\bar{\mu}_N$ of 1,000 time series each of length $N = 5,000$ of a Gaussian ARFIMA(0, 0.3, 0) process X_t with standard deviation $\sigma_X = \frac{1}{2}$ and of the meta-Gaussian processes e^{X_t} , X_t^2 and $|X_t|$ (cp. Sect. 3.3).

5.1. Asymptotics and uncertainty of the sample mean

The focus of Figure 5.1 is the shape of the sample mean distribution. The combinations of the chosen standard deviation $\sigma_X = \frac{1}{2}$ and the example transformations, however, result in highly different variances of the related sample means. These quantitatively specific differences of uncertainties remain without further consideration here.

Remark 5.4: Figure 5.1 involves two stochastic limits. First, the convergence of the sample mean $\bar{\mu}_N$ as a (normalized) sum of N dependent variates by the Theorems 5.1 and 5.3 as the number $N \rightarrow \infty$. And second, the convergence by the CLT of the empirical distribution to the theoretical distribution of the sample mean as the amount $k \in \mathbb{N}$ of empirical sample means approaches infinity. The deliberate choice of the standard deviation $\sigma_X = \frac{1}{2}$ for the examples in Figure 5.1 brings the calculated sample means closer to the first asymptotic regime. As the k different sample time series are independent, the pointwise rate of the second convergence reads $\frac{1}{\sqrt{k}}$ as $k \rightarrow \infty$ by the CLT (cp. Appx. C).

In two ways, we apply Theorem 5.3 to modeling daily precipitation amounts in Chapter 6. First, Section 6.3.5 deals with the daily mean of precipitation amounts and its statistical uncertainty. Second, rescaling the yearly sample mean to annual totals by the multiplication by 365 days allows for model validation in Section 6.3.6.

5.1.2. Sample mean approach to an effective sample size

Measuring the uncertainty of the sample mean by its variance serves as an approach to quantifying the effect of correlations on statistical quantities. For a finite number $N \in \mathbb{N}$ of *iid* samples with mean $\mu_Y \in \mathbb{R}$ and variance $\sigma_Y^2 \in \mathbb{R}_{>0}$ by basic calculations the mean $\mathbb{E}[\bar{\mu}_N]$ and variance $\sigma_{\bar{\mu}_N}^2$ of the sample mean (5.1) read

$$\mathbb{E}[\bar{\mu}_N] = \mu_Y \quad \text{and} \quad \sigma_{\bar{\mu}_N}^2 = \frac{\sigma_Y^2}{N}. \quad (5.4)$$

The square root $\sigma_{\bar{\mu}_N} = \frac{\sigma_Y}{\sqrt{N}}$ is well known as the *standard error of the mean*. Lemma 5.5 calculates the variance of the sample mean for *stationary* processes.

Lemma 5.5: *Let Y_t be a stationary process with mean $\mu_Y \in \mathbb{R}$, variance $\sigma_Y^2 \in \mathbb{R}_{>0}$ and nonnegative acf ϱ_Y . Then the mean $\mathbb{E}[\bar{\mu}_N]$ and the variance $\sigma_{\bar{\mu}_N}^2$ of the sample mean (5.1) of $N \in \mathbb{N}$ samples of the process Y_t read*

$$\mathbb{E}[\bar{\mu}_N] = \mu_Y \quad \text{and} \quad \sigma_{\bar{\mu}_N}^2 = \frac{\sigma_Y^2}{N} \tau_D(N), \quad (5.5)$$

where

$$\tau_D(N) := 1 + 2 \sum_{k=1}^{N-1} \left(1 - \frac{k}{N}\right) \varrho_Y(k) \geq 1. \quad (5.6)$$

5. Effective sample size of stationary processes

Proof. The linearity of the mean implies $\mathbb{E}[\bar{\mu}_N] = \mu_Y$. By the definition of the acf (2.3) for stationarity processes the variance (5.1) of the sample mean of N samples reads

$$\sigma_{\bar{\mu}_N}^2 = \frac{1}{N^2} \sum_{i,j=1}^N \text{Cov}(Y_i, Y_j) = \frac{\sigma_Y^2}{N^2} \sum_{i,j=1}^N \varrho_Y(|i-j|) \quad (5.7)$$

$$= \frac{\sigma_Y^2}{N^2} \left(N \varrho_Y(0) + 2 \sum_{k:=|i-j|=1}^{N-1} (N-k) \varrho_Y(k) \right). \quad (5.8)$$

Division of equation (5.8) by the number N of samples together with $\varrho_Y(0) = 1$ implies equation (5.5) using definition (5.6). We further have $\tau_D(N) \geq 1$ by the sum in definition (5.6) and the nonnegativity of the acf ϱ_Y . \square

As a direct consequence of equation (5.5) the variance of the sample mean is larger in the presence of positive autocorrelations than for an equal number of iid samples with the same marginal variance (cp. the concept of the *associated independent sequence* in Sect. 7.2). The value $\tau_D(N)$ quantifies this increase of statistical uncertainty for finite sample sizes.

Definition (Finite-size decorrelation time): The value $\tau_D(N)$ defined by equation (5.6) is called the *finite-size decorrelation time* of $N \in \mathbb{N}$ samples of a stationary process.

Remark 5.6: The choice of the term “finite-size decorrelation time” $\tau_D(N)$ is according to the definition of a “decorrelation time” by its asymptotics as the number N of samples increases. If finite, this limit serves as a characteristic time scale (cp. Sect. 5.1.3).

Lemma 5.5 allows for the definition of an effective sample size by comparing the variances of the sample mean in the iid setting (5.4) and in the presence of correlations (5.5).

Definition (Effective sample size): Using the finite-size decorrelation time $\tau_D(N)$, the *effective sample size* of $N \in \mathbb{N}$ samples of a stationary process is defined by

$$N_{\text{eff}} := \frac{N}{\tau_D(N)}. \quad (5.9)$$

Note that by equation (5.6) in the presence of positive correlations the finite-size decorrelation time satisfies $\tau_D(N) > 1$, so that the effective sample size N_{eff} is always smaller than the sample size N . The interpretation of the number N_{eff} as an effective sample size bases on the observation that the uncertainty

$$\sigma_{\bar{\mu}_N}^2 = \frac{\sigma_Y^2}{N_{\text{eff}}} \quad (5.10)$$

of the sample mean (5.5) of N samples of a stationary process coincides with the uncertainty of the sample mean (5.4) of N_{eff} iid samples with the same marginal variance σ_Y^2 .

5.1. Asymptotics and uncertainty of the sample mean

Effective sample sizes for AR(1) and ARFIMA processes follow straightforwardly by plugging in their analytically known acfs (2.12) and (2.22) or (2.23), respectively, into definition (5.9) using equation (5.6).

Example 5.7 (Effective sample size of AR(1) processes): Let X_t be a Gaussian AR(1) process (2.11) with AR parameter $\varphi \in (0, 1)$ and marginal standard deviation $\sigma_X \in \mathbb{R}_{>0}$. Its acf reads $\varrho_X(k) = \varphi^k$ for time lags $k \in \mathbb{N}$ by equation (2.12). For $N \in \mathbb{N}$ samples definition (5.6) gives the finite-size decorrelation time

$$\tau_D(N) = \frac{1 + \varphi + \frac{2\varphi}{N}(\varphi^N - 1)}{1 - \varphi} \quad (5.11)$$

with an effective sample size of

$$N_{\text{eff}} = \frac{1 - \varphi}{\frac{1}{N}(1 + \varphi) + \frac{2\varphi}{N^2}(\varphi^N - 1)} \quad (5.12)$$

by equation (5.9) and variance $\sigma_{\bar{\mu}_N}^2 = \frac{\sigma_X^2}{N} \tau_D(N)$ of the sample mean by equation (5.5). In the white-noise limit of a vanishing AR parameter $\varphi \rightarrow 0$, equation (5.12) implies $N_{\text{eff}} = N$. The Examples 5.10 and 5.12 consider the asymptotics of equations (5.11) and (5.12).

The above definition of an effective sample size as a finite-sample property accounts for an application of this concept to LRC processes, as well, even in the absence of a finite correlation time (2.8).

Example 5.8 (Effective sample size of ARFIMA processes): Let X_t be a Gaussian ARFIMA process (2.11) with LRC parameter $d \in (0, \frac{1}{2})$, acf $\varrho_X(k)$ for time lags $k \in \mathbb{N}$ and marginal standard deviation $\sigma_X \in \mathbb{R}_{>0}$. For ARFIMA(0, d , 0) or ARFIMA(1, d , 0) processes the acf is given by equation (2.22) and (2.23), respectively. Using definition (5.6) then yields the finite-size decorrelation time $\tau_D(N)$ for $N \in \mathbb{N}$ samples. As an example, for ARFIMA(0, d , 0) processes it reads⁵

$$\tau_D(N) = 1 + 2 \cdot \frac{\Gamma(1-d)}{\Gamma(d)} \sum_{k=1}^{N-1} \left(1 - \frac{k}{N}\right) \frac{\Gamma(k+d)}{\Gamma(k-d+1)} \quad (5.13)$$

$$= 1 + \frac{1}{(2d+1)N} \left(\frac{(N^2 - d^2)\Gamma(1-d)\Gamma(N+d)}{d\Gamma(N-d+1)\Gamma(d)} - (2d+1)N + d \right). \quad (5.14)$$

Analogously to Example 5.7, the effective sample size N_{eff} and the variance $\sigma_{\bar{\mu}_N}^2$ of the sample mean can then be determined by equation (5.9) and (5.5), respectively, by plugging in the result (5.14). Note that in the white-noise limit of a vanishing LRC parameter $d \rightarrow 0$, equation (5.14) yields $\tau_D(N) = 1$, so that $N_{\text{eff}} = N$. Example 5.13 considers the asymptotics of equation (5.14).

⁵by the software WOLFRAM MATHEMATICA 11.3.0.0 for rearrangement of symbolic formulae

5. Effective sample size of stationary processes

Figure 5.2 shows the distribution of the sample mean and the effective sample sizes for 400 iid, AR(1) and ARFIMA(0,0.2,0) time series with standard Gaussian marginal distribution. The variances $\bar{\sigma}_{\bar{\mu}_N}^2$ estimated by a ML-fitted normal distribution under the assumption of independence well coincide with the variances $\sigma_{\bar{\mu}_N}^2$ calculated by equation (5.11) and (5.14), respectively, using the effective sample sizes.

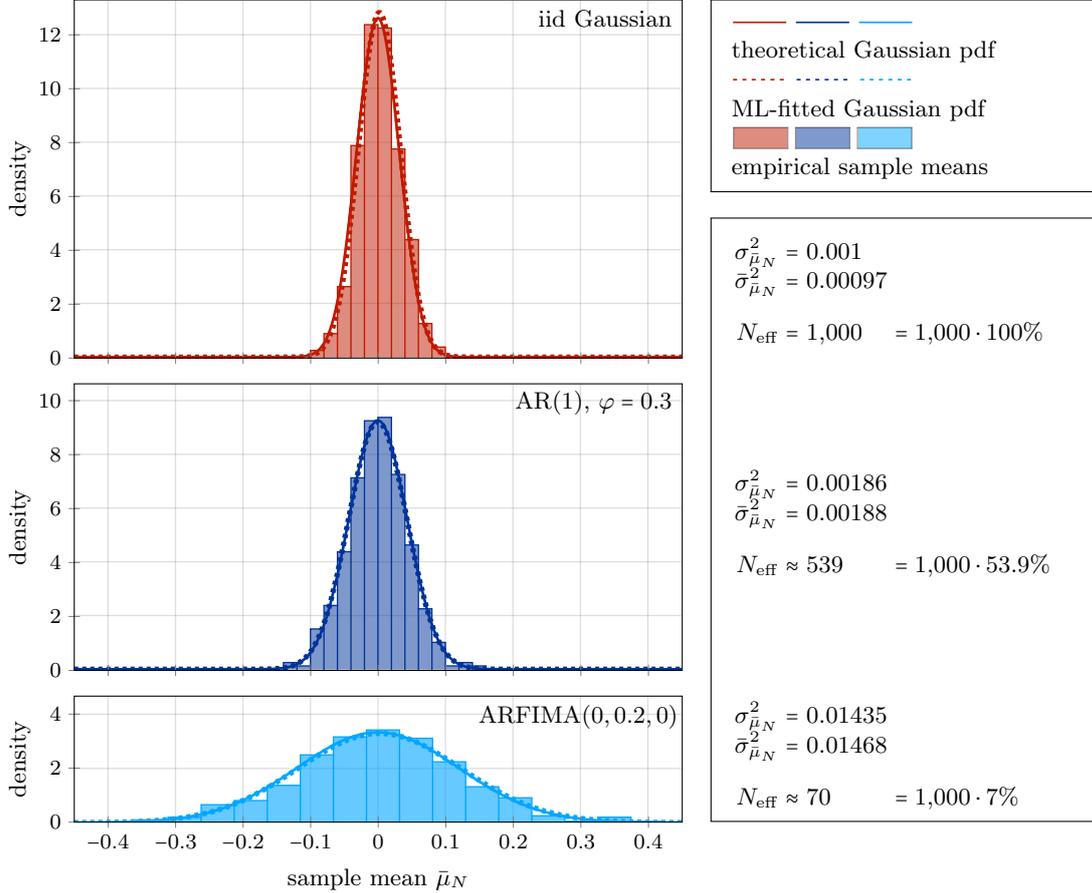


Fig. 5.2 Empirical and theoretical distributions of the sample mean of 400 sample time series of a Gaussian iid, an AR(1) with AR parameter $\varphi = 0.3$ and an ARFIMA(0,0.2,0) process each with standard Gaussian marginal distribution each of *fixed length* $N = 1,000$. The ML-fitted Gaussian densities yield estimated variances $\bar{\sigma}_{\bar{\mu}_N}^2$ of the sample mean $\bar{\mu}_N$. The theoretical Gaussian densities obey the variance $\sigma_{\bar{\mu}_N}^2$ calculated by the effective sample size N_{eff} using equation (5.10).

Positive correlations are capable of increasing the uncertainty of statistical quantities such as the sample mean by orders of magnitudes (see Fig. 5.2). The effective number of samples N_{eff} provides a quantitative impression of the strength of this effect. Additionally, the inverse value $\frac{1}{\tau_D(N)}$ quantifies the influence of correlations relatively by equation (5.9). The two extreme settings of independence and of fully correlated data yield $\tau_D(N) = 1$ and $\tau_D(N) = N$ by definition (5.6) resulting in $N_{\text{eff}} = N$ and $N_{\text{eff}} = 1$, respectively. In Section 5.2 we determine effective sample sizes for meta-Gaussian processes.

5.1.3. Sample mean approach to a characteristic time scale

Accompanying the correlation time (2.8) of stationary processes, according to [174, Chap. 17] we shall define a decorrelation time using the description (5.6) of a finite-size decorrelation time.

Definition (Decorrelation time): The *decorrelation time* τ_D of a stationary process is defined as the limit

$$\tau_D := \lim_{N \rightarrow \infty} \tau_D(N) \quad (5.15)$$

of the finite-size decorrelation time $\tau_D(N)$ as the number $N \in \mathbb{N}$ of samples increases to infinity.

Storch and Zwiers [174] use the equivalent definition $\tau_D := \lim_{N \rightarrow \infty} \frac{N}{N_{\text{eff}}}$ to determine a characteristic time scale if the limit exists. Lemma 5.9 immediately concludes by basic calculations that SRC processes exhibit a finite decorrelation time, other than LRC processes.

Lemma 5.9: *Let Y_t be a stationary process with acf ϱ_Y . If the process Y_t obeys SRC with finite correlation time $\tau < \infty$ according to definition (2.8), then its decorrelation time τ_D defined as the limit (5.15) satisfies*

$$\tau_D = 2\tau - 1. \quad (5.16)$$

If the process Y_t obeys LRC such that $\tau = \infty$, then also $\tau_D = \infty$ and using the finite-size correlation time $\tau(N) := \sum_{k=0}^{N-1} \varrho_Y(k)$, we find the asymptotic equivalence

$$\frac{\tau_D(N)}{\tau(N)} \sim 2 \quad (N \rightarrow \infty). \quad (5.17)$$

Proof. Using the definitions (5.6) of the finite-size decorrelation time $\tau_D(N)$ for $N \in \mathbb{N}$ samples and (2.8) of the correlation time τ yields

$$\tau_D = \lim_{N \rightarrow \infty} 1 + 2 \sum_{k=1}^{N-1} \left(1 - \frac{k}{N}\right) \varrho_Y(k) = 1 + 2 \sum_{k=1}^{\infty} \varrho_Y(k) \quad (5.18)$$

$$= 1 + 2 \left(\sum_{k=0}^{\infty} \varrho_Y(k) - 1 \right) = 1 + 2\tau - 2 = 2\tau - 1 \quad (5.19)$$

for SRC processes. For LRC processes we have an infinite correlation time $\tau = \infty$, so that the decorrelation time $\tau_D = \infty$, the limit in equation (5.18), is infinite as well. In this case, equation (5.19) implies relation (5.17) as the growth rates of the decorrelation time and twice the finite-size correlation time $\tau(N)$ coincide as $N \rightarrow \infty$. \square

5. Effective sample size of stationary processes

The implications of Lemma 5.9 are at least twofold. First, considering the variance of the sample mean (5.1) provides an alternative approach to a characteristic time scale for stationary processes besides the correlation time (2.8). Then equation (5.16) describes the relation between these two time scales if they exist.

Second, Lemma 5.9 illustrates that the concrete values of characteristic time scales depend on the perspective of approaching them. The factor of two in equation (5.18) stems from the symmetry of the acf in the calculation (5.8). Involving statistical quantities different from the sample mean such as the sample variance or even certain cross-correlations [174] would yield relations between characteristic time scales similar but different to equation (5.16).

Example 5.10 (Decorrelation time of AR(1) processes): The decorrelation time (5.20) for AR(1) processes is well known [120,174,177]. Consider an AR(1) process X_t as defined in Example 5.7. By definition (5.15) the finite-size decorrelation time in equation (5.11) implies the decorrelation time

$$\tau_D = \lim_{N \rightarrow \infty} \frac{1 + \varphi + \frac{2\varphi}{N}(\varphi^N - 1)}{1 - \varphi} = \frac{1 + \varphi}{1 - \varphi} \quad (5.20)$$

and confirms Lemma 5.9 by

$$\tau_D = 2\tau - 1 = \frac{2}{1 - \varphi} - 1 = \frac{1 + \varphi}{1 - \varphi}. \quad (5.21)$$

Mind that the correlation time (2.8) of AR(1) processes reads $\tau = \sum_{k=0}^{\infty} \varphi^k = \frac{1}{1 - \varphi}$.

Remark 5.11: For AR(1) processes Massah et al. [120] determined the variance $\sigma_{\bar{\mu}_N}^2$ of the sample mean directly by plugging in the autoregression (2.11) into the definition (5.1) of the variance $\sigma_{\bar{\mu}_N}^2$ including the autocovariances (5.7). As a result, by the decorrelation time (5.20) the asymptotic equivalence

$$\frac{\tau_D}{\tau} = \frac{\frac{1 + \varphi}{1 - \varphi}}{\frac{1}{1 - \varphi}} = 1 + \varphi \rightarrow 2 \quad (\varphi \rightarrow 1) \quad (5.22)$$

implies $\tau_D \sim 2\tau$ as $\varphi \rightarrow 1$ as a special limit case of Lemma 5.9.

Lemma 5.9 generalizes the asymptotic observation of Remark 5.11 for AR(1) processes to general stationary processes. By its very definition the decorrelation time satisfies the asymptotic equivalence (5.17) to twice the correlation time in case of an infinite (de)correlation time, as well. Section 5.1.4 draws inference about the scaling of the effective sample size for large numbers of samples by considering the asymptotics of the decorrelation time.

5.1.4. Asymptotics of the effective sample size

Previous definitions of effective sample sizes [174, 179] strongly relate to the existence of a finite characteristic time scale. Considering the asymptotic behavior of the decorrelation time (5.15), however, allows for comparing the uncertainty of the sample mean of iid and not only SRC but also LRC samples. If the decorrelation time (5.15) is infinite, then its asymptotic behavior is of interest as the speed of the growth of the effective sample size reduced due to the presence of correlations allows for a descriptive comparison to the iid setting. Additionally to the MSD or fluctuation functions (cp. Section 4.4.1), also the asymptotic behavior of the effective sample size serves as an indicator of anomalous scaling.

Example 5.12 (Asymptotics of the effective sample size of AR(1) processes): Consider an AR(1) process as defined in Example 5.7. Such processes exhibit a characteristic time scale by the decorrelation time (5.20). Hence, we find

$$N_{\text{eff}} \sim N \cdot \frac{1-\varphi}{1+\varphi} \quad \text{and} \quad \sigma_{\bar{\mu}_N}^2 \sim \frac{\sigma_X^2}{N} \cdot \frac{1+\varphi}{1-\varphi} = \frac{1}{N(1-\varphi)^2} \quad (N \rightarrow \infty) \quad (5.23)$$

as the number N of samples increases. Asymptotically the effective sample size N_{eff} grows linearly with a slope less than unity, and the variance (5.5) of the sample mean of Gaussian zero-mean standard AR(1) processes with variance $\sigma_X^2 = \frac{1}{1-\varphi^2}$ by equation (2.13) decreases at an $\frac{1}{N}$ -rate just as in the classical CLT setting.

Example 5.13 (Asymptotics of the effective sample size of ARFIMA(0, d , 0) processes): Consider an ARFIMA(0, d , 0) process X_t as defined in Example 5.8. The asymptotics of equation (5.14) describe the growth rate of the finite-size decorrelation time $\tau_D(N)$ as the number N of samples increases. Taking into account leading orders only and applying the property $\Gamma(x + \alpha) \sim \Gamma(x)x^\alpha$ for all $\alpha \in \mathbb{R}$ as $x \rightarrow \infty$ of the gamma function yields

$$\tau_D(N) \sim \frac{\Gamma(1-d)}{d(2d+1)\Gamma(d)} \cdot \frac{N^2}{N} \cdot \frac{\Gamma(N+d)}{\Gamma(N-d+1)} \sim \frac{\Gamma(1-d)}{d(2d+1)\Gamma(d)} \cdot N^{2d} \quad (N \rightarrow \infty). \quad (5.24)$$

Then the effective sample size (5.9) and the variance (5.5) of the sample mean imply (cp. [137, p. 49])

$$N_{\text{eff}} \sim \frac{d(2d+1)\Gamma(d)}{\Gamma(1-d)} N^{1-2d} \quad \text{and} \quad \sigma_{\bar{\mu}_N}^2 \sim \frac{\Gamma(1-d)\sigma_X^2}{d(2d+1)\Gamma(d)} \cdot N^{2d-1} \quad (N \rightarrow \infty), \quad (5.25)$$

where σ_X^2 denotes the marginal variance of the ARFIMA(0, d , 0) process X_t .

5. Effective sample size of stationary processes

Remark 5.14: Euler's reflection formula $\Gamma(1-z)\Gamma(z) = \frac{\pi}{\sin(\pi z)}$ for all $z \in \mathbb{C}$ together with $\Gamma(1) = 1$ implies

$$\frac{d(2d+1)\Gamma(d)}{\Gamma(1-d)} = \frac{(2d+1)}{\Gamma(1-d)^2} \cdot \frac{\pi d}{\sin(\pi d)} \sim \frac{\pi d}{\pi d} = 1 \quad (d \rightarrow 0). \quad (5.26)$$

The variance (5.25) of the sample mean thus coincides with its behavior (5.4) in the white-noise limit as the LRC parameter d vanishes. We find $0.9 < \frac{\Gamma(1-d)}{d(2d+1)\Gamma(d)} \leq 1$ for all $d \in (0, \frac{1}{2})$, so that predominantly the power-law (5.25) asymptotically determines the effective sample size of ARFIMA(0, d , 0) processes. However, this is not true in general (cp. Ex. 5.23).

Equation (5.6) describing the finite-size decorrelation time directly by the acf allows for general inference about the asymptotic behavior of the effective sample size for stationary processes with power-law decaying acf.

Lemma 5.15: *Let Y_t be a stationary process with power-law decaying acf $\varrho_Y(k) \sim ck^{-\gamma}$ as time lags $k \in \mathbb{N}$ approach infinity with a constant $c \in \mathbb{R}_{>0}$ and an exponent $\gamma \in (0, 1)$. Then the asymptotic behavior of the effective sample size N_{eff} of $N \in \mathbb{N}$ samples of the process Y_t reads*

$$N_{\text{eff}} \sim \frac{(1-\gamma)(2-\gamma)}{2c} \cdot N^\gamma \quad (N \rightarrow \infty). \quad (5.27)$$

Proof. We shall approximate the countable sum in definition (5.6) of the finite-size decorrelation time by a continuous integral. Let $\beta \in (-1, 0)$ and $N \in \mathbb{N}$, then asymptotically $\sum_{k=1}^{N-1} k^\beta \sim \int_{k=1}^{N-1} k^\beta dk \sim \frac{N^{\beta+1}}{\beta+1}$ as $N \rightarrow \infty$. Note that these two asymptotic equivalences are only valid in case of divergence of the sum (and the integral), otherwise, sum and integral differ and the limit integral reads $\frac{1}{\beta+1}$. Setting $\beta = -\gamma$, we conclude the leading order

$$\tau_D(N) = 1 + 2 \sum_{k=1}^{N-1} \left(1 - \frac{k}{N}\right) ck^{-\gamma} = 1 + 2c \left(\sum_{k=1}^{N-1} k^{-\gamma} - \frac{1}{N} \sum_{k=1}^{N-1} k^{1-\gamma} \right) \quad (5.28)$$

$$\sim 2c \left(\frac{N^{1-\gamma}}{1-\gamma} - \frac{1}{N} \cdot \frac{N^{2-\gamma}}{2-\gamma} \right) = \frac{2c}{(1-\gamma)(2-\gamma)} \cdot N^{1-\gamma} \quad (N \rightarrow \infty). \quad (5.29)$$

Plugging in the result (5.29) into definition (5.9) yields the asymptotic growth rate (5.27) for the effective sample size. \square

The asymptotic equivalence (5.25) in Example 5.13 calculated for ARFIMA(0, d , 0) processes directly by the asymptotics (5.24) of their finite-size decorrelation time (5.14) is a special case of Lemma 5.15 with $\gamma = 1 - 2d$ and $c = \frac{\Gamma(1-d)}{\Gamma(d)}$. In Section 5.2 we determine these two parameter values for meta-Gaussian processes based on Lemma 5.15.

5.2. Effective sample size for meta-Gaussian processes

The Hermite polynomials introduced in Chapter 3 provide a convenient approach to assessing the sample mean of meta-Gaussian processes. As a remark, alternative approaches to answering related questions such as functional CLTs for non-Gaussian LRC processes apply other basis functions like for example Appell polynomials [57]. Joining the concepts of an effective sample size based on the sample mean (cp. Sect. 5.1.2) and the Hermite expansion of the acf of meta-Gaussian processes (cp. Sect. 3.1) allows for an analytical calculation of their effective sample sizes in Section 5.2.1. Additionally, in Section 5.2.2 the asymptotics of the effective sample size adds to the research area of anomalous scaling (cp. Sect. 4.4.1) by concrete values of previously unknown prefactors in asymptotic relations (cp. Sect. 5.3).

5.2.1. Finite-sample interpretation

The effective sample size provides an assimilable description of the effect of correlations to statistical quantities. Theorem 3.2 on the calculation of the acf of meta-Gaussian processes applied to Lemma 5.5 on the uncertainty of the sample mean for stationary processes directly yields a closed form formula for the effective sample size of meta-Gaussian processes.

Theorem 5.16: *Let $Y_t = g(X_t)$ be a meta-Gaussian process with acf ϱ_Y and with a stationary original Gaussian process X_t with nonnegative acf ϱ_X . Then for $N \in \mathbb{N}$ samples of the process Y_t the effective sample size reads*

$$N_{\text{eff}} = \frac{N}{\tau_D(N)}, \quad \text{where} \quad \tau_D(N) = 1 + \sum_{k=1}^{N-1} \left(1 - \frac{k}{N}\right) \varrho_Y(k) \quad (5.30)$$

denotes the finite-size decorrelation time (5.6) of the process Y_t . Its acf

$$\varrho_Y(k) = \frac{1}{\sigma_Y^2} \sum_{j=1}^{\infty} \frac{\alpha_j^2}{\sigma_X^{2j} j!} \varrho_X(k)^j \quad (5.31)$$

for time lags $k \in \mathbb{N}$ is given by Theorem 3.2. In equation (5.31) α_j denotes the j -th Hermite coefficient (3.8) for indices $j \in \mathbb{N}$ and σ_X^2 and σ_Y^2 the marginal variance of the process X_t and Y_t , respectively.

Proof. The exact formula (5.30) for the effective sample size of meta-Gaussian processes is a direct consequence of plugging in the Hermite expansion (5.31) of the acf ϱ_Y from Theorem 3.2 into definition (5.6) of the finite-size decorrelation time $\tau_D(N)$. \square

5. Effective sample size of stationary processes

Figure 5.3 shows effective sample sizes for the example transformations of Section 3.3 by applying Theorem 5.16. A numerical approximation of effective sample sizes by equation (5.30) shall truncate the infinite Hermite expansion therein. Alternatively, a closed-form representation of the relation $\varrho_X(k) \mapsto \varrho_Y(k)$ for time lags $k \in \mathbb{N}$ might exist as for the above examples given by the results (3.38), (3.39) and (3.40). Just as for the Gaussian toy models in Figure 5.2, for these meta-Gaussian examples the theoretically determined effective sample sizes and thus variances (5.5) of the sample mean well coincide with the ML-fitted Gaussian densities.

The effective sample sizes of meta-Gaussian processes strongly depend on the transformation (3.1). The lognormal example exhibits a significantly reduced effective sample size to less than 5% of the number of correlated lognormal samples other than to about 60 – 70% for the square and absolute-value process for an LRC parameter of $d = 0.2$. Mind that as discussed in Section 5.1, the distribution of the sample mean is not necessarily Gaussian in the case of Hermite ranks larger than unity. This deviation from normality becomes the more prominent the stronger the correlations are, resulting in a slightly worse agreement of the ML-estimated and theoretical Gaussian for $d = 0.3$. The estimated effective sample size nonetheless provides an intuition of the magnitude of increased uncertainty also in such cases of strong LRC. The variance (5.5) of the sample mean remains a measure of uncertainty, independently of whether the asymptotic shape of its distribution is already reached.

A notable result about the effective sample size stems from the inferences drawn about the correlations of meta-Gaussian processes in Section 3.2. The effective sample size of a meta-Gaussian process is always larger than the one of its original Gaussian process.

Corollary 5.17: *Let $Y_t = g(X_t)$ be a meta-Gaussian process with a stationary original Gaussian process X_t with nonnegative acf. Denote by N_{eff}^Y and N_{eff}^X the effective sample size of the process Y_t and X_t , respectively. Then we have*

$$N_{\text{eff}}^Y \geq N_{\text{eff}}^X \quad (5.32)$$

with equality in the estimation (5.32) if the transformation $g \in \mathcal{L}^2$ is the identity or if the process X_t is iid.

Proof. For $N \in \mathbb{N}$ samples denote by $\tau_D^X(N)$ and $\tau_D^Y(N)$ the finite-size decorrelation time of the original process X_t and the process Y_t , respectively. Then the monotonicity (3.16) applied to definition (5.6) implies $\tau_D^X(N) \geq \tau_D^Y(N)$, and thus

$$N_{\text{eff}}^X = \frac{N}{\tau_D^X(N)} \leq \frac{N}{\tau_D^Y(N)} = N_{\text{eff}}^Y. \quad (5.33)$$

If the process Y_t is iid, we find $\tau_D^Y(N) = \tau_D^X(N) = 1$ and thus $N_{\text{eff}}^Y = N_{\text{eff}}^X$. If the function g is the identity, then we have $\varrho_Y = \varrho_X$, so that $\tau_D^Y(N) = \tau_D^X(N)$ in relation (5.33). \square

Figure 5.3 confirms the observation of Corollary 5.17 for the examples of Section 3.3. In Section 6.3.1 we apply the effective sample size approach for meta-Gaussian processes to the assessment of the uncertainty of the daily mean precipitation amount and to the distribution of annual totals.

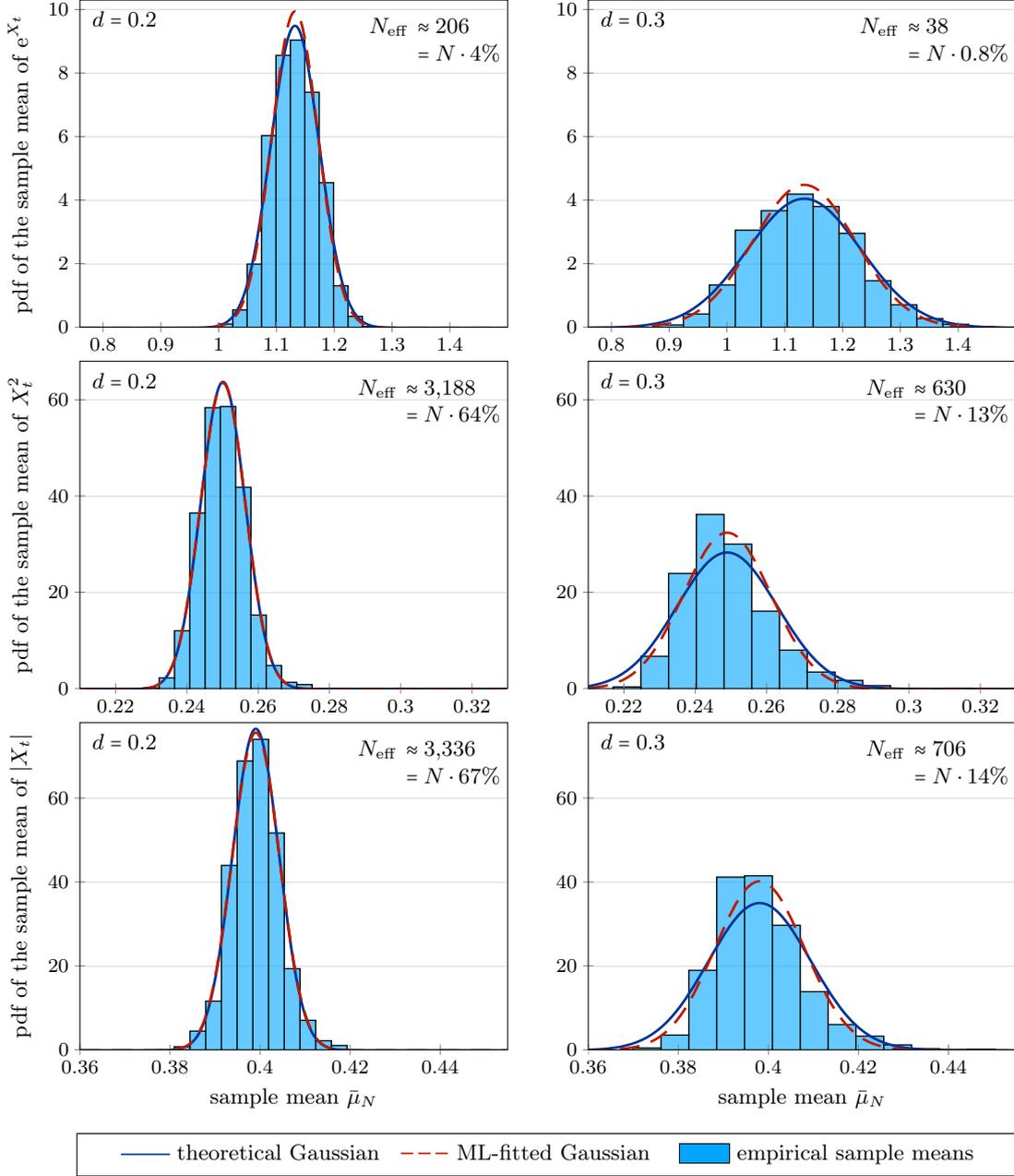


Fig. 5.3 Histograms and ML-fitted Gaussian densities of the empirical sample means $\bar{\mu}_N$ of 1,000 time series each of length $N = 5,000$ of the meta-Gaussian processes e^{X_t} , X_t^2 and $|X_t|$ (cp. Sect. 3.3) for the original Gaussian ARFIMA(0, 0.2, 0) (left) and ARFIMA(0, 0.3, 0) (right) processes X_t with standard deviation $\sigma_X = \frac{1}{2}$ (cp. Fig. 5.1). The effective sample sizes of the original processes X_t read $N_{\text{eff}} \approx 183 = N \cdot 3.7\%$ and $N_{\text{eff}} \approx 33 = N \cdot 0.7\%$ for $d = 0.2$ and $d = 0.3$, respectively.

5. Effective sample size of stationary processes

The inequality (5.32) in Corollary 5.17 states that the effective sample size of a meta-Gaussian process is always larger than the one of its original Gaussian process except in trivial cases of equality (i.e., iid processes or the identity transformation). What appears as a gain of information at the first glance, is a matter of comparing appropriate quantities. The effective sample size compares the uncertainty of the sample mean of stationary correlated data to iid data with the same marginal distribution. It thus provides a description of the strength of the correlations in the data. A pointwise transformations of a Gaussian process generically reduces the strength of correlations by Lemma 3.6, so that the effective sample size of meta-Gaussian processes exceeds the one of the original Gaussian process.

In general, however, this effect does not improve estimations of the sample mean of the original Gaussian process. The sample mean of nonlinearly transformed data does not coincide with the same transformation applied to the sample mean of the original data. Hence, expecting smaller confidence intervals from transforming Gaussian data and inverting back the obtained sample mean is not reasonable. As an example, for the lognormal transformation of standard Gaussian samples we have

$$1 = e^0 = e^{\mathbb{E}[\frac{1}{N} \sum_{i=1}^N X_i]} \neq \mathbb{E} \left[\frac{1}{N} \sum_{i=1}^N e^{X_i} \right] = e^{\frac{1}{2}} \quad (5.34)$$

in case of iid data X_1, \dots, X_N already.

5.2.2. Large sample behavior

By Lemma 5.9 meta-Gaussian SRC processes exhibit a finite decorrelation time and thus a characteristic time scale. Theorem 5.16 directly implies an analytical formula for approximating this quantity.

Corollary 5.18: *Let a transformation $g \in \mathcal{L}^2$ be such that $Y_t = g(X_t)$ is a meta-Gaussian SRC process with marginal variance $\sigma_Y^2 \in \mathbb{R}_{>0}$ with a stationary Gaussian original process X_t with marginal variance $\sigma_X^2 \in \mathbb{R}_{>0}$ and nonnegative acf ϱ_X . Denote by $\tau_D^Y < \infty$ and $\tau_D^X \in \mathbb{R}_{>0} \cup \{\infty\}$ the decorrelation time and by $\tau_Y < \infty$ and $\tau_X \in \mathbb{R}_{>0} \cup \{\infty\}$ the correlation time of the process Y_t and X_t , respectively. Then we have*

$$\tau_D^Y = 1 + \frac{2}{\sigma_Y^2} \sum_{j,k=1}^{\infty} \frac{\alpha_j^2}{\sigma_X^{2j} j!} \varrho_X(k)^j = 2\tau_Y - 1 \quad \text{and} \quad (5.35)$$

$$\tau_D^Y \leq \tau_D^X = 2\tau_X - 1 \quad (5.36)$$

with equality in the estimation (5.36) if the function g is the identity or if the process X_t is iid. In equation (5.35) α_j denotes the j -th Hermite coefficient (3.8) for indices $j \in \mathbb{N}$.

Proof. The equalities on the right-hand side of the relations (5.35) and (5.36) are straightforward translations of Lemma 5.9. For the derivation of the left-hand side equation (5.35) let ϱ_Y be the acf of the process Y_t . Then by the formulae (5.30) and (5.31) the decorrelation time of the process Y_t reads

$$\begin{aligned}\tau_D^Y &= 1 + 2 \lim_{N \rightarrow \infty} \sum_{k=1}^{N-1} \left(1 - \frac{k}{N}\right) \varrho_Y(k) \\ &= 1 + \frac{2}{\sigma_Y^2} \lim_{K, N \rightarrow \infty} \sum_{k=1}^{N-1} \left(1 - \frac{k}{N}\right) \sum_{j=1}^K \frac{\alpha_j^2}{\sigma_X^{2j} j!} \varrho_X(k)^j = 1 + \frac{2}{\sigma_Y^2} \sum_{j,k=1}^{\infty} \frac{\alpha_j^2}{\sigma_X^{2j} j!} \varrho_X(k)^j.\end{aligned}\tag{5.37}$$

Note that the sums and limits are interchangeable due to the finite correlation time of the process Y_t and the convergence of the Hermite series for every time lag $k \in \mathbb{N}$. The inequality (5.36) follows from the monotonicity (3.16) analogously to the proof of Corollary 5.17. \square

If a closed-form representation of the acf ϱ_Y as a function of the original acf ϱ_X is known, than by equation (5.30) the decorrelation time τ_D^Y of a meta-Gaussian SRC process Y_t reads

$$\tau_D^Y = 1 + 2 \sum_{k=1}^{\infty} \varrho_Y(k).\tag{5.38}$$

As an example, for the square of an AR(1) process equation (5.38) yields a closed-form representation of its decorrelation time as calculated in Example 5.19.

Example 5.19 (Effective sample size of transformed AR(1) processes): Let X_t be an AR(1) process with AR parameter $\varphi \in (0, 1)$ and acf $\varrho_X(k) = \varphi^k$ for time lags $k \in \mathbb{N}$. Then the decorrelation times τ_D^Y of the meta-Gaussian example processes of Section 3.3 can be approximated by equation (5.38). For the squared process we obtain the closed-form result

$$\tau_D^Y = 1 + 2 \sum_{k=1}^{\infty} \varphi^{2k} = \frac{1 + \varphi^2}{1 - \varphi^2} \quad \text{and}\tag{5.39}$$

$$\tau_D^Y = 2\tau_Y - 1, \quad \text{where} \quad \tau_Y = \sum_{k=0}^{\infty} \varphi^{2k} = \frac{1}{1 - \varphi^2}\tag{5.40}$$

is the correlation time of the process Y_t since by equation (3.38) we have $\varrho_Y = \varrho_X^2$. Figure 5.4 depicts the asymptotically linear relation (5.42) between the sample sizes and the effective samples sizes of transformed AR(1) processes with different AR parameters. Table 5.1 collects the inverted decorrelation times of the example processes that are the slopes of this linear relation.

5. Effective sample size of stationary processes

The finite-sample computation of effective samples sizes of meta-Gaussian processes given by Theorem 5.16 implies the asymptotics for meta-Gaussian both SRC and LRC processes with a power-law decaying acf. Theorem 5.20 provides direct access to not only the growth exponent of the effective sample size as the number of samples increases but also to the prefactors of these proportionalities.

Theorem 5.20: *Let $Y_t = g(X_t)$ be a meta-Gaussian process with marginal variance $\sigma_Y^2 \in \mathbb{R}_{>0}$, acf ϱ_Y and with a stationary original Gaussian process X_t with marginal variance $\sigma_X^2 \in \mathbb{R}_{>0}$ and nonnegative acf ϱ_X . Denote by $J \in \mathbb{N}_{>0}$ the Hermite rank of the transformation $g \in \mathcal{L}^2$ and by $\alpha_J \in \mathbb{R}_{>0}$ the corresponding J -th Hermite coefficient.*

If the acf of the process X_t satisfies $\varrho_X(k) \sim c_X k^{-\gamma_X}$ as time lags $k \rightarrow \infty$ with a constant $c_X \in \mathbb{R}_{>0}$ and an exponent $\gamma_X \in (0, \frac{1}{J}]$, then the effective sample size N_{eff} of the process Y_t fulfills the asymptotic equivalence

$$N_{\text{eff}} \sim a_Y N^{\gamma_Y} \quad (N \rightarrow \infty), \quad \text{where} \quad (5.41)$$

$$a_Y = \frac{(1 - \gamma_Y)(2 - \gamma_Y)}{2c_Y} \quad \text{with} \quad c_Y = \frac{c_X^J \alpha_J^2}{\sigma_Y^2 \sigma_X^{2J} J!} \quad \text{and} \quad \gamma_Y = \gamma_X J.$$

If the acf of the process X_t satisfies $\varrho_X(k) \propto k^{-\gamma_X}$ as time lags $k \rightarrow \infty$ with an exponent $\gamma_X \in (\frac{1}{J}, 1]$ or decays faster than that, then the effective sample size N_{eff} of the process Y_t asymptotically grows as

$$N_{\text{eff}} \sim a_Y N \quad (N \rightarrow \infty), \quad \text{where} \quad (5.42)$$

$$a_Y = \frac{1}{\tau_D^Y}$$

and τ_D^Y denotes the decorrelation time (5.35) of the process Y_t .

Proof. If the exponent γ_X of the acf of the original Gaussian process satisfies $0 < \gamma_X \leq \frac{1}{J}$, then both the original process X_t and the transformed process Y_t obey LRC (cp. Sect. 3.2). The formulae of the prefactor a_Y and the exponent γ_Y in the descriptions (5.41) are direct translations of Lemma 5.15. Note that the Hermite rank J determines the leading (i.e., slowest) order of the Hermite series (5.31) of the acf $\varrho_Y(k)$ as time lags $k \rightarrow \infty$. It remains to derive the factor c_Y as the prefactor of the asymptotics $\varrho_Y(k) \sim c_Y k^{\gamma_Y}$ as time lags $k \rightarrow \infty$. Its representation (5.41) follows from plugging in the asymptotics of the acf $\varrho_X(k) \sim c_X k^{-\gamma_X}$ as $k \rightarrow \infty$ into the above Hermite series and calculating the prefactor c_Y of the J -th summand $c_Y k^{-\gamma_X J}$.

In the intermediate setting of $\frac{1}{J} < \gamma_X < 1$ or a faster decay of the acf ϱ_X of the original process, the transformed process Y_t is an SRC process. Hence, this process exhibits a finite decorrelation time and Corollary 5.18 applies. Definition (5.9) of the effective sample size implies the asymptotic equivalence (5.42). \square

5.2. Effective sample size for meta-Gaussian processes

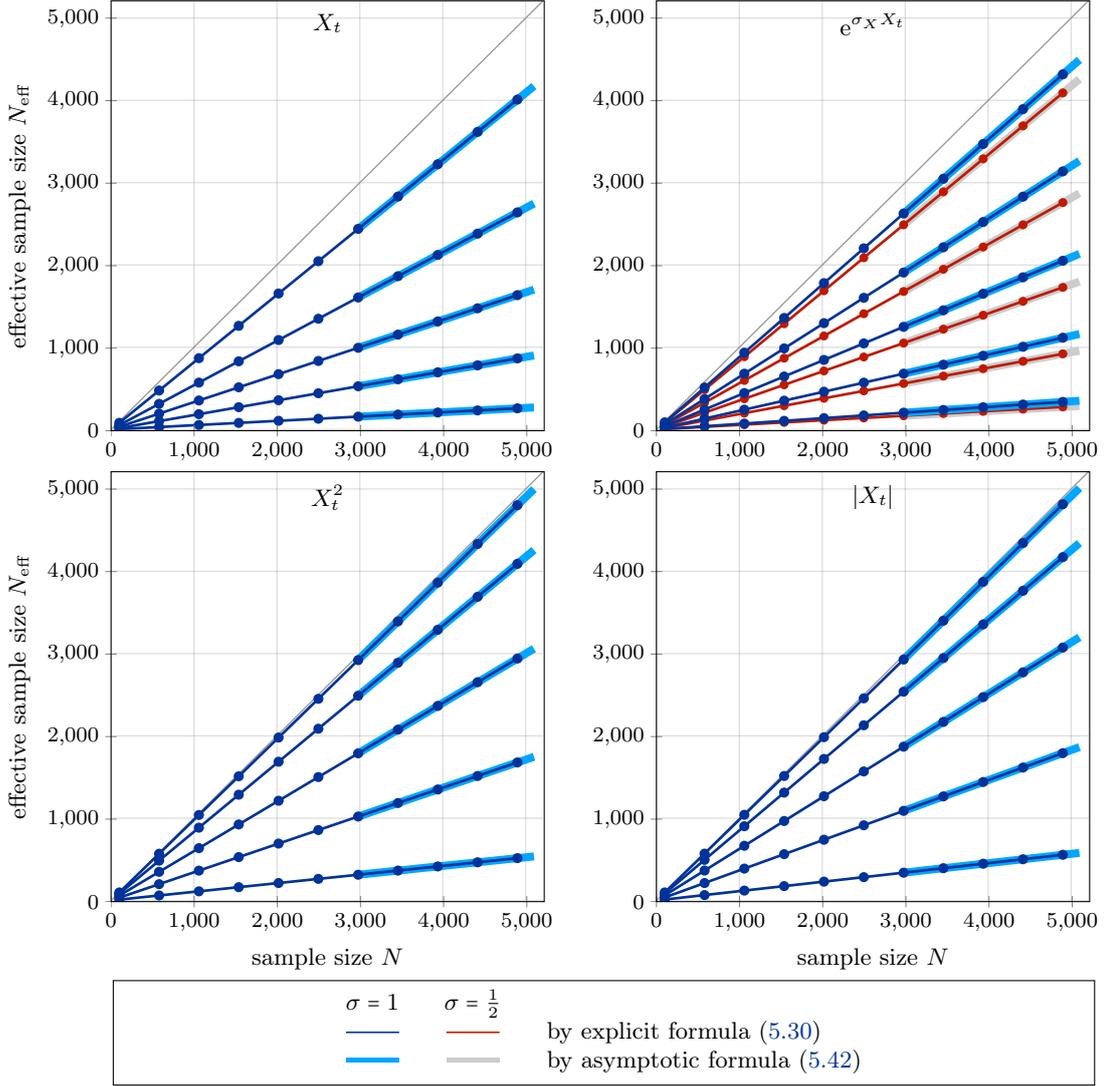


Fig. 5.4 Finite-sample and asymptotic effective sample sizes N_{eff} of the meta-Gaussian processes $e^{\sigma_X X_t}$, X_t^2 and $|X_t|$ (cp. Sect. 3.3) for original Gaussian AR(1) processes X_t with AR parameters $\varphi \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ (from top to bottom each) and marginal standard deviation $\sigma_X \in \{1, \frac{1}{2}\}$. Table 5.1 collects the values of the depicted slopes.

φ	a_X	$a_{e^{\sigma_X X}}$	a_{X^2}	$a_{ X }$
0.1	0.82	0.88 (0.83)	0.98	0.98
0.3	0.54	0.64 (0.56)	0.83	0.85
0.5	0.33	0.42 (0.35)	0.60	0.63
0.7	0.18	0.23 (0.19)	0.34	0.37
0.9	0.05	0.07 (0.06)	0.11	0.11

Table 5.1 Prefactors $a_Y = \frac{1}{\tau_Y^D}$ of the asymptotics $N_{\text{eff}} \sim a_Y N$ as $N \rightarrow \infty$ by equation (5.42) of the effective sample sizes of the example processes $Y_t = g(X_t)$ of Figure 5.4 with standard deviation $\sigma_X = 1$ (and additionally $\sigma_X = \frac{1}{2}$ for the process e^{X_t}).

5. Effective sample size of stationary processes

Figure 5.5 shows the asymptotic behavior of the effective sample size of the example meta-Gaussian processes $Y_t = g(X_t)$ of Section 3.3, namely the lognormal, the square and the absolute value of ARFIMA(0, d , 0) processes, for different LRC parameters $d \in (0, \frac{1}{2})$ and of their original Gaussian process X_t . Table 5.3 collects the related exponents γ_Y and prefactors a_Y according to relation (5.41), which determine the slopes and the vertical position of the straight lines in the double-logarithmic scaling. The prefactors c_X for ARFIMA(0, d , 0) and ARFIMA(1, d , 0) processes are given by relation (2.25) and (2.26), respectively. The slopes of the effective sample size read $\gamma_Y = 1$ in case of LRC parameters $d \in [\frac{1}{2} - \frac{1}{2J}, \frac{1}{2})$ and $\gamma_Y = \gamma_X J$ for LRC parameters $d \in (0, \frac{1}{2} - \frac{1}{2J})$ with the exponent

$$\gamma_X = 1 - 2d \quad (5.43)$$

in the asymptotics (2.20) of the acf of ARFIMA processes and Hermite rank J of the function $g \in \mathcal{L}^2$ (cp. Sect. 3.2). In Figure 5.5 the finite-sample approximation (5.30) of the effective sample size for meta-Gaussian processes coincides with the above asymptotics (5.41) and confirm Theorem 5.20.

Example 5.21 (Effective sample size of transformed ARFIMA(0, d , 0) processes): For the example transformations of Section 3.3 and Figure 5.3 the prefactors a_Y in the asymptotics (5.41) of the effective sample size are analytically known by the closed-form results (3.38), (3.39) and (3.40) for the acf of the transformed processes (cp. Tab. 5.2). The identity and exponential function obey Hermite rank $J = 1$, whereas the square and absolute value exhibit Hermite rank $J = 2$. For LRC parameters $0 < d < \frac{1}{4}$ the transformed processes thus obey SRC and their effective sample size follows the linear asymptotics (5.42).

Y_t	X_t	$e^{\sigma_X X_t}$	X_t^2	$ X_t $
J	1	1	2	2
α_J	$\alpha_1 = \sigma_X^2$	$\alpha_1 = \sigma_X^2 e^{\frac{\sigma_X^2}{2}}$	$\alpha_2 = 2\sigma_X^4$	$\alpha_2 = \sqrt{\frac{2}{\pi}} \sigma_X^3$
σ_Y^2	σ_X^2	$e^{\sigma_X^2} (e^{\sigma_X^2} - 1)$	$2\sigma_X^4$	$\frac{\pi-2}{\pi} \sigma_X^2$
c_Y	c_X	$\frac{c_X \sigma_X^2}{e^{\sigma_X^2} - 1}$	c_X^2	$\frac{c_X^2}{\pi-2}$

Table 5.2 Calculation of the factor c_Y for the prefactors of the asymptotics of the effective sample sizes by formula (5.41) using the results of Section 3.3 for the Hermite coefficients J , the Hermite coefficients α_J (Sect. 3.3.2) and the variance σ_Y^2 of the transformed processes Y_t (Sect. 3.3.1). The resulting factors c_Y coincide with the asymptotic expansions (3.38), (3.39) and (3.40).

5.2. Effective sample size for meta-Gaussian processes

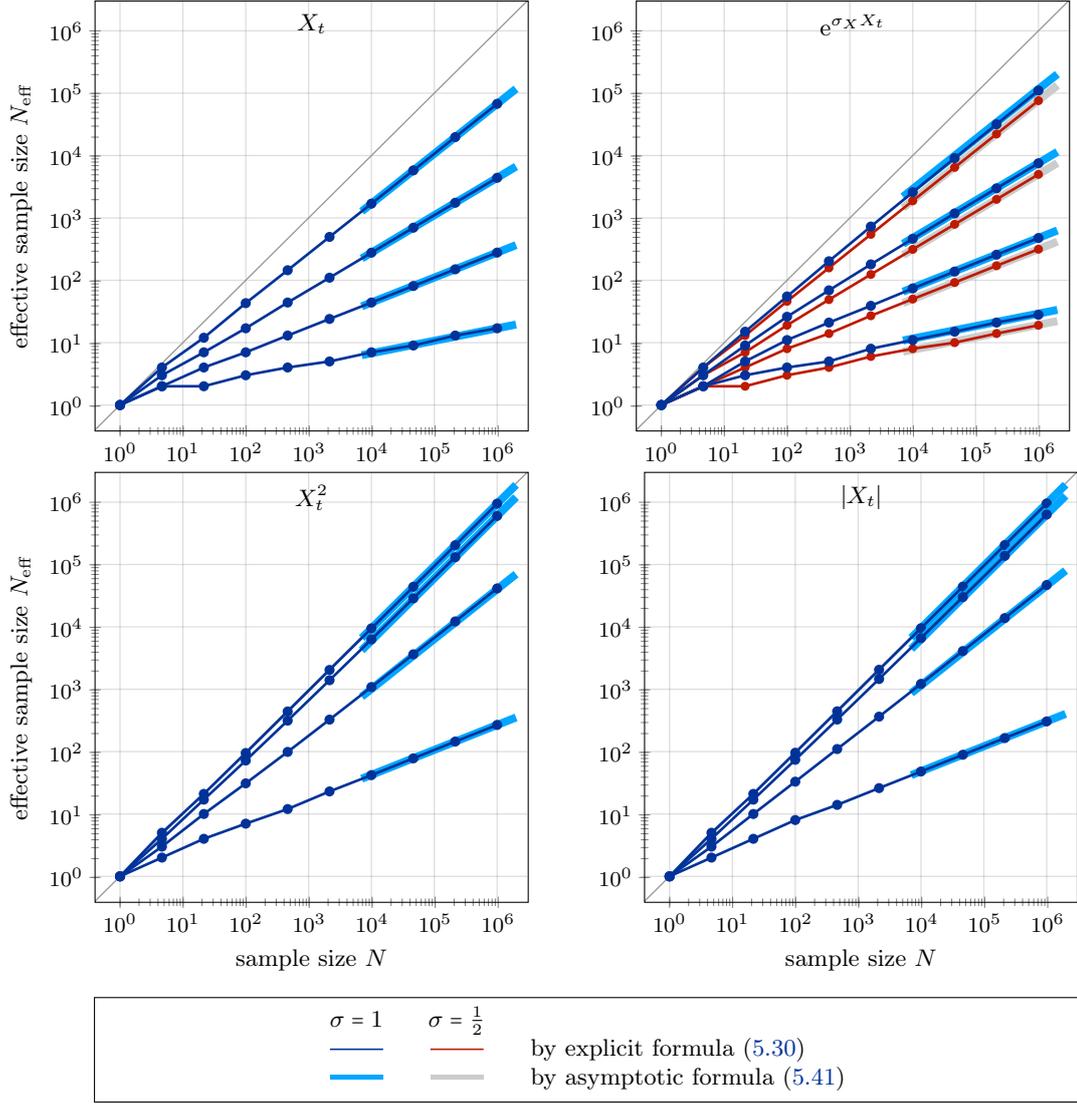


Fig. 5.5 Finite-sample and asymptotic effective sample sizes N_{eff} of the meta-Gaussian processes $e^{\sigma_X X_t}$, X_t^2 and $|X_t|$ (cp. Sect. 3.3) for original Gaussian ARFIMA(0, d , 0) processes X_t with LRC parameters $d \in \{0.1, 0.2, 0.3, 0.4\}$ (from top to bottom each) and marginal standard deviation $\sigma_X = 1$ and $\sigma_X = \frac{1}{2}$. Table 5.3 collects the values of the depicted slopes and prefactors.

d	a_X	$a_{e^{\sigma_X X}}$	$\gamma_X = \gamma_{e^{\sigma_X X}}$ $= 1 - 2d$	a_{X^2}	$a_{ X }$	$\gamma_{X^2} = \gamma_{ X }$ $\in \{1, 2 - 4d\}$
0.1	1.07	1.84 (1.21)	0.8	0.95	0.95	1
0.2	1.10	1.90 (1.25)	0.6	0.63	0.66	1
0.3	1.11	1.90 (1.26)	0.4	0.64	0.73	0.8
0.4	1.07	1.84 (1.22)	0.2	1.06	1.22	0.4

Table 5.3 Prefactors a_Y by relations (5.41) and (5.42) of the asymptotics $N_{\text{eff}} \sim a_Y N^{\gamma_Y}$ as $N \rightarrow \infty$ of the effective sample sizes for the example processes $Y_t = g(X_t)$ of Figure 5.5 with standard deviation $\sigma_X = 1$ and $\sigma_X = \frac{1}{2}$.

5. Effective sample size of stationary processes

Remark 5.22: Note that by equation (3.39), the smaller is the standard deviation σ_X of the original process, the larger is the acf $\varrho_{e^{\sigma_X X}}(k)$ of the meta-Gaussian process $e^{\sigma_X X_t}$ of a stationary standard Gaussian process X_t and marginal standard deviation $\sigma_X \in \mathbb{R}_{>0}$ at time lag $k \in \mathbb{N}$ and the other way round. In other words, we find

$$\varrho_{e^{\sigma_X X}} \rightarrow \varrho_{e^X} \quad (\sigma_X \rightarrow 0) \quad \text{and} \quad \varrho_{e^{\sigma_X X}} \rightarrow 0 \quad (\sigma_X \rightarrow \infty). \quad (5.44)$$

Remarkably, the effective sample size of the lognormal process thus is the larger, the larger is the variance of the original Gaussian process. For the square and absolute-value process, however, the effective sample size does not depend on the variance of the original process by equation (3.38) and (3.40), respectively. Both Figure 5.4 and Figure 5.5 reflect these observations for transformed AR(1) and ARFIMA processes.

Example 5.23 (Range of the prefactor in the asymptotics of N_{eff}): Table 5.3 confirms Remark 5.14 as the prefactor in the asymptotic scaling relations of Theorem 5.20 is close to unity for ARFIMA(0, d , 0) processes. In general, however, the prefactor may be clearly distinct from unity. Figure 5.6 depicts the range of the prefactor a_X and a_{e^X} of an ARFIMA(1, d , 0) process X_t and of the lognormal process e^{X_t} of an ARFIMA(0, d , 0) process X_t with standard deviation σ_X for different AR parameters $\varphi \in (0, 1)$ and standard deviations $\sigma_X \in \mathbb{R}_{>0}$, respectively.

In the limit of a vanishing LRC parameter $d \rightarrow 0$, by equation (5.45) the prefactor a_X for an ARFIMA(1, d , 0) process X_t resembles the characteristic time scale of AR(1) processes in Example 5.12 just as the prefactor in relation (5.25) for an ARFIMA(0, d , 0) for vanishing AR parameters $\varphi \rightarrow 0$. For AR parameters $\varphi \rightarrow 1$ the prefactor a_X is arbitrarily small whereas approaches unity (fully correlated with $N_{\text{eff}} = 1$) as $d \rightarrow \frac{1}{2}$ by

$$\lim_{d \rightarrow 0} a_X = \frac{1 - \varphi}{1 + \varphi} \quad \text{and} \quad \lim_{d \rightarrow 1/2} a_X = 1 \quad \text{for all } \varphi \in (0, 1) \quad \text{and} \quad (5.45)$$

$$\lim_{\varphi \rightarrow 0} a_X = \frac{d(2d+1)\Gamma(d)}{\Gamma(1-d)} \quad \text{and} \quad \lim_{\varphi \rightarrow 1} a_X = 0 \quad \text{for all } d \in (0, 1/2) \quad (5.46)$$

due to the asymptotics of the acf (2.26) of ARFIMA(1, d , 0) processes⁶.

⁶by the software WOLFRAM MATHEMATICA 11.3.0.0 for rearrangement of symbolic formulae

5.2. Effective sample size for meta-Gaussian processes

Let X_t be a standard Gaussian ARFIMA(0, d , 0) process and consider different standard deviations $\sigma_X \in \mathbb{R}_{>0}$. Then the effective sample size of the meta-Gaussian processes $(\sigma_X X_t)^2$ and $|\sigma_X X_t|$ does not depend on the scale σ_X , which cancels out in the acfs (3.38) and (3.40) of the transformed processes and so in their finite-size decorrelation time (5.6). For the lognormal transformation, however, the variance of the original Gaussian processes affects the acf $\varrho_{e^{\sigma_X X}}$ by equation (3.39). The prefactor satisfies⁶

$$\lim_{d \rightarrow 0} a_{e^{\sigma_X X}} = \frac{e^{\sigma_X^2} - 1}{\sigma_X^2} = \lim_{d \rightarrow 1/2} a_{e^{\sigma_X X}} \quad \text{for all } \sigma_X \in \mathbb{R}_{>0} \quad \text{and} \quad (5.47)$$

$$\lim_{\sigma_X \rightarrow 0} a_{e^{\sigma_X X}} = 1 \quad \text{and} \quad \lim_{\sigma_X \rightarrow \infty} a_{e^{\sigma_X X}} = \infty \quad \text{for all } d \in (0, 1/2). \quad (5.48)$$

The increase of the prefactor as the standard deviation increases reflects that the asymptotic scaling (5.41) of the effective sample size becomes valid for the larger numbers of samples the larger the value σ_X is (cp. Rem. 5.22). As a remark, the magnitude of the prefactor appears axially symmetric to the vertical line $d = 0.25$ in Figure 5.6 (right panel) but is not.

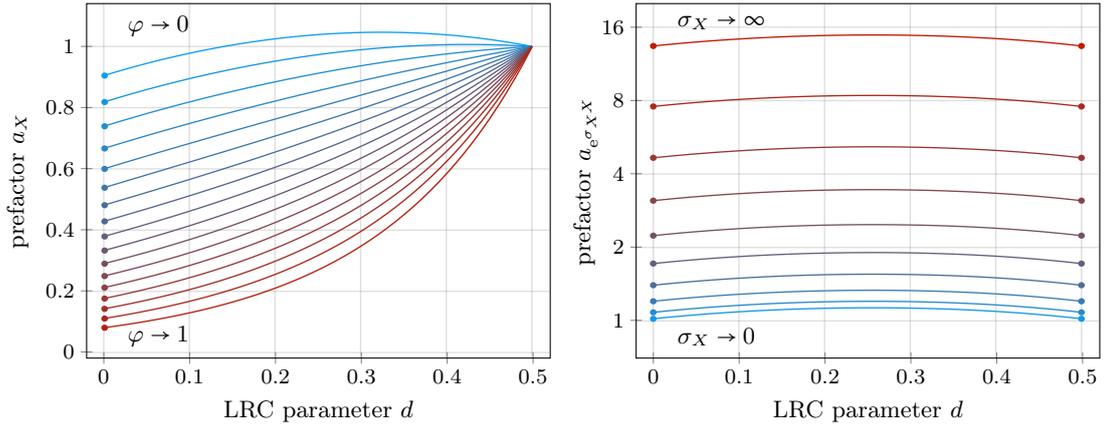


Fig. 5.6 Prefactors a_X and $a_{e^{\sigma_X X}}$ of the asymptotic scaling (5.41) of the effective sample sizes of an ARFIMA(1, d , 0) process X_t (left) and the lognormal process $e^{\sigma_X X_t}$ of an ARFIMA(0, d , 0) process X_t with standard deviation $\sigma_X \in \mathbb{R}_{>0}$ (right) and with LRC parameters $d \in (0, 1/2)$ each. The filled circles mark the limit cases (5.45) and (5.47). The depicted AR parameters φ and standard deviations σ_X are $\varphi = 0.05, 0.1, \dots, 0.85$ and $\sigma_X = 0.2, 0.4, \dots, 2$, respectively.

5.3. Closing remarks and outlook

The Sections 5.3.1, 5.3.2 and 5.3.3 establish relations between the results of Chapter 5 about effective sample sizes and contemporary research in the field of anomalous scaling [42, 119, 120]. The concept of effective sample sizes specifies anomalous scaling for stationary processes such as the asymptotics of both the variance and large deviations of the sample mean.

5.3.1. Interpretation of the effective sample size

The definition of an effective sample size by the variance of the sample mean in Section 5.1.2 allows for different interpretations. It is a matter of perspective whether to consider the variance of the sample mean or the effective sample size itself. By relation (5.10) direct numerical estimations of effective sample sizes [179] are an approach to statistical uncertainty and the other way round. More precisely, given the variance of the sample mean, the corresponding effective sample size follows immediately.

Massah et al. [120] state a closed-form calculation of the variance $\sigma_{\bar{\mu}_N}^2$ of the sample mean of a finite number $N \in \mathbb{N}$ of samples for Gaussian AR(1) processes along with the asymptotic scaling of the related effective sample size (cp. Rem. 5.11). Section 5.1.2 generalizes this idea to stationary processes and provides effective sample sizes as a finite-sample property also for LRC processes by Lemma 5.5 and specifically for meta-Gaussian processes by Theorem 5.16. The finite-size decorrelation time τ_D defined by equation (5.6) quantifies the variance (5.5) of the sample mean independently of whether the asymptotic regimes of the acf or of the distribution of the sample mean (cp. Sect. 5.1.1) is already valid by involving the known correlations for time lags up to the sample size.

5.3.2. Anomalous scaling and the effective sample size

The asymptotic scaling of the effective sample size serves as an alternative approach to anomalous scaling besides diffusion measured by the MSD (cp. Section 4.4.1) or LRC by the Hurst exponent (cp. Section 2.4). For Gaussian ARFIMA(0, d , 0) processes with LRC parameter $d \in (0, \frac{1}{2})$ Massah et al. [119, 120] reason the asymptotic scaling exponents (5.25) as $1-2d$ of the effective sample size and $2d-1$ of the variance of the sample mean, respectively, by the MSD of fBm (cp. Sect. 2.5). It is well known [42] that using definition (4.30), the MSD $\langle B_H^2 \rangle$ of fBm B_H asymptotically scales as $\langle B_H(T)^2 \rangle \propto T^{2d+1}$ as time $T \rightarrow \infty$. Considering the sample mean $\bar{\mu}_N$ of $N \in \mathbb{N}$ zero-mean ARFIMA samples as a discrete-time version of fBm scaled by $\frac{1}{N}$ implies the asymptotic scaling

$$\sigma_{\bar{\mu}_N}^2 = \langle \bar{\mu}_N^2 \rangle = \frac{1}{N^2} \left\langle \left(\sum_{i=1}^N X_i \right)^2 \right\rangle \propto \frac{1}{N^2} \langle B_H(N)^2 \rangle \sim \alpha N^{2d-1} \quad (N \rightarrow \infty) \quad (5.49)$$

of the variance $\sigma_{\bar{\mu}_N}^2$ of the sample mean. This heuristic approach applies to ARFIMA processes and the prefactor α in the scaling relation (5.49) remains unknown. The effective sample size approach of Chapter 5 generalizes the observation (5.49) to stationary processes with power-law decaying acf and provides exact formulae for both the scaling exponents and the prefactors α (Lem. 5.15 and Thm. 5.20).

Using the concept of a finite-size decorrelation time moreover allows for applying effective sample sizes as a finite-sample property of SRC and LRC processes by Lemma 5.5. This approach thus captures and generalizes either concepts studied in [119, 120] for ARFIMA processes, namely effective sample sizes and their asymptotics.

5.3.3. Link to large deviation theory

Large deviation theory deals with the probability of outliers in the estimation of time averages [108, 120]. Using the notion of the sample mean (5.1), large deviation probabilities (LDP) for stationary processes Y_t with mean $\mu_Y \in \mathbb{R}$ are described by

$$\text{LDP}(N, \varepsilon) := \mathbb{P}(|\bar{\mu}_N - \mu_Y| > \varepsilon) \quad (5.50)$$

for $N \in \mathbb{N}$ samples and deviations $\varepsilon \in \mathbb{R}_{>0}$. If the distribution of the sample mean is symmetric, then its LDP simplifies to $\text{LDP}(N, \varepsilon) = 2\mathbb{P}(\bar{\mu}_N - \mu_Y > \varepsilon)$. Otherwise, the scaling of the two probabilities $\mathbb{P}(\bar{\mu}_N - \mu_Y > \varepsilon)$ and $\mathbb{P}(\bar{\mu}_N - \mu_Y < -\varepsilon)$ as $\varepsilon \rightarrow \infty$ might differ. For stationary Gaussian processes and meta-Gaussian processes with unit Hermite rank the sample mean is asymptotically Gaussian with $\bar{\mu}_N \sim \mathbb{N}(\mu_Y, \sigma_{\bar{\mu}_N}^2)$ (cp. Sect. 5.1.1). In the classical setting of iid variates, under mild conditions on the marginal distribution [120] the asymptotic decrease of LDP is known as

$$\text{LDP}(N, \varepsilon) \propto e^{-I(\varepsilon)N} \quad (N \rightarrow \infty) \quad (5.51)$$

with a rate function $I(\varepsilon) > 0$ for $\varepsilon \neq 0$ independent of the number N of samples. For Gaussian samples the rate function reads $I(\varepsilon) = \frac{\varepsilon^2}{2}$. Massah et al. [120] apply the series expansion

$$\text{erfc}(x) = \frac{e^{-x^2}}{x\sqrt{\pi}} \sum_{k=0}^{K-1} (-1)^k \frac{(2k-1)!!}{(2x^2)^k} + R_K(x) = \frac{e^{-x^2}}{x\sqrt{\pi}} + \mathcal{O}\left(\frac{e^{-x^2}}{x^3}\right) \quad (x \rightarrow \infty) \quad (5.52)$$

of the inverse error function erfc involving $K \in \mathbb{N}$ summands. Note that for every fixed value $x \in \mathbb{R}$ the error term $R_K(x) \rightarrow \infty$ as $K \rightarrow \infty$ in the expansion (5.52). For a fixed number K of summands, however, the asymptotic scaling is independent from this number K as $x \rightarrow \infty$.

5. Effective sample size of stationary processes

Then for zero-mean stationary samples with an asymptotically Gaussian sample mean the LDP reads

$$\mathbb{P}(\bar{\mu}_N > \varepsilon) = \frac{1}{2} \operatorname{erfc}\left(\frac{\varepsilon}{\sigma_{\bar{\mu}_N} \sqrt{2}}\right) \sim \frac{\sigma_{\bar{\mu}_N}}{\varepsilon \sqrt{2\pi}} e^{-\frac{\varepsilon^2}{2\sigma_{\bar{\mu}_N}^2}} \quad (\varepsilon \rightarrow \infty), \quad (5.53)$$

where the variance $\sigma_{\bar{\mu}_N}^2$ of the sample mean depends on the variance σ_Y^2 of the samples. By plugging in the variances (5.23) and (5.49) of the sample mean into relation (5.53) Massah et al. derive exponential rates of decay of the LDP (5.51) for AR(1) and ARFIMA processes, respectively. Using the asymptotics of effective sample sizes obtained by Theorem 5.20 and relation (5.10) for the variance $\sigma_{\bar{\mu}_N}^2 = \frac{\sigma_Y^2}{N_{\text{eff}}}$ captures these two results for Gaussian stationary processes and moreover specifies the unknown prefactor in the asymptotic scaling (5.49) for meta-Gaussian processes with unit Hermite rank by

$$\mathbb{P}(\bar{\mu}_N > \varepsilon) \sim \frac{\sigma_Y}{\varepsilon \sqrt{2\pi N_{\text{eff}}}} e^{-\frac{\varepsilon^2}{2\sigma_Y^2} N_{\text{eff}}} \quad (\varepsilon, N \rightarrow \infty). \quad (5.54)$$

For AR(1) and ARFIMA(0, d , 0) processes these LDP thus read

$$\text{AR}(1) \quad \mathbb{P}(\bar{\mu}_N > \varepsilon) \sim \frac{1}{\varepsilon(1-\varphi)\sqrt{2\pi N}} e^{-\frac{\varepsilon^2(1-\varphi)^2}{2} N} \quad (N, \varepsilon \rightarrow \infty) \quad (5.55)$$

$$\text{ARFIMA}(0, d, 0) \quad \mathbb{P}(\bar{\mu}_N > \varepsilon) \sim \frac{\sqrt{\alpha}}{\varepsilon \sqrt{2\pi}} N^{d-\frac{1}{2}} e^{-\frac{\varepsilon^2}{2\alpha} N^{1-2d}} \quad (N, \varepsilon \rightarrow \infty) \quad (5.56)$$

and reproduce the results of [120] along with adding exact values for the proportionality factor α by Theorem 5.20. Example 5.23 visualizes the range of this factor for ARFIMA(0, d , 0) processes and the lognormal process e^{X_t} for ARFIMA(0, d , 0) processes X_t with different marginal variances. Note that the asymptotic scaling (5.56) is a non-standard LDP result due to the exponent $1-2d < 1$. The standard LDP assumption (5.51) depends linearly on the number N of samples with unit exponent and aims at calculating the rate function.

6. Data model for local daily precipitation amounts

Local records of precipitation are typically measured by the height of collected water in rain gauges with a precision of 0.1 mm (cp. Sect. 1.4) and obey two essential properties. First, the distribution of daily precipitation amounts is highly non-Gaussian (see Fig. 1.1). Second, the presence of LRC in such and other geophysical time series have been discussed intensively before (cp. Sect. 1.4).

Our model for mid-latitude daily precipitation amounts established in Section 6.1 is intended to reproduce both the marginal distribution and the temporal correlations of observed mid-latitude precipitation data by applying a stationary meta-Gaussian process (cp. Chap. 3). Five parameters in total suffice to describe the marginal distribution and both SRC and LRC of the empirical data. Section 6.2 describes a procedure for an appropriate estimation of the model parameters. Section 6.3 tests this modeling approach for daily precipitation records [95] of land-based locations in Europe and provides the results for 20 examples. These data sets are a random selection of the databases [43], [49] by [95] and [124] satisfying the criterion that the observed period spans more than 40 years, the data sets are nearly complete and satisfy the condition of weak seasonality (cp. Sect. 6.1.1).

6.1. Model design

Our meta-Gaussian model combines generating correlations by an ARFIMA process (2.5) and reproducing the distribution of daily precipitation amounts by powers of a truncated Gaussian. In earlier studies such truncated-Gaussian-power (tGp) distributions proved to be an appropriate choice for the distribution of daily amounts (cp. Sect. 1.4). Applying a tGp transformation to an ARFIMA process provides direct access to SRC and LRC of the model.

6.1.1. Stationarity and annual cycle

European mid-latitude daily precipitation amounts typically exhibit only a moderate annual cycle and essentially no dominant trend over the years, unlike for example temperature measurements with their strong seasonality and their positive trend in times of climate change. For such precipitation records a stationary model is thus an appropriate choice if only weak nonstationarities are assured. Figure 6.1 shows six recent years of recorded daily precipitation amounts and daily mean temperatures for the station at the city of Potsdam, Germany [49]. The shape of this temperature time series clearly reflects its annual cycle. Figure 6.2 is dedicated to the trend behavior of both entire 125-year time series.

6. Data model for local daily precipitation amounts

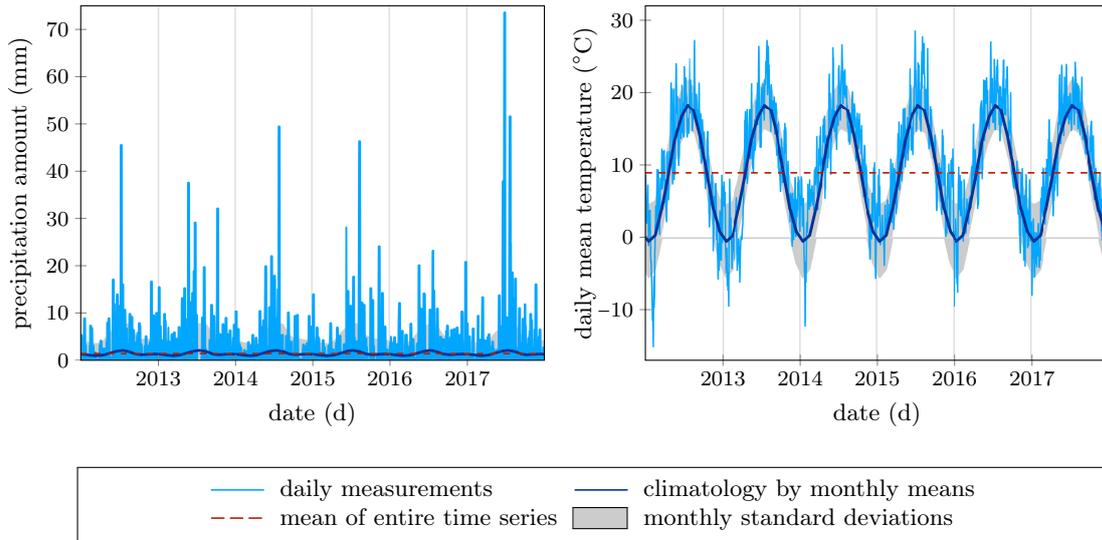


Fig. 6.1 Strength of the seasonality of daily precipitation amounts (left) and daily mean temperatures (right) in station data for the city of Potsdam, Germany. The depicted climatology considers monthly means of each day of a certain month over the entire record from 1893 until 2018.

Considering monthly means and variances of the daily data each of the entire empirical horizon of 126 years allows for preliminarily estimating the strength of seasonality in Figure 6.1 as follows. The ratio of the difference between the maximal and the minimal monthly mean and the standard deviation of the *anomalies* of the time series are a natural measure for this purpose. Removing the seasonality by subtracting the climatology from the data by day gives the anomalies. For Potsdam's temperatures this ratio measures circa 4.7, whereas circa 0.29 for its precipitation amounts. The annual cycle thus determines the daily temperatures about five times more than the fluctuations around this seasonality. The other way round, the variance of Potsdam's daily precipitation affects these amounts about 3.5 times stronger than its annual cycle.

As an example, Figure 6.2 provides a rough impression of the evolution of the annual means of Potsdam's daily precipitation and temperature measurements and the corresponding annual standard deviations over time. During the 20-th century at this station annual mean precipitation exhibited no apparent trend (i), whereas annual mean temperature anomalies rose by about 1.5°C (ii). In contrast, the variance of the annual mean temperature anomalies visually slightly decreased (iv). The variance of the annual mean daily precipitation amounts, however, reached both higher and lower values more often over the years (iii). As a remark, both observations are in compliance with assumed consequences of climate change as a consequence of, for example, increased persistence of macro weather situations (IPCC 2021 [121]).

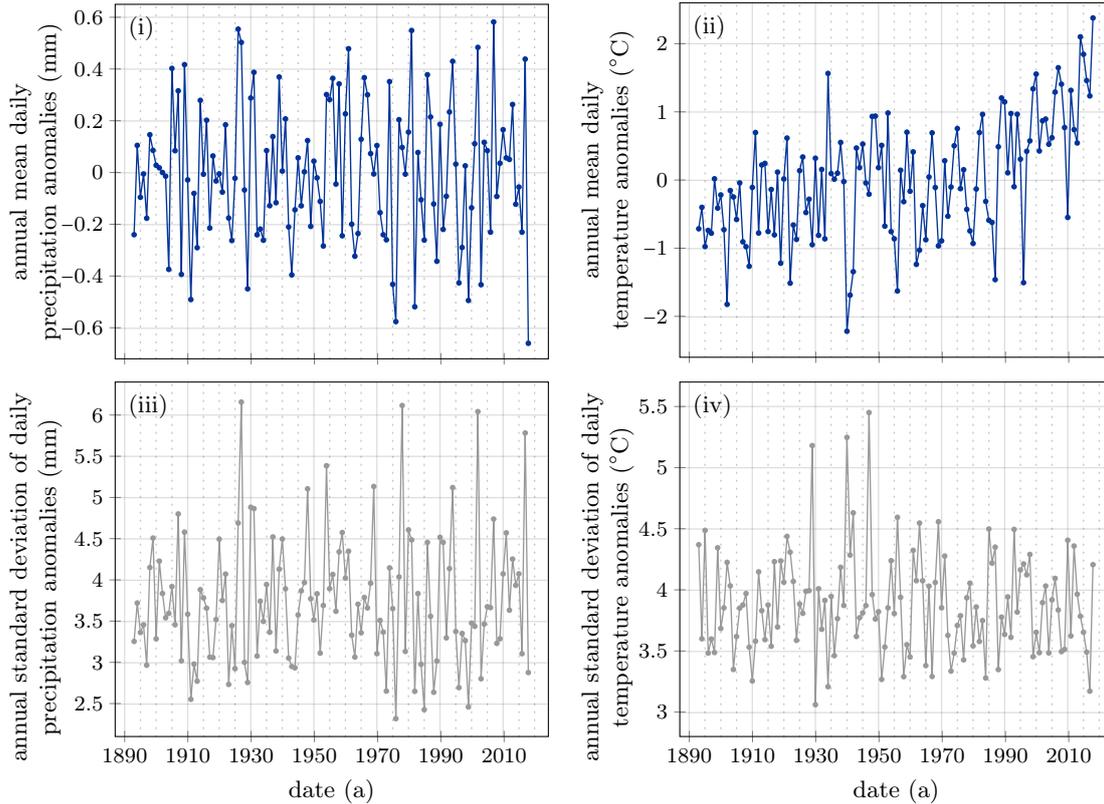


Fig. 6.2 Presence of trends in the annual mean anomalies (i), (ii) and their standard deviation (iii), (iv) for the daily precipitation amounts and daily mean temperatures measured at the Potsdam station, Germany, respectively.

An alternative approach to the assessment of stationarity bases on the scaling of fluctuations (cp. Sect. 4.3). For stationary LRC processes they grow linearly in a double-logarithmic scaling with Hurst exponents less than 1. Deviations from linearity serve as an indicator of prominent (annual) cyclicity and Hurst exponents above 1 of non-stationarity. Regression values describe the goodness of a linear fit and allow for an automatized evaluation of the stationarity of time series by linear regression of method-dependent fluctuation functions.

Neglecting SRC for small time lags, for all the data sets the regression values for the scaling of the R/s -statistics, the fluctuation function of DFA(3) and the wavelet coefficients (cp. Sect. 4.2) are close to 1 and thus induce an asymptotic straight-line behavior each (see Tab. E.3). In compliance with stationarity all corresponding Hurst exponents are larger than $\frac{1}{2}$ and smaller than 1. Further, the detrending of DFA and implicitly contained in wavelet analysis but not in R/s -statistics does not alter much the obtained Hurst exponents. This agreement of the results of all three methods further substantiates the significance of the at most weak nonstationarity of the considered precipitation time series. Nonetheless, their fluctuations exhibit a crossover on shorter time scales from a steeper to their asymptotic linear increase (cp. Fig. 6.8). Modeling of such a behavior requires the additional inclusion of flexibly adjustable SRC in Section 6.2.3.

6. Data model for local daily precipitation amounts

6.1.2. Long-range correlations

By applying R/S -statistics, DFA, and wavelet transforms, in all our precipitation time series we consistently observe LRC with Hurst exponents larger than $\frac{1}{2}$ and smaller than 1 (see Tab. E.3) and with high agreement of these values obtained by the three different methods. As LRC in daily precipitation amounts are typically weak (cp. Sect. 1.2), repeating the analysis for several randomly shuffled versions of the time series helps infer the significance of this finding. For these data sets with removed correlations the estimated Hurst exponents are close to $\frac{1}{2}$ in compliance with their enforced independence (cp. Sect. 2.4) and closer to $\frac{1}{2}$ than the results for the original data. As a result, LRC can be prominent in daily precipitation time series on the observed time horizons (cp. Sect. 4.3). Figure 6.3 shows the estimated Hurst exponents for three of the data sets and 30 randomly shuffled versions each.

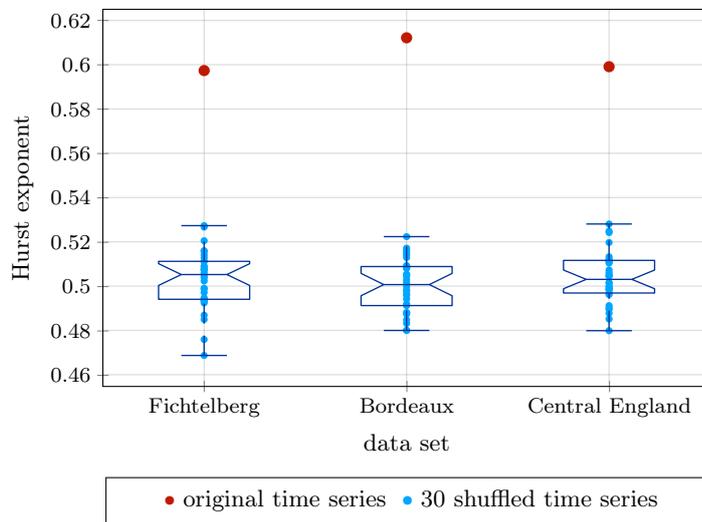


Fig. 6.3 Significance of LRC in measurements of daily precipitation amounts by comparing estimated Hurst exponents of the original time series and 30 randomly shuffled versions each for three example data sets. The horizontal line in the box plot mark the median of the latter Hurst exponents each.

6.1.3. Truncated-Gaussian-power model

Modeling nearly stationary mid-latitude daily precipitation amounts requires including SRC and LRC along with adequately reproducing their highly skewed marginal distribution illustrated by the Sections 6.1.1, 6.1.2 and Figure 1.1, respectively. Let $(X_t)_{t \in \mathbb{N}}$ be a stationary ARFIMA(1, d , 0) process as defined in equation (2.21) with AR parameter $|\varphi| < 1$ describing SRC, an LRC parameter $0 < d < \frac{1}{2}$ and a Gaussian marginal distribution $N(0, \sigma^2)$ with standard deviation $\sigma \in \mathbb{R}_{>0}$. We obtain a meta-Gaussian model

$$Y_t := g(X_t) \quad (6.1)$$

with a tGp marginal distribution by applying the transformation

$$g(x) := (x + \nu)_+^\eta, \quad (6.2)$$

where $x_+ := \max(x, 0)$ projects onto the positive part for arguments $x \in \mathbb{R}$. The transformation (6.2) shifts the zero-mean ARFIMA process X_t to a mean $\nu \in \mathbb{R}$, then truncates the marginal distribution by mapping negative values to zero and rises all values of this process to a power $\eta \in \mathbb{R}_{>0}$. This procedure creates a point mass in zero that accounts for the probability of the absence of precipitation. The zero values in time series of such a model are crucial for the reproduction of intermittency and the study of correlations. The model employs five parameters in total: η , ν and σ for the marginal distribution and φ and d for the short- and long-range correlations, respectively. This modeling approach is semi-parametric due to the clearly identifiable but not directly physical role of the parameters.

Let f_X and F_X denote the Gaussian pdf and cdf, respectively, of the marginal distribution $N(0, \sigma^2)$ of the underlying Gaussian process X_t . Mind that in case of unit variance $\sigma = 1$, we have $f_X = \phi$ and $F_X = \Phi$. By a coordinate transform the pdf f_Y and cdf F_Y of the stationary marginal distribution of the transformed process Y_t read

$$f_Y(y) = \delta(y)F_X(-\nu) + \frac{f_X(\sqrt[\eta]{y} - \nu)}{\eta y^{\frac{\eta-1}{\eta}}} \chi_{(0, \infty)}(y) \quad (6.3)$$

$$F_Y(y) = F_X(\sqrt[\eta]{y} - \nu) \chi_{[0, \infty)}(y), \quad (6.4)$$

where δ is the Dirac delta function and χ_A denotes the indicator function that equals 1 on a set $A \subseteq \mathbb{R}$ and vanishes outside. By equation (6.3) the tail of the pdf of the tGp process Y_t decreases as

$$f_Y(y) \propto e^{-\frac{y^{2/\eta}}{2\sigma^2} y^{\frac{1-\eta}{\eta}}} \quad (y \rightarrow \infty) \quad (6.5)$$

in leading order, so that the stretched exponential factor quickly dominates the shape of the product (6.5).

6. Data model for local daily precipitation amounts

The tGp model reflects the marginal Gaussian distribution if $\eta = 1$, an exponential marginal if $\eta = 2$ and generates a point mass in $y = 1$ as the exponent $\eta \rightarrow 0$. By the leading order (6.5) we find

$$\mathbb{E}[e^{sY_t}] = \int_{\mathbb{R}} e^{sy} f_Y(y) dy \begin{cases} < \infty, & \forall s \in \mathbb{R}_{>0}, \text{ if } 0 < \eta < 2 \\ < \infty, & \forall s < \frac{1}{2\sigma^2}, \text{ if } \eta = 2 \\ = \infty, & \forall s \in \mathbb{R}_{>0}, \text{ if } \eta > 2. \end{cases} \quad (6.6)$$

In the sense of Section 2.3, the tGp model is thus light tailed in case of $0 < \eta \leq 2$ as probabilities decay at least exponentially. For exponents $\eta > 2$ the model is heavy tailed since the tail (6.5) of the model pdf f_Y is asymptotically stretched exponential and decays slower than exponentially. The tGp distribution is, however, not fat tailed as for all exponents $\eta \in \mathbb{R}_{>0}$ all its moments are finite due to

$$\mathbb{E}[Y_t^m] = \int_{\mathbb{R}} y^m f_Y(y) dy = \int_{-\nu}^{\infty} (x + \nu)^{m \cdot \eta} f_X(x) dx < \infty \quad \text{for all } m \in \mathbb{N}, \quad (6.7)$$

in other words, the model pdf decays faster than every power law. Note that in equation (6.7) we have $\int_{\mathbb{R}} y^m \delta(y) F_X(-\nu) dy = 0$ for all power exponents $m \in \mathbb{N}$. The model mean and variance are analytically known by using equation (6.7) for the power exponents $m = 1$ and $m = 2$. As $\eta \rightarrow \infty$, the skewness of the tGp distribution grows since its tail becomes continually heavier together with an increasing peak at $y = 0$. Hence, at the same time both the probabilities of very small and very large values increase.

The parameter ν and the underlying variance σ^2 not only determine the probability $F_X(-\nu)$ of the absence of precipitation but also influence the location and shape of the tail of the model pdf (6.3). The exponent η , however, adjusts the tail of the distribution only.

6.2. Model estimation

Proper modeling of measurement data not only depends on the model choice but also on appropriate parameter estimation. Different estimation approaches emphasize, neglect or assume different properties of the empirical samples. The aim of a particular model determines the choice of the modeling procedure. Note that the marginal distribution of mid-latitude daily precipitation is highly skewed with a huge point mass at zero precipitation (see Fig. 1.1).

A day is commonly considered as *rainy* if its recorded amount exceeds 0.1 mm [104] reflecting the measurement precision (cp. Sect. 1.4). Zero measurements thus tend to overestimate the occurrence of dry days by capturing all amounts in the range of 0 mm to 0.1 mm. Stochastic parameter estimation nevertheless typically [20] focusses on precisely representing the probability of zero precipitation. The functional form of a tGp distribution in our model, however, particularly aims at properly representing extremes in the data. Section 6.2.1 alternatively treats these empirical properties by prioritizing the mean daily precipitation and extreme events over the probability of zero precipitation.

Applying a meta-Gaussian approach additionally aims at appropriately reproducing the correlations and fluctuations in precipitation data. The tGp transformation obeys Hermite rank 1 by Remark 3.5, so that the estimated Hurst exponent of the empirical data can be used directly to adapt LRC in the underlying ARFIMA(1, d , 0) model in Section 6.2.2. What remains is adjusting SRC through the AR part of the ARFIMA model in Section 6.2.3. Section 6.2.4 assembles the individual steps to a holistic modeling procedure.

6.2.1. Tail-oriented marginal distribution

Precipitation amounts span several magnitudes from less than one up to low three-digit numbers of liters per square meter a day (cp. Sect. 1.4). These different scales are unlikely to be modeled equally well by a single tGp distribution. In terms of risk assessment, quantities such as the mean precipitation amount, the occurrence and duration of droughts or extremal events are of particular interest. For this purpose precise values of *very little* precipitation are less important than approximately capturing the occurrence of such situations. Having in mind that historical records carry uncertainty in particular for small values anyway (cp. Sect. 1.4), we aim at properly reproducing very large and mean daily amounts rather than very small ones.

Maximum likelihood (ML) estimation or the method of moments (MM) are typical techniques for the adjustment of parameters in stochastic models. ML fitting focuses on regions of the distribution with high probability, whereas the MM aims at matching the mean and variance of the empirical data along with the exact probability of nonzero precipitation [20]. The tail of a distribution, however, is naturally only rarely sampled with low impact on such estimators. Hence, very small amounts are emphasized with the cost of a worse representation of the tail of the distribution, so that typically high-frequency amounts are represented well with deviations in low-frequency amounts. To emphasize the tail more than the small values of high probability we apply an alternative procedure.

6. Data model for local daily precipitation amounts

Accepting that our model might be slightly less accurate for very small amounts, our approach is dedicated to modeling accurately precipitation that exceeds about 4 mm a day. An accumulated amount of such magnitude at one day lies in the lower range of actually noticeable precipitation (cp. Sect. 1.4). Section 6.3.1 contains further comments on the choice of this arbitrary threshold. For mid-latitude daily precipitation, however, about 75 to 85 percent of the daily records are smaller than 4 mm (cp. Tab. E.1), so that we face the issue of modeling statistics while allowing for deviations in the probabilities for the majority of the measurements.

Inspired by a generalized Kolmogorov-Smirnov test [118] we adjust the parameters η , ν and σ of the tGp model described in Section 6.1.3 by a least-square fit of the model survival function $1 - F_Y$ to the empirical survival function $1 - \bar{F}$ in semi-logarithmic scaling, where F_Y is the model cdf (6.4) and \bar{F} is the empirical distribution function (6.8) defined as follows.

Definition (empirical distribution function (ecdf) [37, cp. Def. 2.4]): Let Y_1, \dots, Y_N be $N \in \mathbb{N}$ random variates with common cdf F_Y . Then for values $y \in \mathbb{R}$ the *empirical (cumulative) distribution function* \bar{F} is defined as

$$\bar{F}(y) := \frac{1}{N+1} \sum_{i=1}^N \chi_{\{Y_i \leq y\}}. \quad (6.8)$$

Note that the ecdf employs order statistics and is a random variable itself.

Remark 6.1: Alternative definitions of the ecdf use a division by N instead of $N+1$ in equation (6.8). Makkonen [109] provides a profound elaboration on the preference of the latter normalization in regard of properly estimating extremes. Independently from the distribution the probability of not exceeding the i -th value of N ordered observations reads $\frac{i}{N+1}$ for iid variates. The ecdf should hence coincide with these values at the order-ranked sample data. This reasoning also bases the choice of plotting positions for q-q and p-p plots in Section 6.3.1.

The above approach discriminates high probabilities for small amounts and highlights low probabilities for large amounts in the tail. As a result, deviations might occur in the estimated probability of zero precipitation. Including an additional parameter for this quantity could eliminate such modeling errors. As argued above and in Section 1.4 these errors are negligible for our purpose and we abstain from another parameter for the sake of parsimony. Due to the measurement precision, the distribution of amounts less than 0.1 mm remains unknown. The zero measurements carry their entire probability and are thus over-represented and shall not be included into the least-square estimate.

Remark 6.2: To match the tail of the model cdf (6.4) to the empirical data also certain quantiles $q \in \mathbb{R}_{\geq 0}$ with probability $p \in (0, 1)$ close to unity can be fixed by numerically solving

$$\nu \stackrel{!}{=} \sqrt[q]{q} - F_X^{-1}(p) \quad (6.9)$$

due to the equality

$$p = F_Y(q) = F_X(\sqrt[q]{q} - \nu). \quad (6.10)$$

Whether reducing the number of parameters by claiming such specific quantities for the model is feasible depends on the existence of parameters satisfying equalities as above.

Section 6.3.1 applies the above estimation of the model parameters for the marginal distribution to measurement data for 20 European locations (cp. Tab. E.1).

6.2.2. Long-range correlations by detrended fluctuation analysis

Since the tGp transformation (6.2) has Hermite rank 1 (cp. Rem. 3.5), the asymptotic power-law decay (2.20) of the acf of the underlying ARFIMA(1, d , 0) process X_t yields

$$\varrho_Y(k) \propto k^{-\gamma} = k^{2d-1} \quad (k \rightarrow \infty) \quad (6.11)$$

for the acf ϱ_Y of the tGp process Y_t for time lags $k \in \mathbb{N}$ by the power law (3.15). The Hurst exponent α of the data estimated by the methods introduced in Section 4.2 provides the LRC parameter d of the process X_t . Based on relation (6.11) we obtain this parameter value as

$$d \stackrel{!}{=} \alpha - \frac{1}{2} \quad (6.12)$$

by applying equation (4.11) and generalizing equation (4.12) to meta-Gaussian processes with Hermite rank 1. Section 6.3.2 shows the results of R/S -analysis, DFA(3) and wavelet transform and the estimated model parameters for 20 example data sets (cp. Tab. E.3).

6.2.3. Short-range correlations by conditional probabilities

Violating the linear long-term scaling of the strength (4.3) of fluctuations for small window sizes indicates the presence of prominent SRC in the data as visualized in Figure 4.10 (cp. Sect. 4.3).

The identification of an appropriate AR parameter φ in equation (2.21) for the underlying ARFIMA(1, d , 0) process X_t from the data is not straightforward though. Generating the model process Y_t , the tGp transformation (6.2) changes the SRC albeit not the power-law exponent of the LRC of the process X_t . The Hermite polynomial approach indeed provides a relation between the two acfs also for small time lags by the series expansion (3.9). There is, however, no closed form of the acf (3.9) of the transformed process, so that it cannot be inverted easily. Our approach below for data with non-Gaussian, strongly asymmetric marginal distributions turns away from the Hermite series and further differs from common earlier techniques [27] for Gaussian SRC models (cp. Sect. 1.4). We gain insight into the short-range dependencies in our daily precipitation data by exploring conditional probabilities instead.

Definition (conditional follow-up probability (cfp)): Let $(Y_t)_{t \in \mathbb{N}}$ be a stationary process. Then the *conditional probability* of $c \in \mathbb{R}_{\geq 0}$ follow-up $k \in \mathbb{N}$ days after is defined as

$$p_c(k) := \mathbb{P}(Y_t > c \mid Y_{t-k} > c). \quad (6.13)$$

In the setting of the process Y_t in definition (6.13) describing a daily precipitation time series, the cfp $p_c(k)$ describes the conditional probability of the occurrence of a day with an accumulated precipitation amount larger than c millimeter k days after a day of such kind. A natural estimator of the cfp $p_c(k)$ for a time series of length $N \in \mathbb{N}$ and time lag $k \in \{0, \dots, N-1\}$ is given by

$$\bar{p}_c(k) := \frac{|\{(Y_t, Y_{t-k}) \mid Y_t, Y_{t-k} > c\}|}{|\{(Y_t, Y_{t-k}) \mid Y_{t-k} > c\}|}. \quad (6.14)$$

We estimate the AR parameter φ by equating the empirical cfp (6.14) and the respective cfp (6.13) of the model for time lag $k = 1$ and a daily precipitation amount of more than $c = 4$ mm (cp. Sect. 6.2.1). For that purpose we numerically solve the equation

$$\bar{p}_4(1) - p_4(1) \stackrel{!}{=} 0 \quad (6.15)$$

for the parameter φ by applying an optimization algorithm to obtain as much agreement among these conditional probabilities as possible. The estimator (6.14) determines the empirical cfp $\bar{p}_4(1)$ directly from the data by dividing the number of pairs of consecutive days with either amount larger than 4 mm by the overall number of days with an accumulated amount larger than 4 mm neglecting the last sample of the time series by following the definition of conditional probabilities.

Let η, ν and σ be the tGp parameters and d the model LRC parameter estimated according to Section 6.2.1 and Section 6.2.2, respectively. Then for time lags k the model cfp (6.13) is analytically known by

$$p_c(k) = \frac{\iint_{\mathbb{R}^2} f_{(X_t, X_{t-k})}(x, y) d(x, y)}{\int_{\mathbb{R}} f_X(x) dx}, \quad (6.16)$$

where $f_{(X_t, X_{t-k})}$ denotes the joint pdf of the two variates X_t and X_{t-k} , which follow a zero-mean bivariate Gaussian distribution $N(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma)$ with covariance matrix

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \varrho_X(1) \\ \varrho_X(1) & 1 \end{pmatrix} \quad (6.17)$$

and lag-1 autocorrelation

$$\varrho_X(1) = \frac{(1 + \varphi^2) \cdot {}_2F_1(1, d, 1 - d; \varphi) - 1}{\varphi ({}_2F_1(1, d, 1 - d; \varphi) - 1)} \quad (6.18)$$

of the process X_t by equation (2.24). Since for given parameters η, ν, σ and d by the covariance matrix (6.17) the cfp (6.16) depends on the AR parameter φ only, the solution to equation (6.15) serves as an estimator of the SRC parameter φ . Section 6.3.3 exemplifies estimated AR parameters for 20 example data sets (cp. Tab. E.1).

6.2.4. Step-by-step modeling procedure

Assembling the Sections 6.2.1, 6.2.2 and 6.2.3 yields a ready-to-use procedure for the application of our tGp model to precipitation records. Section 6.3 provides results for real world measurement data.

Our *algorithm for modeling mid-latitude daily precipitation* reads:

- (I) Estimation of the parameters η, ν and σ of the tGp distribution in virtue of the distribution (6.3) by a least-square adjustment of the model survival function to the empirical survival function
- (II) Estimation of the LRC parameter $d = \alpha - \frac{1}{2}$ with Hurst exponent α in the asymptotics of the fluctuation function (4.3) by applying *R/S* analysis, DFA or a wavelet analysis to the empirical data
- (III) Estimation of the SRC parameter φ in equation (6.18) by the cfp in equation (6.15) given the estimated values of the parameters η, ν, σ and d
- (IV) Synthesis of model time series by the generation of an ARFIMA(1, d , 0) time series with variance σ^2 and AR parameter φ and transformation of these series by the tGp transformation (6.2) with parameters ν and η

6. Data model for local daily precipitation amounts

Section 2.6 describes the synthesis of ARFIMA time series for given SRC and LRC parameters. The variance σ_ε^2 of the input white noise can be calculated by the identity

$$\frac{\sigma^2}{\sigma_\varepsilon^2} = \frac{\Gamma(1-2d)}{\Gamma(1-d)^2} \cdot \frac{{}_2F_1(1, 1+d, 1-d; \varphi)}{1+\varphi} \quad (6.19)$$

using the estimated values σ^2 , d and φ . The right hand side of equation (6.19) equals the variance $\varrho_X(0)$ for $\sigma_\varepsilon^2 = 1$ by plugging in the time lag $k = 0$ into the acf (2.23) of the ARFIMA(1, d , 0) process X_t . Note that for an ARFIMA(0, d , 0) process with AR parameter $\varphi = 0$ the equation (6.19) reduces to $\frac{\sigma^2}{\sigma_\varepsilon^2} = \frac{\Gamma(1-2d)}{\Gamma(1-d)^2}$.

6.3. Model validation

In the modeling of geophysical data, model validation typically bases on comparing empirical properties of particular interest to their synthetic pendants generated by the model [20, 104]. Aiming at the generation of realistic precipitation time series as an input of weather generators and at fluctuations and extremes for risk assessment (cp. Chap. 1), we focus on

- the distribution of daily amounts (Sect. 6.3.1),
- LRC (Sect. 6.3.2) and SRC (Sect. 6.3.3) among daily amounts,
- the occurrence and duration of wet and dry spells (Sect. 6.3.4),
- the mean daily amount (Sect. 6.3.5),
- the distribution of annual totals (Sect. 6.3.6) and
- the distribution of annual daily maxima (Sect. 6.3.7).

Table 6.1 lists three data sets, for which the Sections 6.3.1 to 6.3.7 visualize the validation of our modeling procedure established in Section 6.2.4. The data sets (a) and (b) contain station data, whereas the data set (c) provides regional records accumulated over several stations. Graphical presentations of the results are similar for all the 20 data sets in Table E.1, so that based on their statistics in Table E.2 any of these examples would illustrate well our modeling approach and so do the three chosen ones.

	location of data set	country	measurement period
(a)	Fichtelberg	Germany	1916 - 2018
(b)	Bordeaux	France	1946 - 2018
(c)	Central England	England	1931 - 2018

Table 6.1 List of example data sets of daily precipitation amounts for model validation.

6.3.1. Daily amounts

By applying the tail-oriented parameter estimation introduced in Section 6.2.1 both the model survival function (sf) and pdf properly capture the distribution of large amounts (see Fig. 6.4). The semi-logarithmic scaling emphasizes the heavy tail of the marginal distribution of mid-latitude daily precipitation amounts. Visualizing the cdf highlights the steepness of the distribution for small amounts (see Fig. 6.5).

Remark 6.3: Accounting for the steepness of the distribution close to zero, we shall map model data below 0.1 to zero in consistency with the measurement precision. Considering only amounts above a prescribed threshold of $r \in \mathbb{R}_{\geq 0}$ millimeters as rainy changes the model probability of zero precipitation to $F_X(\sqrt[r]{r} - \nu)$ and the model mean μ_Y and variance σ_Y^2 from equation (6.7) to

$$\begin{aligned}\mu_Y = \mathbb{E}[Y_t] &= \int_r^\infty y f_Y(y) dy \quad \text{and} \\ \sigma_Y^2 = \text{Var}(Y_t) &= \int_r^\infty y^2 f_Y(y) dy - \mathbb{E}[Y_t]^2.\end{aligned}\tag{6.20}$$

The results presented below and in Appendix E apply the threshold $r = 0.1$ millimeters in equations (6.20) when estimating the model parameters. Table 6.2 collects the adjusted model parameters for the three example data sets of Table 6.1.

data set	η	ν	σ	d	φ
(a)	4.083	1.048	0.521	0.096	0.284
(b)	2.330	0.430	1.569	0.111	0.356
(c)	2.967	0.843	0.696	0.099	0.222

Table 6.2 Parameter values of the tGp model adjusted to the data sets in Table 6.1.

The empirical pdfs in Figure 6.4 base on centers and normalized heights of the bars in a histogram of the daily measurements. As the area of all bars sums up to 1, the height of a bin approximates the pdf at the center of that bin and is given by the number of samples within divided by the overall number of samples and by the bin width. Minimizing the least-square distance between the logarithm of the empirical and the model sf is, however, a more robust approach. A least-square approximation by an empirical pdf in semi-logarithmic scaling is highly inaccurate, in particular, for extreme events, as bars of zero height exhibit the value of negative infinity. Empirical pdfs thus crucially depend on the choice of the location and size of the histogram bins.

6. Data model for local daily precipitation amounts

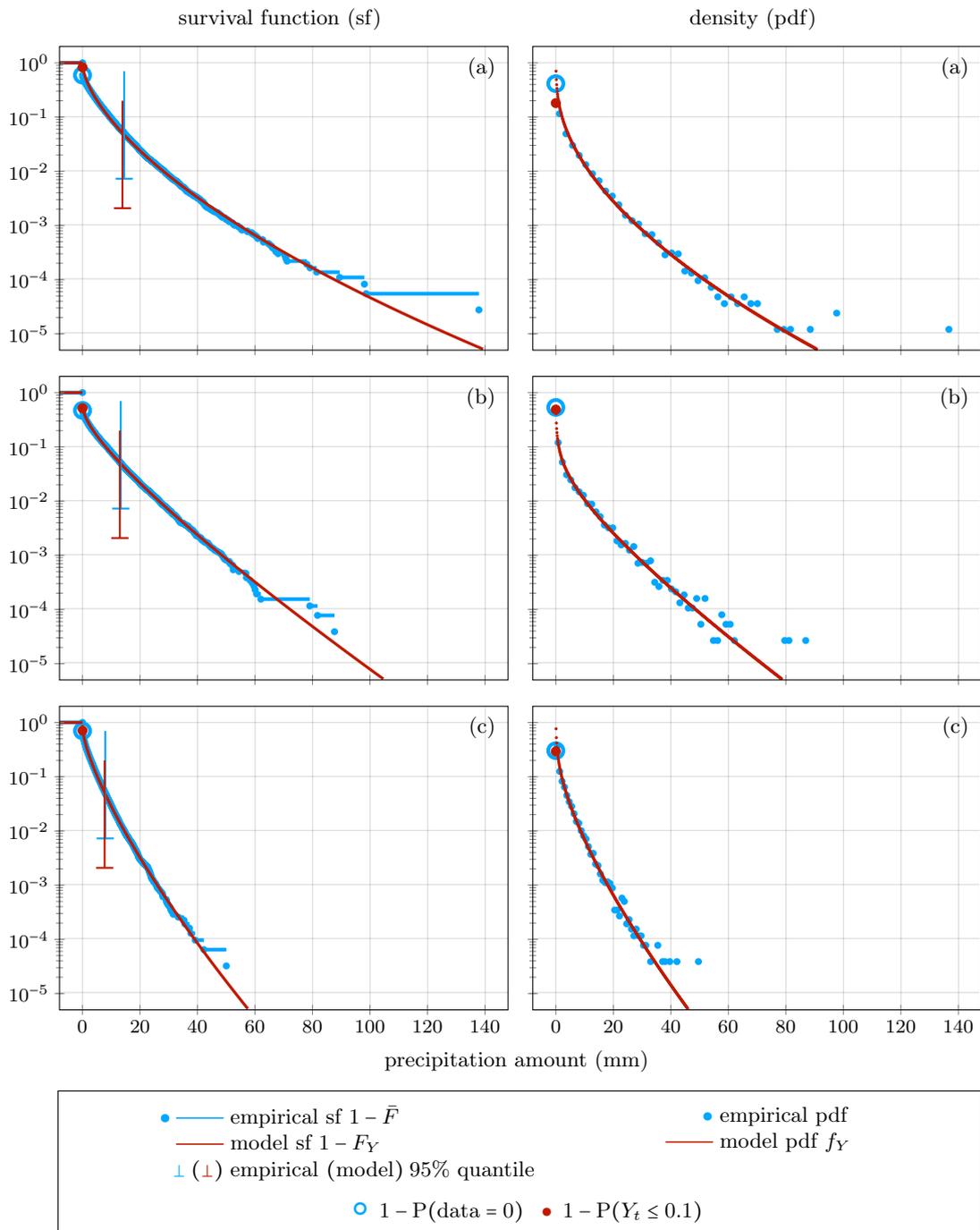


Fig. 6.4 Empirical and model sf and pdf for the three data sets in Table 6.1.

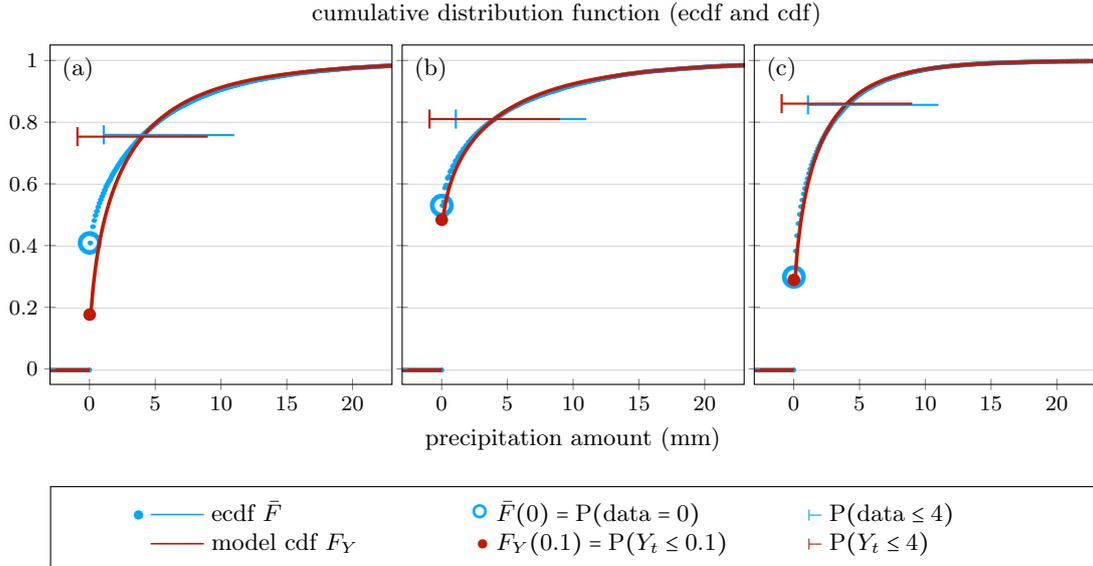


Fig. 6.5 Empirical and model cdf for the three data sets in Table 6.1.

Comparing the statistics of the empirical data and our fitted model in Table 6.3, we see great agreement in the daily mean, the daily variance and the probability of our 4 mm benchmark of perceptible precipitation (cp. Sect. 6.2.1) with deviations between the measured data and the model for small amounts close to zero. The smallest daily amount for which the deviation between the empirical and the model quantile is smaller than a certain prescribed error, can be determined precisely as apparent from Figure 6.5. If the modeling focusses more on small amounts for specific applications, a more detailed elaboration on the model properties in this data range is feasible. Our focus lies on the larger events, so that for simplicity we keep the threshold of 4 mm for model validation. Figure 6.6 depicts that the difference between the empirical and model daily mean amounts is below or in the range of the data precision of 0.1 mm.

data set	mean (data)	mean (model)	variance (data)	variance (model)
(a)	3.141	3.349	36.003	33.939
(b)	2.508	2.540	30.501	29.943
(c)	1.783	1.805	9.595	9.295
	$P(\text{data} = 0)$	$P(Y_t \leq 0.1)$	$P(\text{data} \leq 4)$	$P(Y_t \leq 4)$
(a)	0.411	0.179	0.760	0.753
(b)	0.531	0.485	0.813	0.811
(c)	0.301	0.291	0.856	0.860

Table 6.3 Central empirical and model statistics for the three data sets in Table 6.1.

6. Data model for local daily precipitation amounts

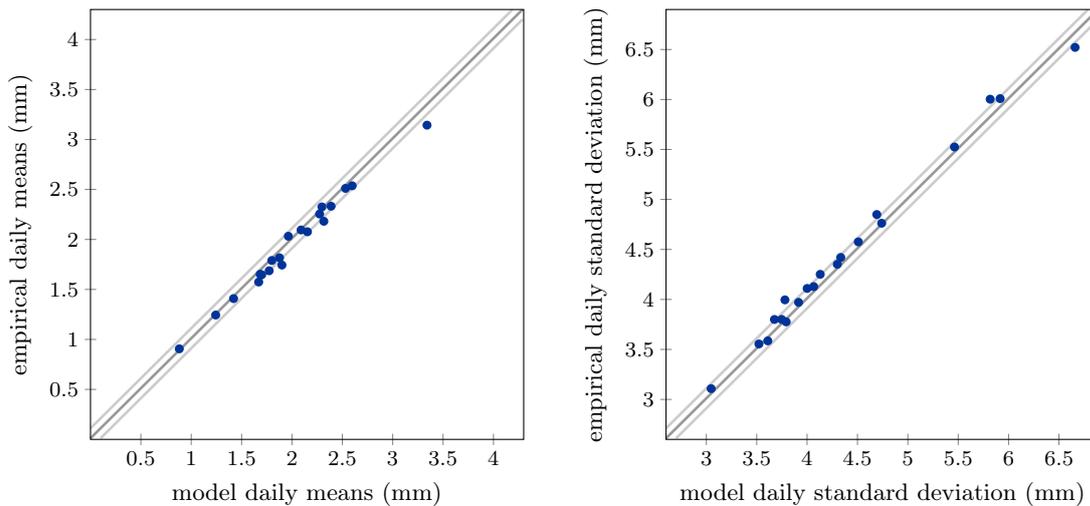


Fig. 6.6 Comparison of the empirical and model daily means and standard deviations for all 20 data sets in Table E.2. The (shifted) diagonals mark equality along with the measurement precision of 0.1 mm, respectively.

Probability plots provide visual comparisons of the empirical and model distribution. A q-q plot for assessing the empirical and model quantiles shows the strong coincidence of the tails of the distributions (Fig. 6.7, left) achieved by the tail-oriented parameter estimation. A closer look at the probabilities by a p-p plot (Fig. 6.7, right) reveals the difference between the data and the model for small amounts. Depending on the data set, the probability $P(Y_t \leq 0.1)$ for the absence of precipitation in the model can highly differ from the one in the data (cp. Tab. 6.3 and Tab. E.2). Note that the p-p plots particularly highlight the deviations for very small amounts as indicated by the empirical and model cdf in Figure 6.5.

The prominent staircase shape of the p-p plots arises from the data accuracy of the empirical data. Due to the low precision of 0.1 mm compared to the high steepness of the distribution for small values, roughly 50% of the entire probability mass are accompanied by only about ten empirical data points. Hence, the deviations in the p-p plot could be decreased by discretizing the model distribution to the same precision of 0.1. Considering all model values of at most 0.1 mm as no precipitation by mapping them to zero changes the p-p plot as exemplified by the arrow in Figure 6.7. Using synthetic data that obeys the same resolution as the empirical data is crucial for the generation of confidence intervals in Section 9.3.2 when applying the tail-oriented fit of the model distribution.

As a remark, due to the very skewness of daily precipitation amounts, a ML adjustment of the model to the data would generate highly accurate p-p plots but poor q-q plots as in studies like [20, 104]. The same occurs when applying the MM. Section 6.3.6 and Section 6.3.7 assess the impact of the chosen estimation procedure on the model statistics of annual total and annual maximal amounts, respectively.

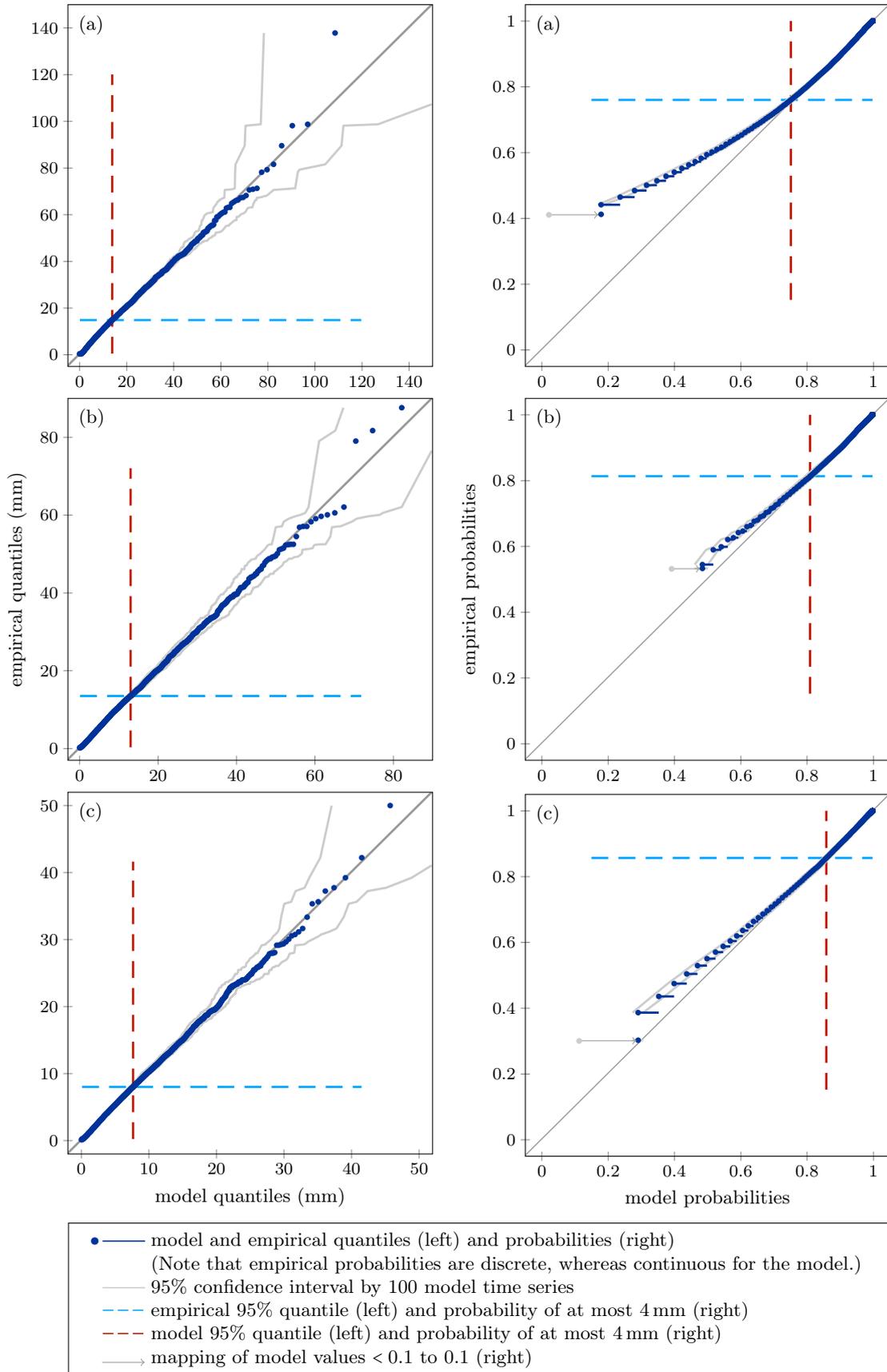


Fig. 6.7 q-q and p-p plots for the three data sets in Table 6.1.

6.3.2. Long-range correlations

Figure 6.8 shows the results of the application of DFA(3), R/s analysis and the wavelet transform introduced in Section 4.2 to the three data sets in Table 6.1. The estimated Hurst exponents obtained by the three different methods for the same data set are very similar, while there are variations from data set to data set. We obtain the LRC parameter d by the relation (6.12) based on the exponent α estimated by DFA(3) (cp. Sect. 6.1.2). The implications of Figure 6.8 are twofold. First, the spread reflecting the statistical uncertainty is rather small. Second, the estimated Hurst exponents for the measured precipitation data are well within the spread of the synthetic data, which validates that our model is able to reproduce the temporal correlations of the observed data well.

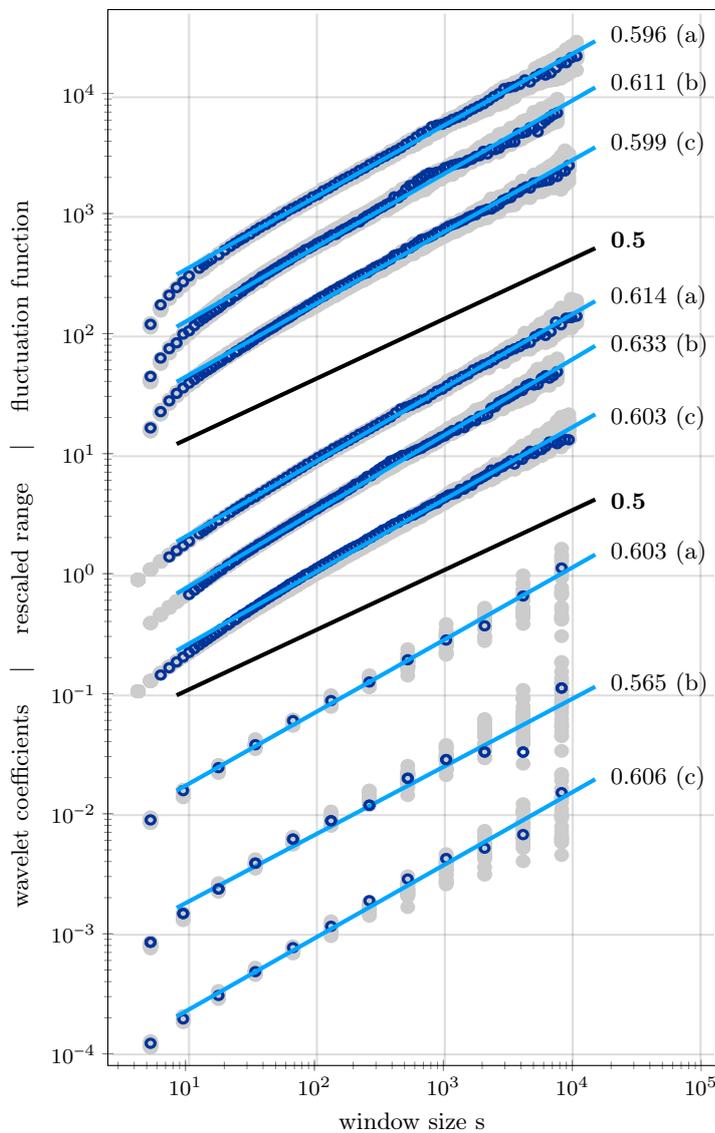


Fig. 6.8 Estimation of the Hurst exponent for the three data sets in Table 6.1 together with straight lines with slope $\frac{1}{2}$ for comparison. The shadow visualizes the results for 25 model time series, each synthesized by the adjusted models according to Table 6.2.

6.3.3. Short-range correlations

Figure 6.9 shows the conditional probabilities cfp established in Section 6.2.3 for the three data sets in Table 6.1. Given the estimated model parameters η , ν , σ and d by Section 6.2.1 and Section 6.2.2, solving equation (6.15) for the threshold of $c = 4$ mm yields the AR parameter φ for the underlying ARFIMA(1, d , 0) process.

For all empirical data sets the cfp $p_c(1)$ defined by equation (6.13) of a day with a precipitation amount larger than 4 mm right after a day suchlike noticeably exceeds the unconditioned probability of a single day with an amount above 4 mm. We find good agreement in the cfp also for time lags $k > 1$ for the ‘‘Fichtelberg’’ data set (a). For the other two examples, ‘‘Bordeaux’’ (b) and ‘‘Central England’’ (c), visual deviations occur for time lags $k \geq 2$ already. An improved representation of the cfp for larger time lags than $k = 1$ can be achieved by increasing the number p of AR components and using an ARFIMA($p, d, 0$) (cp. Sect. 2.5) process as the underlying Gaussian process for the tGp model.

For comparison Figure 6.9 includes the analytically known cfp (6.13) of a tGp-transformed ARFIMA(0, $d, 0$) process with the same parameters η , ν , σ and d as above. Even though SRC are still inherent to such a process, the short-range dependence we observe in the empirical data is not entirely captured by fractional differencing only. For the benchmark $c = 4$ and small time lags the cfp of such a model evidently falls below the empirical values.

Figure 6.9 provides a closer look on the long-term behavior of the cfp (6.13) as well. The covariance matrix (6.17) of the underlying Gaussian process is the key ingredient of the representation (6.16) of the cfp. The joint probability in (6.16) asymptotically factorizes since the acf $\varrho_Y(k)$ asymptotically vanishes. For the model Y_t we thus have $p_c(k) \rightarrow P(Y_t > c)$ as time lags $k \rightarrow \infty$, so that the cfp decreases slowly to the unconditioned probability as time lags increase. Comparing the difference

$$|P(Y_t > 4) - P(Y_t > 4 | Y_{t-k} > 4)| \quad (6.21)$$

calculate for the model by equation (6.13) and estimated for the data by equation (6.14), we observe the same decrease to zero as time lags k increase. Moreover, the decay of the cfp p_c to the probability $P(Y_t > c)$ of the model follows a power law alike the model acf ϱ_Y . The cfp of the empirical data sets show the same scaling behavior, although, we only implicitly include their autocorrelations by the estimator (6.14). Other than the empirical acf (cp. Sect. 4.1), the estimated cfp are not subject to negative values in a logarithmic scaling and are thus capable of approximating the asymptotic scaling of the acf.

6. Data model for local daily precipitation amounts

As a remark, the estimated values φ do not correspond to a typical correlation time other than in AR or ARMA (cp. Sect. 2.5) models with a finite sum (2.8). The impact of the auto-regression in an ARFIMA(1, d , 0) model (2.21) on the acf decays exponentially, though, the correlation time (2.8) remains infinite, due to the LRC of the model.

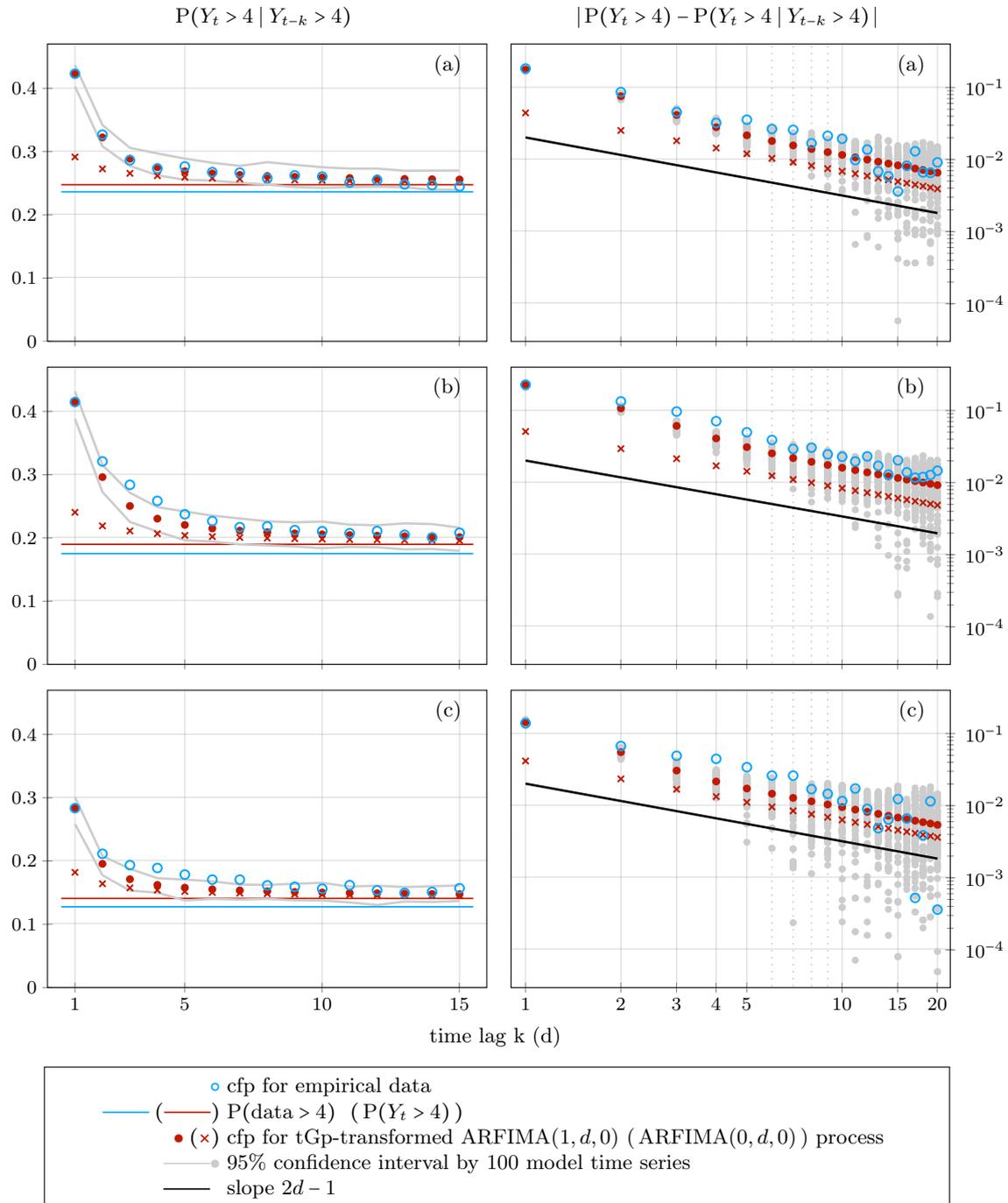


Fig. 6.9 Visualization of the cfp (6.13) (left) and of their decay rates by the difference (6.21) (right) for the three data sets in Table 6.1.

6.3.4. Wet and dry spells

A noticeable statistical effect of LRC in time series is a change of the distribution of waiting times [29]. For white noise or SRC data the waiting times between events of a common type is exponentially distributed. In the presence of LRC stretched exponential tails of such waiting times distribution occur [8].

Waiting times between two days with an accumulated precipitation amount of $c > 0$ mm shall be interpreted as periods of daily amounts of at most c mm. For $c = 0$ they describe dry spells. Studying wet and dry spells in measurement data based on the strict threshold of zero is not appropriate as argued in Section 6.3.1. Applying the tail-oriented parameter estimation described in Section 6.2.1, we do not aim at precisely reproducing the probability of zero daily precipitation that we find in the empirical data.

Figure 6.10 gives a visual impression of the effect of LRC on the waiting times for the more practicable benchmark of $c = 4$. Further detailed investigation of dry and wet spells beyond these semi-logarithmic histograms is required. For such an analysis considering waiting times with respect to small values of $c > 0$ as a measure of the duration of dry periods in terms of applications accounts more for the measurement uncertainty of zero precipitation.

For comparison Figure 6.10 depicts the waiting times of both a tGp-transformed AR-FIMA(1, d , 0) and an AR(1) process with the marginal distribution and LRC and AR parameters estimated as described in the Sections 6.2.1 to 6.2.3. For the threshold $c = 0$ both models underestimate the occurrence of dry spells for all the three data sets because our tGp model still tends to underestimate the probability of a dry day even when mapping all values less than 0.1 mm to zero (see Table E.2). For the threshold $c = 4$ the AR(1) based process still fails to reproduce the distribution of long dry spells in the sense that periods longer than about 45 days are visually much more unlikely than in the empirical data. The LRC tGp model, however, is capable of reproducing a higher number of such long dry periods in accordance with the statistics of the waiting times in the original data of the three examples. This visualization does not test the significance of a stretched exponential decay of the waiting time densities here. Nevertheless, Figure 6.10 illustrates that introducing LRC in our data model is a promising approach to modeling the tails of the waiting time distributions of daily precipitation time series.

As a remark, for both depicted thresholds in Figure 6.10 the waiting time distribution of a randomly shuffled version of the originally observed time series clearly differs from the one of the original data. As expected for uncorrelated data (correlations are destroyed by the shuffling) the density of its waiting times decays exponentially and visibly significantly faster than the original waiting times.

6. Data model for local daily precipitation amounts

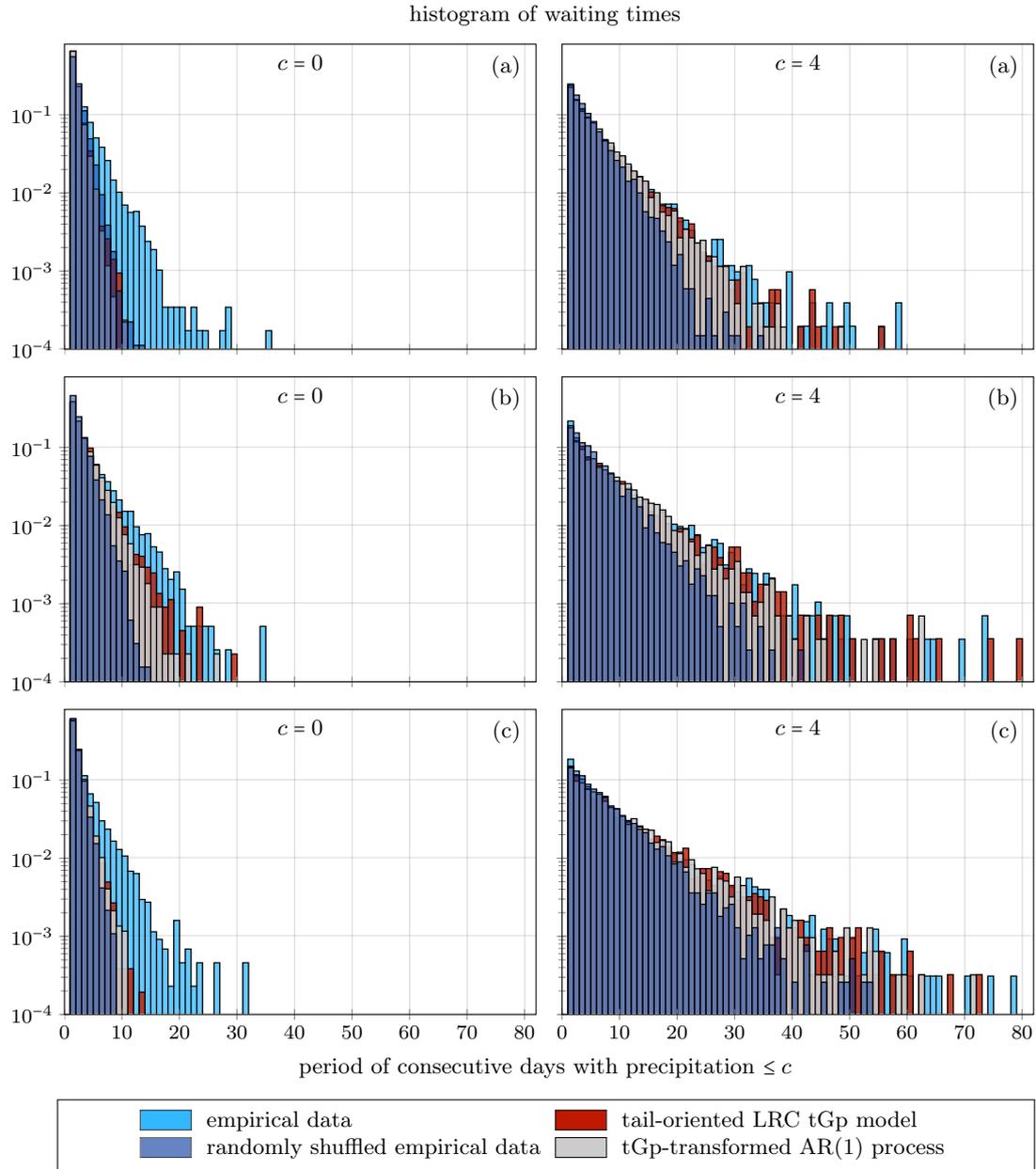


Fig. 6.10 Comparison between empirical waiting time distributions of the empirical and different synthetic time series for the three data sets in Table 6.1.

6.3.5. Effective sample size and daily mean

Definition (5.9) together with the finite-size decorrelation time (5.6) provide effective sample sizes as established in Chapter 5 for stationary empirical data by the empirical acf. Given the estimated tGp model with LRC in the data captured by an underlying AR-FIMA process yields a more robust approach to the correlation structure in the data and so to its effective sample size. Table 6.4 shows an excerpt containing the three data sets in Table 6.1 from the effective sample sizes for all example data sets in Table E.1. Figure 6.11 (left) compares the effective (N_{eff}) to the empirical (N) sample size for all examples. With respect to the variance of the sample mean (cp. Sect. 5.1.2) these effective sample sizes measure from about 9% to 44% of the empirical sample size with most ratios between 10% and 30%.

data set	N	N_{eff}	N_{eff}/N
(a)	37,621	5,074	0.135
(b)	26,641	2,778	0.104
(c)	32,142	4,472	0.139

Table 6.4 Comparison between the empirical (N) and effective (N_{eff}) sample sizes by their ratio $\frac{N_{\text{eff}}}{N}$ for the three data sets in Table 6.1.

Involving the model mean μ_Y and variance σ_Y^2 along with the effective sample size, Lemma 5.5 provides an approximation of the distribution of the sample mean for the tGp model. By Theorem 5.3 the sum of the meta-Gaussian LRC process Y_t generated by the tGp transformation with Hermite rank 1 asymptotically follows a Gaussian $\text{N}(\mu_Y, \frac{\sigma_Y^2}{N_{\text{eff}}})$ distribution. The model mean and variance are analytically known by the equations (6.20). For all data sets the relative deviation, defined by $\frac{\bar{\sigma}_{\mu_N} - \sigma_Y}{\bar{\sigma}_{\mu_N}}$, between the empirical $\bar{\sigma}_N$ (cp. equ. (4.2)) and model σ_Y daily standard deviation is less than 5%. Hence, the above Gaussian distribution describes the uncertainty of the sample mean such as for example the 68% confidence interval

$$I_N := \left(\mu_Y - \frac{\sigma_Y}{\sqrt{N_{\text{eff}}}}, \mu_Y + \frac{\sigma_Y}{\sqrt{N_{\text{eff}}}} \right) \quad (6.22)$$

capturing one standard deviation around the model mean. By favoring the tail of the distribution of daily amounts the tGp model lacks exact reproduction of the empirical daily mean, though, this deviation is in the range of the data precision of 0.1 mm for all the example data sets (see Fig. 6.6, left). Figure 6.11 (right) shows that nonetheless for about half of the 20 data sets the empirical mean lies inside one standard deviation of the model sample mean. The tendency of our model towards an underestimation of the probability of zero daily rainfall (cp. Sect. 6.3.1) translates to a possible positive bias in the annual totals in Section 6.3.6. As a remark, the empirical and model mean $\bar{\mu}_N$ and μ_Y , respectively, coincide, when estimating the tGp distribution by the MM.

6. Data model for local daily precipitation amounts

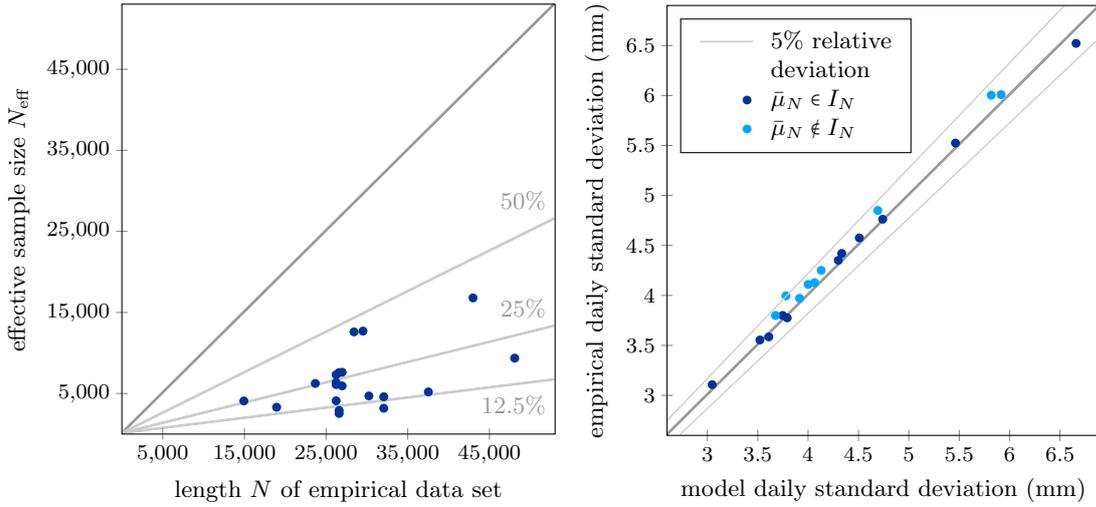


Fig. 6.11 Comparison of the empirical and model daily means $\bar{\mu}_N$ and μ_Y and standard deviations $\bar{\sigma}_N$ and σ_Y , respectively, for all 20 data sets in Table E.2. For about half of the data sets the empirical mean lies within the 68% confidence interval (6.22).

6.3.6. Annual totals

Analogously to the daily sample mean in Section 6.3.5, Theorem 5.3 yields a Gaussian approximation of the distribution of annual totals in the tGp model. For annual total amounts let $K = 365$, so that $A := \sum_{i=1}^K Y_i$ denotes the annual sum over windows (Y_1, \dots, Y_K) of the tGp model process. We shall assume the annual sum A asymptotically Gaussian $N(\mu_A, \sigma_A^2)$ with mean $\mu_A = K\mu_Y$. By definition (5.9), we approximate the variance σ_A^2 of the annual sum A by

$$\sigma_A^2 = K\sigma_Y^2\tau_D(K) \quad (6.23)$$

with the model variance σ_Y^2 of daily amounts and the finite-size decorrelation time $\tau_D(K)$ defined by equation (5.6). In Figure 6.12 (bottom right), we find coincidence between the standard deviation of the empirical and model annual totals. Due the measurement precision of 0.1 mm of the daily data, the precision of the annual sum is limited to 36.5 mm, so that values that differ at this magnitude are practically indistinguishable. The sample variance is further known for the tendency to underestimate the true variance in the presence of LRC [22], which could explain that the empirical standard deviations slightly fall below model standard deviations. Elaborating the same procedure for the tGp model with its marginal distribution estimated by the MM, we find very small differences in the representation of the statistics of annual totals.

Figure 6.12 visualizes that the Gaussian pdf adjusted to the empirical annual totals only slightly differs from the proposed $N(\mu_A, \sigma_A^2)$ distribution given by equation (6.23). For all data sets the empirical mean of the annual totals lies within one standard deviation σ_A of the respective model mean μ_A with little differences between the two methods. Using the MM for model estimation yields coincidence between the empirical

and model mean of annual totals alike for the daily mean amount. By the tail-oriented model estimation, deviations of the daily and annual mean above measurement precision are generally possible, since we do not explicitly adjust for these quantities.

For comparison, Figure 6.12 shows the standard deviation of annual totals of an iid model with finite-size decorrelation time $\tau_D(K) = 1$ in equation (6.23), which clearly underestimates the empirical fluctuations. Involving the strength of correlations is thus crucial for proper modeling of daily precipitation measurements.

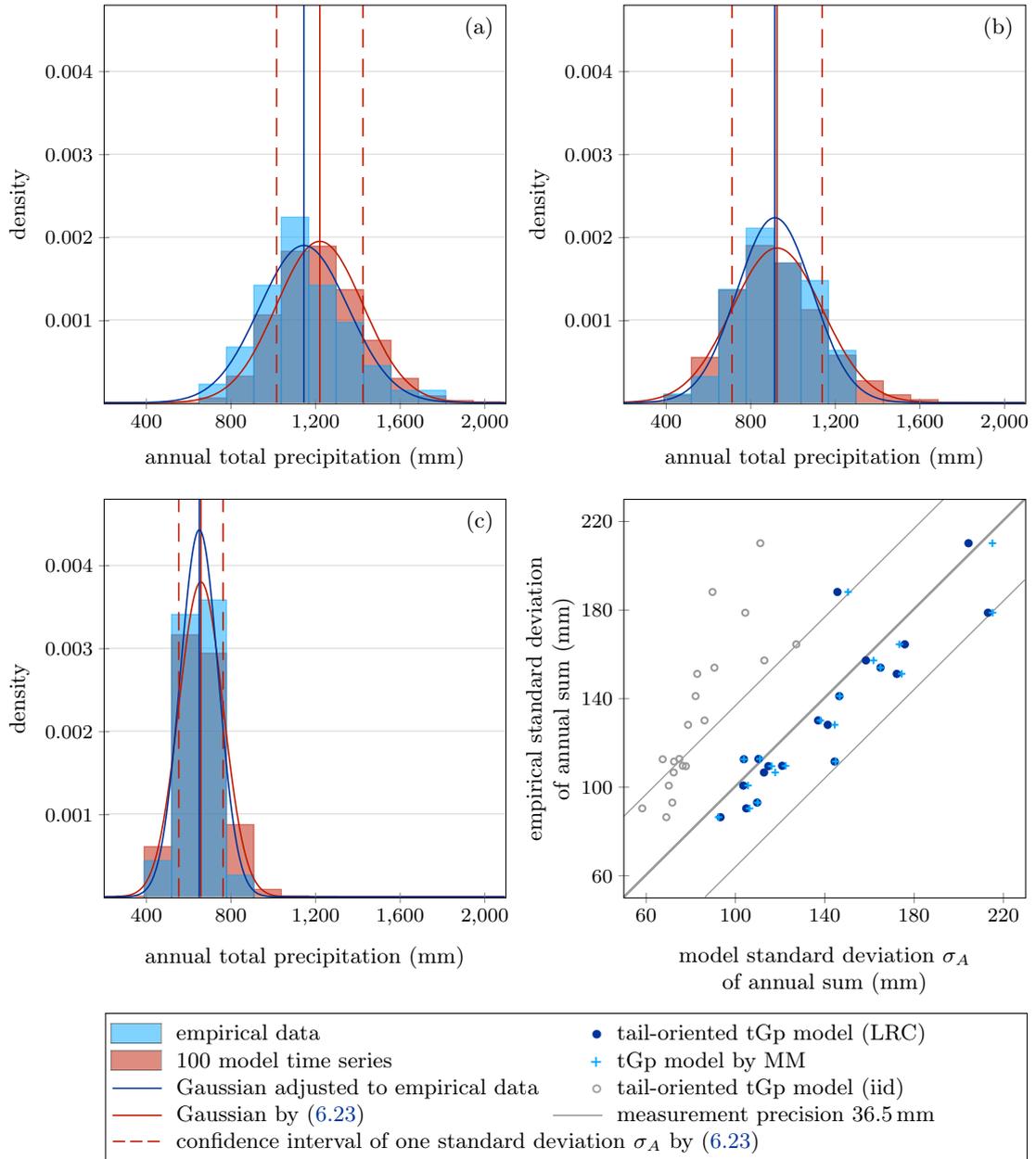


Fig. 6.12 Histogram of annual totals for the three data sets (a), (b) and (c) in Table 6.1 along with histograms and Gaussian pdfs with variance σ_A^2 (6.23) for annual totals of 100 model time series estimated by the tGp sf. Comparison of the empirical to the standard deviation σ_A for both the tail-oriented and MM model estimation (bottom right).

6. Data model for local daily precipitation amounts

6.3.7. Annual maxima

For risk assessment extreme precipitation events are a statistical quantity of great interest. The *Mahalanobis distance* [7] provides a measure for the representation of the statistics of annual daily maximum amounts by the tGp model. Applying this distance measure in a multidimensional event space allows for including the joint distribution of multiple observables into an error evaluation at once.

Given a covariance matrix $\Sigma \in \mathbb{R}^{s \times s}$ in $s \in \mathbb{N}$ dimensions the *Mahalanobis distance* $d_M(x, y)$ between two points $x, y \in \mathbb{R}^s$ is defined by

$$d_M(x, y) := \left((x - y)^T \Sigma^{-1} (x - y) \right)^{1/2}. \quad (6.24)$$

Points of equal distance from a mean $\mu \in \mathbb{R}^s$ with respect to the distance d_M (6.24) form ellipses in two dimensions and multidimensional ellipsoids in higher dimensions, respectively. These level sets serve as probability limits with respect to the parent multivariate Gaussian $N(\mu, \Sigma)$ distribution. By the principal axis theorem any multivariate Gaussian originates from the multivariate standard Gaussian with zero mean and the identity covariance matrix by a rotation and stretching or compression. Inverting these affine transformation, the distance d_M from the mean μ determines the radius of a sphere of equal probability mass like the ellipsoid prior to rotation and scaling. ‘Confidence regions’ of a prescribed probability are spheres with a radius given by the $\chi^2(s)$ -quantile function with s degrees of freedom as the squared distance $d_M^2 \sim \chi^2(s)$ (cp. Sect. 3.4.3).

A powerful feature of the Mahalanobis distance is that the value $d_M(x, y)$ directly translates to distances in terms of standard deviations of the $N(\mu, \Sigma)$ distribution. Analogously to unit dimension, we consider the 68% confidence regions (‘one standard deviation around the mean’). Note that the radius $r = d_M$ of the sphere with a prescribed probability mass depends on the dimension s (cp. Fig. 6.13). For $s = 2$ the radius r of the 68% sphere reads $r = \sqrt{F_{\chi^2(2)}^{-1}(0.68)} \approx 1.52$, while $r = \sqrt{F_{\chi^2(2)}^{-1}(0.68)} \approx 1.88$ for $s = 3$, where $F_{\chi^2(s)}^{-1}$ denotes the quantile function of the $\chi^2(s)$ distribution. The sphere with radius $r = 1$, however, obeys the probability $F_{\chi^2(2)}(1) \approx 0.39$ in two and $F_{\chi^2(3)}(1) \approx 0.2$ in three dimensions with the cdf $F_{\chi^2(s)}$ of the $\chi^2(s)$ -distribution.

For the Fichtelberg data set (a) Figure 6.13 shows the two-dimensional distribution of pairs of the mean and the standard deviation of annual maxima of different model time series. The depicted probability limits base on the bivariate Gaussian adjusted to 100 synthetic pairs. For comparison Figure 6.13 involves LRC and iid tGp model time series estimated by both the sf and the MM. For the tail-oriented method the point pairing the mean and variance of the empirical annual maxima lies ‘within one standard deviation’ with respect to $d_M = 1$ for LRC and inside the 95% confidence region for iid model time series, whereas even outside the 95% limit for the both kinds of MM model time series. As a conclusion, the MM possibly clearly underestimates the mean and variance of annual maxima.

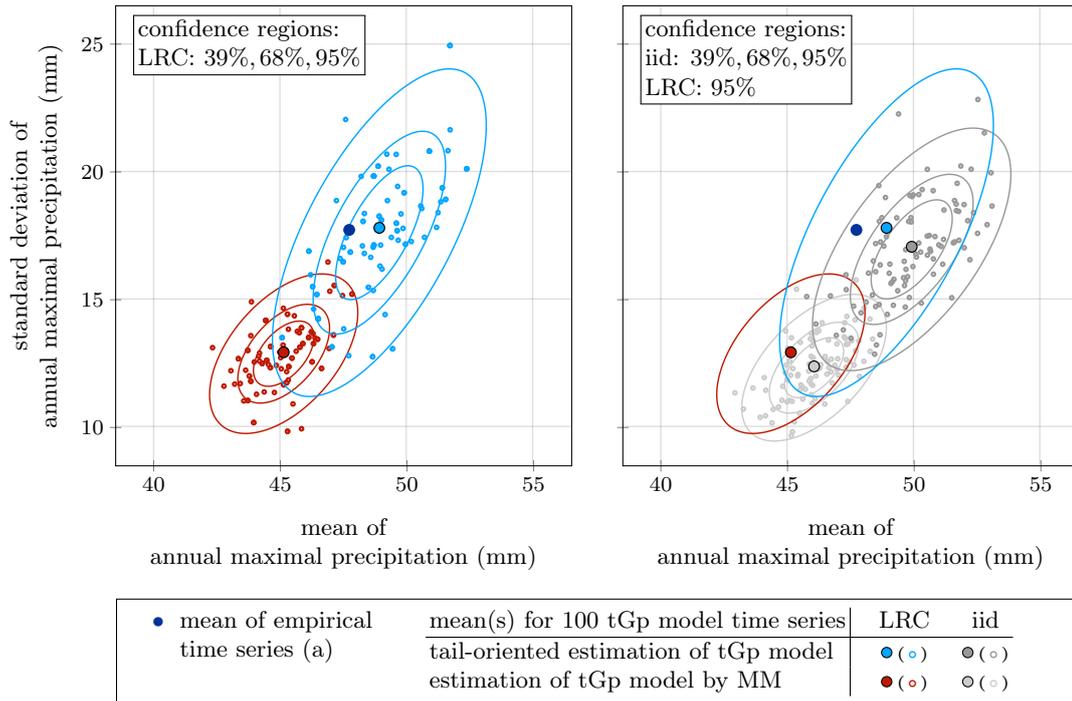


Fig. 6.13 Visualization of probability limits by the Mahalanobis distance estimated based on 100 tGp model time series adjusted by the tail-oriented approach of Section 6.2.1 and by the MM both with LRC (left) and for comparison additionally iid (right) for the Fichtelberg station (a) of Table 6.1. Each empty circle marks the pair of the mean and the standard deviation of the annual maxima of a model time series of length $N = 37,621$ like the empirical data set (a). The filled circles depict the mean of these pairs for all 100 synthetic time series for each of the four model variants.

Figure 6.14 visualizes a parallel assessment of three statistical quantities for all 20 data sets in Table E.2, by considering a three-dimensional Mahalanobis distance between triples consisting of the mean, the standard deviation and the 100-year return level (cp. Sect. 1.3 and Sect. 9.1) of annual maxima. The latter is estimated by the 99%-quantile of a GEV distribution adjusted⁷ to the annual maxima of the empirical and synthetic time series (cp. Sect. 7.4 and Sect. 7.4.4).

We find that measured in the Mahalanobis distance the mean triple as introduced above of the empirical annual maxima predominantly lies within the 68% limit and for about half of the data sets within one standard deviation ($= F_{\chi^2(3)(1)}$ confidence region) around the mean of the LRC tGpmodel time series with tail-oriented parameter estimation. When adjusted by the MM, we observe larger errors with more than half of the data sets outside of 68% probability.

⁷by the software package Climex [134]

6. Data model for local daily precipitation amounts

The presence of correlations compared to an iid model is indeed conceivable in Figure 6.14 but the effect of properly estimating the tail of the marginal distribution has larger influence on how well the statistics of annual maxima are represented. Assessed by probability limits of the Mahalanobis distance, the tail-oriented parameter estimation of Section 6.2.1 excels the MM in the representation of annual extremes, whereas correlations appear less crucial for these statistics. Mind that the opposite applies to annual mean precipitation amounts by Figure 6.12.

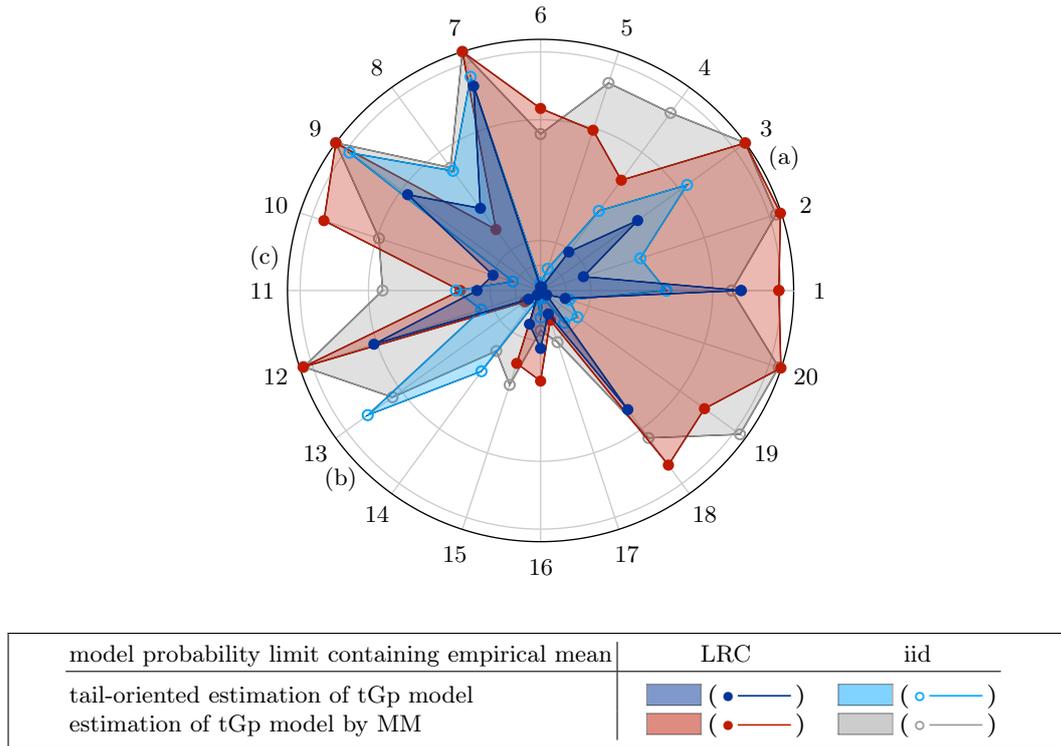


Fig. 6.14 Comparison of the Mahalanobis distance of the empirical and model mean triple of the mean, the standard deviation and the estimated 100-year return level of annual maximal amounts described by probability limits adjusted to 100 tGp model time series generated by the four approaches introduced in Figure 6.13 for the 20 data sets of Table E.2. The smaller the probability is of the model confidence region that contains the empirical mean the more accurate is the representation of the data by the model. The outer circle of the radar chart marks unit probability, the center probability zero and the inner circles probabilities 10%, 20%, 68% and 95%.

6.4. Closing remarks and outlook

The detailed statistical analysis of the marginal distribution of daily precipitation amounts in Section 6.3 adds to the ongoing discussion about the shape of the tail of the marginal pdf (see Sect. 1.4). Section 6.4.1 summarizes the successful representation of the empirical statistics by the LRC tGp model (cp. Chap. 10) and Section 6.4.2 sets these results in contemporary context of research.

Section 6.3.6 assesses the uncertainty of mean annual total precipitation amounts by assuming a Gaussian distribution of this quantity and Section 6.4.3 provides a justification of this approach by the skewness of daily and annual amounts.

An essential property of the LRC tGp model established in Section 6.1 is stationarity. An outlook in Section 6.4.4 provides approaches to generalizations of the model by describing its limitations due to adjustment to mid-latitude climate.

6.4.1. Statistical representation of empirical data

The main result of the statistical analysis in the Sections 6.3.1 to 6.3.7 is that focussing on large precipitation amounts, the tGp model well represents statistical key quantities of mid-latitude daily precipitation by applying both LRC and a tail-oriented estimation of the marginal distribution. For capturing the statistics of annual totals including LRC is highly crucial, whereas exact modeling of small amounts less. The other way round, for representing well the statistics of annual extremes appropriate estimation of the tail of the marginal distribution is more important than LRC.

The tail-oriented parameter adjustment allows for deviations for small amounts visualized in p-p plots for some data sets in Section 6.3.1. The model mean daily amount might thus deviate from the empirical one. Nonetheless, the empirical daily standard deviation for all data sets measures within 5% relative deviation from the model, so that for about half of the data sets the daily mean lies within one standard deviation of the sample mean considering LRC via the effective sample size (Fig. 6.11). The deviation between the empirical and model standard deviation of annual totals is even below measurements precision and not significantly altered when estimating the marginal distribution by the MM instead of with emphasis on the tail. An iid and tail-oriented tGp model, however, clearly underestimates the fluctuations of annual totals (Fig. 6.12). For all data sets the mean annual sum lies within one standard deviation (6.23) assuming Gaussianity and considering correlations by applying the finite-size decorrelation time.

6. Data model for local daily precipitation amounts

Note that the MM yields high agreement in p-p plots but poor q-q plots as the match of the mean, the variance and the probability of zero precipitation is exact but disregards the tail of the distribution. The proper representation of empirical data in q-q plots (Fig. 6.7) substantiates the appropriateness of the tail-oriented tGp distribution for mid-latitude daily precipitation amounts, in agreement with heavy-tailed [106, 142] and contrary to light-tailed [104] or fat-tailed [140, 186] models.

In Section 6.3.7 the Mahalanobis distance allows for parallel assessment of the mean, the standard deviation and the 100-year return level of annual maxima. The tail-oriented tGp model excels the tGp model estimated by the MM in reproducing the statistics of annual maxima. Applying an LRC or an iid model does not change this goodness significantly (Fig. 6.14).

The LRC tGp well represents also SRC measured by conditional follow-up probabilities (6.13) (Fig. 6.9) and wet and dry spells (Fig. 6.10) assessed by waiting times between days with an accumulated amount of at least a certain size. Comparisons with ARFIMA(0, d , 0), AR(1) and iid models emphasize that properly adjusting both LRC and SRC by the underlying ARFIMA(1, d , 0) process is crucial for achieving these statistics.

6.4.2. The tail of the marginal distribution

The model parameter η , which determines the asymptotic stretched exponential decay (6.5) of the tail of the tGp distribution, depends on the particular data set (cp. Table E.1 and Fig. 6.15).

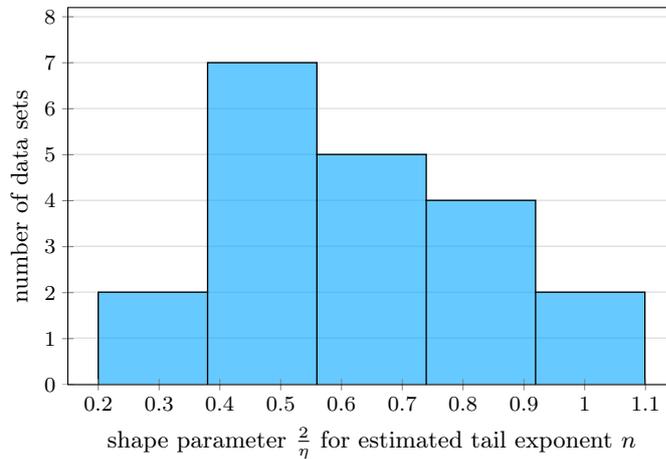


Fig. 6.15 Shape parameters $\frac{2}{\eta}$ according to [185] by the estimated tail exponents η of the tGp distribution for all 20 data sets in Table E.2. Note that the smallest entry in the leftmost bin is 0.28 and the largest in the rightmost is 1.08.

Incorporating physical dynamics, Wilson and Toumi [185] recently derived a universal approximate stretched exponential $e^{-(y/R_0)^c}$ ($y \rightarrow \infty$) tail behavior of daily rainfall amounts $y \in \mathbb{R}_{\geq 0}$ with a scale parameter $R_0 \in \mathbb{R}_{>0}$ and shape $c \in \mathbb{R}_{>0}$ with $c \approx \frac{2}{3}$ with a standard deviation of 0.12. In contrast, we include not only the asymptotically large but the entire range of the samples into our parameter estimation albeit with different emphasis. By equation (6.3), the parameter η controls the shape of the pdf for the power-law part of small amounts and the stretched exponential tail of the tGp at once. Specific geographical conditions thus possibly influence the parameter values. Nevertheless, we predominantly observe powers η in the limited range of roughly 2 to 7 at the utmost, which accords with shape parameters $c = \frac{2}{\eta}$ in the range of 0.64 ± 0.2 in agreement with [185]. Wilson and Toumi focus on the tail of the distribution, so that only two parameters suffice in their study for an appropriate description of large rainfall events. Our additional third model parameter helps adjust for smaller values as well with the reward of representing statistically well both extreme precipitation events and the daily mean and variance at once.

6.4.3. The skewness of precipitation amounts

Assuming a Gaussian distribution for annual total precipitation amounts in Section 6.3.6 bases on the Hermite rank 1 of the tGp distribution (cp. Sect. 6.2) and Theorem 5.3. Figure 6.12 visually confirms this assumption for model time series for the three data sets of Table 6.1. For the 100 synthetic time series considered therein the skewnesses of their annual totals read approximately (a) 0.35, (b) 0.47 and (c) 0.31, which are slightly larger than zero, other than the assumption.

The skewness of empirical mean annual totals might as well slightly differ from zero ((a) 0.49, (b) 0.04, (c) -0.01). Model time series might thus have a negligible tendency of underestimating this skewness as seen in Figure 6.12 (a). Compared to the widespread skewness of daily amounts, however, the skewness of annual totals is moderate to low. Figure 6.16 visualizes this comparison and the overall proper agreement of empirical and model skewnesses for either amounts daily and annual. The assumption of Gaussianity for annual totals yields access to their statistical uncertainty regardless of the small deviation from zero skewness for measured or synthetic data.

6. Data model for local daily precipitation amounts

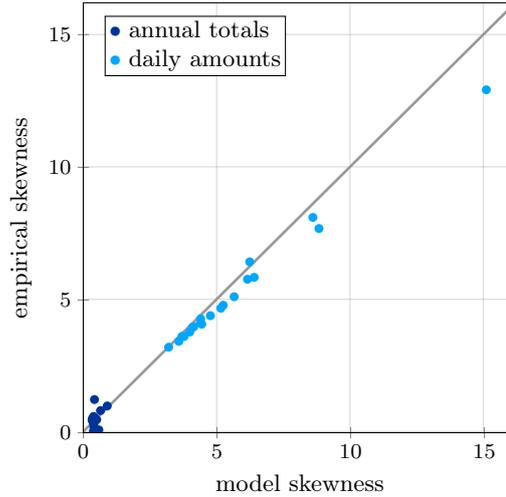


Fig. 6.16 Comparison of the empirical and model skewness of daily amounts and annual totals for all 20 data sets in Table E.2. The model skewness of daily amounts is analytically known by $\mathbb{E}\left[\left(\frac{Y_t - \mu_Y}{\sigma_Y}\right)^3\right]$ using the model mean μ_Y and variance σ_Y^2 given by equation (6.20) and the model density (6.3) for integration. The model skewness of annuals totals bases on the mean empirical skewness for 100 model time series.

6.4.4. Limitations of the model and outlook

Due to their very stationarity, ARFIMA processes cannot generate any kind of seasonality. If desired, generalizations are nonetheless readily accessible. A time-dependent periodic shift of the tGp marginal distribution can capture an annual cycle in daily precipitation amounts such as observed for Indian or tropical regions for example. Possible non-stationary behavior like described in [142] and indicated by Figure 6.2 for the variance of daily precipitation amounts can find representation in a generalized tGp model by adding appropriate trends to the parameters or to the transformation of the underlying Gaussian process in general.

The stationary model yields potential beyond the applications described in Chapter 1 as it serves as a benchmark for example for analyses of phenomena like climate change. Generating stationary extrapolations of previous precipitation statistics provides points of comparison against recent and near future measurements and help identify systematic changes.

7. Extreme value theory

Based on the textbooks [101], [48] and [37], we gain an overview of the fundamentals of *extreme value theory* (EVT). Classical EVT deals with the distribution of the maximum of iid random variables. The main result, the *Fisher-Tippet-Gedenko theorem* or *extremal types theorem*, derives the generalized extreme value distribution as the universal limit distribution of (appropriately shifted and scaled) maxima of iid variates. Under certain mild conditions this observation remains valid for stationary processes such as ARFIMA processes (cp. Sect. 2.5) or meta-Gaussian processes (cp. Chap. 3). The application of EVT primarily aims at the estimation of return levels given return periods or the other way round (cp. Sect. 1.3).

Sections 7.1 and 7.2 collect essential concepts and results from the EVT for iid processes and its generalizations to stationary process, respectively, along with an introduction to statistical inference regarding return levels in Section 7.4.

For an application of EVT to meta-Gaussian processes in Chapters 8 and 9, our introduction to EVT is with a special focus on norming constants and rates of convergence in broad detail in Section 7.3. These results base our assessment of return levels in Section 9.2 and the formulation of our method for the estimation of return levels in Section 9.3 with application to precipitation data each.

7.1. Classical extreme value theory

Classical EVT deals with the maxima of random samples that are drawn *independently* from a *common* distribution. For a number $n \in \mathbb{N}$ let

$$M_n := \max \{X_1, \dots, X_n\} \tag{7.1}$$

denote the maximum of an n -tuple of iid random variables X_i , $i = 1, \dots, n$, with common cdf F_X (and pdf $f_X : \text{dom}(f_X) =: U \subseteq \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ if existent). Since the variates X_i are independent, the cdf F_{M_n} and, if so, the pdf $f_{M_n} : U \rightarrow \mathbb{R}_{\geq 0}$ of M_n follow immediately as

$$F_{M_n}(x) = \text{P}(M_n \leq x) = \text{P}(X_1 \leq x, \dots, X_n \leq x) = F_X(x)^n, \quad x \in \mathbb{R}, \quad \text{and} \tag{7.2}$$

$$f_{M_n}(x) = \frac{d}{dx} F_{M_n}(x) = n F_X(x)^{n-1} f_X(x), \quad x \in U, \tag{7.3}$$

in the points $x \in U$ where F_X is differentiable. As a remark, the existence of a pdf f_X is not necessary for equation (7.2) and EVT in general.

7. Extreme value theory

The primary interest of the field lies in the asymptotic stochastic properties of the maximum M_n (7.1) in the regime of large numbers n of samples. In the limit of $n \rightarrow \infty$, $\lim_{n \rightarrow \infty} F_{M_n}(x) \in \{0, 1\}$ is apparent for all $x \in \mathbb{R}$ by $F_X(x) \in [0, 1]$ and equation (7.2). Note that $F_{M_n}(x) \rightarrow 1$ ($n \rightarrow \infty$) is only possible if the distribution of the samples X_i features a finite *upper endpoint* $x_m < \infty$, defined by

$$x_m := \sup \{x \in \mathbb{R} \mid F_X(x) < 1\}, \quad (7.4)$$

with $F_X(x_m) = 1$. For distributions with $x_m = \infty$ we find the trivial limit $F_{M_n}(x) \rightarrow 0$ as $n \rightarrow \infty$ for all $x \in \mathbb{R}$. Figure 7.1 exemplifies the evolution of the cdf F_{M_n} (7.2) for standard uniformly and normally distributed samples as $n \rightarrow \infty$.

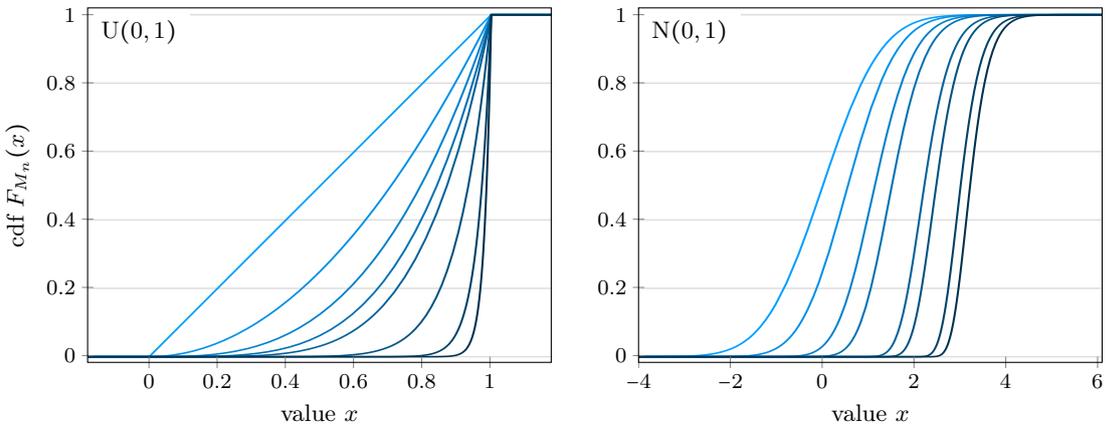


Fig. 7.1 Evolution of the cdfs (7.3) of the maximum M_n (7.1) of $n \in \mathbb{N}$ iid samples of the standard uniform ($x_m = 1$) and standard Gaussian ($x_m = \infty$) distribution (cp. Appendix B) for $n \in \{1, 2, 3, 4, 5, 10, 25, 50\}$ (left) and $n \in \{1, 2, 5, 10, 50, 100, 500, 1,000\}$ with darkening lines as n increases.

Example 7.1 visualizes the evolution of the probability mass of the maximum M_n as n increases by depicting the pdf f_{M_n} (7.3) for some frequently used distributions.

Example 7.1: Figure 7.2 exemplifies the distribution of the maximum M_n for the four distributions $\text{Exp}(1)$, $\text{Par}(1, 3)$ and $\text{U}(0, 1)$ and $\text{N}(0, 1)$ (cp. Appendix B). The shaping of the pdf from the original pdf ($n = 1$) of the samples to the pdf (7.3) of the maximum of $n \in \mathbb{N}$ samples differs amongst these examples. As to the exponential distribution, for large numbers n the pdf retains its shape and keeps shifting rightwards as n increases. In case of Pareto distributed samples, the pdf flattens out with increasing n , while for the uniform distribution it accumulates in its upper endpoint $x_m = 1$. The pdf of the maximum of iid Gaussian samples keeps shifting rightwards while being stretched vertically more and more.

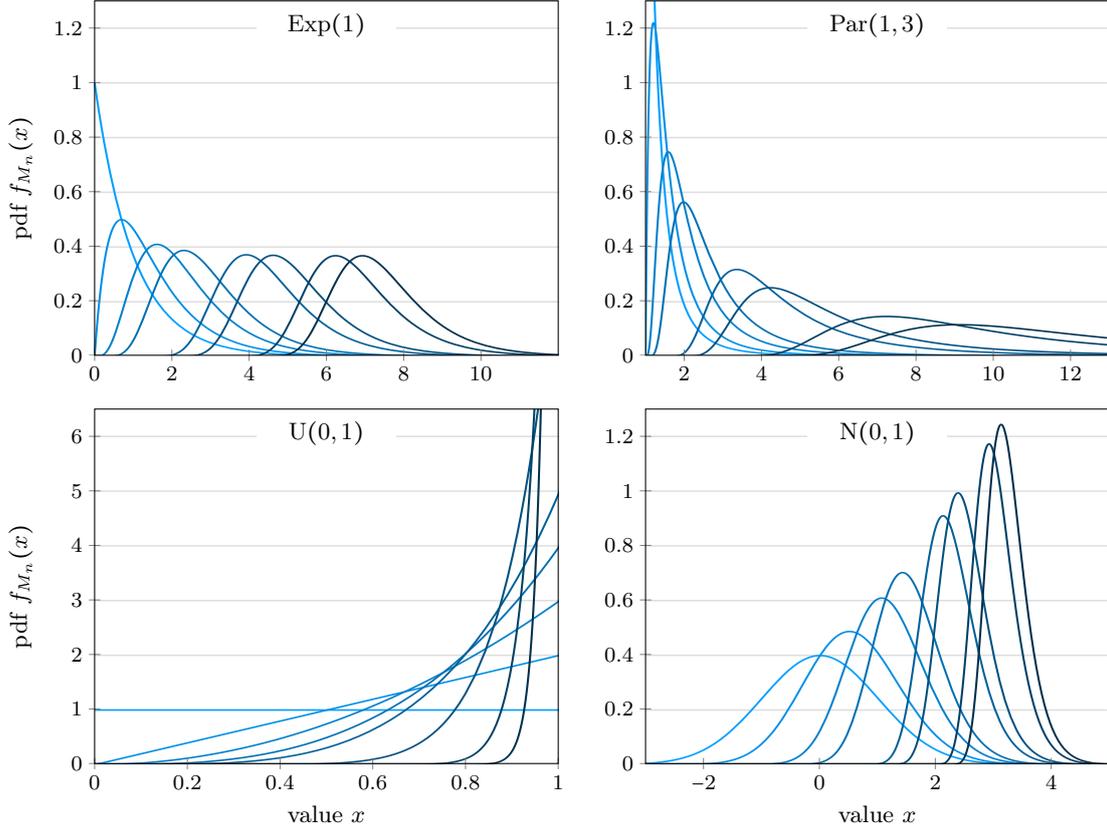


Fig. 7.2 Evolution of the pdfs (7.3) of the maximum M_n (7.1) of $n \in \mathbb{N}$ iid samples of the standard exponential, a Pareto ($x_{min} = 1, \alpha = 3$), the standard uniform and standard Gaussian distribution (cp. Appendix B) for $n \in \{1, 2, 3, 4, 5, 10, 25, 50\}$ (bottom left) and $n \in \{1, 2, 5, 10, 50, 100, 500, 1,000\}$ (otherwise) with darkening lines as n increases.

Norming constants A non-trivial limit distribution for the maximum M_n is only expected under shifting and scaling in equation (7.2). Let $a_n \in \mathbb{R}_{>0}$ and $b_n \in \mathbb{R}$, $n \in \mathbb{N}$. In what follows, we refer to a shifted and scaled maximum $a_n(M_n - b_n)$ as *normed* to distinguish the denotation from *normalized* random variables with typically zero mean and unit variance. According to [48] we call a_n and b_n *norming constants*. Occasionally, the constants a_n and b_n are referred to as the *scaling constant* and the *centering constant*, respectively. For $x \in \mathbb{R}$ the cdf of a normed maximum reads

$$F_{a_n(M_n - b_n)}(x) = \mathbb{P}(a_n(M_n - b_n) \leq x) = F_X \left(\frac{x}{a_n} + b_n \right)^n. \quad (7.5)$$

For appropriately chosen norming constants possible limits as $n \rightarrow \infty$ in (7.5) narrow down to extreme value distributions as we recap in Section 7.1.1. If a pdf f_X exists, then description (7.5) implies that the pdf $f_{a_n(M_n - b_n)}$ of normed maxima of iid samples reads

$$f_{a_n(M_n - b_n)}(x) = \frac{n}{a_n} F_X \left(\frac{x}{a_n} + b_n \right)^{n-1} f_X \left(\frac{x}{a_n} + b_n \right) \quad (7.6)$$

for all values $x \in a_n(U - b_n)$ in that $F \left(\frac{x}{a_n} + b_n \right)$ is differentiable (U is the domain of the pdf f_X in equation (7.3)).

7. Extreme value theory

7.1.1. Extreme value distributions

The basis of EVT is the crucial observation by Fisher and Tippett [52] that if a limit distribution exists for appropriately normed maxima M_n (7.5), then it is stable under maximization.

Heuristics. For numbers $m, n \in \mathbb{N}$ let $X_{i,j}$, $i = 1, \dots, n$, $j = 1, \dots, m$, be iid random variables and let $M_i := \max\{X_{i,1}, \dots, X_{i,m}\}$, $i = 1, \dots, n$, be n maxima of m -tuples of these variables. Assuming a common cdf F for the distribution of each maximum M_i , we have $P(M_i \leq x) = F(x)$, $x \in \mathbb{R}$. Then by equation (7.2), $P(M_{nm} \leq x) \approx F(x)^n$ for large numbers n , in which $M_{nm} := \max\{M_1, \dots, M_n\}$ denotes the maximum (7.7) of n maxima of m samples each and likewise the maximum of $n \cdot m$ samples. If the distributions of both all M_i and M_{nm} share their functional form, then their cdfs F and F^n , respectively, coincide up to a linear rescaling. (\square)

The concept of *max-stability* describes the property

$$M_{nm} = \max \left\{ \underbrace{\max\{X_{1,1}, \dots, X_{1,m}\}}_{\sim F}, \dots, \underbrace{\max\{X_{n,1}, \dots, X_{n,m}\}}_{\sim F} \right\} \sim F^n \quad (7.7)$$

of stability under maximization. The symbol \sim in scheme (7.7) denotes ‘following a distribution with cdf’.

Definition (Max-stability): A distribution with cdf F is called *max-stable* if for all exponents $n \in \mathbb{N}$ there are constants $c_n \in \mathbb{R}_{>0}$ and $d_n \in \mathbb{R}$ such that

$$F(x)^n = F(c_n x + d_n). \quad (7.8)$$

By equation (7.2) distributions satisfying equation (7.8) are stable under maximization up to an affine transformation. As a remark, max-stability relates to the well-known concept of α -stability of random variables under summation (cp. Sect. 5.1) by considering maximization instead. Solving (7.8) for the cdf F , a classification of max-stable distributions into three types emerges [52]. Usually they are enumerated as *Type-I*, *Type-II* and *Type-III extreme value distributions* (EVD) and typically named the *Gumbel*, *Fréchet* and *Weibull* class, respectively. Table 7.1 states a corresponding standard cdf G for each of the types. Including location and scale parameters $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}_{>0}$ and applying $G\left(\frac{x-\mu}{\sigma}\right)$ to these cdfs generates the full variety of EVDs.

extremal type		standard cdfs ($\alpha > 0$)	characteristics ($n \in \mathbb{N}$)
Type I	(Gumbel)	$\Lambda(x) = e^{-e^{-x}}, \quad x \in \mathbb{R}$	$\Lambda(x)^n = \Lambda(x - \ln n)$
Type II	(Fréchet)	$\Phi_\alpha(x) = \begin{cases} 0, & x < 0 \\ e^{-x^{-\alpha}}, & x \geq 0 \end{cases}$	$\Phi_\alpha(x)^n = \Phi_\alpha(n^{-1/\alpha}x)$
Type III	(Weibull)	$\Psi_\alpha(x) = \begin{cases} e^{-(-x)^\alpha}, & x \leq 0 \\ 1, & x > 0 \end{cases}$	$\Psi_\alpha(x)^n = \Psi_\alpha(n^{1/\alpha}x)$

Table 7.1 The standard cdfs of the three extreme value distributions.

The key difference between the three types of EVDs lies in the tails of their pdfs, that capture the probability of the occurrence of high extremes. The Gumbel class contains light-tailed EVDs with exponentially decaying tails, while the Fréchet class comprises fat-tailed EVDs with power-law tails. In either case, we have $x_m = \infty$ in definition (7.4), whereas EVDs of Weibull type exhibit a finite upper endpoint $x_m < \infty$.

Considering the maximum M_n for an EVD by applying relation (7.2) reflects further differences in their characteristics. As Table 7.1 (right column) illustrates, pdfs of maxima of Type-I EVDs are shifted rightwards as n increases. In contrast, Type-II and Type-III distributions undergo an increasing and decreasing scale, respectively.

Generalized extreme value distribution An efficient representation of the three types of EVDs is provided by the *generalized extreme value* (GEV) distribution, which captures them all in a single formula.

Definition (Generalized extreme value distribution): Let $\mu, \xi \in \mathbb{R}$ and $\sigma \in \mathbb{R}_{>0}$. Then members of the family of distributions with cdfs of the form

$$G_\xi(x) = \begin{cases} e^{-(1 + \xi(\frac{x-\mu}{\sigma}))^{-1/\xi}}, & \xi \neq 0 \\ e^{-e^{-\frac{x-\mu}{\sigma}}}, & \xi = 0 \end{cases} \quad (7.9)$$

are called *generalized extreme value* (GEV) *distributions*. The parameters denote the *location* (μ), *scale* (σ) and *shape* (ξ) of the distribution. Depending on these parameters the domain of a GEV distribution is $\{x \in \mathbb{R} \mid 1 + \xi(\frac{x-\mu}{\sigma}) > 0\}$. A GEV distribution with parameters μ, σ and ξ is denoted by $\text{GEV}(\mu, \sigma, \xi)$.

7. Extreme value theory

The shape parameter ξ of a GEV distribution is occasionally referred to as the *extreme value index* [19, 32], which is not to be confused with the *extremal index* of a time series (cp. Sect. 7.2.1). Shape parameters $\xi > 0$ and $\xi < 0$ yield Type-II and Type-III distributions, respectively, with $\alpha = \frac{1}{\xi}$ and $\text{GEV}(\mu + \sigma, \frac{\sigma}{\alpha}, \frac{1}{\alpha})$ representing the Fréchet and Weibull family each with location $\mu \in \mathbb{R}$ and scale $\sigma \in \mathbb{R}_{>0}$. The limit case of $\xi \rightarrow 0$ results in the Type-I or Gumbel class. Figure 7.3 shows an example pdf for each regime of the shape parameter ξ .

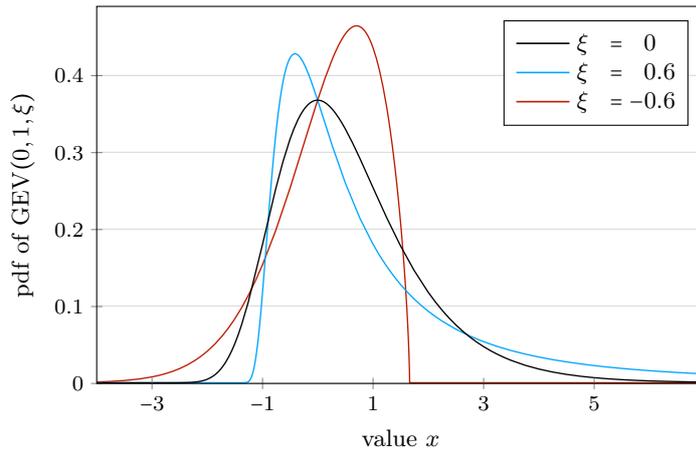


Fig. 7.3 Example Gumbel ($\xi = 0$), Fréchet ($\xi > 0$) and Weibull ($\xi < 0$) type pdfs exhibiting an exponential, power-law and finite upper tail, respectively.

GEV distributions are max-stable as a short calculation validates property (7.8). Moreover, max-stable distributions are precisely the GEV distributions.

Theorem 7.2 (Thm. 1.4.1 in [101]): *A distribution is max-stable if and only if it is a member of the GEV family.*

The central conclusion of classical EVT is the *Fisher-Tippett-Gnedenko* or *extremal types theorem* Theorem 7.3. It states that non-degenerate limit distributions of appropriately normed maxima are of GEV type. An example of a *degenerate* limit is given by distributions with finite upper endpoint (7.4) $x_m < \infty$ and a jump in that point [101, p. 12]. For any choice of norming constants the cdf (7.5) of the normed maximum of variates with such a distribution approaches a degenerate limit cdf, which attains only the two values zero and one. Another example with a degenerate limit but more prominent in application is the Poisson distribution (cp. Rem. 7.8, Appx. F and Ex. 1.7.14 in [101]).

Theorem 7.3 (Fisher-Tippett-Gnedenko or extremal types theorem [101]): *For a number $n \in \mathbb{N}$ of iid random variables X_i , $i = 1, \dots, n$, let $M_n = \max\{X_1, \dots, X_n\}$ denote their maximum. If for some constants $a_n \in \mathbb{R}_{>0}$ and $b_n \in \mathbb{R}$ we have*

$$P(a_n(M_n - b_n) \leq x) \xrightarrow{d} G(x) \quad (n \rightarrow \infty) \quad (7.10)$$

for all $x \in \mathbb{R}$ and a non-degenerate cdf G , then G belongs to the GEV family (7.9). Conversely, each cdf G of a GEV distribution appears as a limit in (7.10).

(The symbol \xrightarrow{d} in (7.10) denotes convergence in distribution, i.e., pointwise convergence of the cdf (7.5) of the normed maximum to the cdf G in all the continuity points of G .)

Remarkable about Theorem 7.3 is that if norming constants exist such that a non-degenerate limit cdf is attained in (7.10), then it will be of GEV type whatever the distribution of the samples X_i . This observation originates from [52] and was proved in full generality in [58]. Due to its universality, Theorem 7.3 readily provides a tool for statistical inference about the distribution of extreme events. Section 7.4 delineates the application of EVT. By Khintchine's theorem Theorem 7.4 the limit EVD in the extremal types theorem is unique up to an affine transformation.

Theorem 7.4 (Khintchine's theorem): *Let $(F_n)_{n \in \mathbb{N}}$ be a sequence of cdfs F_n and G a non-degenerate cdf and let $a_n \in \mathbb{R}_{>0}$ and $b_n \in \mathbb{R}$ be norming constants such that in all continuity points $x \in \mathbb{R}$ of G*

$$F_n\left(\frac{x}{a_n} + b_n\right) \xrightarrow{d} G(x) \quad (n \rightarrow \infty). \quad (7.11)$$

Then for a non-degenerate cdf G_* and norming constants $\alpha_n \in \mathbb{R}_{>0}$ and $\beta_n \in \mathbb{R}$ we find

$$F_n\left(\frac{x}{\alpha_n} + \beta_n\right) \xrightarrow{d} G_*(x) \quad (n \rightarrow \infty) \quad (7.12)$$

if and only if

$$\frac{a_n}{\alpha_n} \longrightarrow a \quad \text{and} \quad a_n(\beta_n - b_n) \longrightarrow b \quad (n \rightarrow \infty) \quad (7.13)$$

for some $a \in \mathbb{R}_{>0}$ and $b \in \mathbb{R}$, and then

$$G_*(x) = G\left(\frac{x}{a} + b\right). \quad (7.14)$$

7. Extreme value theory

Khinchine's theorem Theorem 7.4 implies that if norming constants exist such that convergence (7.10) is valid, then they are unique up to the relations (7.13). Therefore, each possible choice of norming constants yields the same type of limit EVD by the max-stability of the GEV family (Thm. 7.2). The exact type of the limit EVD G in convergence (7.10) is uniquely determined by the tail of the distribution of the samples X_i in Theorem 7.3. Section 7.1.2 introduces specific characterizations of the tail that identify the limit type. Furthermore, the choice of norming constants crucially influences the pointwise and uniform rate of the convergence (7.10) as the Sections 7.3 describes. A deliberate choice of norming constants is the essence of the method for the estimation of return levels established in Section 9.3.

Example 7.5: The exponential, Pareto and uniform distributions in Example 7.1 and Figure 7.2 are prototypes of the three classes of EVD (Tab. 7.1). Let F denote the cdf of the $\text{Exp}(\lambda)$ (with rate $\lambda \in \mathbb{R}_{>0}$), $\text{Par}(1, \alpha)$ (with shape $\alpha \in \mathbb{R}_{>0}$) or the uniform distribution, respectively. Taking arguments $x \in \mathbb{R}$ in the support of these distributions and using the sequence representation $(1 + \frac{x}{n})^n \rightarrow e^x$ ($n \rightarrow \infty$), for large numbers $n \in \mathbb{N}$ of samples the cdfs F^n of their maxima approach

$$\begin{aligned}
 \text{Exp}(\lambda) \quad F(x)^n &= (1 - e^{-\lambda x})^n \approx e^{-ne^{-\lambda x}} = e^{-e^{-\lambda(x - \frac{1}{\lambda} \ln n)}} && \text{(Type I)} \\
 \text{Par}(1, \alpha) \quad F(x)^n &= (1 - x^{-\alpha})^n \approx e^{-nx^{-\alpha}} = e^{-\left(n^{-1/\alpha} x\right)^{-\alpha}} && \text{(Type II)} \\
 \text{U}(0, 1) \quad F(x)^n &= (1 - (1 - x))^n \approx e^{-n(1-x)} = e^{-(-n(x-1))} && \text{(Type III)}.
 \end{aligned} \tag{7.15}$$

With norming constants $a_n = \lambda$ and $b_n = \frac{1}{\lambda} \ln n$ the exponential distribution is of Gumbel type. The uniform distribution normed by $a_n = n$ and $b_n = 1$ is of Weibull type with parameter $\alpha = 1$. Note that for the power-law Pareto distribution $\text{Par}(x_{\min}, \alpha)$ with $x_{\min} \in \mathbb{R}_{>0}$, the tail parameter α directly translates to the tail parameter of the Fréchet class by using $b_n = 0$ and $a_n = n^{-1/\alpha}$.

7.1.2. Domains of attraction

A distribution with cdf F is said to belong to the *domain of attraction* of one of the three types of EVDs (Tab. 7.1) if norming constants exist such that the limit GEV distribution of its normed maxima in the extremal types theorem Theorem 7.3 is of that type. In the iid setting considering the tail of the survival function $1 - F$ suffices to determine to which of the classes the distribution belongs to [48, 108]. Lemma 7.6 describes this effect.

Lemma 7.6 (Lem. 1.2.2 in [102]): *Let $u_n \in \mathbb{R}$, $n \in \mathbb{N}$, be constants and $\tau \in \mathbb{R}_{\geq 0}$. If X_i , $i = 1, \dots, n$, are iid random variables with common cdf F and $M_n = \max\{X_1, \dots, X_n\}$, then*

$$\mathbb{P}(M_n \leq u_n) \rightarrow e^{-\tau} \quad (n \rightarrow \infty) \quad (7.16)$$

if and only if

$$n(1 - F(u_n)) \rightarrow \tau \quad (n \rightarrow \infty). \quad (7.17)$$

Heuristics. The convergence (7.17) is equivalent to $1 - \frac{\tau}{n} \sim F(u_n)$ as $n \rightarrow \infty$. Then the equivalence of the asymptotics (7.16) and (7.17) is evident by

$$\mathbb{P}(M_n \leq u_n) = F(u_n)^n \sim \left(1 - \frac{\tau}{n}\right)^n \rightarrow e^{-\tau} \quad (7.18)$$

due to the sequence representation of the exponential function. (□)

Beyond the general conditions of Lemma 7.6 and Lemma F.1 there are specific criteria for the rate of the decrease of the survival function $1 - F$, such that a distribution belongs to the domain of attraction of one of three EVD types. Table F.1 in Appendix F gives an overview of these conditions. If the distribution is absolutely continuous with respect to the Lebesgue measure, then sufficient conditions involving the pdf for belonging to a GEV class are available and called *von Mises conditions* [48]. Norming constants along with the related GEV class are analytically known for many common distributions. For a detailed list of conditions see [48, Sect. 3.3]. Table 7.2 gives some examples for each class.

shape of GEV class	example distributions
$\xi = 0$	Gaussian, Lognormal, Weibull, Gamma, exponential and truncated-Gaussian-power distribution, distributions with exponential tail and finite upper endpoint
$\xi > 0$	Pareto, Cauchy and Burr distribution, α -stable distributions with $\alpha < 2$
$\xi < 0$	uniform and Beta distribution, distributions with power-law tail and finite upper endpoint

Table 7.2 Example distributions for each of the three domains of attraction of EVDs.

7. Extreme value theory

Note that the convergence (7.10) in the extremal types theorem Theorem 7.3 is a special case of Lemma 7.6 for the choice of linearly normed thresholds

$$u_n(x) = \frac{x}{a_n} + b_n \quad (7.19)$$

and tail behavior $\tau(x) = -\ln(G(x))$, $x \in \mathbb{R}$, of a GEV cdf G . For cdfs F in the domain of attraction of each of the three types of EVDs it is possible to choose a sequence of linearly transformed thresholds $u_n = \gamma_n \in \mathbb{R}$ in equation (7.19) such that $n(1 - F(\gamma_n)) \rightarrow 1$ as $n \rightarrow \infty$ [101, p. 17]. The related norming constants clearly depend on the specific type of the EVD (cp. Tab. 7.3) with $\gamma_n \rightarrow x_m$ as $n \rightarrow \infty$ each. In either case, such a choice implies

$$F(\gamma_n) \approx 1 - \frac{1}{n} \quad (7.20)$$

for large numbers $n \in \mathbb{N}$. Note that the $(1 - \frac{1}{n})$ -quantile represents the empirical maximum of n samples [48, p. 129]. We modify this convenient choice of thresholds when establishing our method for the estimation of return levels of tGp processes in Section 9.3.

Type II	$a_n = \frac{1}{\gamma_n}$	$b_n = 0$
Type III	$a_n = \frac{1}{x_m - \gamma_n}$	$b_n = x_m$
Type I	$a_n = \frac{1}{h(\gamma_n)}$	$b_n = \gamma_n$

Table 7.3 Possible choices of norming constants $a_n \in \mathbb{R}_{>0}$ and $b_n \in \mathbb{R}$ for each of the three EVDs by defining $\gamma_n = F^{-1}(1 - \frac{1}{n})$ using relation (7.20). The function h is the auxiliary function of a cdf F in the Gumbel domain of attraction [48, p. 141]. For an appropriate choice of the function h see Table F.1 [101, Thm. 1.6.2].

Example 7.7: It is well known that the Gaussian distribution belongs to the Gumbel domain of attraction [52]. For the standard Gaussian with pdf ϕ one appropriate choice of norming constants is implicitly defined by the equations

$$b_n = n\phi(b_n) \quad \text{and} \quad a_n = b_n \quad (7.21)$$

for $n \in \mathbb{N}$ iid samples [69]. These norming constants are the first summands of a series expansion of the constants proposed in Table 7.3 and discussed in more detail in Section 7.3. Applying equation (7.6), Figure 7.4 shows how the pdf of the normed maxima of standard Gaussian samples collapses to the pdf of the Gumbel distribution as the number of samples increases.

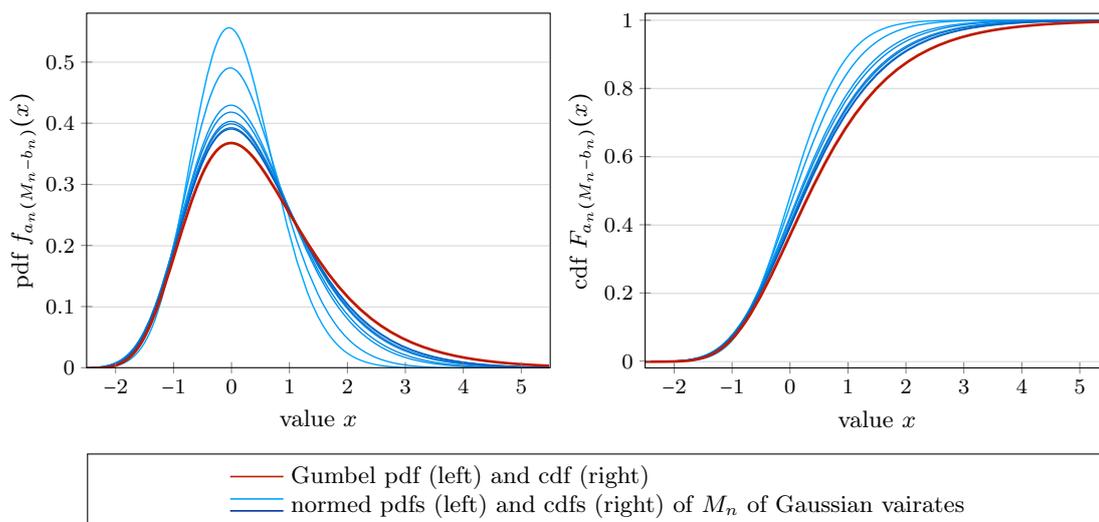


Fig. 7.4 Evolution of the pdfs $f_{a_n(M_n-b_n)}$ (7.6) (left) and cdfs $F_{a_n(M_n-b_n)}$ (7.5) (right) of the normed maxima $a_n(M_n - b_n)$ with norming constants (7.21) of $n \in \mathbb{N}$ iid samples of the standard Gaussian distribution for numbers $n \in \{1, 2, 5, 10, 50, 100, 500, 1,000\}$ with darkening lines as n increases.

The *error* between the cdf (7.5) of the normed maximum $a_n(M_n - b_n)$ and the limit GEV distribution in the extremal types theorem Theorem 7.3 for finite numbers $n \in \mathbb{N}$ of samples depends on the specific choice of the norming constants a_n and b_n . In Sections 7.3 we discuss different choices of norming constants for maxima of Gaussian samples. Section 8.1 on norming constants for certain meta-Gaussian processes prepares the assessment of the *rate* of the convergence (7.10) for tGp processes in Section 8.2. A specific choice of norming constants in Section 9.3.1 helps decrease the statistical uncertainty when estimating return levels of tGp processes.

Remark 7.8: There are distributions that are not in any of the three domains of attraction of EVDs. In particular, for discrete distributions with infinite upper endpoint every valid choice of exceedances u_n in Lemma 7.6 may yield a degenerate limit distribution for their transformed maxima. The Poisson and geometric distribution are such examples with broad relevance to application (cp. Exs. 1.7.14 and 1.7.15 in [101]). The speed of the decay of the jump heights of their survival functions implies the violation of the assumption of non-degeneracy of the limit distribution in Theorem 7.3 (see Lem. F.1 in Appx. F). A different violation of Theorem 7.3 occurs if only nonlinearly transformed thresholds u_n induce the convergence (7.17). A related example according to von Mises [102, p. 435] is the distribution with cdf $F(x) = 1 - e^{-x - \sin x}$. For this cdf and continuous cdfs F in general the typically nonlinear choice

$$u_n(x) = F^{-1} \left(e^{-\frac{\tau(x)}{n}} \right) \quad (7.22)$$

7. Extreme value theory

always yields pointwise convergence of $P(M_n \leq u_n)$ to the limit cdf $e^{-\tau(x)}$ [101, p. 36]. Note that for large $n \in \mathbb{N}$

$$e^{-\frac{\tau(x)}{n}} = \left(e^{-\tau(x)}\right)^{\frac{1}{n}} \approx \left(1 - \frac{\tau(x)}{n}\right)^{n \cdot \frac{1}{n}} = 1 - \frac{\tau(x)}{n}, \quad (7.23)$$

so that $n(1 - F(u_n(x))) = \tau(x)$ for all $x \in \mathbb{R}$. The choice (7.22), however, is only of the linear form (7.19) if and only if F is the cdf of a GEV distribution itself even though Lemma 7.6 applies [156]. We study equation (7.23) in more detail in Section 7.3.

7.2. Extreme value theory for stationary processes

Section 7.1 above introduces classical EVT for the maxima of iid samples. Now let $(X_t)_{t \in \mathbb{N}}$ be a stationary process (cp. Sect. 2.2) with marginal cdf F in the domain of attraction of one of the three EVD distributions in Table 7.1. Under certain weak conditions on the dependence among the different samples X_t the extremal types theorem Theorem 7.3 generalizes to stationary processes in the sense that their maxima follow a GEV distribution asymptotically. There are several concepts of describing dependencies between different variates of a stochastic process, such as m -dependence, the Markov property and strong or distributional mixing (cp. Appx. D). For each of them generalizations of the classical EVT are available [101, Sect. 3.1].

7.2.1. The conditions $D(u_n)$ and $D'(u_n)$

For a number $k \in \mathbb{N}$ of indices $i_1, \dots, i_k \in \mathbb{N}$ and a threshold $u \in \mathbb{R}$ let

$$F_{i_1, \dots, i_k}(u) := F_{(X_{i_1}, \dots, X_{i_k})}(u) := P(\{X_{i_j} \leq u \mid j = 1, \dots, k\}) \quad (7.24)$$

denote the joint cdf of random variates X_{i_1}, \dots, X_{i_k} .

Let $(u_n)_{n \in \mathbb{N}}$ be a sequence of thresholds $u_n \in \mathbb{R}$. As a relaxation of strong mixing (cp. Appx. D) Leadbetter [100] introduced the following *distributional mixing* condition for such sequences of thresholds.

Definition (Condition $D(u_n)$): The condition $D(u_n)$ is said to hold if for any set of indices $i_1, \dots, i_p, j_1, \dots, j_q \in \mathbb{N}$ with $p, q \in \mathbb{N}$ such that

$$1 \leq i_1 < \dots < i_p < j_1 < \dots < j_q \leq n \quad (7.25)$$

and $j_1 - i_p \geq l \in \mathbb{N}$ we have

$$|F_{i_1, \dots, i_p, j_1, \dots, j_q}(u_n) - F_{i_1, \dots, i_p}(u_n) \cdot F_{j_1, \dots, j_q}(u_n)| \leq c_{n,l} \quad (7.26)$$

for a function $c_{n,l}$ such that $c_{n,l_n} \rightarrow 0$ as $n \rightarrow \infty$ for some sequence $(l_n)_{n \in \mathbb{N}}$ with $l_n = o(n)$.

Condition $D(u_n)$ implies a degree of dependence between the maxima of the process X_t on distinct subintervals of the (temporal) indices $t \in \mathbb{N}$ [101, p. 54]. If the condition $D(u_n)$ is satisfied for thresholds u_n of affine linear form, then the conclusion of the extremal types theorem Theorem 7.3 transfers to stationary processes as captured by Theorem 7.9.

Theorem 7.9 (cp. [101]): For a number $n \in \mathbb{N}$ let $M_n := \{X_1, \dots, X_n\}$ denote the maximum of n samples of a stationary process $(X_t)_{t \in \mathbb{N}}$. Suppose there exist norming constants $a_n \in \mathbb{R}_{>0}$ and $b_n \in \mathbb{R}$ such that

$$\mathbb{P}(a_n(M_n - b_n) \leq x) \xrightarrow{d} G(x) \quad (n \rightarrow \infty) \quad (7.27)$$

for all $x \in \mathbb{R}$ and a non-degenerate cdf G . If for all values $x \in \mathbb{R}$ the condition $D(u_n)$ is satisfied for the sequence $(u_n)_{n \in \mathbb{N}}$ of affine linear thresholds $u_n = \frac{x}{a_n} + b_n$, then the cdf G is an EVD of one of the three types in Table 7.1.

The domains of attraction for stationary sequences trace back to the results for the iid setting of Section 7.1.2 by considering an iid sequence with the same marginal cdf F like the stationary process X_t .

Definition (Associated independent sequence): For a stationary process $(X_t)_{t \in \mathbb{N}}$ with marginal cdf F an iid sequence $(X_t^*)_{t \in \mathbb{N}}$ with the same marginal cdf F but without dependence among the different samples X_t^* is called the *associated independent sequence* of the process $(X_t)_{t \in \mathbb{N}}$. For a number $n \in \mathbb{N}$ we denote the related maximum of an n -tuple of samples X_1^*, \dots, X_n^* by

$$M_n^* := \max \{X_1^*, \dots, X_n^*\}. \quad (7.28)$$

7. Extreme value theory

Theorem 7.10 concludes that assuming condition $D(u_n)$, the domains of attraction of a stationary process and its associated iid sequence coincide and thus depend on the tail of their common marginal distribution only.

Theorem 7.10 (cp. [37, Thm. 5.2] and [101, Thm. 3.7.1]): *Let $(X_t)_{t \in \mathbb{N}}$ be a stationary process and $(X_t^*)_{t \in \mathbb{N}}$ the associated iid sequence. Suppose there exists norming constants $a_n \in \mathbb{R}_{>0}$ and $b_n \in \mathbb{R}$ such that the assumptions of Theorem 7.9 are valid. If the condition $D(u_n)$ is satisfied for the sequence of thresholds $u_n = \frac{x}{a_n} + b_n$ for all $x \in \mathbb{R}$, then*

$$\mathbb{P}(a_n(M_n^* - b_n) \leq x) \xrightarrow{d} G_1(x) \quad (n \rightarrow \infty) \quad (7.29)$$

if and only if

$$\mathbb{P}(a_n(M_n - b_n) \leq x) \xrightarrow{d} G_2(x) \quad (n \rightarrow \infty), \quad (7.30)$$

where G_1 and G_2 are non-degenerate cdfs of GEV distributions and

$$G_2(x) = G_1^\theta(x) \quad (7.31)$$

for a constant $\theta \in (0, 1]$.

Note that by the max-stability (7.8) the power G^θ of a $\text{GEV}(\mu, \sigma, \xi)$ cdf G is an EVD of the same type, namely $\text{GEV}\left(\mu - \frac{\sigma}{\xi}(1 - \theta^{-\xi}), \sigma\theta^\xi, \xi\right)$ for shape parameters $\xi \neq 0$ and a $\text{GEV}(\mu + \sigma \ln \theta, \sigma, 0)$ if $\xi = 0$, respectively. Moreover, for maxima of stationary processes the same norming constants like for the associated iid sequence in convergence (7.29) may be used to achieve the limit GEV distribution (7.30) [101, Cor. 3.7.3].

Extremal index The constant θ in Theorem 7.10 is referred to as the *extremal index* of the process X_t and is not to be confused with the *extreme value index* ξ of a GEV distribution (cp. Sect. 7.1.1).

Definition (Extremal index): Let $(X_t)_{t \in \mathbb{N}}$ be a stationary process with marginal cdf F and let $\theta \in \mathbb{R}_{>0}$ be a constant. Suppose for all $\tau \in \mathbb{R}_{>0}$ there exists a sequence $(u_n)_{n \in \mathbb{N}}$ of thresholds u_n such that

$$\lim_{n \rightarrow \infty} \mathbb{P}(M_n \leq u_n) = e^{-\theta\tau} \quad \text{and} \quad (7.32)$$

$$\lim_{n \rightarrow \infty} n(1 - F(u_n)) = \tau. \quad (7.33)$$

Then the constant θ is called the *extremal index* of the process $(X_t)_{t \in \mathbb{N}}$.

Relation (7.32) bears resemblance to relation (7.16) of Lemma 7.6. In contrast to the iid setting, however, condition $D(u_n)$ implies just $\liminf_{n \rightarrow \infty} \mathbb{P}(M_n \leq u_n) \geq e^{-\tau}$ [101, p. 66], whereas only a scaling of the tail variable τ by the extremal index θ yields the unique limit (7.32). Due to demanding only a certain decay of the dependence between maxima of different separated parts of the processes by condition $D(u_n)$ [101, p. 54], clustering of extremes within these parts is still allowed. From a point processes perspective equation (7.33) implies that there are approximately τ exceedances of u_n among an n -tuple X_1, \dots, X_n of the process X_t [48, p. 243]. An interpretation of the extremal index is the inverse of the mean asymptotic size of such clusters of exceedances [37, p. 97]. Example 7.11 visualizes a stationary process with extremal index $\theta = \frac{1}{2}$ and mean cluster size $\frac{1}{\theta} = 2$.

Example 7.11: Let $(Y_i)_{i \in \mathbb{N}}$ be an iid sequence of standard Gaussian random variables $Y_i \sim \mathcal{N}(0, 1)$ and define a stationary process $(X_i)_{i \in \mathbb{N}}$ by $X_i := \max\{Y_i, Y_{i+1}\}$. Due to the independence of the variates Y_i , the marginal cdf of both the process X_t and its associated iid sequence X_t^* is Φ^2 . For a value $x \in \mathbb{R}$ consider the sequence of thresholds $u_n = \frac{x}{a_n} + b_n$ with the norming constants a_n and b_n defined by the equations (7.21) for $n \in \mathbb{N}$. Then the maxima M_n^* (7.28) and M_n of the two processes satisfy

$$\begin{aligned} \mathbb{P}(M_n^* \leq u_n) &= \mathbb{P}(\max\{X_1^*, \dots, X_n^*\} \leq u_n) \\ &= (\Phi(u_n))^n \quad \rightarrow \Lambda(x)^2 \quad (n \rightarrow \infty). \end{aligned} \quad (7.34)$$

$$\begin{aligned} \mathbb{P}(M_n \leq u_n) &= \mathbb{P}(\max\{X_1, \dots, X_n\} \leq u_n) \\ &= \mathbb{P}(\max\{Y_1, Y_2\} \leq u_n, \dots, \max\{Y_n, Y_{n+1}\} \leq u_n) \\ &= \mathbb{P}(\max\{Y_1, \dots, Y_{n+1}\} \leq u_n) \\ &= \mathbb{P}(\max\{Y_1, \dots, Y_n\} \leq u_n) \mathbb{P}(Y_{n+1} \leq u_n) \\ &= \Phi(u_n)^n \mathbb{P}(Y_{n+1} \leq u_n) \quad \rightarrow \Lambda(x) \quad (n \rightarrow \infty). \end{aligned} \quad (7.35)$$

Note that $\Phi(u_n)^n \rightarrow \Lambda(x)$ in the limit (7.34) and $\mathbb{P}(Y_{n+1} \leq u_n) \rightarrow 1$ in the limit (7.35) as $n \rightarrow \infty$ each. In the language of equation (7.31) in Theorem 7.10 the limits (7.34) and (7.35) yield $G_1 = \Lambda^2$ and $G_2 = \Lambda$, so that $G_2 = G_1^{1/2}$ and the extremal index of the process X_i reads $\theta = \frac{1}{2}$. Figure 7.5 visualizes the mean cluster size of $\frac{1}{\theta} = 2$ of this process.

Defining $X_i := \max\{cY_i, Y_{i+1}\}$ for $i \in \mathbb{N}$ with $c \in [0, 1]$ and iid Fréchet variates $Y_i \sim \Phi_1$ with shape and scale parameters $\alpha = 1$ and $\sigma = \frac{1}{c+1}$, respectively, and norming constants $a_n = \frac{1}{n}$ and $b_n = 0$ according to Table 7.1 provides another example of a stationary process $(X_i)_{i \in \mathbb{N}}$ with extremal index $\theta \neq 1$. Similar arguments as for the derivations (7.34) and (7.35) with $u_n = nx$ yield $\mathbb{P}(M_n^* \leq nx) = e^{-x^{-1}}$ for all $n \in \mathbb{N}$ and $\mathbb{P}(M_n \leq nx) \rightarrow e^{-((c+1)x)^{-1}}$ as $n \rightarrow \infty$, so that $\theta = \frac{1}{c+1}$ [37, Ex. 5.1].

7. Extreme value theory

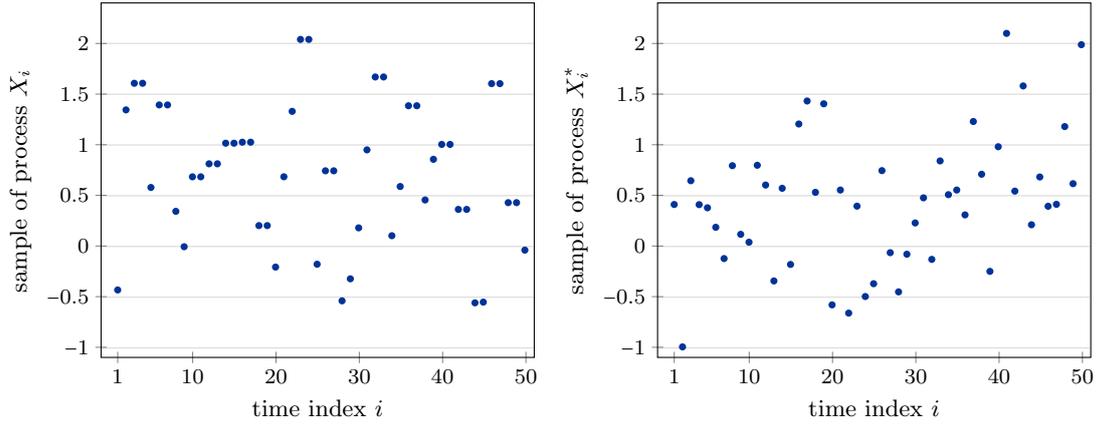


Fig. 7.5 Sample time series of length $N = 50$ the process X_i (left) in Example 7.11 and its associated iid sequence X_i^* (right) with common marginal cdf Φ^2 . The mean cluster size is $\frac{1}{\theta} = 2$, so that, in particular, extremes of the process X_i occur in pairs.

Theorem 7.10 states that using the same norming constants, in general, the limit EVD distributions of the normed maxima of the stationary process X_t and its associated iid sequence X_t^* differ in their parameters yet agree in their GEV type. It is clearly possible to adjust the norming constants of the iid process appropriately for the stationary process, such that the limit GEV distributions of the normed maxima of either processes are identical. Further conditions on the dependence structure of the process X_t imply such an equivalence of the limit distributions even with extremal index $\theta = 1$ in both relations (7.31) and (7.32), using the same norming constants for both processes.

Definition (Condition $D'(u_n)$): The condition $D'(u_n)$ is said to hold if

$$\lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} n \sum_{i=2}^{\lfloor n/k \rfloor} \mathbb{P}(X_1 > u_n, X_i > u_n) \rightarrow 0, \quad (7.36)$$

where $\lfloor \cdot \rfloor$ denotes the integer part.

Assuming condition $D(u_n)$ and following the aforementioned argumentation, by equation (7.33), there are on average $\frac{\tau}{k}$ exceedances among the variates $X_1, \dots, X_{\lfloor n/k \rfloor}$. Condition $D'(u_n)$ limits the asymptotic probability of more than one exceedance among such an $\lfloor \frac{n}{k} \rfloor$ -tuple for high thresholds [48, p. 243]. In Example 7.11 condition $D(u_n)$ is satisfied while $D'(u_n)$ is violated (cp. [48, Ex. 4.4.4] with $F = \Phi^2$). For Gaussian processes these conditions can be deduced straightforwardly from the acf (Sect. 7.2.2). In Chapter 8 we discuss the validity of the two conditions for meta-Gaussian processes.

Remark 7.12: The verification of the conditions $D(u_n)$ and $D'(u_n)$ for specific stochastic processes might be highly non-trivial. In the context of dynamical systems Lucarini et al. [108] formulate a number of alternative conditions on the dependence structure of a system, which are weaker and more easily applicable but still yield EVDs as limit laws. For example a condition $D^{(m)}(u_n)$ [108, Equ. (3.2.2)] with $m \in \mathbb{N}$ generalizes condition $D'(u_n)$ by starting the sum in definition (7.36) at $j = m + 1$ instead of $j = 2$ ($m = 1$).

Lemma 7.13 (Proposition 4.4.3 in [48]): *Let $(X_t)_{t \in \mathbb{N}}$ be a stationary process with marginal cdf F and suppose a sequence of thresholds $u_n \in \mathbb{R}$, $n \in \mathbb{N}$, satisfies conditions $D(u_n)$ and $D'(u_n)$. Then for a tail variable $\tau \in \mathbb{R}_{\geq 0}$ we find*

$$\mathbb{P}(M_n \leq u_n) \rightarrow e^{-\tau} \quad (n \rightarrow \infty) \quad (7.37)$$

if and only if

$$n(1 - F(u_n)) \rightarrow \tau \quad (n \rightarrow \infty). \quad (7.38)$$

From Theorem 7.10 and Lemma 7.13 it becomes apparent that iid processes exhibit an extremal index $\theta = 1$. The reverse statement, however, is false in general. There are non-iid stationary processes with unit extremal index. Long-range correlated ARFIMA processes (Sect. 2.5) are prominent representatives of this kind. The presence of LRC causes clustering of extremes in such time series [29] yet just such that the asymptotic extremal behavior still conforms to the associated iid sequence. Section 9.2 discusses the influence of strong short-range dependencies on the estimation of return levels.

Remark 7.14: There are generalizations of EVT for stationary to certain nonstationary processes, for example, processes of the form $X_t = \varepsilon_t + m_t$ with a Gaussian process ε_t and deterministic components $m_t \in \mathbb{R}$. Then under certain conditions on the correlations of the process ε_t and the summands m_t using an additive redefinition of the constants b_n (7.48) yields the Gumbel law for the extremal behavior of the process just as in the iid setting [101, Thm. 6.2.1].

7.2.2. The conditions $D(u_n)$ and $D'(u_n)$ for Gaussian processes

The acf provides direct access to the dependence structure of stationary Gaussian processes. Berman [23] formulated an approach to verifying the conditions $D(u_n)$ and $D'(u_n)$ for such processes solely based on their correlations.

Lemma 7.15 (cp. [48, Lem. 4.4.7(c)]): *Assume $(X_t)_{t \in \mathbb{N}}$ is a stationary Gaussian process with covariance function $r(n) := \text{Cov}(X_t, X_{t+n})$. Let $(u_n)_{n \in \mathbb{N}}$ be a sequence of thresholds $u_n \in \mathbb{R}$. If $\limsup_{n \rightarrow \infty} n(1 - \Phi(u_n)) < \infty$ and*

$$r(n) \ln n \rightarrow 0 \quad (n \rightarrow \infty), \quad (7.39)$$

then both conditions $D(u_n)$ and $D'(u_n)$ are satisfied.

The relation (7.39) is called *Berman's condition* and is equivalent to $r(n) = o\left(\frac{1}{\ln n}\right)$ as $n \rightarrow \infty$. Hence, if the correlations of a stationary Gaussian process decay faster than logarithmically, then Lemma 7.13 is valid along with Theorem 7.10 with extremal index $\theta = 1$. A variety of well-known Gaussian processes features an acf suchlike, in particular, AR and ARFIMA (Sect. 2.5) processes as members of the class of linear stationary Gaussian processes with finite variance [48, Ex. 4.4.9].

Remark 7.16 (cp. [101, Sect. 6.5]): Berman's condition (7.39) provides the almost best (i.e., weakest) possible requirement on the dependence structure of a stationary Gaussian process such that the extremal behavior is GEV-like by Lemma 7.13 [101, p. 133]. If the covariance function satisfies $r(n) \ln n \rightarrow c > 0$ as $n \rightarrow \infty$, then the distribution of the normed maxima approaches a convolution of the Gumbel and a Gaussian instead of the pure Gumbel distribution. If $r(n) \ln n \rightarrow \infty$ as $n \rightarrow \infty$, this limit distribution is even Gaussian (under certain continuity conditions on the covariance function r and using different norming constants).

7.3. Rates of convergence

Statistical inference based on the extremal types theorem Theorem 7.3 requires information about the speed of the convergence (7.10) and (7.9) of appropriately normed maxima of an increasing number of samples of an iid or stationary process. Many authors addressed this question in general settings [59, 66, 153, 156, 173] and for specific distributions [31, 45, 54, 67, 69, 105]. For example, it is well known that in case of Gaussian samples the convergence of their normed maxima is particularly slow (see Thm. 7.19) with consequences on the statistics of their extremes as described in Section 7.4.

There are different measures for the speed of convergence of random variables such as almost sure convergence and convergence in probability or distribution. The textbook [153] is dedicated to a detailed analysis of the speed of the different types of convergence in EVT including convergence of moments and densities. Note that pointwise convergence of cdfs is equivalent to convergence in distribution. For the application of the extremal types theorem Theorem 7.3 validation of pointwise limits of the normed cdfs suffices. Knowledge about theoretical rates of convergence with respect to any of the aforementioned kinds allows for error assessment when estimating return levels. In general, a slow rate indicates high inaccuracy of statistical estimates.

7.3.1. Rate of convergence in the classical setting

In case of continuous limit cdfs, such as GEV distributions, pointwise convergence implies uniform convergence (which is not true in general). Appendix C provides a proof of this essential statement. Uniform rates of convergence address a broad error assessment as they capture all quantiles at once. For our method in Section 9.3 we leave such a global view on error estimation and focus on convergence in specific quantiles.

Let F be a cdf in the domain of attraction of a GEV distribution with cdf G using norming constants $a_n \in \mathbb{R}_{>0}$ and $b_n \in \mathbb{R}$. For a finite number $n \in \mathbb{N}$ of samples and a value $x \in \mathbb{R}$, using equation (7.5), we define the *local error* between the cdf of the normed maxima $a_n(M_n - b_n)$ and the limit GEV distribution by

$$\Delta_n(x) := \left| F\left(\frac{x}{a_n} + b_n\right)^n - G(x) \right|. \quad (7.40)$$

The *global error* we denote by

$$\Delta_n := \sup_{x \in \mathbb{R}} \Delta_n(x). \quad (7.41)$$

It is convenient to perceive Lemma 7.6 as an approach to the estimation of the local error (7.40) [101, p. 36]. Given a sequence of thresholds $(u_n)_{n \in \mathbb{N}}$ for an $x \in \mathbb{R}$ with $u_n := u_n(x) = \frac{x}{a_n} + b_n$ such that $n(1 - F(u_n)) \rightarrow \tau(x) \in \mathbb{R}_{\geq 0}$ as $n \rightarrow \infty$, define a tail variable $\tau_n = \tau_n(x) := n(1 - F(u_n))$. Then by equation (7.2) we find the pointwise limit

$$\mathbb{P}(M_n \leq u_n) = F(u_n)^n = \left(1 - \frac{\tau_n}{n}\right)^n \rightarrow e^{-\tau(x)} \quad (n \rightarrow \infty). \quad (7.42)$$

Mind that the function $\tau = \tau(x)$ specifies the tail of the limit GEV distribution. As a remark, thresholds u_n more general than affine linear transformations of values x are possible but not required for our purpose.

7. Extreme value theory

Using notion (7.42) for a fixed number $n \in \mathbb{N}$ of samples, we define two different local errors in $x \in \mathbb{R}$ by

$$\Delta'_n(x) := \left| \left(1 - \frac{\tau_n(x)}{n} \right)^n - e^{-\tau_n(x)} \right| \quad \text{and} \quad (7.43)$$

$$\Delta''_n(x) := \left| e^{-\tau_n(x)} - e^{-\tau(x)} \right| \quad (7.44)$$

according to [101]. The related global errors Δ'_n and Δ''_n are defined analogously to (7.41). Then we shall distinguish between the two origins (7.43) and (7.44) of deviations in the approximation (7.10) of the limit EVD by

$$\Delta_n(x) \leq \Delta'_n(x) + \Delta''_n(x). \quad (7.45)$$

Elementary arguments imply $\left| \left(1 - \frac{z}{n} \right)^n - e^{-z} \right| = \frac{z^2 e^z}{2n} + \mathcal{O}\left(\frac{1}{n}\right)$ for $z \in \mathbb{R}$. Therefore, by definition (7.43) and inequality (7.45) the convergence in the extremal types theorem Theorem 7.3 is not faster than linear of order $\frac{1}{n}$ in general. Higher speeds of convergence, however, are possible in specific situations [65, p. 179]. If and only if F is already the cdf of a GEV distribution itself, the global error (7.41) vanishes uniformly and for all sample sizes as $\Delta_n = 0$ for all $n \in \mathbb{N}$ [156].

Example 7.17: For the exponential, Pareto and uniform distribution (cp. Ex. 7.1) with norming constants chosen alike equations (7.15) in Example 7.5 the error (7.44) vanishes as $\Delta''_n = 0$ since $\tau_n(x) = \tau(x)$ for all $n \in \mathbb{N}$. In other words, for all finite numbers n of samples we find equality in relation (7.17) of Lemma 7.6 between the tail behavior τ of the limit GEV distribution and its approximate description τ_n . Hence, in either case, the pointwise convergence is precisely linear [156]. In fact, the cdf of normed maxima of samples drawn from an exponential distribution $\text{Exp}(\lambda)$, $\lambda \in \mathbb{R}_{>0}$, converges uniformly to the Gumbel cdf at rate $\frac{1}{n} \left(2 + \frac{1}{n} \right) \frac{1}{e^2}$ independent of the rate parameter λ [69].

Remark 7.18: On the one hand every exponential rate of decay of the error (7.41) occurs in the extremal types theorem Theorem 7.3. Moreover, any rate faster than exponential implies a uniformly vanishing error $\Delta_n = 0$ for all $n \in \mathbb{N}$. On the other hand the convergence can also be arbitrarily slow [156].

The error $\Delta_n''(x)$ in (7.44) is governed by the way how the cdf $P(a_n(M_n - b_n) \leq x)$ approaches its limit $e^{-\tau(x)}$. One can show that

$$\Delta_n'' = e^{-\tau} \left((\tau - \tau_n) + \zeta(\tau - \tau_n)^2 \right) \quad (7.46)$$

for a constant $\zeta \in (0, 1)$ if $\tau - \tau_n \leq \ln 2$ [101, Thm. 2.4.2]. The choice of the norming constants determines the specific rate of the decrease of the error (7.45) of convergence (7.10) as the tail variable τ_n comprises the shape of the survival function of the normed maxima. Moreover, the local deviation $\Delta_n''(x)$ in a quantile $x \in \mathbb{R}$ crucially depends on the similarity between the tail of the distribution of the normed maxima and of the limit GEV distribution. The central idea for our method in Section 9.3 is increasing this similarity in the neighborhood of a desired quantile by a specific choice of norming constants.

Rate of convergence for iid Gaussian processes Already Fisher and Tippett observed a remarkable slowness of the convergence of the distribution of the maxima of iid Gaussian samples to their limit Gumbel shape by numerical experiments [52]. Several different pairs a_n and b_n of norming constants for the Gaussian case along with pointwise or uniform rates of convergence have been proposed ever since each with a different accuracy in the approximation. Table 7.4 lists some of these norming constants. In his seminal article [67] Hall proved that the optimal uniform rate of convergence is logarithmic of order $\frac{1}{\ln n}$ as $n \rightarrow \infty$ (see Thm. 7.19) using the norming constants (7.48). The equations (7.49) to (7.52) provide further optimal choices of norming constants.

$$a_n = \sqrt{2 \ln n} \quad b_n = \sqrt{2 \ln n} - \frac{\ln(\ln n) + \ln(4\pi)}{2\sqrt{2 \ln n}} \quad (7.47)$$

$$a_n = b_n \quad b_n = n \phi(b_n) \quad (7.48)$$

$$a_n = -\frac{\phi(b_n)}{\Phi(b_n) \ln \phi(b_n)} \quad b_n = \Phi^{-1}(e^{-1/n}) \quad (7.49)$$

$$a_n = \frac{e^{-1/n}}{n \phi(b_n)} \quad b_n = \Phi^{-1}(e^{-1/n}) \quad (7.50)$$

$$a_n = \frac{b_n}{1 + b_n^2} \quad b_n = \Phi^{-1}\left(1 - \frac{1}{n}\right) \quad (7.51)$$

$$a_n = b_n \quad b_n = \sqrt{W_0\left(\frac{n^2}{2\pi}\right)}, \quad (7.52)$$

Table 7.4 Selection of norming constants for the maxima of $n \in \mathbb{N}$ iid Gaussian samples from [101], [67], [31], [66], [54] and [188], respectively, in this order. The function W_0 in definition (7.52) denotes the principal branch of the *Lambert W function* that solves the equation $W(z)e^{W(z)} = z$ for $z \in \mathbb{R}$ [38].

7. Extreme value theory

As an introduction to the methods in the proof Theorem 7.19 below we discuss some preliminary properties of the norming constants in Table 7.4. In Section 9.3 we apply this methodology for determining asymptotic properties of norming constants for maxima of tGp samples. All the centering constants b_n defined in Table 7.4 are different approximations of the centering constants b_n in equations (7.49) and (7.50). The definition of b_n in (7.51) involves the series expansion of the exponential distribution and realizes the idea introduced in definition (7.20). The centering constant (7.48) results from applying the approximation

$$1 - \Phi(x) = \frac{\phi(x)}{x} \left(1 - \frac{1}{x^2} + \frac{1 \cdot 3}{x^4} + \dots + \frac{(-1)^k \cdot 1 \cdot 3 \cdot \dots \cdot (2k-1)}{x^{2k}} \right) + R_k(x) \quad (7.53)$$

of the survival function of the normal distribution [2, Equ. 26.2.12, p. 932] for $x \in \mathbb{R}$ and $k \in \mathbb{N}$ to definition (7.51). In absolute value the error term in equation (7.53) is less than the first neglected summand and satisfies $R_k(x) = (-1)^{k+1} \cdot 1 \cdot 3 \cdot \dots \cdot (2k+1) \int_x^\infty \frac{\phi(t)}{t^{2k+2}} dt$. Definition (7.52) provides an exact solution of equation (7.48) for b_n by the Lambert W function. The centering constant b_n (7.47), however, is a leading order approximation of the function W as follows.

Heuristics. An approach to generating an approximate solution of equation (7.48) for the centering constant b_n , $n \in \mathbb{N}$, bases on the asymptotic dominance of the exponential function [38] in

$$n^2 = 2\pi b_n^2 e^{b_n^2} \sim e^{b_n^2} \quad (n \rightarrow \infty). \quad (7.54)$$

Note that equation (7.54) results from plugging in the Gaussian pdf ϕ into definition (7.48). Taking logarithms of both sides of approximation (7.54) yields

$$b_n^2 \sim 2 \ln n + u \quad (n \rightarrow \infty) \quad (7.55)$$

for some $u \in \mathbb{R}_{>0}$ with $u \ll \ln n$. Plugging in approximation (7.55) into relation (7.54), we obtain

$$n^2 \sim 2\pi(2 \ln n + u)e^{2 \ln n + u} = 2\pi(2 \ln n + u)n^2 e^u \quad (n \rightarrow \infty). \quad (7.56)$$

Dividing both sides of approximation (7.56) by $2\pi \cdot n^2 \cdot 2 \ln n$ and a short calculation imply

$$\begin{aligned} \frac{1}{4\pi \ln n} &\sim \left(1 + \underbrace{\frac{u}{2 \ln n}}_{\ll 1} \right) e^u \sim e^u \quad (n \rightarrow \infty) \quad \text{and, thus,} \\ u &\sim \ln \frac{1}{4\pi \ln n} = -\ln 4\pi - \ln(\ln n) \quad (n \rightarrow \infty). \end{aligned} \quad (7.57)$$

In a last step we turn back to assumption (7.55), plug in the result (7.57) for the constant u and obtain

$$b_n^2 \sim 2 \ln n - \ln(\ln n) - \ln 4\pi \quad (n \rightarrow \infty). \quad (7.58)$$

Completing the squares and using the equality $\sqrt{A^2 - B^2} = A\sqrt{1 - \frac{B^2}{A^2}}$ for $A, B \in \mathbb{R}$ yields the leading order approximation of the centering constant b_n (7.48)

$$b_n = \sqrt{2 \ln n} - \frac{\ln(\ln n) + \ln(4\pi)}{2\sqrt{2 \ln n}} + \mathcal{O}\left(\frac{1}{\ln n}\right) \quad (n \rightarrow \infty) \quad (7.59)$$

as given in definition (7.47). A rigorous derivation of the result (7.59) is formulated in [40, p. 374]. (□)

In Section 9.3 we define alternative norming constants for distributions in the Gumbel domain of attraction, aiming at involving them in the estimation of return levels. The heuristics above help deduce the rate of the convergence these constants imply for the normed maxima of Gaussian and tGp samples in Section 9.3.1.

For the sake of gaining more insight into techniques for the derivation of bounds of pointwise and uniform deviations in EVT we transfer the proof of Theorem 7.19 given in [67] to our notion (7.5) of the cdf of normed maxima. In Section 8.2 we summarize existing results based on similar approaches to rates of convergence for certain meta-Gaussian processes.

Theorem 7.19 (cp. [69]): *Let a_n and b_n be the norming constants defined by the equations (7.48). Then there exist constants $C_1, C_2 \in \mathbb{R}_{>0}$ such that for every number $n \in \mathbb{N}$ the global error (7.41) between the cdf (7.5) of the normed maximum of n iid standard Gaussian samples and the limit Gumbel cdf Λ (Tab. 7.1) satisfies*

$$\frac{C_1}{\ln n} \leq \Delta_n \leq \frac{C_2}{\ln n}. \quad (7.60)$$

The upper bound in the inequality (7.60) may be specified by the constant $C_2 = 3$. The uniform rate of the convergence (7.60) cannot be improved beyond logarithmic by choosing a different sequence of norming constants.

7. Extreme value theory

Sketch of proof. The reasoning splits into two separate argumentations of the uniform bounds in the inequalities (7.60). For the *lower bound* we deduce that the rate of the convergence cannot be better than $\frac{1}{\ln n}$ as $n \rightarrow \infty$ whatever the norming constants. By Khintchine's theorem Theorem 7.4 (with $G = G_*$) all suitable norming constants $\alpha_n \in \mathbb{R}_{>0}$ and $\beta_n \in \mathbb{R}$ satisfy $\frac{a_n}{\alpha_n} \rightarrow 1$ and $a_n(\beta_n - b_n) \rightarrow 0$ as $n \rightarrow \infty$, so that

$$\alpha_n = \frac{r_n}{a_n} \quad \text{and} \quad \beta_n = b_n + \frac{\delta_n}{a_n} \quad (7.61)$$

for some real sequences $(r_n)_{n \in \mathbb{N}}, (\delta_n)_{n \in \mathbb{N}}$ such that $r_n \rightarrow 1$ and $\delta_n \rightarrow 0$ as $n \rightarrow \infty$. Hence, for all values $x \in \mathbb{R}$ by equations (7.61) the threshold $u_n := u_n(x) = \frac{x}{\alpha_n} + \beta_n$ reads

$$u_n = b_n \left(1 + \frac{1}{b_n^2} (r_n x + \delta_n) \right). \quad (7.62)$$

Approximation (7.53) of the survival function $1 - \Phi$ of the standard Gaussian distribution implies

$$\Phi(u_n)^n = \left(1 - \frac{\phi(u_n)}{u_n} \left(1 - \frac{1}{u_n^2} \right) + R_1(u_n) \right)^n \quad (7.63)$$

with $0 < |R_1(u_n)| < \frac{3\phi(u_n)}{u_n^5} = \mathcal{O}(b_n^{-4})$ as $n \rightarrow \infty$ by equations (7.62) and (7.63). Elementary but extensive calculations starting from the approximation (7.63) and involving expansions such as $\frac{1}{1+z} = 1 - z + \mathcal{O}(z^2)$ and $e^z = 1 + z + \mathcal{O}(z^2)$ as $z \rightarrow 0$, $z \in \mathbb{R}$, yield

$$\begin{aligned} & \Phi(u_n)^n - \Lambda(x) \\ &= \Lambda(x) e^{-x} \left(\frac{1}{b_n^2} (1 + x + x^2/2) + (r_n - 1)x + \delta_n + \mathcal{O}(b_n^{-4} + (r_n - 1)^2 + \delta_n^2 + 1/n) \right). \end{aligned} \quad (7.64)$$

The leading order estimate (7.64) illustrates that the pointwise convergence of the local error $\Delta_n(x)$ (7.40) cannot be faster than of rate $\frac{1}{\ln n}$ and so does the global error Δ_n (7.41). Note that by approximation (7.58) of the constants b_n

$$\frac{1}{2 \ln n} \leq \frac{1}{b_n^2} \leq \frac{1}{2 \ln n - (\ln(\ln n) + \ln 4\pi)}. \quad (7.65)$$

Setting $r_n = 1$ and $\delta_n = 0$ in equation (7.64) yields the left hand side in inequality (7.60), where for example choosing $C_1 = 0.65$ is valid since $\sup_{x \in \mathbb{R}} \Lambda(x) e^{-x} (1 + x + x^2/2) \approx 0.6527$.

Aiming at the constant $C_2 = 3$ for the *upper bound* in inequality (7.60), let $n \geq 21$ since otherwise $\frac{3}{\ln n} > 1$. Then it suffices to estimate the global error (7.41) by $\Delta_n \leq \frac{2.4}{b_n^2}$ because $b_n^2 > 0.8 \ln n$ for all $n \in \mathbb{N}_{\geq 21}$ by approximation (7.58). For $x \in \mathbb{R}$ consider the threshold $u_n = \frac{x}{a_n} + b_n$. Choosing $c_n := \ln(\ln b_n^2)$ allows for separate estimates of a uniform bound of the error $\Delta_n(x)$ for values x from one of three subintervals

$$I_1 := (-\infty, -c_n], \quad I_2 := (-c_n, 0) \quad \text{and} \quad I_3 := [0, \infty). \quad (7.66)$$

By the choice of the constants c_n and the monotonicity of both the Gumbel cdf Λ and the cdf Φ^n for quantiles $x \in I_1$ one can show that

$$\Lambda(x) \leq \Lambda(-c_n) = \frac{1}{b_n^2} \quad \text{and} \quad \Phi(u_n)^n \leq \Phi\left(-\frac{c_n}{a_n} + b_n\right)^n \leq \frac{2.08}{b_n^2}. \quad (7.67)$$

Hence, the local error on the interval I_1 is uniformly bounded by $\sup_{x \in I_1} \Delta_n(x) \leq \frac{2.08}{b_n^2}$.

For $x \in I_2 \cup I_3$ we consider the function $\Psi_n(x) := 1 - \Phi(u_n)$ and

$$\Phi(u_n)^n = e^{n \ln \Phi(u_n)} = e^{n \ln(1 - \Psi_n(x))}. \quad (7.68)$$

The monotonicity of the function Ψ_n implies $\Psi_n(x) < \Psi_n(-c_n) < 0.097$. By the series expansion $\ln(1 - z) = -z - \frac{z^2}{2} - \frac{z^3}{3} + \mathcal{O}(z^4)$ ($z \rightarrow 0$), $z \in \mathbb{R}_{<1}$, and the definition (7.48) of the constants b_n one can show that

$$n \ln(1 - \Psi_n(x)) = -n\Psi_n(x) - \mathcal{R}_n(x) \quad \text{with} \quad (7.69)$$

$$0 < \mathcal{R}_n := \mathcal{R}_n(x) \leq \frac{n\Psi_n(x)^2}{2(1 - \Psi_n(x))} \leq \frac{0.18}{b_n^2}. \quad (7.70)$$

For the sake of perceptibility we write $A_n(x) := e^{-n\Psi_n(x) + e^{-x}}$ and estimate the local error (7.40) by

$$\begin{aligned} \Delta_n(x) &= \Lambda(x) \left| e^{e^{-x}} \Phi(u_n)^n - 1 \right| \\ &= \Lambda(x) \left| e^{-n\Psi_n(x) + e^{-x}} e^{-\mathcal{R}_n} - 1 + (e^{-\mathcal{R}_n} - e^{-\mathcal{R}_n}) \right| \quad (\text{by (7.68) and (7.69)}) \\ &\leq \underbrace{\Lambda(x)}_{\nearrow 1 \text{ (} n \rightarrow \infty)} \left| A_n(x) - 1 \right| + \underbrace{\Lambda(x)}_{\in (0,1)} \left| e^{-\mathcal{R}_n} - 1 \right| \\ &\leq \Lambda(x) |A_n(x) - 1| + \mathcal{R}_n \quad (\text{by expansion of } e^{-\mathcal{R}_n} \text{ (} \mathcal{R}_n \rightarrow 0)) \\ &\leq \frac{\kappa}{b_n^2} + \frac{0.18}{b_n^2} \quad (\text{by (7.70)}). \end{aligned} \quad (7.71)$$

One can show that a valid choice of the constant κ in estimate (7.71) is 0.84 if $x \in I_2$ and 1.39 if $x \in I_3$. Hence, the local error (7.41) on the intervals I_2 and I_3 satisfies $\sup_{x \in I_2 \cup I_3} \Delta_n(x) \leq \frac{1.57}{b_n^2}$ uniformly. \square

7. Extreme value theory

Remark 7.20: The rate of the convergence in the extremal types theorem Theorem 7.3 indeed depends on the choice of norming constants. There exist norming constants that are not optimal and cause lower rates of convergence. For maxima of iid Gaussian samples the constants a_n and b_n defined by equations (7.47) yield

$$\Delta_n(x) \sim \frac{\Lambda(x)e^{-x}}{16} \frac{(\ln(\ln n))^2}{\ln n} \quad (n \rightarrow \infty) \quad (7.72)$$

for the local error (7.40) in $x \in \mathbb{R}$ [101, Equ. (2.4.8)], which is significantly slower than $\frac{1}{\ln n}$.

The central idea of the proof of Theorem 7.19 on the best possible rate of the convergence (7.10) in the extremal types theorem Theorem 7.3 in the Gaussian setting is tracing back the global error Δ_n (7.41) with respect to any choice of norming constants α_n and β_n to the norming constants a_n and b_n defined by equations (7.48) using Khintchine's theorem Theorem 7.4. The latter are optimal constants in the sense that using them as norming constants for the maxima of iid Gaussian samples results in the optimal rate of convergence of $\frac{1}{\ln n}$. Figure 7.6 visualizes the pointwise convergence in the value $x = 0$ for these norming constants. The derivation of the uniform upper bound, however, mainly bases on estimates of specific functions such as $z^{-1} \ln(\ln z)$, $z^{-1} \ln z$, $z^{-1} (\ln z)^2$ and $z^3 e^{-x^2/2}$ evaluated at certain positions $z \in \mathbb{R}$ depending on the constants b_n [67]. Similar estimates show that also the pairs of norming constants given by the equations (7.49) to (7.52) are optimal [31, 54, 66, 188] along with a sharper uniform upper bound by improving the generously chosen constant $C_2 = 3$ in inequality (7.60).

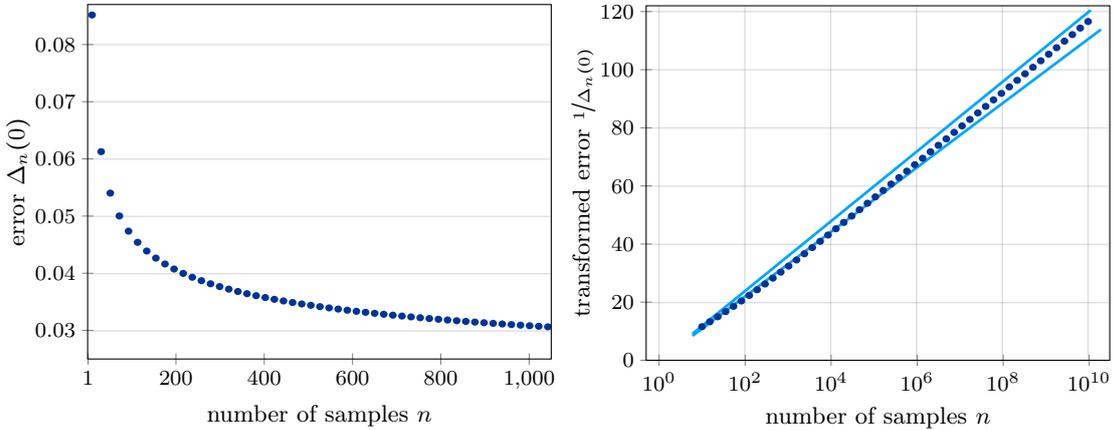


Fig. 7.6 Rate of decay of the pointwise error $\Delta_n(x)$ (7.40) in the value $x = 0$ as the number n of iid standard Gaussian samples increases (left). For the sake of perceptibility the right panel depicts the transformed error $\frac{1}{\Delta_n(x)}$. For comparison, an envelope of two straight lines of logarithmic growth is added (namely $c \ln n$ with $c \in \{4.8, 5.2\}$). Note that by estimate (7.65) the shape of the error is logarithmic only in first order with deviations on higher orders.

Remark 7.21: When using the norming constants (7.48) the global error Δ_n between the cdf of the normed maxima of iid Gaussian samples and the limit Gumbel cdf is fairly small for moderate numbers $n \in \mathbb{N}$ already (Fig. 7.6, left panel). The slow convergence at rate $\frac{1}{\ln n}$ as $n \rightarrow \infty$ causes nevertheless that a not much better approximation is gained as the number n of samples increases. More precisely, with a constant $C_2 = 3$ in inequality (7.60) the error estimate ranges from 0.65 to 0.13 for n between 100 and 10^{10} , which remains the same order of magnitude (cp. [101, p. 40]).

7.3.2. Rate of convergence for stationary processes

Section 7.3.1 separates the origins of the pointwise speed of the convergence in the extremal types theorem Theorem 7.3 for iid stochastic processes. For a finite number $n \in \mathbb{N}$ and a value $x \in \mathbb{R}$ the rather technical error $\Delta'_n(x)$ (7.43) relates the exponential of the tail of the n -time survival function of the normed maxima to their cdf $F(u_n)^n$ (7.5) with $u_n = \frac{x}{a_n} + b_n$. The error $\Delta''_n(x)$ (7.44) captures the deviation of this exponential to the limit GEV cdf. For stationary processes with extremal index $\theta = 1$ we extend this approach to the rate of the convergence in Theorem 7.10 and Lemma 7.13 by including a further influencing factor according to [102]. For appropriate norming constants $a_n \in \mathbb{R}_{>0}$ and $b_n \in \mathbb{R}$ let

$$\Delta'''_n(x) := \left| \mathbb{P}(a_n(M_n - b_n) \leq x) - F\left(\frac{x}{a_n} + b_n\right)^n \right| = |\mathbb{P}(M_n \leq u_n) - F(u_n)^n| \quad (7.73)$$

denote the error between the cdfs of the normed maxima of the stationary process and its associated iid sequence in a quantile $x \in \mathbb{R}$. Analogously to definition (7.41), the related global error shall be denoted by Δ'''_n . Then for stationary processes an estimate of the pointwise error

$$\Delta_n(x) := |\mathbb{P}(a_n(M_n - b_n) \leq x) - G(x)| \quad (7.74)$$

(denoted by the same symbol as for the iid setting in definition (7.40) with the uniform error Δ_n defined accordingly) of the convergence (7.30) to a limit GEV cdf G splits up into

$$\Delta_n(x) \leq \Delta'_n(x) + \Delta''_n(x) + \Delta'''_n(x). \quad (7.75)$$

7. Extreme value theory

The restriction to processes with extremal index $\theta = 1$ allows for the split-up (7.75) as the limit GEV cdf G remains even in the presence of dependencies by Theorem 7.10. Note that for iid processes we uniformly find $\Delta_n''' = 0$ in definition (7.73) by equation (7.2). For a choice of valid norming constants of the associated iid sequence in convergence (7.29) the division of the rate of convergence (7.30) by inequality (7.75) provides a reasonable interpretation of the origins of the speed of convergence. Deriving an intuitive description of the behavior of the error $\Delta_n'''(x)$ as the number of samples n approaches infinity is in general delicate since the probability $P(a_n(M_n - b_n) \leq x)$ represents a multivariate distribution with dependencies among the single dimensions. For stationary Gaussian processes, however, the acf provides direct access to the rate of the convergence (7.73).

Rate of convergence for stationary Gaussian processes Leadbetter et al. [101, Sect. 4.6] provide a discussion of the asymptotics of extreme events of stationary Gaussian processes. Aiming at an application to meta-Gaussian processes obtained from ARFIMA models in Section 8.2.2, we focus on Gaussian processes that exhibit LRC in the sense of definition (2.20). Due to their power-law decaying acf (2.10), ARFIMA processes satisfy Berman's condition (7.39). As a remark, Berman [23, Thm. 3.1] showed that also a square-summable acf function ϱ , namely $\sum_{k=0}^{\infty} \varrho(k)^2 < \infty$, implies the conditions $D(u_n)$ and $D'(u_n)$ and thus extremal behavior of Gumbel type with extremal index $\theta = 1$ by Lemma 7.13. The maxima of Gaussian ARFIMA processes, hence, asymptotically follow the same Gumbel distribution like those of their iid counterparts. Moreover, by Theorem 7.10 the same norming constants that are appropriate for their associated iid sequence in Theorem 7.3 are valid for norming the maxima of ARFIMA processes.

A concept of quantifying the error $\Delta_n'''(x)$ (7.73) for stationary Gaussian processes originates from measuring the distance between multivariate Gaussian vectors by comparing their covariance matrices [48, p. 216]. The *normal comparison lemma* bounds the pointwise distance between the cdfs of two general multivariate Gaussian distributions by a convenient function of their covariances. Lemma 7.22 is a special case of this lemma for comparing a stationary Gaussian process and its associated iid sequence using their acfs.

Lemma 7.22 (cp. Lem. [101, 4.2.4]): *Let $(X_t)_{t \in \mathbb{N}}$ be a stationary Gaussian process with zero mean, unit variance and acf ϱ_X and let $|\varrho_X(k)| \leq \delta := \max_{k=1, \dots, n} |\varrho_X(k)| < 1$ for an $n \in \mathbb{N}$ and $k = 1, \dots, n$. Then for each threshold $u \in \mathbb{R}$ we find*

$$|P(M_n \leq u) - \Phi(u)^n| \leq Kn \sum_{k=1}^n |\varrho_X(k)| e^{-\frac{u^2}{1+|\varrho_X(k)|}}, \quad (7.76)$$

where the constant $K \in \mathbb{R}_{>0}$ depends on the bound δ .

Defining $\rho := \sup \{0, \varrho_X(k) \mid k \in \mathbb{N}_{>0}\}$ (note that negative autocorrelations are not excluded), based on Lemma 7.22 one can deduce [155] that the rate of the decay of the error (7.73) is at least of order

$$\Delta_n''' \propto n^{-\frac{1-\rho}{1+\rho}} (\ln n)^{-\frac{\rho}{1+\rho}} \quad (n \rightarrow \infty) \quad (7.77)$$

for stationary Gaussian processes under suitable conditions [101, Thm. 4.6.3] on the growth of the sum $\sum_{k=1}^n |\varrho_X(k)|$ as $n \rightarrow \infty$ and so, in particular, for ARFIMA processes. The rate (7.77) is in general faster than the optimal logarithmic rate (7.60) of the convergence in the extremal types theorem Theorem 7.3 for Gaussian iid processes. Hence, by the relation (7.75) the rate of the convergence of the normed maxima of ARFIMA processes is dominated by, yet not worse than, the slow logarithmic convergence to the Gumbel limit for its associated iid sequence.

Figure 7.7 gives an impression of the pointwise error (7.74) using the norming constants (7.48) for the maxima of standard ARFIMA(0, d , 0) samples for different LRC parameters d . The slope of the transformed error $\frac{1}{\Delta_n(x)}$ in a logarithmic-linear scaling of the errors $\Delta_n(x)$ as the number n of samples increases depends on both the position x and the LRC parameter d . Note that the value $d = 0$ reproduces Figure 7.6. For $d > 0$ a crossover from a first faster and then logarithmic decrease is visible. The calculation of the error (7.74) is of high numerical effort⁸. The joint cdf (7.24) needed to determine $P(a_n(M_n - b_n) \leq x)$ involves the fully occupied covariance matrix (6.17) of dimension n for asymptotically large numbers n of samples.

7.3.3. Ultimate and penultimate approximations

Fisher and Tippett [52] identified the particularly slow convergence of the maxima distributions of iid Gaussian samples to their limit Gumbel shape. According to Hall [67] this convergence is logarithmic of order $\frac{1}{\ln n}$ as the number $n \in \mathbb{N}$ of samples increases to infinity (cp. Thm 7.19). Approaching the limit standard Gumbel cdf in the extremal types theorem Theorem 7.3 by a sequence of Type-I cdfs with different location and scale parameters is typically called an *ultimate approximation*. Any Type-I EVD can, however, be approximated arbitrarily well by Type-II or Type-III EVDs [102, p. 455]. Such an approach to the limit Gumbel EVD by Fréchet or Weibull EVDs (cp. Tab. 7.1) is commonly referred to as an *penultimate approximation* according to [52].

⁸done by the minimax exponentially tilted (MET) estimator in MATLAB software package mvncdf [25]

7. Extreme value theory

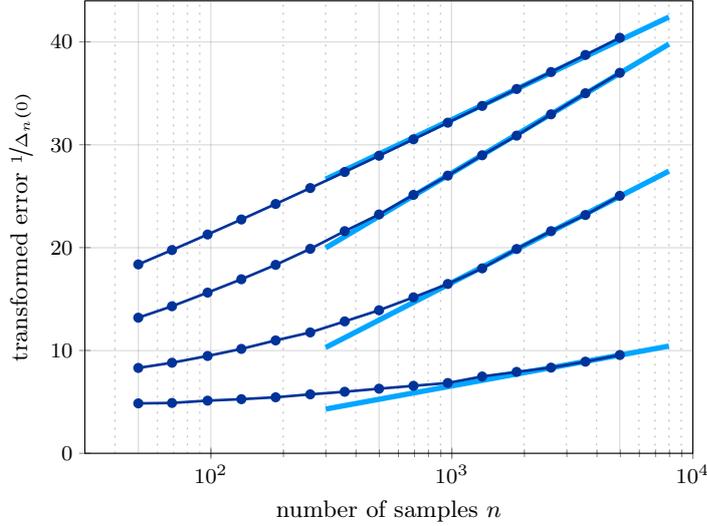


Fig. 7.7 Rate of decay of the pointwise error $\Delta_n(x)$ (7.74) in the value $x = 0$ as the number n of standard ARFIMA(0, d , 0) samples increases. For the sake of perceptibility the transformed error $\frac{1}{\Delta_n(x)}$ is depicted. The slopes estimated by linear regression are 4.8, 6.0, 5.2 and 1.9 for LRC parameters $d \in \{0, 0.1, 0.2, 0.3\}$ (from top to bottom).

Gomes and de Haan [59] provide a second-order theory for penultimate approximations of any limit GEV shape: let the cdf F of iid random samples be in the domain of attraction of a GEV distribution with cdf G_ξ (7.9) and shape parameter $\xi \in \mathbb{R}$. Under appropriate conditions (von Mises type first and second order conditions) on the tail of the survival function $1 - F$ in either case of the type of the limit GEV distribution, the cdf of the normed maxima can be better approximated by a sequence of cdfs G_{ξ_n} of different EVDs with shape $\xi_n \in \mathbb{R}$ than by the final G_ξ [59]. Defining a function $v(t) := F^{-1}(e^{-e^{-t}})$ for $t \in \mathbb{R}$, norming constants $a_n := \frac{1}{v'(\ln n)}$ and $b_n := v(\ln n)$ and a sequence of shape parameters $\xi_n := \tilde{\xi}(\ln n)$ with $\tilde{\xi}(t) := v''(t)/v'(t)$ yields

$$\lim_{n \rightarrow \infty} \frac{F^n\left(\frac{x}{a_n} + b_n\right) - G_{\xi_n}(x)}{\tilde{\xi}'(\ln n)} = c_x G'_\xi(x), \quad (7.78)$$

where $c_x \in \mathbb{R}$ is a constant dependent on the value $x \in \mathbb{R}$ and uniformly bounded on bounded intervals. Note that by von Mises' first order condition we have $\tilde{\xi}(t) \rightarrow \xi$ as $t \rightarrow \infty$, so that $\xi_n \rightarrow \xi$ and $G_{\xi_n}(x) \rightarrow G_\xi(x)$ as $n \rightarrow \infty$. Under further mild conditions on the cdf F the rate $\tilde{\xi}'(\ln n)$ in the convergence (7.78) can be replaced by $\tilde{\xi}(\ln n)^\beta$ for some exponent $\beta \in \mathbb{R}_{>0}$. For the lognormal distribution we have $\beta = 3$ (cp. Sect. 8.2.2), whereas the standard Gaussian features $\beta = 2$ as derived by Cohen [36] in Lemma 7.23. As a remark, exponential and Pareto distributions do not exhibit penultimate behavior [59].

Lemma 7.23 (cp. Thm. 2 in [36]): *For a number $n \in \mathbb{N}$ let b_n be the norming constant for the maxima of iid Gaussian samples defined by equation (7.48). Then there exists sequences $(\alpha_n)_{n \in \mathbb{N}}$, $(A_n)_{n \in \mathbb{N}}$ and $(B_n)_{n \in \mathbb{N}}$ of Weibull exponents $\alpha_n \in \mathbb{R}_{>0}$ and norming constants $A_n \in \mathbb{R}_{>0}$ and $B_n \in \mathbb{R}_{>0}$ such that for all values $x \in \mathbb{R}$*

$$|\Phi(x)^n - \Psi_{\alpha_n}(A_n(x - B_n))| = \mathcal{O}\left(\frac{1}{b_n^4}\right) = \mathcal{O}\left(\frac{1}{(\ln n)^2}\right) \quad (n \rightarrow \infty). \quad (7.79)$$

Moreover, there do not exist such sequences, so that a Fréchet approximation would yield $|\Phi(x)^n - \Phi_{\alpha_n}(A_n(x - B_n))| = \mathcal{O}\left(\frac{1}{(\ln n)^2}\right)$ as $n \rightarrow \infty$. The functions Ψ_{α_n} and Φ_{α_n} denote the cdf of Weibull and Fréchet type, respectively, with exponent α_n (cp. Tab. 7.1).

Lemma 7.23 states that a penultimate approximation of the limit Gumbel shape of the maxima of iid Gaussian samples by Weibull distributions decreases the error compared to the ultimate approximation by Type-I distributions by improving the rate of convergence from $\frac{1}{\ln n}$ in relation (7.10) to $\frac{1}{(\ln n)^2}$ in relation (7.79). A Fréchet approximation does not result in a better rate of convergence. In [35] Cohen generalizes suchlike improvements to a wide class of distributions in the Gumbel domain of attraction.

Remark 7.24: The Hill estimator $H_{k,n} := \frac{1}{k} \sum_{i=1}^k \ln\left(\frac{X_{n-i+1,n}}{X_{n-k,n}}\right)$, for a number $k \in \mathbb{N}$ with $1 \leq k \leq n$, and the upper order statistics $X_{i,n}$, $i = 1, \dots, n$, of $n \in \mathbb{N}$ iid samples, mentioned in Remark 7.26 is designed to provide an estimate of the shape parameter of a penultimate approximation of the limit GEV distribution. Conceiving this relation requires a specific definition of the penultimate shape parameters ξ_n [66, Prop. 2.2].

Figure 7.8 depicts the shape parameters ξ estimated for the maxima of several time series of iid standard Gaussian samples. Due to the improved approximation of the limit Gumbel shape by Weibull type EVDs (Lem. 7.23), the majority of the estimates obtained by a maximum likelihood are negative. For a description of the statistical inference by extreme value theory see Section 7.4 and Section 7.4.4. In Section 9.2 we apply ultimate and penultimate approaches to the estimation and comparison of return levels.

7. Extreme value theory

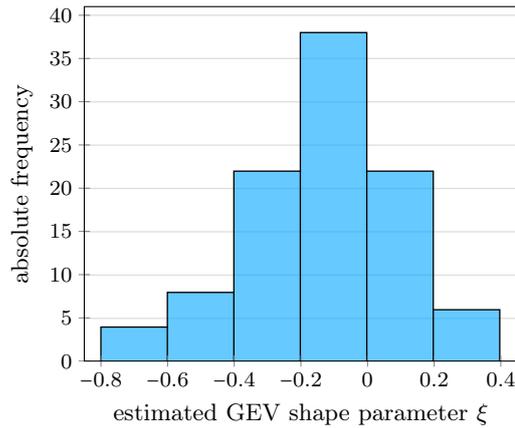


Fig. 7.8 Absolute frequencies of the penultimate GEV estimates of the shape parameter ξ for 100 maxima of $n = 365$ (cp. Sect.7.4) iid standard Gaussian samples each. The histogram bases on 100 of such estimates.

In what follows we call adjusting a Gumbel distribution with shape parameter $\xi = 0$ to the empirical maxima of time series a *Gumbel estimation*. Allowing shape parameters $\xi \neq 0$ in such a representation of the maxima distribution by Fréchet or Weibull distributions we refer to as a (*full*) *GEV estimation*.

7.4. Statistical inference by extreme value theory

The extremal types theorem Theorem 7.3 and its generalization Theorem 7.9 allow for statistical inference about the occurrence of extremes in both iid and stationary processes. Return levels and return periods are statistical quantities of interest regarding extreme events and introduced in Section 7.4.1. The universality of the distribution of maxima given by GEV distributions provides direct statistical access to the estimation of return levels.

Geophysical data sets typically consist of a single time series of measurements, for example of precipitation amounts or temperatures. Statistics for the single maximum of the entire scope of the record are accordingly unfeasible. We obtain an empirical distribution of the extremes by blocking the data into segments of a fixed lengths of typically one year and considering the maximum of each block. This concept gives rise to the notion of the *block maxima approach* in Section 7.4.2. Involving all high values above a prescribed threshold for the statistics of extremes instead of only block maxima allows for an alternative usage of EVT in Section 7.4.3. Such a *peaks over threshold* approach involves the full data if more measurements than maximal values only are available. Figure 7.9 visualizes the two concepts. We recap below how both methods base on the extremal types theorem Theorem 7.3. Theorem 7.9 implies the validity of these procedures for stationary LRC processes in Section 7.4.4.

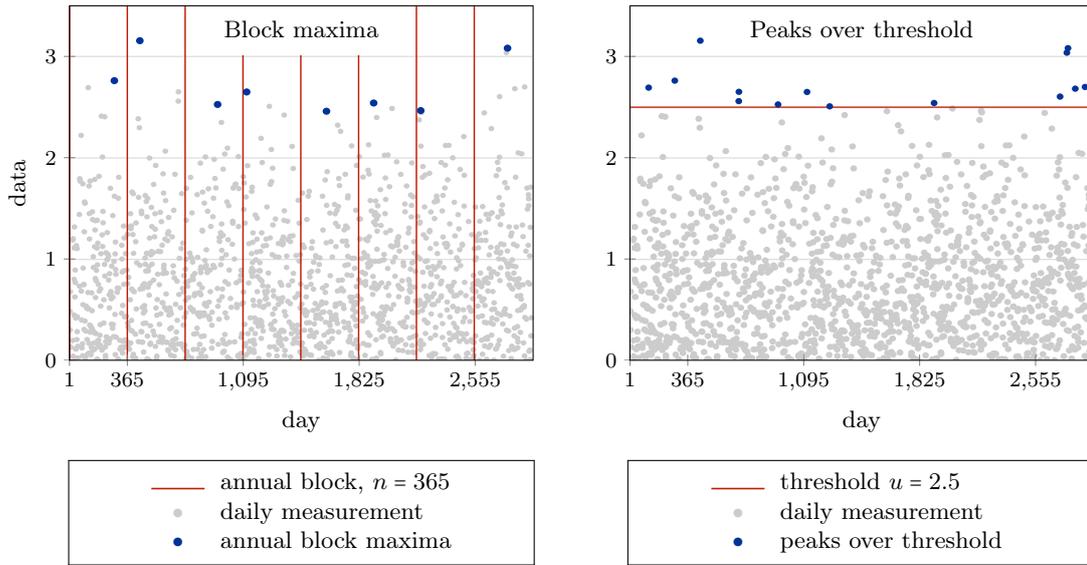


Fig. 7.9 Visualization of the block maxima (left) and peaks over threshold (right) approach for eight years of a Gaussian $N(0, 1)$ example data (only positive values depicted).

7.4.1. Return levels and return periods of extremes

A return level denotes the quantile of a distribution that is exceeded on average once during a corresponding return period.

Definition (Return level and return period): Let X be a random variable with cdf F . Given an *excess probability* $p \in [0, 1]$, a quantile $q \in \mathbb{R}$ such that

$$F(q) = P(X \leq q) = 1 - p \quad (7.80)$$

is called a *return level* with *return period* $\frac{1}{p}$.

Relation (7.2) implies the interpretation of quantiles of the maximum distribution as return levels (7.80) of extreme events of stochastic processes. For a fixed number $n \in \mathbb{N}$ a return level $q \in \mathbb{R}$ of the maximum M_n (7.1) of an n -tuple of iid samples with excess probability $p \in [0, 1]$ describes the magnitude of an extreme event that occurs every $\frac{1}{p}$ tuples on average.

7. Extreme value theory

In practical terms, the size n of the considered tuples or *blocks* of data accounts for the interpretation of the return periods of corresponding return levels. Let F be the cdf of a random variable that describes data with daily resolution. Then for a number $m \in \mathbb{N}$ of years we denote an m -year return level of its annual maximum distribution (7.2) with block size $n = 365$ by

$$Q_m := F^{-1} \left(-\sqrt[n]{1 - \frac{1}{m}} \right), \quad (7.81)$$

using an upper-case letter. In contrast, we denote the m -year return level of the distribution of the daily data by lower-case

$$q_m := F^{-1} \left(1 - \frac{1}{n \cdot m} \right) \quad (7.82)$$

in view of definition (7.80) of return levels. On average one out of $n \cdot m$ daily measurements exceeds a return level q_m with return period $n \cdot m$ and frequency $\frac{1}{n \cdot m}$. A return level Q_m occurs as a maximum on average once every m annual blocks. Section 9.1 compares these two approaches to return levels. The 100-year return levels Q_{100} and q_{100} are of particular interest for applications (cp. Chap. 1). Considering different temporal ranges is yet possible by variation of the block size n or the return period. Definition (7.81) of annual return levels bases on the iid setting (7.2). Section 7.4.4 discusses the application of EVT to stationary LRC processes.

The strength of Theorem 7.3 for the estimation of return levels is trifold. First, sporadically missing daily data does not significantly influence the statistical inference by EVT since only rare events enter the analyses. Second, the GEV assumption is an appropriate approximation of the distribution of extremes independently from the distribution of the individual iid samples. And third, it allows for the extrapolation of the distribution of extreme events to return periods beyond the temporal scope of the empirical data.

Remark 7.25: In case of empirical data, for example from geophysics or economics, daily records typically range from about 30 up to less than 100 years. The *empirical* cdf of annual maxima or threshold exceedances thus can not distinguish any return period larger than the measurement range (cp. also Rem. 9.2). Considering return levels of an empirical maxima cdf is only meaningful in theoretical settings with synthetic time series of arbitrary length as done in [47] but in general futile for statistical analyses of real world data.

7.4.2. Block maxima approach

In regard of definition (7.1) the maxima M_n denote the maxima of blocks of samples of size $n \in \mathbb{N}$. Aiming at estimating 100-year return levels (7.81), we consider annual blocks of size $n = 365$ days. The distribution of a maximum M_n then describes annual daily maxima as the units of the maximum and the samples themselves coincide.

Based on the extremal types theorem Theorem 7.3 for iid and Theorem 7.9 for stationary processes, we assess the statistics of extremes and their return levels by adjusting a GEV distribution to empirical block maxima. If the marginal cdf of the process is in the domain of attraction of a GEV distribution with cdf G , then this distribution describes the normed maximum $a_n(M_n - b_n)$ via

$$P(a_n(M_n - b_n) \leq q) \approx G(q) \quad (7.83)$$

for quantiles $q \in \mathbb{R}$ and appropriately chosen norming constants $a_n \in \mathbb{R}_{>0}$ and $b_n \in \mathbb{R}$. Mind that approximation (7.83) is appropriate for numbers n of samples that are large enough such that the asymptotic regime of the convergence (7.10) is valid [37]. Due to the max-stability (7.8) of the GEV distributions, we equivalently have

$$P(M_n \leq q) \approx G(a_n(q - b_n)) = G_*(q), \quad (7.84)$$

where G_* is another member of the same GEV family. Estimating the GEV parameters $\mu, \xi \in \mathbb{R}$ and $\sigma \in \mathbb{R}_{>0}$ of the cdf G_* in approximation (7.84) readily captures both the type of the limit EVD (cp. Sect. 7.3.3) and the norming constants without explicitly requiring their calculation. Inverting the cdf (7.9) of the estimated $\text{GEV}(\mu, \sigma, \xi)$ distribution yields formulae for the extraction of the return level Q_m for a given return period $m = \frac{1}{p}$ by

$$Q_m = \begin{cases} \mu - \frac{\sigma}{\xi} (1 - (-\ln(1-p))^{-\xi}), & \xi \neq 0 \\ \mu - \sigma \ln(-\ln(1-p)), & \xi = 0. \end{cases} \quad (7.85)$$

7. Extreme value theory

For the application of the equations (7.85) the parameters of the GEV distribution are typically estimated by the maximum likelihood (ML) method. Then the parameters are multivariate Gaussian and nonlinear transformations of respective confidence regions (cp. Mahalanobis distance in Sect. 6.3.7) allow for error assessment of the estimated return levels. Every contemporary software package includes procedures for ML estimation of GEV distributions. The optimization of the likelihood function for GEV distributions is typically non-trivial due to a highly complex parameter landscape. For the estimates of return levels in Section 6.3 and Section 9.2 we apply the ML estimation of the software package `climex`⁹, which provides stable numerical optimization algorithms for this purpose. Amongst others the textbook [37] gives details on the ML estimation of return levels. Figure 7.10 shows a GEV estimation for block maxima of a Gaussian iid time series.

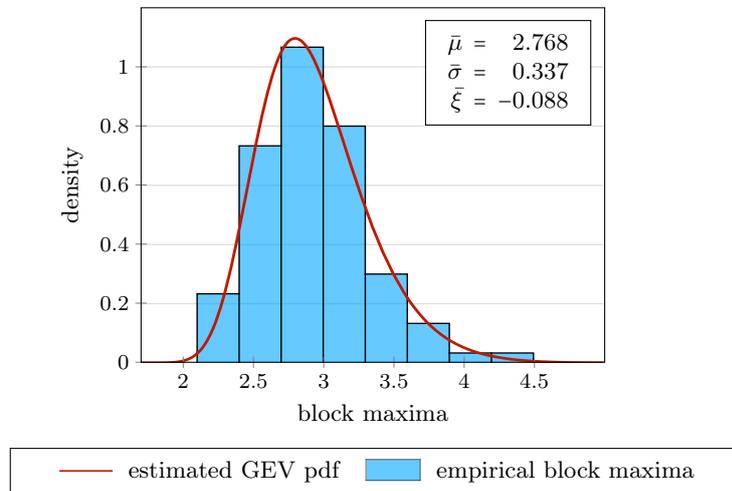


Fig. 7.10 ML-estimated $\text{GEV}(\bar{\mu}, \bar{\sigma}, \bar{\xi})$ pdf with estimated parameters $\bar{\mu}, \bar{\sigma}$ and $\bar{\xi}$ for 100 annual block maxima ($n = 365$) of iid standard Gaussian samples.

If additional information about the distribution of the samples beyond only the empirical block maximum distribution is available, then a-priori knowledge about the shape parameter might help reduce the number of parameters in the ML estimation. If the domain of attraction is known to be the Gumbel class, one might zeroize the shape parameter for the ML estimation by setting $\xi = 0$. Such a constraint, however, might impose further inaccuracy in the estimation of return levels due to a slow rate of the convergence (7.10) as introduced in Section 7.3. Figure 7.11 and Figure 7.14 below comment on this effect. In Section 9.3 involving norming constants, we establish a way of still benefitting from knowledge about validity of the Gumbel domain of attraction for the estimation of return levels.

⁹open source R software package and web application by Philipp Müller [134]

Remark 7.26: Several estimators for the shape parameter were proposed such as the *Hill estimator* [75], *Pickand's estimator* [150] or the *Dekkers-Einmahl-de Haan estimator* [41]. These estimators are of special interest in case of distributions in the Fréchet and Weibull classes with a shape parameter clearly distinct from zero. Note that for slowly varying tails or in the presence of correlations the quality of such estimates of the tail index worsens drastically. For related "Hill horror plots", for example for AR(1) processes, see [48, Fig. 4.1.13 and Fig. 5.5.4].

Statistical uncertainty inheres in the estimation of return levels by applying the block maxima approach due to several sources. A typical violation of the assumptions of the extremal types theorem Theorem 7.3 in practical terms is the lack of independency among empirical data. Section 7.4.4 recaps the role of dependencies in the application of EVT. Another well-known origin of uncertainty is the sampling error. A ML estimation of the GEV distribution is subject to a *finite* number of empirical block maxima, whereas the empirical distribution coincides with the limit GEV distribution only asymptotically by the strong law of large numbers. The application of Theorem 7.3, however, involves two kinds of theoretical limits. Availability of even an *infinite* number of sample maxima would not guarantee full statistical reliability. The extremal types theorem Theorem 7.3 ensures the validity of a GEV shape for the distribution of the maximum M_n only in the limit of an *infinite* block size $n = \infty$. Any *finite* block size $n < \infty$, such as $n = 365$ for annual return levels, may induce significant deviation from the asymptotic GEV shape.

This phenomenon becomes apparent, in particular, in the case of Gaussian samples (cp. Fig. 7.11). The presence of different stochastic limits contains in itself the danger of mistaking one with the other. This problem gains in importance when the appropriateness of adjusting the Gumbel distribution ($\xi = 0$) to annual maxima is reasoned by the presence of samples with a distribution in the Gumbel domain of attraction along with a high number of sample maxima [96, 158]. The *rate* of the convergence (7.10) discussed in Sections 7.3.1 and 7.3.2 is a way towards an evaluation of this source of statistical uncertainty. A penultimate approximation (cp. Sect. 7.3.3) by estimating the shape parameter despite knowing about the belonging to the Gumbel class might improve the estimation of return levels under the cost of a higher uncertainty due to larger confidence intervals. Example 7.27 compares ultimate and penultimate estimates of the 100-year return levels Q_{100} (7.81) of iid standard normal and exponentially distributed samples using the block maxima approach.

7. Extreme value theory

Example 7.27: Figure 7.11 shows both ultimate and penultimate estimates of the 100-year return level Q_{100} (7.81) of iid standard Gaussian (left panel) and exponentially distributed (right panel) samples. Both distributions are in the Gumbel domain of attraction. The sampling error is larger for the penultimate than for the ultimate approach due to a higher number of estimated parameters. In either case, the variance of the estimate decreases as the number m of included block maxima increases.

The mean Gumbel estimate \bar{Q}_{100} of the analytically known $Q_{100} \approx 4.033$ for the annual maxima of standard Gaussians does not converge to the correct return level as the sample size m increases because the asymptotic Gumbel regime is not yet dominant for block size $n = 365$ in convergence (7.10) (cp. Thm. 7.19). The mean GEV estimate \bar{Q}_{100} of Q_{100} by including the shape parameter to the estimation excels the Gumbel estimate due to an increased speed of convergence by Lemma 7.23.

For block maxima of iid samples from the exponential distribution, the convergence to the Gumbel shape is faster (cp. Ex. 7.17), so that zeroizing the shape parameter still yields a proper estimation of $Q_{100} \approx 10.5$ with reduced statistical uncertainty compared to the full GEV estimation.

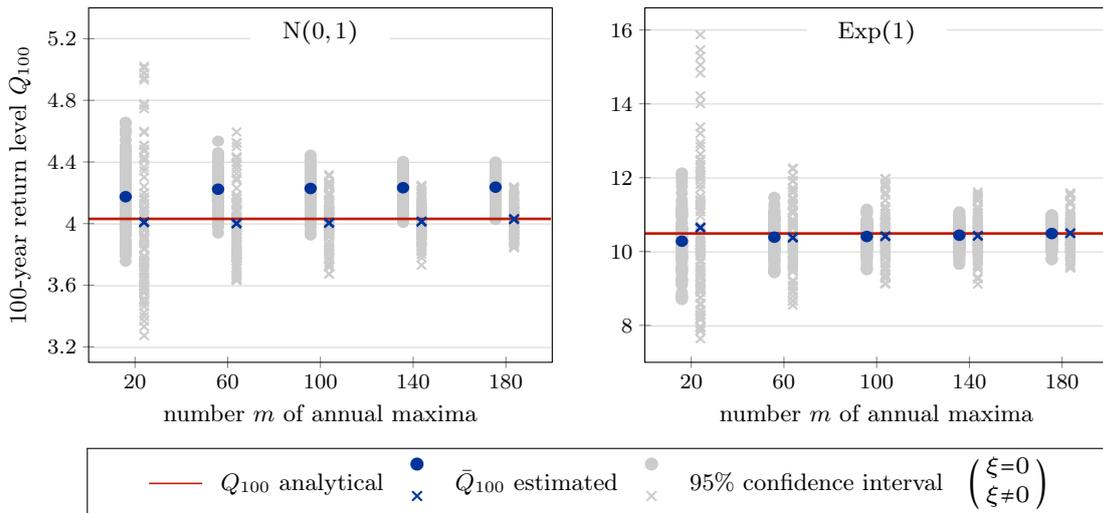


Fig. 7.11 Estimation of the 100-year return level Q_{100} by the block maxima approach for 100 time series of length $N = 100 \cdot 365$ days for iid standard Gaussian (left) and exponentially distributed (right) samples by a ML-estimated Gumbel distribution ($\xi = 0$) or a general GEV distribution ($\xi \neq 0$ estimated) to different numbers m of block maxima in approximation (7.84). The estimated return level \bar{Q}_{100} is the mean of the estimates for each of the time series. The confidence intervals depict the inner 95% of these estimates and are slightly shifted horizontally in the graphic for the sake of perceptibility.

7.4.3. Peaks over threshold approach

In case of the availability of empirical maximum values only, the block maxima approach of Section 7.4.2 is an appropriate method for the estimation of return levels of extreme events. In the presence of measurements with a higher (temporal) resolution, however, the block maximum procedure is wasteful because of neglected data when considering only the maximum of each block [19, 108]. Including this additional information about the right tail of the distribution instead is promising for decreasing the sampling error. The method of choice for this purpose runs under the name *point process of exceedances* [99] or *peaks over threshold* (POT) approach, in particular, in applied science like hydrology [48, p. 340].

We recap the essence of considering exceedances of high thresholds instead of block maxima according to [37, Sect. 4.2.1]. This methodology provides one way of using additional statistical information beyond only block maxima for the estimation of return levels. Section 9.3 presents an alternative method to benefit from additional information about the distribution of the data by involving norming constants.

Definition: Let X be a random variable. The *probability of exceeding a given threshold* $u \in \mathbb{R}$ by an excess value $x \in \mathbb{R}_{>0}$ shall be defined by

$$T_u(x) := P(X > u + x \mid X > u). \quad (7.86)$$

If the cdf F of the variate X is known, then the cdf of the conditional probability (7.86) of threshold exceedance by $x \in \mathbb{R}_{>0}$ reads

$$T_u(x) = \frac{1 - F(u + x)}{1 - F(u)}. \quad (7.87)$$

Similar to the block maxima approach the essence of the POT approach is an asymptotically universal description of conditional threshold excesses by one family of distributions for large thresholds [12].

Generalized Pareto distribution The extremal types theorem Theorem 7.3 implies that if a non-degenerate limit distribution exists for the distribution of asymptotically high threshold exceedances (7.87), then it is a *generalized Pareto distribution* (see Thm. 7.28).

7. Extreme value theory

Definition (Generalized Pareto distribution): Let $\mu, \xi \in \mathbb{R}$ and $\sigma \in \mathbb{R}_{\geq 0}$. Then a member of the family of distributions with a cdf of the form

$$G_P(x) = \begin{cases} 1 - \left(1 + \xi \left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi}, & \xi \neq 0 \\ 1 - e^{-\frac{x-\mu}{\sigma}}, & \xi = 0 \end{cases} \quad (7.88)$$

is called a *generalized Pareto distribution* (GPD). The parameters denote the *location* (μ), *scale* (σ) and *shape* (ξ) of the distribution. Depending on these parameters the domain of a GPD is $\{x \in \mathbb{R} \mid 1 + \xi \left(\frac{x-\mu}{\sigma}\right) > 0\}$. A GPD with parameters μ, σ and ξ is denoted by $\text{GPD}(\mu, \sigma, \xi)$.

We consider a heuristic introduction of Theorem 7.28 below. A detailed discussion of the point process of exceedances can be found in [48, Sect. 6.5] amongst others.

Theorem 7.28 (cp. Thm. 4.1 in [37]): *Let X be a random variable with a cdf in the domain of attraction of a $\text{GEV}(\mu, \sigma, \xi)$ distribution with $\mu, \xi \in \mathbb{R}$ and $\sigma \in \mathbb{R}_{> 0}$. Then for large enough thresholds $u \in \mathbb{R}$ the distribution of the random variable $X - u$ conditioned on $X > u$ is a $\text{GPD}(0, \tilde{\sigma}, \xi)$ approximately, where $\tilde{\sigma} := \sigma + \xi(u - \mu)$ for threshold exceedances in the domain $\{x \in \mathbb{R}_{> 0} \mid 1 + \xi \frac{x}{\tilde{\sigma}} > 0\}$.*

Heuristics. Let F denote the cdf of the random variable X . By the extremal types theorem Theorem 7.3 for quantiles $y \in \mathbb{R}$ and large enough numbers $n \in \mathbb{N}$ of independently drawn samples X_1, \dots, X_n with common cdf F we have

$$\text{P}(\max\{X_1, \dots, X_n\} \leq y) = F(y)^n \approx e^{-(1 + \xi \left(\frac{y-\mu}{\sigma}\right))^{-1/\xi}}. \quad (7.89)$$

The right hand side in approximation (7.89) is the cdf of a $\text{GEV}(\mu, \sigma, \xi)$ distribution with shape parameter $\xi \neq 0$. For cdfs F in the Gumbel domain of attraction ($\xi = 0$) the derivation is valid analogously. Since $F(y) \rightarrow 1$ as $y \rightarrow \infty$, the Taylor expansion of the natural logarithm for large enough quantiles y implies $1 - F(y) \approx -\frac{1}{n} \ln(F(y)^n)$. Applying this approximation to relation (7.89), we obtain

$$1 - F(y) \approx \frac{1}{n} \left(1 + \xi \left(\frac{y-\mu}{\sigma}\right)\right)^{-1/\xi}. \quad (7.90)$$

Plugging in $y = u + x$ and $y = u$ into relation (7.90) and involving notion (7.87) of the probability $T_u(x)$ of threshold excesses $x \in \mathbb{R}_{> 0}$, a short calculation yields the cdf F_u of $X - u$ conditioned on $X > u$ as

$$F_u(x) := \text{P}(X \leq u + x \mid X > u) = 1 - T_u(x) \approx 1 - \left(1 + \xi \frac{x}{\tilde{\sigma}}\right)^{-1/\xi} = G_P(x) \quad (7.91)$$

with $\tilde{\sigma} = \sigma + \xi(u - \mu)$ for large enough thresholds $u \in \mathbb{R}$. In the Gumbel setting of shape $\xi = 0$, we obtain $\tilde{\sigma} = \sigma$ analogously. Note that the domain of the cdf F_u is $[0, \infty)$. The approximation (7.91) is the cdf G_P in definition (7.88) of the $\text{GPD}(0, \tilde{\sigma}, \xi)$ with the scale $\tilde{\sigma}$ depending on the threshold u . \square

Like the GEV family captures the asymptotic distributions of block maxima, the GPDs describe the asymptotic distributions of conditional threshold excesses. Note that by relation (7.91) the shape parameter ξ of a distribution in the domain of attraction of $\text{GEV}(\mu, \sigma, \xi)$ coincides with the shape parameter of the resulting $\text{GPD}(0, \tilde{\sigma}, \xi)$ in Theorem 7.28. Hence, threshold exceedances exhibit a similar classification into exponentially ($\xi = 0$) or power-law ($\xi > 0$) decaying probability of extremes or having a finite upper bound ($\xi < 0$).

The POT approach assesses extremes of a random variate X by the survival probability $\text{P}(X > u + x)$ of rare events using the conditional probability $T_u(x)$ of threshold exceedances (7.86). From approximation (7.91) we deduce

$$1 - F_u(x) = \text{P}(X > u + x \mid X > u) = \frac{\text{P}(X > u + x)}{\text{P}(X > u)} \approx 1 - G_P(x). \quad (7.92)$$

Given a return period $\tilde{m} \in \mathbb{N}$, by relation (7.92) the associated return level $q \in \mathbb{R}_{\geq 0}$ satisfies

$$\frac{1}{\tilde{m}} = \text{P}(X > q) \approx \text{P}(X > u) (1 - G_P(q - u)). \quad (7.93)$$

A natural estimator for the probability $\text{P}(X > u)$ is given by the empirical cdf via $\frac{N_u}{N}$, where $N \in \mathbb{N}$ denotes the empirical sample size and $N_u \in \mathbb{N}$ the number $N_u < N$ of samples larger than the threshold u [48]. If the temporal resolution of the empirical samples is daily, then an excess of a return level q in approximation (7.93) occurs every \tilde{m} days on average. Analogously to the procedure (7.85) in the block maxima method, inverting relation (7.93) yields estimates of *annual* return levels $q_m \in \mathbb{R}_{\geq 0}$ with a prescribed *annual* return period $m \in \mathbb{N}$ and excess probability $p := \frac{1}{365m}$ as

$$q_m = \begin{cases} u + \frac{\sigma}{\xi} \left(\left(\frac{N_u}{Np} \right)^\xi - 1 \right), & \xi \neq 0 \\ u + \sigma \ln \left(\frac{N_u}{Np} \right) & \xi = 0. \end{cases} \quad (7.94)$$

The return level q_m is exceeded once in m years on average. Choosing $p = \frac{1}{\tilde{m}}$ for a *daily* return period $\tilde{m} \in \mathbb{N}$ yields *daily* return levels in equation (7.94). For a comparison of the two kinds of annual return levels in the block maxima and the POT approach see Section 9.1.

7. Extreme value theory

In relation (7.93) the location zero of the cdf G_P in approximation (7.91) is shifted by the threshold u . Thus, ML estimation of a GPD with location zero to the empirical excesses $X - u$ for all data points $X > u$ above a given threshold u yields estimates of the scale and shape parameters σ and ξ , respectively, for the calculation of return levels by equations (7.94).

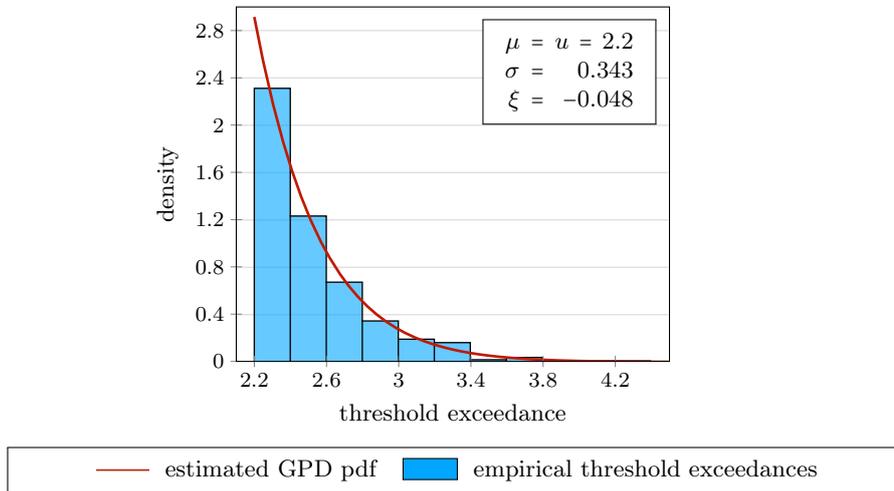


Fig. 7.12 ML estimated GPD($\bar{\mu}, \bar{\sigma}, \bar{\xi}$) pdf with estimated parameters $\bar{\sigma}$ and $\bar{\xi}$ for threshold exceedances ($\mu = u = 2.2$ in equation (7.94)) of a 100-year ($n = 365$ in approximation (7.89)) sample time series of daily iid standard Gaussian samples.

Statistical inferences by the POT approach are subject to three sources of uncertainty compared to the two (namely the block size and the number of empirical block maxima) for the block maxima method. Including more information about the sample distribution than only block maxima still improves the statistical accuracy. First, the sampling error is now governed by a higher number of samples obtained as threshold excesses. The asymptotic GPD regime (7.91), though, is valid for large thresholds only, which is the second source of statistical uncertainty. An appropriate choice of the specific threshold u for the ML estimation of the probability (7.91) is delicate. It requires balancing between the bias by violating the asymptotics (7.90) for low thresholds and an increased variance due to less exceedances for high thresholds [37, Sect. 4.3.1]. A *mean residual life plot* helps evaluate the appropriateness of a specific threshold (see Ex. 7.29, Fig. 7.13 and Appx. G). Third, the block size, or more precisely, the sample size n bases the proof of the asymptotic GPD distribution in Theorem 7.28 by applying the extremal types theorem Theorem 7.3, such that the approximation (7.89) by a GEV distribution is subject to the rate of the convergence (7.10) just as for the block maxima approach.

In practical terms, the POT approach leads to an improved error in the convergence of the normed block maxima to their limit GEV distribution as follows. As introduced above the block maxima method for an empirical data set of $m \in \mathbb{N}$ years of daily records bases statistical inference about annual return levels on a block size of $n = 365$ days in the extremal types theorem 7.3. The higher number of $n = 365m$, however, underlies the POT approach, so that the approximation (7.89) is closer to the asymptotic GEV regime. In case of a slow convergence (7.10) like for example for Gaussian time series (cp. Thm. 7.19), the aforementioned increase of the block size to the size of the empirical data might still not significantly improve the estimation of return levels (cp. Rem. 7.21). For comparison see Figure 7.14 on the rate of the convergence in the extremal types theorem Theorem 7.3.

Example 7.29: We apply the POT method to iid sample data from a standard Gaussian and exponential distribution interpreted as daily measurements. Aiming at estimating the 100-year return level q_{100} (7.82) for these time series by equation (7.93), we consider a return period of $\tilde{m} = 36,500$ days in definition (7.80), which corresponds to $m = 100$ years in equation (7.94). Linear regimes in a mean residual life plot help identify appropriate choices of thresholds u for the adjustment of a GPD in approximation (7.92) (cp. Appx. G).

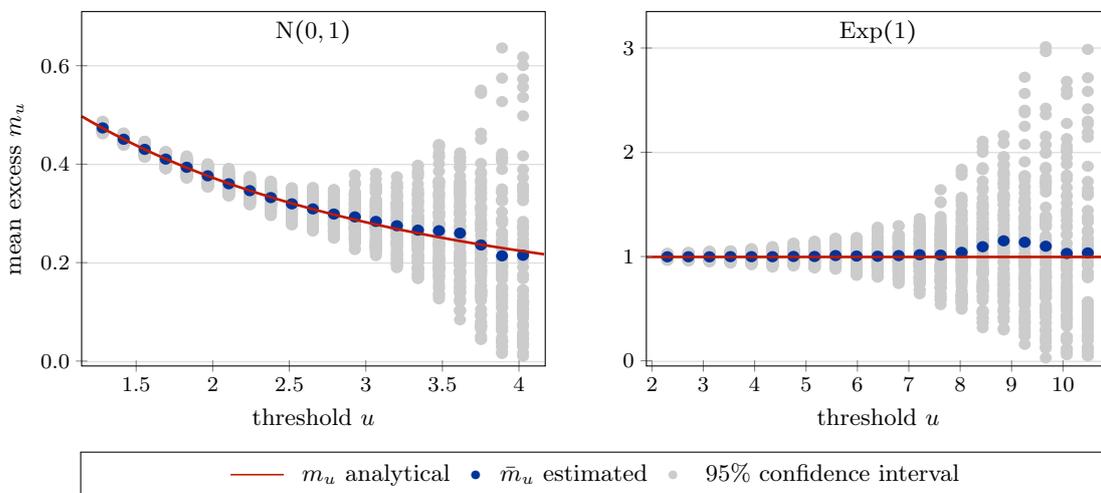


Fig. 7.13 Mean residual life plot for 100 time series of length $N = 36,500$ of iid samples of a standard Gaussian (left) and exponential (right) distribution. The analytical solutions for the mean excess m_u are given in the equations (G.4) and (G.3), respectively. The estimated excesses \tilde{m}_u are the mean of the estimates for each of the time series with the inner 95% values giving the depicted confidence intervals.

7. Extreme value theory

For the exponential distribution Figure 7.13 (right panel) reveals a linear mean residual life plot for thresholds $u \in [2, 7]$ roughly, so that POT estimates of the quantile q_{100} are fairly reliable when deploying thresholds suchlike. Note that for the exponential distribution $\text{Exp}(\lambda)$ with $\lambda \in \mathbb{R}_{>0}$, the conditional excess probability T_u in definition (7.86) is independent of the specific threshold u , or more precisely, $T_u(x) = e^{-\lambda x}$ for all values $x \in \mathbb{R}_{>0}$. As a remark, this property is called *memorylessness* (not to be confused with the memoryless transformations of Chap. 3) of the exponential distribution and describes that when interpreted as a waiting time distribution the waiting time until the occurrence of a next event does not depend on the time already passed [53, p. 33]. The mean excess $m_u := \mathbb{E}[X - u \mid X > u] = \sigma$ is thus constant for all thresholds u , where $\sigma = a_n = \lambda$ is the scale of the approximate GEV distribution (7.15) of the related block maxima (cp. eq. (G.3)). The mean estimate of the 100-year return level of the exponential samples does not depend on the choice of the threshold in Figure 7.14 (right panel). The variance of the estimate, however, increases as the height of the threshold increases to the point of a lack of availability of exceedances for too high thresholds.

For the Gaussian samples, however, the mean residual life plot is inconclusive as the relation between the thresholds and the related mean conditional excesses appears curved all over. Indeed, the mean excess reads m_u is linear only asymptotically as $u \rightarrow \infty$ at a slow rate by equation (G.4). The extreme slowness of the convergence (7.10) to the Gumbel regime in the Gaussian case provides a reasoning of this effect (see Sect. 7.3).

Accordingly, the POT estimates of the 100-year return level q_{100} of Gaussian samples in Figure 7.14 (left panel) are inaccurate when assuming a shape parameter of $\xi = 0$ for the adjustment of the GPD in approximation (7.91). Like for the penultimate block maxima approach in Figure 7.11 including the shape parameter into the GPD estimation improves the estimate.

Comparing Figure 7.11 and Figure 7.14 for $m = 100$ block maxima reveals an improved statistical uncertainty by the POT approach over the block maxima method for the Gaussian samples. The GPD estimates without assuming zero shape for thresholds between about $u = 2$ and $u = 2.5$ exhibit a lower variance than the full GEV estimate. Nevertheless the uncertainty of the choice of a threshold remains. We observe the same reduction of the variance of the estimate for the exponential distribution for thresholds below about $u = 5$ for either assumption of the shape parameter.

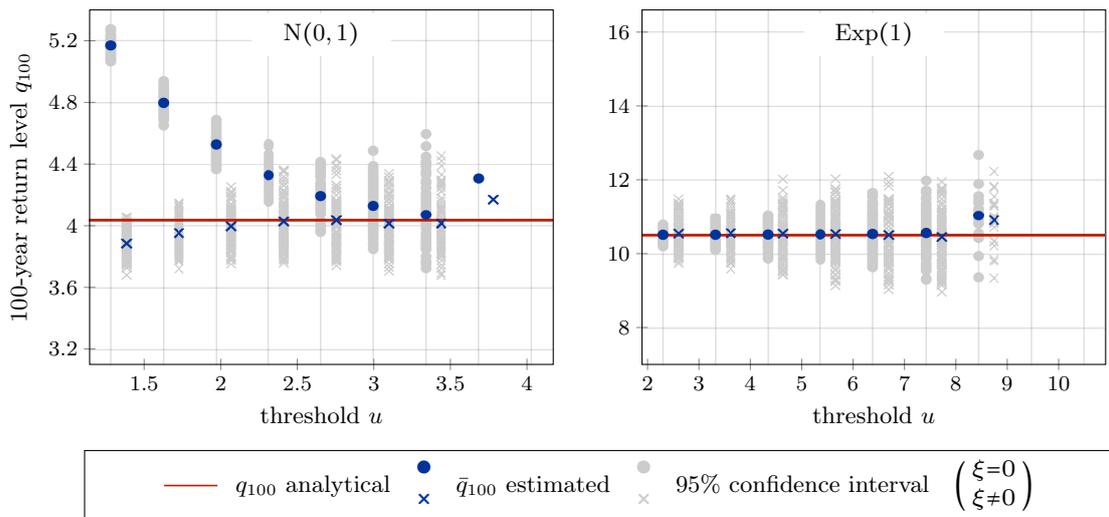


Fig. 7.14 Estimation of the 100-year return level q_{100} by the peaks over threshold approach for 100 time series of length $N = 36,500$ days for iid standard Gaussian (left) and exponentially distributed (right) samples by a ML-adjusted GPD assuming $\xi = 0$ or estimating $\xi \neq 0$ and using different thresholds u . The chosen threshold positions are nine equally spaced values from the 90% quantile (≈ 1.28 and ≈ 2.3) up to the 100-year return level (≈ 4.03 and ≈ 10.5 , respectively). For better perceptibility, the estimates with $\xi \neq 0$ are slightly shifted horizontally in the graphics while conducted for the same thresholds as in the $\xi = 0$ setting. Note that the number of threshold exceedances per time series depends on the height of the threshold u , so that even no estimate might be feasible for large thresholds. This occurs for most of the sample time series of both example distributions for the three highest choices of thresholds. In either case, the sampling error is larger if the shape parameter ξ is estimated instead of assumed zero similar to the block maxima method in Figure 7.11.

Example 7.30: The Pareto distribution is in the Fréchet domain of attraction (Tab. 7.2), so that a Gumbel estimation is not reasonable. The relation between thresholds u and mean excesses m_u is linear by equation (G.5). Figure 7.15 emphasizes that for large thresholds the linearity of the mean residual life plot is violated. This effect stems from the large variance of the sample mean (G.1) of the excesses due to the large variance of the samples compared to their mean value.

7. Extreme value theory

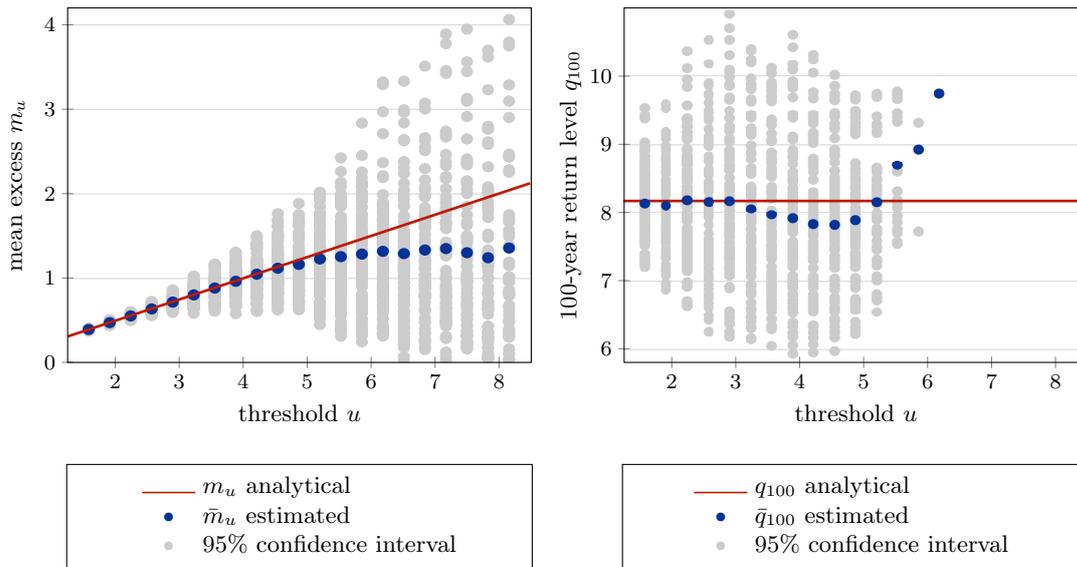


Fig. 7.15 Mean residual life plot (left) and POT estimate of the 100-year return level q_{100} (right) for 100 time series of length $N = 36,500$ of iid samples of a Pareto $\text{Par}(1, 5)$ distribution. The chosen threshold positions are 20 equally spaced values from the 95% quantile (≈ 1.58) up to the 100-year return level $q_{100} \approx 8.17$. For thresholds above about $u = 5$ excesses very likely do not exist in the times series, so that no estimate is possible.

The Examples 7.29 and 7.30 visualize the strengths and weaknesses of the POT approach. Like the block maxima approach the appropriateness of the estimates of return levels by the POT method depend on the quality of the approximation (7.10) in the extremal types theorem Theorem 7.3. In case of slow convergence for a cdf in the Gumbel domain of attraction, penultimate approximations might improve the estimate under the cost of an increased variance. The uncertainty further stems from the choice of the threshold. Example 7.30 exemplifies for Pareto distributed samples that higher uncertainty in this choice might be balanced out by a fast convergence (7.10) (cp. Ex. 7.17) with accurate estimates of return levels despite lower certainty in the threshold.

Dependencies in the data are capable of introducing further uncertainty into the estimation of return levels of extremes. Section 7.4.4 summarizes the implications for statistical inference by EVT for stationary processes introduced in Section 7.2.

7.4.4. Statistical inference for stationary processes

Theorem 7.9 implies that for stationary processes satisfying the condition $D(u_n)$ the limit maxima distribution is an EVD similar to the iid setting. The strength of the dependence among the samples of a stationary time series influences the specific shape of the limit GEV distribution. Theorem 7.10 describes the relation between these two limit distributions by using the extremal index θ to power-transform one into the other.

Due to the max-stability of the EVDs, the two distributions obey to the same extremal type just with possibly different location, scale and shape parameters. Estimation of these GEV parameters by ML-adjustment to empirical block maxima thus captures both settings of independence and dependence, so that the same statistical procedure like for iid data is applicable.

The presence of dependencies in time series is capable of influencing both the specific parameters of the limit GEV distribution and the rate of the convergence in the generalized extremal types theorem Theorem 7.10. An extremal index $\theta = 1$ implies that asymptotically the statistics of extremes equal the behavior of extremes of the associated iid process and result in the same limit GEV distribution as the block size increases to infinity. In practical terms with finite block sizes, however, the dependencies in the data possibly highly influence the statistics of extremes [37, p. 97]. Rust [157] provides a numerical study on the effect of LRC on the statistical inference by EVT. Example 7.31 visualizes such effects for different ARFIMA(0, d , 0) processes. The applicability of the POT approach to stationary processes bases on the above reasoning for the block maxima approach by Theorem 7.28 in the asymptotic limit of large thresholds. For finite thresholds, in particular, in case of extremal indices $\theta > 1$, clustering of extremes [29] increases the statistical uncertainty of GPD estimates.

For Gaussian LRC processes with extremal index $\theta = 1$ the slow logarithmic rate of convergence nonetheless dominates the error (7.75) when approaching the limit Gumbel distribution, so that the impact of the correlations on extreme value statistics for ARFIMA(0, d , 0) becomes visually apparent only for strong LRC with parameters $d \geq 0.3$ in Figure 7.16.

Example 7.31: Figure 7.16 shows GEV and Gumbel estimates of the 100-year return level of standard Gaussian ARFIMA(0, d , 0) time series for different LRC parameters d . Note that the value $d = 0$ reproduces Figure 7.11 (left panel). On average, the full GEV estimate excels the Gumbel estimate with the cost of larger confidence intervals. The appropriateness of the Gumbel distribution is still dominated by the slow logarithmic convergence of the normed maxima distribution for Gaussian data, so that the penultimate approximation by a GEV distribution overcomes better the impacts of slow convergence in general and of power-law decaying autocorrelations for ARFIMA time series. Remarkably, for a low number of annual maxima involved into the estimation the increased sampling error improves the Gumbel estimate by actually underestimating the 100-year return level of the limit Gumbel distribution.

7. Extreme value theory

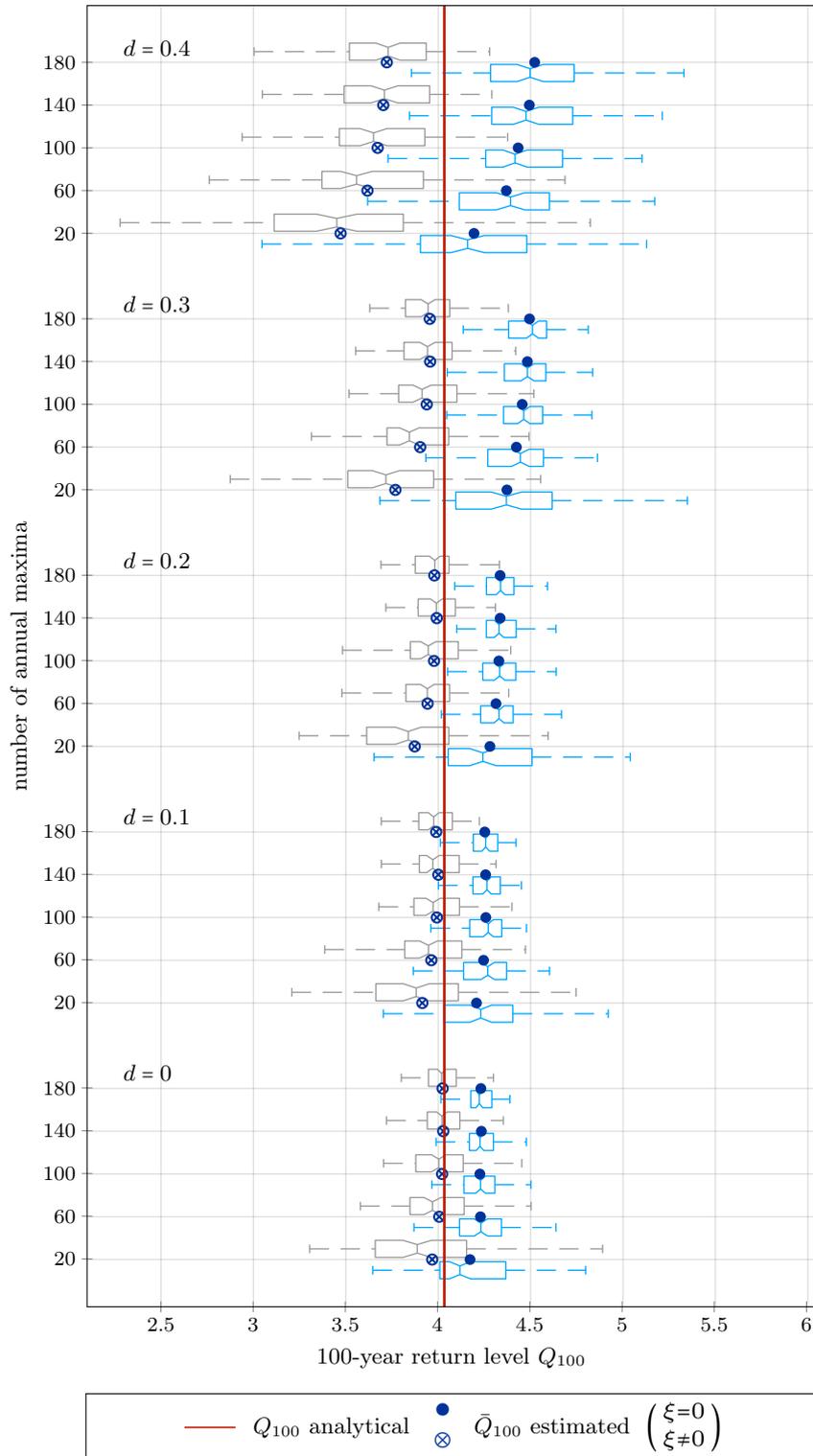


Fig. 7.16 GEV and Gumbel estimates of the 100-year return level $Q_{100} \approx 4.03$ of ARFI-MA(0, d , 0) time series for different LRC parameters d and different numbers of annual maxima included. The estimate \bar{Q}_{100} depicts the mean of the estimates of 100 sample time series visualized by the box plots. The estimate \tilde{Q}_{100} depicts the mean of the estimates of 100 sample time series visualized by the box plots. The boxes show their median and the 25% and 75% quantiles, the whiskers extend to the adjacent value each.

8. Extreme value theory for meta-Gaussian processes

Aiming at an application to precipitation measurements, we consider EVT for meta-Gaussian LRC processes $Y_t = g(X_t)$ with power-law decaying acfs and generated by a strictly monotonic \mathcal{L}^2 transformation g (cp. Sect. 3.1). On the positive real line the tGp transformation of ARFIMA processes introduced in Section 6.1.3 for the modeling of precipitation time series satisfies these conditions in Example 8.3. The transformation $F^{-1} \circ \Phi$ typically used for inverse sampling of Gaussian processes (cp. Sect. 1.1 and Chap. 3) is strictly monotonic for most frequently desired prescribed marginal distributions with cdf F as well and applied in Example 8.2 and Example 8.4. Let

$$\tilde{M}_n := \max\{X_1, \dots, X_n\} \quad \text{and} \quad M_n := \max\{Y_1, \dots, Y_n\} \quad (8.1)$$

denote the maximum of a window or block of size $n \in \mathbb{N}$ of an underlying Gaussian process X_t and the transformed process Y_t with marginal cdf F , respectively. The monotonicity of the transformation g preserves the ordering of the members of the process, so that we have

$$\mathbb{P}(\tilde{M}_n \leq x) = \mathbb{P}(M_n \leq g(x)) \quad (8.2)$$

for all $x \in \mathbb{R}$ with $F(x)$ in the domain of the target cdf F . For the tGp distribution we consider the cdf (6.4) for values $x \geq 0$. Besides the Gumbel class of the underlying Gaussian, the marginal distribution generated by the transformation might belong to any of the three types of EVDs, which can be determined by the von-Mises conditions (cp. Tab. F.1).

Equation (8.2) directly implies the validity of the conditions $D(g(u_n))$ and $D'(g(u_n))$ for meta-Gaussian processes if the original Gaussian process satisfies the conditions $D(u_n)$ and $D'(u_n)$ for a series of thresholds u_n . Mind that memoryless transformations decrease or at most keep the strength of correlations by Lemma 3.6. The validity of the generalized extremal types theorem Theorem 7.9 for meta-Gaussian processes requires thresholds $g(u_n)$ of affine linear form, which might be a challenging task. Section 8.1 establishes that under certain conditions on the transformation g a Taylor expansion of the transformed thresholds $g(u_n)$ allows for deducing affine linear thresholds and convergence to the Gumbel limit based on the thresholds u_n chosen for the Gaussian process.

Section 8.2 deals with the rate of the convergence of the normed maxima distribution to its limit EVD for meta-Gaussian processes. Applying these observations, Section 8.3 provides an outlook about effective block sizes for EVT.

8.1. Norming constants

By Theorem 7.10 norming constants valid for maxima of its associated iid process are valid as well for maxima of a stationary process if the condition $D(u_n)$ with affine linear thresholds u_n is satisfied. Hence, a differentiation of the dependence structure is not necessary for deriving norming constants for meta-Gaussian processes. For the rate of the convergence (7.27) both the dependencies in the underlying Gaussian processes and the transformation are crucial and subject of Section 8.2.

Let α_n, β_n denote norming constants for the standard Gaussian distribution implying convergence of the normed maxima of standard Gaussian samples to their limit Gumbel distribution as for example given in Table 7.3. For the sake of simplicity, deriving norming constants for meta-Gaussian processes, we use the constants defined by equations (7.48) with $\alpha_n = \beta_n$ for all block sizes $n \in \mathbb{N}$. The methodology below remains analogously valid for alternative choices of norming constants only with more complex calculations.

Using the notion (8.1) and relation (8.2), in case of strictly monotonic and twice continuously differentiable transformations g , the convergence $\mathbb{P}\left(\tilde{M}_n \leq \frac{x}{\alpha_n} + \beta_n\right) \rightarrow \Lambda(x)$ for normed standard Gaussian maxima to the Gumbel cdf Λ as $n \rightarrow \infty$ implies norming constants a_n and b_n for the meta-Gaussian process by the Taylor expansion as

$$\begin{aligned} \mathbb{P}\left(\tilde{M}_n \leq \frac{x}{\alpha_n} + \beta_n\right) &= \mathbb{P}\left(M_n \leq g\left(\frac{x}{\alpha_n} + \beta_n\right)\right) \\ &= \mathbb{P}\left(M_n \leq \frac{x}{a_n} + b_n + \mathcal{R}_n\right) \\ &= \mathbb{P}(a_n(M_n - b_n) \leq x + a_n\mathcal{R}_n) \rightarrow \Lambda(x) \quad \text{as } n \rightarrow \infty \end{aligned} \quad (8.3)$$

$$\text{with } a_n := \frac{\alpha_n}{g'(\beta_n)} \quad \text{and} \quad b_n := g(\beta_n) \quad (8.4)$$

$$\text{if } a_n\mathcal{R}_n \propto a_n g''(\beta_n) \left(\frac{x}{\alpha_n}\right)^2 \propto \frac{g''(\beta_n)}{\alpha_n g'(\beta_n)} \rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (8.5)$$

The representation of the error \mathcal{R}_n used in equation (8.5) deploys the second order mean-value remainder in the neighborhood of the expansion points β_n and is valid pointwise for all x in the domain of the marginal cdf of the transformed process. Note that by the estimate (7.65) we have $\alpha_n^2 \sim 2 \ln n$, so that $\frac{1}{\alpha_n} \rightarrow 0$ as $n \rightarrow \infty$. Firstly, the derivative $g'(\beta_n)$ determines the behavior of the scaling constant a_n and, secondly, both the first and second derivative $g'(\beta_n)$ and $g''(\beta_n)$ decide about the convergence (8.3).

For affine linear transformations $\sigma X + \nu$ with a standard deviation $\sigma \in \mathbb{R}_{>0}$ and a mean $\nu \in \mathbb{R}$ of standard Gaussian variates X equations (8.4) directly imply $a_n = \frac{\alpha_n}{\sigma}$ and $b_n = \sigma\beta_n + \nu$. The Examples 8.1 to 8.4 study deriving norming constants for inverse sampling of standard Gaussian variates to the exponential and uniform distribution and transformation to the lognormal and tGp distribution, respectively.

Example 8.1 (Norming constants for the lognormal distribution): Using the transformation $g(x) = e^x$ with $g'(x) = e^x$, the Taylor approach (8.4) gives the norming constants [101, Ex. 1.7.4]

$$a_n = \beta_n e^{-\beta_n} \quad \text{and} \quad b_n = e^{\beta_n}. \quad (8.6)$$

The convergence (8.5) follows directly as $\frac{g''(\beta_n)}{g'(\beta_n)} = 1$ and $\frac{1}{\alpha_n} \rightarrow 0$ as $n \rightarrow \infty$. Applying a similar method like for approximating the norming constants (7.48) by equations (7.47) in Section 7.3.1 gives the norming constants [105]

$$a_n = \frac{\sqrt{2 \ln n}}{e^{\sqrt{2 \ln n}}} \quad \text{and} \quad b_n = e^{\sqrt{2 \ln n}} \left(1 - \frac{\ln(\ln n) + \ln(4\pi)}{2\sqrt{2 \ln n}} \right), \quad (8.7)$$

which change the rate of convergence analogously to the Gaussian case (cp. Ex. 8.5).

Example 8.2 (Norming constants for the exponential distribution): Let $F(x) = 1 - e^{-x}$ denote the cdf of the exponential distribution $\text{Exp}(1)$ for arguments $x \in \mathbb{R}_{>0}$. The corresponding transformation g for inverse sampling and its derivative read

$$g(x) = (F_1^{-1} \circ \Phi)(x) = -\ln(1 - \Phi(x)) \quad \text{and} \quad g'(x) = \frac{\phi(x)}{1 - \Phi(x)}. \quad (8.8)$$

Plugging in the norming constants $\alpha_n = \beta_n$ defined by equations (7.48) into equations (8.8) and using the asymptotics (7.53) yields

$$\begin{aligned} g(\beta_n) &= -\ln\left(\frac{\phi(\beta_n)}{\beta_n} (1 + \mathcal{O}(1/\beta_n^2))\right) = -\ln\left(\frac{1}{n} + \mathcal{O}(1/\beta_n^2)\right) = \ln n + \mathcal{O}\left(\frac{1}{n\beta_n^2}\right), \\ g'(\beta_n) &= \frac{\phi(\beta_n)}{1 - \Phi(\beta_n)} = \frac{\phi(\beta_n)}{\frac{\phi(\beta_n)}{\beta_n} (1 + \mathcal{O}(1/\beta_n^2))} = \frac{1}{\frac{1}{\beta_n} + \mathcal{O}(1/\beta_n^3)} = \beta_n + \mathcal{O}\left(\frac{1}{\beta_n}\right) \end{aligned} \quad (8.9)$$

as the sample size $n \rightarrow \infty$ and the norming constants by equations (8.4) as

$$a_n = \frac{\beta_n}{\beta_n + \mathcal{O}(1/\beta_n)} = 1 + \mathcal{O}\left(\frac{1}{\beta_n^2}\right) \rightarrow 1 \quad \text{and} \quad b_n = \ln n + \mathcal{O}\left(\frac{1}{n\beta_n^2}\right) \sim \ln n \quad (8.10)$$

as the sample size $n \rightarrow \infty$. Note that these norming constants asymptotically coincide with those given in Example 7.5. It remains to derive relation (8.5) for concluding convergence to the Gumbel cdf in equation (8.3). We have

$$g''(x) = \phi(x) \left(\frac{\phi(x)}{(1 - \Phi(x))^2} - \frac{x}{1 - \Phi(x)} \right) = g'(x)^2 - g'(x)x, \quad (8.11)$$

so that by the asymptotics (8.9) we obtain $\frac{g''(\beta_n)}{g'(\beta_n)} = \mathcal{O}\left(\frac{1}{\beta_n}\right)$ and thus $a_n \mathcal{R}_n \rightarrow 0$ as $n \rightarrow \infty$.

8. Extreme value theory for meta-Gaussian processes

Example 8.3 (Norming constants for the tGp distribution): For exponents $\eta \in \mathbb{R}_{>0}$ consider the function $g(x) = x^\eta$ for values $x \in \mathbb{R}_{\geq 0}$. Considering nonnegative values only suffices as the truncation by the tGp transformation (6.2) concentrates the probability of all negative values at zero. A continuation of the cdf for positive values to zero thus directly gives the cdf for all nonnegative values. Aiming at tGp distributions for powers of all possible truncated Gaussian distributions, that might be the results of the model estimation in Section 6.2, we apply the norming constants $\tilde{\alpha}_n := \frac{\alpha_n}{\sigma}$ and $\tilde{\beta}_n := \sigma\beta_n + \nu$ for a standard deviation $\sigma \in \mathbb{R}_{>0}$ and a shift $\nu \in \mathbb{R}$ using the norming constants α_n and β_n defined by equations (7.48). Using $g'(x) = \eta x^{\eta-1}$, for exponents $\eta \neq 2$ the norming constants by equations (8.4) read

$$a_n = \frac{\tilde{\alpha}_n}{g'(\tilde{\beta}_n)} = \frac{\alpha_n(\sigma\beta_n + \nu)^{1-\eta}}{\sigma\eta} \quad \text{and} \quad b_n = g(\tilde{\beta}_n) = (\sigma\beta_n + \nu)^\eta \quad (8.12)$$

In case of the power of standard Gaussian variates, the norming constants are $a_n = \frac{1}{s}\beta_n^{2-s}$ and $b_n = \beta_n^s$. Hall [68] obtained the same norming constants (8.12) considering the distribution of the maxima of powers of the absolute value of a standard Gaussian variate. The convergence (8.5) of the scaled error is valid as $\frac{g''(\tilde{\beta}_n)}{\tilde{\alpha}_n g'(\tilde{\beta}_n)} = \mathcal{O}\left(\frac{1}{\tilde{\alpha}_n \tilde{\beta}_n}\right) \sim \frac{1}{\beta_n^2} \rightarrow 0$ as $n \rightarrow \infty$. Example 8.6 and Section 8.2.1 discuss the rate of convergence (8.3) and the case of the exponent $\eta = 2$.

Example 8.4 (Norming constants for the uniform distribution): Analogously to Example 8.2 for inverse sampling let $F(x) = x$ denote the cdf of the uniform distribution $U(0, 1)$ and consider

$$g(x) = (F^{-1} \circ \Phi)(x) = \Phi(x) \quad \text{and} \quad g'(x) = \phi(x), \quad (8.13)$$

which by equations (8.4) gives the norming constants

$$a_n = \frac{\beta_n}{\phi(\beta_n)} = n \quad \text{and} \quad b_n = \Phi(\beta_n) \rightarrow 1 \quad \text{as } n \rightarrow \infty. \quad (8.14)$$

Other than in the Examples 8.1 to 8.3 the scaled error (8.5) does not vanish in case of inverse sampling for the uniform distribution as $\frac{g''(\beta_n)}{\alpha_n g'(\beta_n)} = -\frac{\beta_n \phi(\beta_n)}{\beta_n \phi(\beta_n)} = 1$. Missing convergence to the Gumbel cdf is, however, expected, since the uniform distribution belongs to Weibull instead of the Gumbel class. Asymptotically, the norming constants 8.14 coincide with those obtained in Example 7.5 nonetheless but the higher order terms in the Taylor expansion do not cancel out but modify the limit asymptotically.

8.2. Rate of convergence

The optimal rate of the convergence (7.10) in the extremal types theorem Theorem 7.3 for the normed maxima of iid Gaussian variates is $\frac{1}{\ln n}$ as the block size n increases (see Thm. 7.19). Under the assumptions of Chapter 8 let g be a transformation such that condition (8.5) is satisfied and the maxima distribution of the meta-Gaussian process $g(X_t)$ converges to the Gumbel limit. Let a_n and b_n denote the norming constants (8.4) obtained by the Taylor expansion in Section 8.1 with

$$\mathbb{P}\left(M_n \leq \frac{x}{a_n} + b_n\right) \rightarrow \Lambda(x) \quad \text{as } n \rightarrow \infty, \quad (8.15)$$

using the notion (8.1). Inverting the idea (8.2) yields

$$\mathbb{P}\left(M_n \leq \frac{x}{a_n} + b_n\right) = \begin{cases} \Phi\left(g^{-1}\left(\frac{x}{a_n} + b_n\right)\right)^n & \text{if } X_t \text{ is iid and} \\ F_{(X_1, \dots, X_n)}\left(g^{-1}\left(\frac{x}{a_n} + b_n\right)\right) & \text{if } X_t \text{ is stationary,} \end{cases} \quad (8.16)$$

where $F_{(X_1, \dots, X_n)}$ denotes the joint cdf (7.24) of the members X_1, \dots, X_n in a window of size $n \in \mathbb{N}$ of the underlying stationary Gaussian process X_t . Section 8.2.1 studies the rate of the convergence (8.15) for meta-Gaussian iid processes. For stationary meta-Gaussian processes generated from Gaussian LRC processes Section 8.2.2 deploys the results obtained for the iid setting for estimating the rate of the convergence (8.15).

8.2.1. Rate of convergence in the classical setting

Using a Taylor expansion of the inverse function g^{-1} implies that the error (8.5) established when defining the norming constants (8.4) serves for estimating an upper bound of the rate of the convergence (8.15). Let $\alpha_n = \beta_n = n\phi(n)$ denote the optimal norming constants for the Gaussian distribution defined by equations (7.48) for samples sizes $n \in \mathbb{N}$. Then, using the derivative $(g^{-1})'(x) = \frac{1}{g'(g^{-1}(x))}$, the Taylor series for the inverse function g^{-1} in the neighborhood of the point $b_n = g(\beta_n)$ reads

$$g^{-1}\left(\frac{x}{a_n} + b_n\right) = g^{-1}(b_n) + \frac{(g^{-1})'(b_n)}{a_n}x + \mathcal{R}_n = \beta_n + \frac{x}{\alpha_n} + \mathcal{R}_n \quad (8.17)$$

with

$$\mathcal{R}_n \propto \frac{(g^{-1})''(b_n)}{a_n^2} \propto \frac{g''(g^{-1}(b_n))}{(g'(g^{-1}(b_n)))^2} \cdot \frac{1}{g'(g^{-1}(b_n))} \cdot \frac{g'(\beta_n)^2}{\beta_n^2} = \frac{g''(\beta_n)}{g'(\beta_n)\beta_n^2} \quad (8.18)$$

as the number n of samples increases.

8. Extreme value theory for meta-Gaussian processes

As the inverse of a strictly monotonically increasing function is strictly increasing itself, the mean value theorem and using the notion $u_n := \frac{x}{\alpha_n} + \beta_n \sim \beta_n$ as $n \rightarrow \infty$ implies

$$\begin{aligned}
& \left| \mathbb{P}\left(M_n \leq \frac{x}{a_n} + b_n\right) - \Lambda(x) \right| \\
&= \left| \mathbb{P}\left(g^{-1}(M_n) \leq g^{-1}\left(\frac{x}{a_n} + b_n\right)\right) - \Lambda(x) \right| && \text{(by (8.2))} \\
&\leq \left| \mathbb{P}\left(\tilde{M}_n \leq g^{-1}\left(\frac{x}{a_n} + b_n\right)\right) - \mathbb{P}\left(\tilde{M}_n \leq \frac{x}{\alpha_n} + \beta_n\right) \right| + \mathcal{O}(1/\ln n) && \text{(by Thm. 7.19)} \\
&\leq \left| \mathbb{P}\left(\tilde{M}_n \leq u_n + \mathcal{R}_n\right) - \mathbb{P}\left(\tilde{M}_n \leq u_n\right) \right| + \mathcal{O}(1/\ln n) && \text{(by (8.17))} \\
&= \left| \Phi(u_n + \mathcal{R}_n)^n - \Phi(u_n)^n \right| + \mathcal{O}(1/\ln n) && \text{(by (8.16))} \\
&\propto n\Phi(u_n)^{n-1} \phi(u_n) \mathcal{R}_n + \mathcal{O}(1/\ln n) && \propto \beta_n \mathcal{R}_n + \mathcal{O}(1/\ln n)
\end{aligned} \tag{8.19}$$

Note that $\Phi(u_n)^n \rightarrow \Lambda(x)$ and $\Phi(u_n) \rightarrow 1$, so that $\Phi(u_n)^{n-1} \rightarrow \Lambda(x)$ as $n \rightarrow \infty$. Plugging in the asymptotics (8.18) into the estimate (8.19) yields the pointwise error (7.40)

$$\Delta_n(x) = \left| \mathbb{P}\left(M_n \leq \frac{x}{a_n} + b_n\right) - \Lambda(x) \right| = \mathcal{O}\left(\frac{g''(\beta_n)}{g'(\beta_n)\beta_n}\right) + \mathcal{O}\left(\frac{1}{\ln n}\right) \tag{8.20}$$

as the block size $n \rightarrow \infty$. The asymptotic error (8.20) coincides with the condition (8.5) for the convergence (8.3) and provides an upper bound for its rate. The above methodology yields rates of convergence not better than logarithmic. In case of the inverse sampling to the exponential distribution in Example 8.2, by Example 7.17 the optimal rate of convergence is linear using the asymptotic norming constants (8.10). The approximations of these constants by a Taylor series, however, yield the rate $\beta_n \mathcal{R}_n = \mathcal{O}\left(\frac{1}{\beta_n^2}\right) = \mathcal{O}\left(\frac{1}{\ln n}\right)$ only and keep the logarithmic convergence of the maxima of the original Gaussian process. The norming constants (8.10) are thus valid but not optimal for the exponential distribution.

Example 8.5 (Rate of convergence for maxima of iid lognormal variates): Example 8.1 shows that using the norming constants (8.6) yields an upper bound of the rate (8.20) of the convergence (8.15) of $\frac{1}{\beta_n} \sim \frac{1}{\sqrt{\ln n}}$ as $n \rightarrow \infty$ by approximation (7.59). Liao and Peng [105] prove this rate optimal for the ultimate approximation and that the alternative constants (8.7) give a slower convergence rate of $\frac{(\ln(\ln n))^2}{\sqrt{\ln n}}$ as the number n of samples increases, similar to the convergence (7.72) slower than $\frac{1}{\ln n}$ for maxima of iid Gaussian samples. The penultimate approximation of the maxima of lognormal variates is of the faster but still slow order $\frac{1}{(\ln n)^{3/2}}$ [35]. The lognormal process is a meta-Gaussian process with slower rates of convergence than for maxima of the underlying Gaussian iid process in both ultimate and penultimate approximation.

Example 8.6 (Rate of convergence for maxima of tGp variates): Example 8.3 derives that using the norming constants (8.12), the distribution of the maxima of iid tGp variates converges to the Gumbel limit at rate $\frac{1}{\beta_n^2} \sim \frac{1}{\ln n}$ as $n \rightarrow \infty$ by estimate (7.65). For variates $X \sim N(0, 1)$, in an auxiliary calculation Hall [68, Eq. (11)] calculates the optimality of these constants and this rate for maxima of variates X_+^η aiming at the rate of the convergence for maxima of variates $|X|^\eta$.

Figure 8.1 shows the pointwise error (8.20) at the position $x = 1$ of tGp distributions with shift and scale parameters $\nu = 0$ and $\sigma = 1$, respectively, and with different exponents $\eta \neq 2$. Note that the exponent $\eta = 1$ reproduces Figure 7.6 for the position $x = 1$ instead of $x = 0$. The values $\Delta_n(x)$ are directly calculated by equation (8.16) without any approximation of the inverse function $g^{-1}(x) = x^{1/\eta}$ for values $x \in \mathbb{R}_{\geq 0}$ or of the norming constants. The linear shape of the inverse error $\frac{1}{\Delta_n(x)}$ in semi-logarithmic scaling confirms the theoretical result of $\frac{1}{\ln n}$ convergence.

Analogously to the Gaussian distribution (cp. Lem. 7.23), Hall [68] derives the optimal rate of $\frac{1}{(\ln n)^2}$ for the penultimate approximation of the distribution for the maximum of tGp variates. The optimal rates of the ultimate and penultimate approximation for maxima of tGp variates coincide with those for the underlying Gaussian variates. In case of the exponent $\eta = 2$ [35], the rates of the ultimate and penultimate approximation are faster with $\frac{1}{(\ln n)^2}$ and $\frac{1}{(\ln n)^3}$. Optimal norming constants for approaching the Gumbel limit are $a_n = \frac{2(\beta_n^2 - 1)}{\beta_n^2}$ and $b_n = \beta_n^2 - \frac{2}{\beta_n^2}$ [68]. Section 9.3 deals with the effect that distinguishes the maxima of tGp variates with an exponent $\eta = 2$ from the others with $\eta \neq 2$.

8.2.2. Rate of convergence for stationary meta-Gaussian processes

Let $Y_t = g(X_t)$ be a meta-Gaussian process with marginal cdf F_Y and a transformation g satisfying the assumptions of Chapter 8. The pointwise and global error (7.74) of the convergence (8.15) splits up by the estimate (7.75) into the error of the iid setting and the deviation Δ_n''' between the maxima distribution of the dependent process Y_t and its associated independent sequence Y_t^* (cp. Sect. 7.2.1).

8. Extreme value theory for meta-Gaussian processes

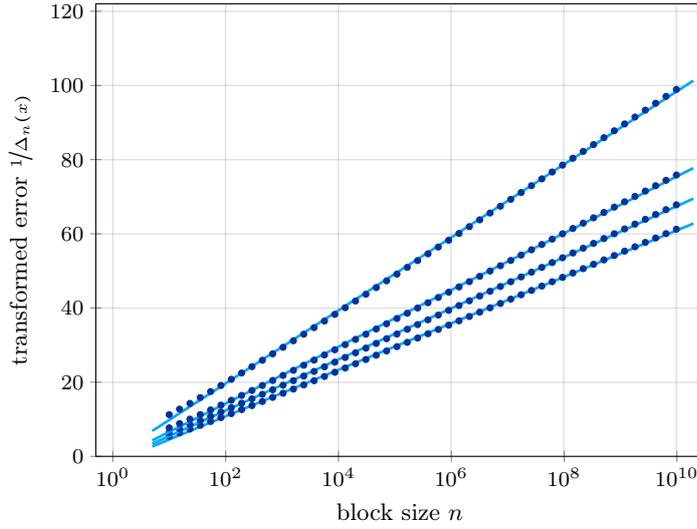


Fig. 8.1 Rate of decay of the pointwise error $\Delta_n(x)$ (7.40) at the position $x = 1$ as the number n of iid tGp samples with shift parameter $\nu = 0$, scale parameter $\sigma = 1$ and different exponents η increases. For the sake of perceptibility the graph depicts the transformed error $\frac{1}{\Delta_n(x)}$. The slopes estimated by linear regression are 2.7, 2.9, 3.3 and 4.3 for the exponents $\eta \in \{0.5, 1, 1.5, 2.5\}$ (from bottom to top).

Using the notions (8.1) and $u_n(x) := \frac{x}{a_n} + b_n$ for a value $x \in \mathbb{R}$ in the support of the cdf F_Y , by relation (8.2) the monotonicity of the inverse transformation g^{-1} preserves the probabilities of both the maxima M_n of the process Y_t and the maxima M_n^* of the associated independent sequence Y_t^* when reconstructing the maxima \tilde{M}_n and \tilde{M}_n^* of the underlying Gaussian LRC and iid process, respectively. Then the error $\Delta_n'''(x)$ satisfies

$$\begin{aligned}
 \Delta_n'''(x) &= |\mathbb{P}(M_n \leq u_n) - F_Y(u_n)^n| \\
 &= |\mathbb{P}(M_n \leq u_n) - \mathbb{P}(M_n^* \leq u_n)| \\
 &= |\mathbb{P}(\tilde{M}_n \leq g^{-1}(u_n)) - \mathbb{P}(\tilde{M}_n^* \leq g^{-1}(u_n))| \\
 &= |\mathbb{P}(\tilde{M}_n \leq g^{-1}(u_n)) - \Phi(g^{-1}(u_n))^n| \quad \propto n^{-c_1} (\ln n)^{-c_2} \quad (n \rightarrow \infty)
 \end{aligned} \tag{8.21}$$

for some constants $c_1, c_2 \in \mathbb{R}_{>0}$ by Lemma 7.22. Note that for ARFIMA processes X_t the power-law decay (7.77) of the deviation (7.76) in Lemma 7.22 is valid whatever the thresholds $g^{-1}(u_n)$ for fixed values x [101, Thm. 4.6.3]. The norming constants for meta-Gaussian processes obtained by the Taylor approach in Section 8.1 are not necessarily optimal. Equation (8.21), however, remains valid for alternative norming constants as well. The rate of the convergence (8.15) is dominated by the slower of the two influencing rates, the one of the iid setting and Δ_n''' (7.73) for the stationary setting.

The Examples 8.7, 8.8 and 8.9 consider the rates of the decay of the error $\Delta_n(x)$ for the meta-Gaussian processes obtained by transforming an ARFIMA process to an LRC process with tGp, lognormal or exponential marginal distribution. Given appropriate norming constants and a position x , the values $\Delta_n(x)$ are directly calculated¹⁰ by the joint cdf in equation (8.16) without any approximation of the inverse function g^{-1} or of the norming constants.

Example 8.7 (Rate of convergence for tGp-transformed ARFIMA processes): Figure 8.2 shows the asymptotic $\frac{1}{\ln n}$ rate of convergence for tGp-transformed ARFIMA(0, d , 0) processes dominated by the error of the iid setting (cp. Ex. 8.3) after a crossover from the polynomial decay (8.21) using the norming constants a_n and b_n given by equations (8.12).

As a remark, equation (8.21) remains valid for meta-Gaussian LRC processes with a Type-II or Type-III marginal distribution. Inverse sampling is capable of generating any marginal distribution by transforming a stationary Gaussian LRC process. An example for this is the uniform distribution discussed in Example 8.4 with linear convergence of the maxima distribution in the iid setting like for the exponential distribution but with Type-III limit.

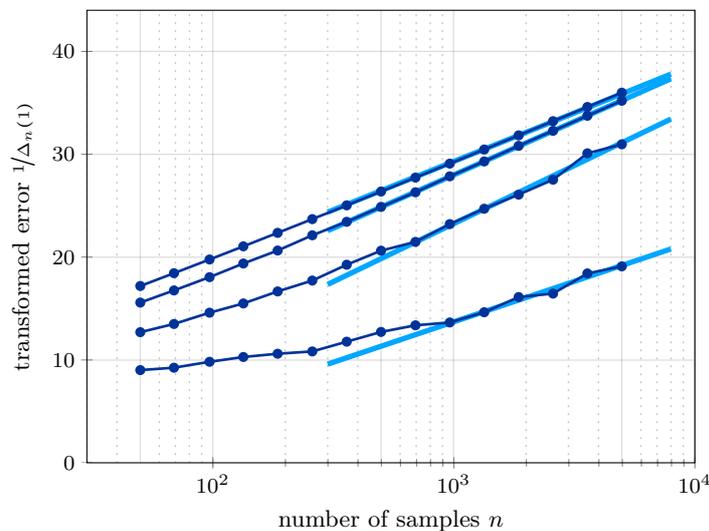


Fig. 8.2 Rate of decay of the pointwise error $\Delta_n(x)$ (7.74) at the position $x = 1$ as the number n of tGp-transformed ARFIMA(0, d , 0) samples with shift parameter $\nu = 0$, scale parameter $\sigma = 1$ and exponent $\eta = 2.5$ increases. For the sake of perceptibility the transformed error $\frac{1}{\Delta_n(x)}$ is depicted. The slopes estimated by linear regression are 4.1, 4.5, 4.9 and 3.4 for LRC parameters $d \in \{0, 0.1, 0.2, 0.3\}$ (from top to bottom).

¹⁰done by the minimax exponentially tilted (MET) estimator in MATLAB software package mvncdf [25]

Example 8.8 (Rate of convergence for ARFIMA processes transformed to lognormal): The rate of the convergence for an iid lognormal process is $\frac{1}{\sqrt{\ln n}}$ and dominates the rate of the error $\frac{1}{\Delta_n(x)}$ for exponentially-transformed ARFIMA processes. Figure 8.3 (right panel) shows the error in double-logarithmic scaling for block sizes up to $n = 5000$. In this range, the power-law decay still dominates the asymptotically slower root-logarithmic decay. Visualizing the rate $\frac{1}{\sqrt{\ln n}}$ by depicting $\frac{1}{\Delta_n(x)^2}$ with logarithmic x-axis requires extensive numerical effort already for the iid process. Note that compared to the error in the left panel of Figure 8.3 for a meta-Gaussian process with exponential marginal distribution from Example 8.9 the decay for the iid process with LRC parameter $d = 0$ is qualitatively slower than a power-law.

Example 8.9 (Rate of convergence for ARFIMA processes transformed to exponential): The optimal rate of the convergence for the maxima of iid exponentially distributed variates is linear for the asymptotic norming constants given in Example 7.17. Using these norming constants in equation (8.21) for an LRC process with exponential marginal distribution obtained by inverse sampling of an ARFIMA(0, d , 0) process implies the power-law decay of the error (8.15) depicted in Figure 8.3 (left panel). In this case, the power-law decay of the error Δ_n''' dominates the linear decay of the iid setting.

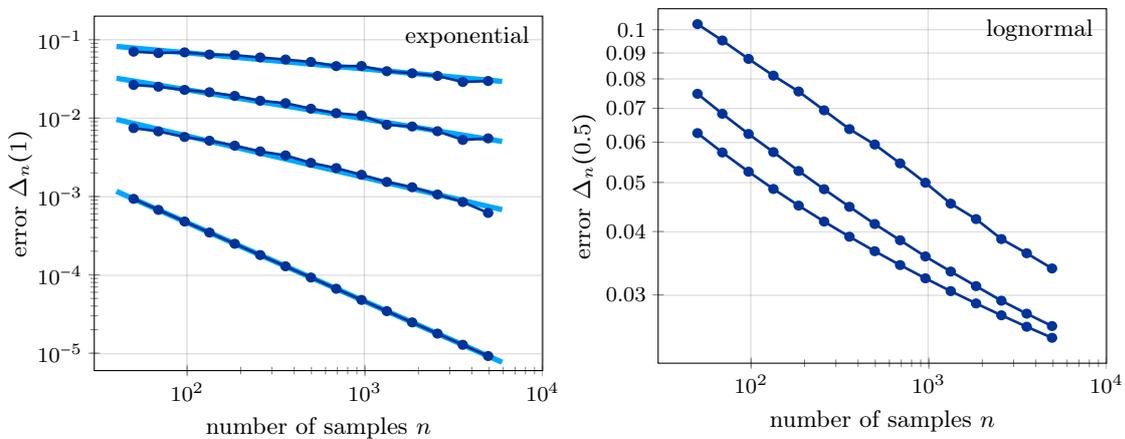


Fig. 8.3 Rate of decay of the pointwise error $\Delta_n(x)$ (7.40) at the position $x = 1$ (left) and $x = 0.5$ (right) as the number n of transformed standard Gaussian ARFIMA samples with exponential (left) and lognormal (right) marginal distribution increases for LRC parameters $d \in \{0, 0.1, 0.2, 0.3\}$ (left) and $d \in \{0, 0.1, 0.2\}$ (right) from bottom to top. The corresponding slopes estimated by linear regression are $-1, -0.5, -0.4$ and -0.2 (left).

The transformation of Gaussian processes is capable of changing both the extremal type of the marginal distribution and the asymptotic rate of the convergence in the extremal types theorem Theorem 7.3 for iid and in the generalized extremal types theorem Theorem 7.9 for stationary processes, respectively. The meta-Gaussian processes considered in the Examples 8.7, 8.8 and 8.9 obey an equally fast, slower and faster asymptotic rate of convergence. Due to the max-stability of the limit EVDs, the statistical application of EVT by adjusting an EVD to an empirical maxima distribution implicitly reflects these rates and the type of the limit EVD by the estimated GEV parameters irrespective of knowledge about precise optimal norming constants or the extremal index. From a theoretical perspective, a transformation of a Gaussian process to a meta-Gaussian process with a faster convergence of its maxima distribution improves the estimation of return levels (cp. Sect. 7.4.1). In practical applications, however, such an approach is meaningless as elaborated in Section 9.2.

8.3. Closing remarks

The presence of dependencies in stochastic processes affects statistical inference. Similar to the idea of an effective sample size based on the sample mean in Chapter 5, Section 8.3.1 provides an outlook on an effective block size in the EVT for stationary processes. The effective sample size has implications on the large deviation theory of the sample mean (cp. Sect. 5.3.3). Analogously, an effective block size serves for the generalization of the large deviation theory for extreme events outlined in Section 8.3.2.

8.3.1. Effective block size of stationary processes

The effective sample size established in Chapter 5 provides a quantitative description of the impact of correlations on the statistics of the sample mean. The Hermite approach yields both asymptotic and finite sample effective sample sizes for meta-Gaussian processes. The generalized extreme value theorem Theorem 7.9 raises the question for an effective block size for stationary processes compared to their associated iid sequence both as an asymptotic and a finite-block-size property.

8. Extreme value theory for meta-Gaussian processes

Comparing the variance of the sample mean for a stationary process to the one of its associated iid sequence, the effective sample size according to the sample mean bases on the (finite-time) decorrelation time (5.6) and (7.23), respectively. A comparison of the global error Δ_n (7.75) between the limit EVD and the distribution of the normed maxima provides an approach to quantifying statistical uncertainty in EVT by effective block sizes for stationary processes. Theorem 7.10 states that for the maxima M_n of a stationary process and the maxima M_n^* of its associated iid sequence the same norming constants are valid. If the process obeys unit extremal index, then comparing the error Δ_n provides an approach to quantifying statistical uncertainty in EVT by effective block sizes for stationary processes similar to the effective sample size in Section 5.1.4. The setting of an extremal index requires a more detailed consideration as discussed below.

Let a stationary process obey unit extremal index and let a_n and b_n be norming constants such that the corresponding normed maxima approach an EVD with cdf G . For meta-Gaussian LRC processes under mild assumptions Section 8.2 derives the two convergence rates in Theorem 7.9 by a Taylor methodology. Denote by \mathcal{R} and \mathcal{R}^* the decay rates of the errors

$$\begin{aligned} |\mathbb{P}(M_n^* \leq u_n) - G(x)| &\propto \mathcal{R}_n^* \\ |\mathbb{P}(M_n \leq u_n) - G(x)| &\propto \mathcal{R}_n \end{aligned} \quad (8.22)$$

as the number n of samples increases to infinity for thresholds $u_n = \frac{x}{a_n} + b_n$. Comparing the asymptotic behavior of the two rates (8.22) yields an asymptotic description of an effective block size n_{eff} by solving

$$c_1 \mathcal{R}_{n_{\text{eff}}}^* \stackrel{!}{=} c_2 \mathcal{R}_n, \quad (8.23)$$

where the constants $c_1, c_2 \in \mathbb{R}_{>0}$ denote the maximal prefactors in the proportionalities (8.22) over all values $x \in \mathbb{R}$. Mind that this maximum exists as pointwise convergence implies uniform convergence in case of continuous limit functions (cp. Appx. C).

Example 8.10: The lognormal and tGp transformation preserve the asymptotic rates of the convergence to the Gumbel limit from iid and LRC lognormal or tGp variates (cp. Exs. 8.8 and 8.7). Asymptotically by $\mathcal{R}_n = \mathcal{R}_n^* = \frac{1}{\ln n}$ (and $\frac{1}{\sqrt{\ln n}}$, respectively) the corresponding effective block sizes are thus of order

$$n_{\text{eff}} \propto n \quad \text{and} \quad n_{\text{eff}} \propto n, \quad (8.24)$$

respectively, as the block size n increases. For the meta-Gaussian LRC process with exponential marginal distribution (cp. Ex. 8.9) both rates are power-laws but with possibly different exponents as $\mathcal{R}_n^* = \frac{1}{n}$ and $\mathcal{R}_n = \frac{1}{n^\kappa}$ for some exponent $\kappa \in \mathbb{R}_{>0}$. The asymptotic effective sample size is thus of order

$$n_{\text{eff}} \propto n^\kappa. \quad (8.25)$$

The precise calculation of the prefactors in the proportionalities (8.22) is delicate and requires extensive estimations as for example elaborated for the Gaussian distribution itself in [67], the lognormal transformation in [105] and powers of the absolute value of a Gaussian in [68]. Such estimates depend on the norming constants and are a matter of current research. Using the norming constants (7.52) based on the Lambert W function recently led to an improved estimate of the uniform error for Gaussian variates [187]. Mind that the logarithmic rate is optimal for maxima of Gaussian variates and cannot be improved further. Using such results for quantifying asymptotic effective block sizes precisely allows for their practical interpretation.

In case of an extremal index other than one (cp. Sect. 7.2), the limit EVDs G^θ and G for the stationary process itself and its associated iid sequence, respectively, coincide in types but differ in their location for Type-I and additionally in their scale for Type-II and -III EVDs. Mind that by the max-stability (7.8) the power G^θ of a $\text{GEV}(\mu, \sigma, \xi)$ cdf G is a cdf of an EVD of the same type with $G^\theta \sim \text{GEV}\left(\mu - \frac{\sigma}{\xi}(1 - \theta^{-\xi}), \sigma\theta^\xi, \xi\right)$ for shape parameters $\xi \neq 0$ and $G^\theta \sim \text{GEV}(\mu + \sigma \ln \theta, \sigma, 0)$ if $\xi = 0$, respectively. Due to different limit EVDs of the normed maxima of the stationary process and its associated iid sequence, defining an effective block size is ambiguous in case of an extremal index other than unity. The deviations Δ_n indeed depend on the choice of the norming constants, which determine how the normed maxima approach their limit EVD. Statistical uncertainty is then not only subject to the rate of the convergence but also to the different ways of how the normed maxima evolve to their asymptotic location. Given norming constants a_n, b_n valid for the associated iid sequence, a reasonable assessment of effective block sizes might be a comparison of the errors

$$\left| \mathbb{P}\left(M_n^* \leq \frac{x}{a_n} + b_n\right) - G(x) \right| \quad \text{and} \quad \left| \mathbb{P}\left(M_n \leq \frac{x}{\tilde{a}_n} + \tilde{b}_n\right) - G(x) \right| \quad (8.26)$$

with norming constants \tilde{a}_n, \tilde{b}_n chosen for compensating the shift. Mind that the normed maxima of the stationary process satisfy $a_n(M_n - b_n) \rightarrow G^\theta$ as the block size $n \rightarrow \infty$. If the limit EVD is the cdf $G = \Lambda$ of the standard Gumbel distribution, then $\tilde{a}_n = a_n$ and $\tilde{b}_n = b_n + \frac{\ln \theta}{a_n}$ satisfy equation (8.26). Example 8.11 depicts the two different choices of norming constants for an example process.

8. Extreme value theory for meta-Gaussian processes

Example 8.11: For the sake of perceptibility consider an alternative of Example 7.11 with extremal index $\theta = \frac{1}{5}$. Define the process $X_i := \{Y_i, Y_{i+1}, \dots, Y_{i+4}\}$ with iid standard Gaussian variates $(Y_i)_{i \in \mathbb{N}}$ with cdf $F = \Phi$. Then the marginal cdf of the process $(X_i)_{i \in \mathbb{N}}$ reads Φ^5 . Analogously to equation (7.34) and (7.35), the distributions of the maxima M_n and M_n^* of the process X_i and its associated iid sequence, respectively, satisfy

$$\begin{aligned} \mathrm{P}(M_n \leq \tilde{u}_n) &= \Phi(\tilde{u}_n)^n \Phi(\tilde{u}_n) \rightarrow \Lambda(x) \\ \mathrm{P}(M_n^* \leq \tilde{u}_n) &= \Phi(\tilde{u}_n)^{5n} \rightarrow \Lambda(x)^5 \end{aligned} \quad \text{as } n \rightarrow \infty \quad (8.27)$$

for normed thresholds $\tilde{u}_n = \frac{x}{\alpha_n} + \beta_n$ with norming constants α_n, β_n for example as defined by equations (7.48). The norming constants α_n, β_n for the cdf Φ yield the norming constants $a_n = \alpha_n$ and $b_n = \frac{\ln(1/\theta)}{\alpha_n} + \beta_n$ for the cdf $\Phi^{1/\theta}$, such that

$$\mathrm{P}(M_n^* \leq u_n) \rightarrow \Lambda(x) \quad \text{and} \quad \mathrm{P}(M_n \leq u_n) \rightarrow \Lambda(x - \ln \theta) \quad \text{as } n \rightarrow \infty \quad (8.28)$$

for normed thresholds $u_n = \frac{x}{a_n} + b_n$. Figure 8.4 shows the empirical and theoretical distributions (8.28) for block sizes $n = 365$ both for the Gaussian distribution and analogously for the exponential distribution with $F(x) = 1 - e^{-x}$.

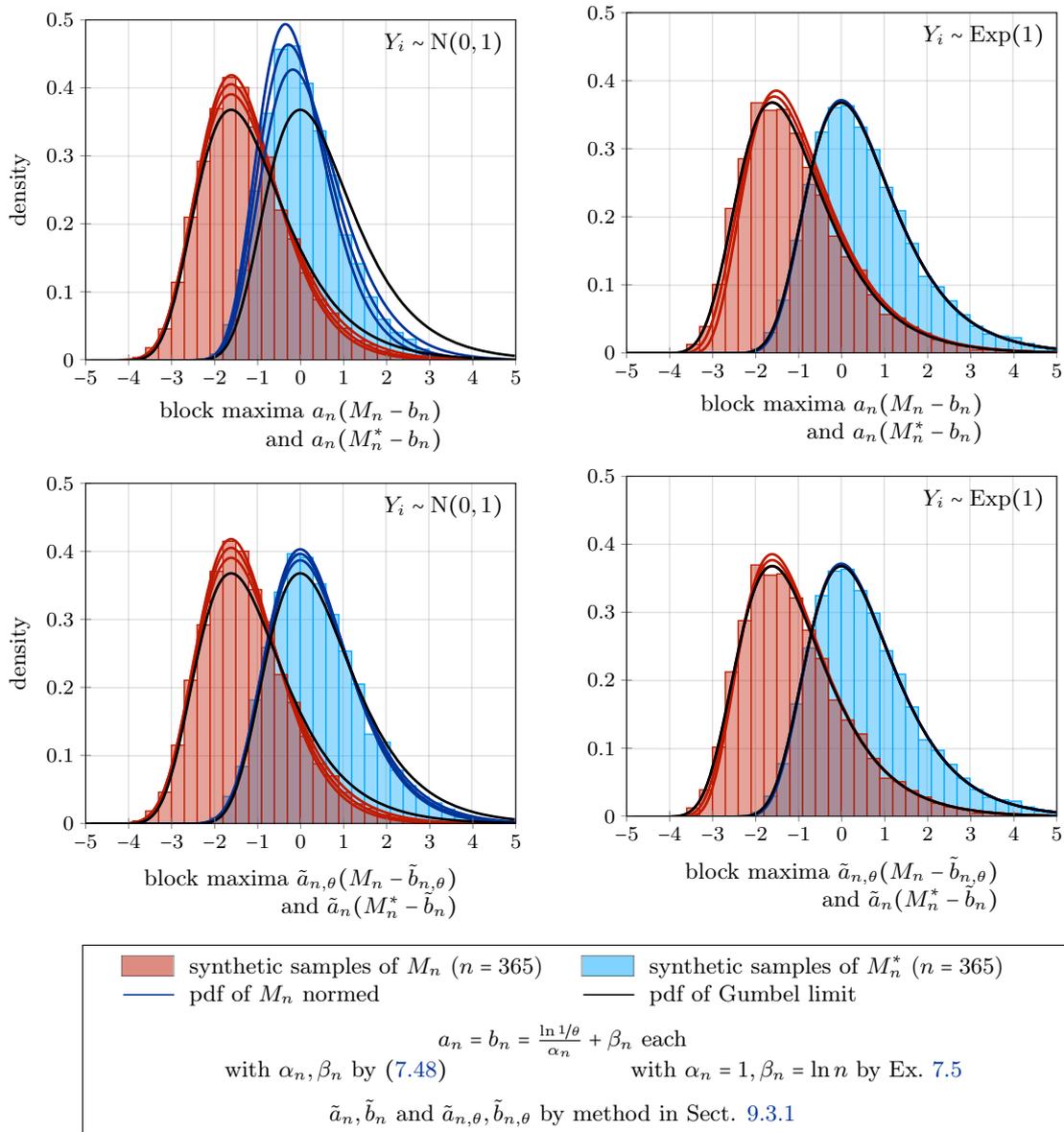


Fig. 8.4 Distribution of the maximum M_n for the process X_i defined in Example 8.11 and of the maximum M_n^* of its associated iid sequence for the marginal cdf of the underlying process Y_i chosen as standard Gaussian (left) and exponentially distributed (right). The colored pdfs depict the distributions (8.28) of the normed maxima for block sizes $n \in \{100, 365, 1,000\}$ (left) and $n \in \{10, 20, 365\}$ (right). The histograms show the empirical distribution for $n = 365$ and 1,000 sample block maxima of the process X_i (left bulk each) and its associated iid sequence (right bulk each). The norming constants a_n, b_n for the cdfs $F^{1/\theta}$ follow directly from the constants α_n, β_n known for the cdf F (top row) and are alternatively chosen by the method in Section 9.3.1 compensating the shift by $\ln \theta$. Note that by the $\frac{1}{n}$ -convergence for maxima of exponentially distributed variates the pdfs of the normed maxima for the finite block sizes n are already visually indistinguishably close to the limit Gumbel pdf.

8. Extreme value theory for meta-Gaussian processes

In case of an extremal index other than unity, the conception of an effective block size by the errors Δ_n in Theorem 7.10 is ambiguous in the following sense. Other than for the effective sample size, where the sample mean calculates equally for both the stationary process and its associated iid sequence, the distributions of the maxima M_n and M_n^* approach their limit GEV shape in a different manner and depending on the chosen norming constants. Example 8.11 visualizes that even though the rates of the convergences (8.28) and thus the effective and chosen block size qualitatively coincide asymptotically, the limit EVD and the error Δ_n when approaching it are quantitatively different. Thus, even when corrected for the location shifted by $\ln \theta$ for a Type-I and $-\frac{1}{\xi}(1 - \theta^{-\xi})$ for a Type-II or -III limit a precise definition of an effective block size with an intuitive, quantitative interpretation analogously to the effective sample size remains subtle and an open question.

In case of a unit extremal index asymptotically the distributions of the maxima of the stationary process and those of its associated iid sequence normed by the same norming constants coincide. As the extremal index is an asymptotic property itself, there are processes with unit extremal index but dependencies among individual samples might still affect the finite-block-size statistics. The comparison of the errors Δ_n then yields a quantitative description of an effective block size as both an asymptotic and finite-block-size property. As an example, Gaussian processes that satisfy Berman's condition (7.39) obey unit extremal index. Strong SRC, nonetheless, yield finite-block-size return levels significantly different from the infinite-block-size limit (cp. Fig. 9.2). Moloney et al. [131] reason an extremal index for finite block sizes by the mean cluster size and infer finite-block-size properties.

Adjusting a GEV distribution to empirical maxima captures both asymptotic and finite-block-size extremal index effects in the optimal model parameters estimated by ML. Section 9.2 deals with how incorporating outside knowledge about the dependence structure of the process helps interpret the results obtained from applying EVT. The object of Section 9.3 is establishing a method for the estimation of return levels by an appropriate choice of norming constants.

8.3.2. Large deviations in extreme value theory

Section 5.3.3 links the effective sample size of a time series measured by the variance of its sample mean to the large deviation theory of the sample mean. Section 8.3.1 allows for applying the idea of an effective block size to the large deviation theory for classical extreme value theory as a generalization to the maxima of stationary processes. For a recap of [182] let M_n denote the maximum of $n \in \mathbb{N}$ iid samples with a cdf in the Gumbel domain of attraction and norming constants $a_n, b_n \rightarrow \infty$ such that $a_n(M_n - b_n)$ convergences in distribution to a Gumbel variate. Then asymptotically the mean of the maximum M_n satisfies $\mathbb{E}[M_n] \sim \frac{\mu_\Lambda}{a_n} + b_n \sim b_n$ as the block size $n \rightarrow \infty$, where μ_Λ denotes the mean of the Gumbel distribution. Similar to large deviations of the sample mean in equation (5.50), classical large deviations for EVT aims at additive deviations of the form

$$\begin{aligned} \text{LDP}_\lambda(N, \varepsilon) &:= \mathbb{P}(M_n - \mathbb{E}[M_n] > \varepsilon) \\ &\sim \mathbb{P}(M_n > b_n + \varepsilon) \sim 1 - e^{-e^{-\varepsilon}} = \mathcal{O}(e^{-\varepsilon}) \end{aligned} \quad (8.29)$$

as $n \rightarrow \infty$. The deviation $\text{LDP}_\lambda(N, \varepsilon)$ hence measures deviations described by the limit Gumbel distribution itself with $\mathcal{O}(e^{-\varepsilon}) = \mathcal{O}(1)$ dependency on the considered deviation ε . Aiming at a rate function $I(\varepsilon)$ analogously to the rate function (5.51) for large deviations of the sample mean, an alternative measure of large deviations in EVT are multiplicative large deviations of the form

$$\text{LDP}_\Lambda(n, \varepsilon) := \mathbb{P}(M_n > b_n \varepsilon). \quad (8.30)$$

For large deviations $\text{LDP}_\Lambda(n, \varepsilon)$ of the maximum of exponential and Gaussian samples Vivo [182] derives the rate functions

$$\text{Exp}(1) \quad \mathbb{P}(M_n > b_n \varepsilon) \sim e^{-(\ln n)(\varepsilon-1)} \quad (n, \varepsilon \rightarrow \infty) \quad (8.31)$$

$$\text{N}(0, 1) \quad \mathbb{P}(M_n > b_n \varepsilon) \sim e^{-(\ln n)(\varepsilon^2-1)} \quad (n, \varepsilon \rightarrow \infty). \quad (8.32)$$

The rate functions (8.31) and (8.32) obtained by multiplicative deviations (8.30) carry more information than the additive large deviations (8.29) as they not only include the limit Gumbel shape of the maxima distribution but also the marginal distribution of the individual samples.

For meta-Gaussian iid processes with appropriate norming constants $a_n, b_n \rightarrow \infty$ the asymptotic behavior of the mean of $\mathbb{E}[M_n] \sim b_n$ potentially differs from the original Gaussian setting yielding non-standard large deviation results of the form

$$\text{LDP}_\Lambda(n, \varepsilon) \propto e^{-(\ln n)I(\varepsilon)} \quad (n, \varepsilon \rightarrow \infty). \quad (8.33)$$

8. Extreme value theory for meta-Gaussian processes

Note that for the exponential and Gaussian marginal distribution with corresponding centering constants b_n each the mean of the maximum M_n in leading order grows at rate $b_n \sim \ln n$ and $b_n \sim \sqrt{2 \ln n}$, respectively, as the block size $n \rightarrow \infty$ with corresponding large deviations

$$\text{Exp}(1) \quad \mathbb{P}(M_n > b_n \varepsilon) \sim e^{-b_n(\varepsilon-1)} \quad (n, \varepsilon \rightarrow \infty) \quad (8.34)$$

$$\text{N}(0, 1) \quad \mathbb{P}(M_n > b_n \varepsilon) \sim e^{-\frac{b_n^2}{2}(\varepsilon^2-1)} \quad (n, \varepsilon \rightarrow \infty). \quad (8.35)$$

The n -dependence in the asymptotic large deviation (8.33) originates in the behavior of the mean $\mathbb{E}[M_n]$ rather than in different rates of convergence to the common Gumbel limit. For stationary processes and, in particular, meta-Gaussian processes the finite and asymptotic effective block sizes provide access to transfer the results (8.34) and (8.35) by reformulating the expression (8.33) using the effective block size n_{eff} similar to Section 5.3.3 for large deviations of the sample mean.

9. Improved estimation of return levels by incorporating outside knowledge

The advantage of EVT lies in the universality of the generalized extremal types theorem Theorem 7.9 for the estimation of return levels (cp. Sect. 7.4). Under mild conditions the limit distribution of the maxima of stationary processes is an EVD irrespective of the process' marginal distribution and its specific dependence structure (cp. Sect. 7.2). Even if empirical data provides (annual) maximum measurements only, an application of EVT by adjusting an EVD to the empirical maxima distribution is possible. The presence of further knowledge on the stochastic properties of the quantity of interest, for example, in form of measurements on higher (temporal) resolution such as daily records, provides a more detailed interpretation of the results obtained by EVT. Our focus lies on stationary processes that satisfy the condition $D(u_n)$. Additional data on higher resolution allows for an assessment of the stationarity of the process that generated recorded extremes and thus of the applicability of the EVT for stationary processes. Section 6.1.1 elaborates such an analysis for daily precipitation amounts.

Remark 9.1: There are generalizations of EVT for certain nonstationary Gaussian processes with additive trends on bounded growth rates [101, Sect. 6.1]. For Gaussian processes with strong dependence due to an acf decaying slower than Berman's condition (7.39), however, the asymptotic limit of properly normed maxima obey even distributions other than EVDs [101, Sect. 6.4 - 6.6] (cp. Rem. 7.16).

Many distributions of high practical relevance such as the Gaussian oder exponential distribution are in the Gumbel domain of attraction. Furthermore, the discription of these Type-I EVDs requires one parameter less than for Type-II or -III distributions. Using a Gumbel estimation of the distribution of extremes is thus a popular approach because of its higher parsimony compared to a full GEV estimation [96]. As an EVD captures the statistical properties of extremes only in the asymptotic limit of infinite block sizes statistical uncertainty applying EVT has at least three origins

- (i) Domain of attraction: The choice of the Gumbel distribution is possibly inappropriate if the marginal distribution is in the Type-II or -III domain of attraction.
- (ii) Rate of convergence: The error of the empirical to the theoretical limit EVD is possibly large for finite block sizes.
- (iii) Extremal index: The limit EVDs of a stationary process and of its associated iid sequence possibly differ due to an extremal index other than unity.

9. Improved estimation of return levels by incorporating outside knowledge

The availability of additional data on the marginal distribution and the dependence structure of the process underlying recorded extremes allows for an assessment of the above sources of uncertainty. Daily records yield an empirical marginal distribution for a verification of its domain of attraction for handling item (i). Possible reasons for slow rates of convergence in item (ii) are the shape of the marginal distribution itself, like in case of the Gaussian (see Sect. 7.3), lognormal (see Ex. 8.5) or tGp (see Ex. 8.6) distribution, or long-term dependencies among individual samples (cp. Ex. 8.9 for an LRC process with exponential marginal). Note that the slow logarithmic convergence of the iid setting dominates also the convergence of the maxima distribution for Gaussian and certain meta-Gaussian processes with power-law decaying acf (cp. Sect. 8.2.2). By Theorem 7.10 the limit EVD of a stationary process with unit extremal index and of its associated iid sequence asymptotically coincide. In case of an extremal index other than unity in item (iii), for example due to strong short-range dependencies among individual records, the asymptotic or finite-block-size distribution of extremes differs in location and scale from the iid setting (cp. Ex. 8.11).

A comparison of return levels obtained from the marginal distribution helps interpret return levels deduced by EVT for treating items (i) and (iii). In any case, a comparison of return levels obtained from a Gumbel and from a full GEV estimation (cp. Sect. 7.4) is a reasonable approach to item (ii) as in some cases it is known that the penultimate approximation by a GEV distribution obeys a faster rate of convergence than the ultimate approximation by a Gumbel distribution (cp. Sect. 7.3.3). If the two estimates are close to each other, then the choice of a Gumbel estimate of return levels is likely to be appropriate and decreases statistical uncertainty.

Section 9.1 discusses different definitions of annual return levels by the marginal and maximum distribution, respectively. Section 9.2 compares return levels obtained from the marginal distribution, a Gumbel and a full GEV estimation for synthetic time series and empirical precipitation data. Section 9.3 establishes an approach to improving the uncertainty of return level estimates by combining the smaller confidence intervals due a marginal distribution in the Gumbel domain and the additional information on return levels given by the marginal distribution for stationary processes with unit extremal index. The object of Section 9.4 is an outlook on a possible generalization of this approach to processes with extremal index other than unity and further techniques for reducing statistical uncertainty in the application of EVT.

9.1. Definitions of annual return levels

For stationary processes there are at least two possible definitions of annual return levels. The quantile Q_m of the distribution of block maxima with probability $1 - \frac{1}{m}$ yields the return level of extremes with a frequency of $\frac{1}{m}$, which occurs once every $m \in \mathbb{N}$ blocks of length $n \in \mathbb{N}$ on average. Alternatively, the quantile q_m of the marginal distribution with probability $1 - \frac{1}{nm}$ describes a return level for the individual measurements underlying the block maxima with frequency $\frac{1}{nm}$, that occurs on average once every $n \cdot m$ samples. The choice of a block size of $n = 365$ allows for an interpretation of the quantiles Q_m and q_m as annual return levels (cp. Sect. 7.4.1).

Let Y_t be a stationary process with marginal cdf F_Y . If the process Y_t is an iid processes, then the two conceptions of return levels base on quantiles of the marginal distribution by equation (7.2) as

$$Q_m = F_Y^{-1} \left(\sqrt[n]{1 - \frac{1}{m}} \right) \quad (9.1)$$

$$q_m = F_Y^{-1} \left(1 - \frac{1}{nm} \right) \quad (9.2)$$

and thus coincide as the deviation of the two probabilities describing the two return levels (9.1) and (9.2) reads

$$\left| \sqrt[n]{1 - \frac{1}{m}} - \left(1 - \frac{1}{nm} \right) \right| = \frac{1}{2nm^2} + \mathcal{O} \left(\frac{1}{nm^3} \right) \quad (n, m \rightarrow \infty). \quad (9.3)$$

For block size $n = 365$ and a return period of $m = 100$ annual blocks the error (9.3) measures about $1.37 \cdot 10^{-7}$. The obtained 100-year return levels Q_{100} and q_{100} thus highly coincide even for heavy-tailed distributions such as the Pareto distribution with infinite first moment. The 100-year return levels are of particular interest in applications (cp. Chap. 1).

Properly normed the Gumbel or GEV distribution approximates the maxima distribution of stationary processes by Theorem 7.9. If the process obeys unit extremal index such as iid processes, then the same limit distribution describes the maxima of the process and of its associated iid sequence. In this case, equation (9.1) and correspondingly equation (9.2) provide access to the return level Q_m and coincide with the return level q_m up to negligible deviations by equation (9.3). If the process Y_t obeys an extremal index other than unity, however, the two limit EVDs differ in location and scale, so that the return levels Q_m of annual maxima and q_m by the marginal distribution no longer agree. For example, strong short-range dependence shifts the return levels of annual extremes compared to those of an iid sequence with the same marginal distribution (cp. Exs. 7.11 and 8.11). In the sense of a finite-block-size extremal index such effects are also significant for process with asymptotically unit extremal index but strong SRC (cp. Sect. 8.3.1 and Fig. 9.2).

9. Improved estimation of return levels by incorporating outside knowledge

Aiming at an application of EVT to risk assessment, the quantiles Q_m are crucial for proper estimation of return levels of annual extremes rather than the quantiles q_m as an extremal index other than unity indicates quantitative differences in the occurrence of annual extremes.

Remark 9.2: Empirical return levels given by both the empirical marginal and maxima distributions are limited to return periods of at most the measurement range. Assuming and adjusting a stochastic model is necessary for the inference of return levels beyond the empirical horizon in either case (cp. Rem. 7.25).

9.2. Comparison of return level estimates

Section 9.1 provides three different approaches to return levels for stationary processes in the presence of an empirical marginal distribution for individual measurements. Comparing estimates for all three quantities together with involving further statistical properties of the individual samples helps properly infer the return levels Q_m of annual maxima by EVT. The three different estimates are

- the return level \bar{Q}_m obtained by a Gumbel estimate,
 - the return level \bar{Q}_m obtained by a full GEV estimate and
 - the return level \bar{q}_m estimated by the marginal distribution.
- (9.4)

Assume the input empirical data stationary. Otherwise, possible nonstationarities such as trends or seasonalities are to be removed from the data (cp. Sect. 6.1.1). The parameters estimated when adjusting an EVD to empirical data implicitly capture all the aforementioned influences on the statistics of extremes. By resolving the individual effects the origins of statistical uncertainty described in items (i) to (iii) above imply a step-by-step procedure to statistical inference for stationary processes by EVT. Coincidences among the different estimates (9.4) help interpret the validity of return level estimates.

For meta-Gaussian processes with a power-law decaying acf the error Δ_n''' (7.73) is a power law as well by Section 8.2. Aiming at improving the estimates of return levels for extreme precipitation amounts in Section 9.2.2, we focus on such meta-Gaussian processes for which the rate of the convergence of the maxima distribution is dominated by the logarithmic rate of their associated iid sequences. This assumption indicates that a possible deviation of the GEV estimate \bar{Q}_m from the estimate \bar{q}_m by the marginal distribution and thus from the return level of annual maxima of the associated iid sequence is not due to slower convergence for the dependent process in the generalized extremal types theorem Theorem 7.9 but rather due to a large sampling error for highly correlated data in general. As a remark, Example 8.9 describes a process with power-law convergence of its maxima distribution and a faster linear convergence for its associated iid sequence.

Our *step-by-step procedure for the comparison of return level estimates* reads

- (A) Comparison of the return levels \bar{Q}_m by a Gumbel and a full GEV estimation:
- If the two estimates are close to each other, the Gumbel distribution is an adequate choice and provides smaller confidence intervals.
 - A discrepancy of the two estimates is possibly due to:
 - (i) Domain of attraction:
The empirical marginal distribution helps identify whether the Gumbel distribution is an appropriate choice.
 - (ii) Rate of convergence:
→ If the marginal distribution is known for slow convergence to the Gumbel limit, a full GEV estimate is to be preferred [35].
- (B) Comparison of the return levels \bar{Q}_m by a full GEV (or if applicable by (A) a Gumbel) estimate and \bar{q}_m by the marginal distribution:
- If the two estimates are close to each other and long-range dependencies in the time series are only weak, then estimate (9.2) is appropriate and possibly provides smaller confidence intervals¹¹
 - A discrepancy of the two estimates is possibly due to:
 - (iii) Extremal index other than unity:
→ The estimate \bar{q}_m is inappropriate and the estimate \bar{Q}_m approximates best the return level of annual maxima among the three possibilities (9.4).

In the presence of strong long-range dependence, the empirical cdf might represent the true statistics of the process only poorly due to sampling errors. Knowledge about the long term dependence structure of the individual measurements helps interpret a coincidence of the estimated return levels \bar{Q}_m by EVT and \bar{q}_m by the marginal distribution. Under weak long-term dependence such a coincidence supports the validity of both, whereas strong long-term dependence implies the opposite. An approach to a validation of the conclusion (B)(iii) is a GEV estimation of the return level \bar{Q}_m for randomly shuffled versions of the original time series. Again sampling errors impede proper representation of the associated iid sequence already in case of strong short-term dependence. Section 9.2.1 gives examples for these two effects.

¹¹Confidence intervals for the estimation (9.2) depend on the method for adjusting a model cdf to the empirical data (cp. Rem. 9.2 and Sect. 9.4).

9. Improved estimation of return levels by incorporating outside knowledge

In any case, the penultimate approximation by a full GEV estimate provides the most reliable approach to return levels due to a flexible adjustment to possibly slow convergence in the extremal types theorem Theorem 7.9 and to a possibly quantitatively different limit distribution by an extremal index other than unity. Even if the marginal distribution is in the domain of attraction of the Gumbel distribution, the penultimate approximation by a GEV distribution possibly exhibits a faster rate of convergence (cp. Exs. 8.7 and 8.8). For some marginal distributions in the domain of attraction of Type-I EVDs the ultimate approximation by the Gumbel distribution provides an estimate of return levels with decreased statistical uncertainty (cp. Ex. 8.9). If the convergence to the Gumbel limit EVD is, however, too slow and the Gumbel distribution does not describe well the finite-block-size maxima distribution, then these two estimates significantly differ as described in comparison (A).

A comparison to the return level \bar{q}_m by the marginal distribution aims at reducing statistical uncertainty by usage of more individual data points. The POT approach (cp. Sect. 7.4.3) aims in the same direction in Section 9.4. Due to strong short-term dependence for finite block sizes or even asymptotically, the annual return level of the marginal distribution not necessarily coincides with the one of the annual maximum since the return level by the marginal distribution corresponds to the one of the associated iid sequence (cp. Sect. 8.3.1).

Section 9.2.1 visualizes the assessment of return level by comparing the different 100-year return levels (9.4) for synthetic time series, namely a tGp-transformed ARFIMA(0, d , 0) process, a Gaussian AR(1) process (both with unit extremal index) and the process from Example 8.11 with extremal index $\frac{1}{5}$. Section 9.2.2 applies the comparison of return levels to empirical data on daily precipitation amounts for the assessment of 100-year return levels of extreme precipitation events.

9.2.1. Application to synthetic data

The maxima distributions for ARFIMA, in particular, Gaussian AR, and tGp-transformed ARFIMA processes and their associated iid sequence asymptotically coincide due to unit extremal index (cp. Sect. 7.3.2 and Sect. 8.2). For finite block sizes, however, prominent clustering of extremes might occur due to strong LRC [29] or SRC [131] and significantly shift the location of the annual return level Q_m compared to return level q_m by the marginal distribution. The process $X_i = \max\{Y_i, Y_{i+1}, \dots, Y_{i+4}\}$ with an underlying iid process $Y_i \sim \text{Exp}(1)$ considered in Example 8.11 obeys an extremal index of $\theta = \frac{1}{5}$, so that asymptotically the maxima of the process X_i are shifted by $\ln \theta$ compared to those of its associated iid sequence.

The Figures 7.16, 9.1, 9.2 and 9.3 show comparisons of the different 100-year return levels Q_{100} and q_{100} (9.4) for these four example processes. All marginal distributions belong to the Gumbel domain of attraction but with different rate of convergence.

- (A) For the AR(1), ARFIMA(0, d , 0) and tGp-transformed ARFIMA(0, d , 0) processes the penultimate approximation of their maxima distribution is of the low order $\frac{1}{\ln n}$ as the block size n increases. All the Figures 7.16, 9.1 and 9.2 visualize a significant discrepancy between the two estimates. The penultimate approximation converges at the faster rate $\frac{1}{(\ln n)^2}$ and thus provides more reliable estimates of the m -year return level Q_m (cp. Sect. 7.3.3 and 8.2).

Figure 9.3 depicts a coincidence of the Gumbel and full GEV estimate for the process X_t with exponential marginal distribution, which is due to the fast linear rate of convergence for the exponential distribution in either case. The Gumbel estimate Q_m is reliable for this example and provides less statistical uncertainty.

- (B) Comparing the estimated return levels Q_m (via Gumbel or full GEV by (A)) to the corresponding return level q_m of the marginal distribution provides insight into the impact of short-range dependencies. Return levels estimated for shuffled versions of the original process serve as an approximation of the return level q_m , which is known for the example models but unknown in applications. Mind that q_m coincides with the return level Q_m of the maxima of the associated iid sequence (cp. Sect. 9.1).

For large AR-parameters $\varphi \geq 0.7$ the maxima of the AR(1) process and of its shuffled versions deviate from each other in Figure 9.2. Strong SRC are known to act like an extremal index other than zero in finite-block-size approximations (cp. Sect. 8.3.1). At the same time the sampling error of the marginal distribution of an AR process vanishes quickly due to an exponentially decaying acf. Hence, the estimate Q_m is reliable for this example and accounts for the finite-block-size effect of a shifted mean for the maximum distribution. For small AR parameters the return level q_m and the estimates Q_m coincide so that q_m provides a reliable estimate and smaller confidence intervals are possible depending on the method¹¹.

The influence of LRC is on average negligible for the estimation of the return level Q_{100} of the original and tGp-transformed ARFIMA(0, d , 0) processes up to very strong LRC with LRC parameter at least $d = 0.4$ by deviating from the known return level q_{100} in Figure 7.16 and Figure 9.1, respectively. Deviations here arise only for small numbers of blocks of about 20 because of the sampling error. The estimated return levels Q_m coincide for the original and shuffled time series for all LRC parameters but deviate from the analytically known return level q_m by the marginal distribution for strong LRC. This observation indicates that the LRC are too strong for any reliable return level estimate due to a significant sampling error describable, for example, by a drastically reduced effective sample size (cp. Sect. 5.1.2).

The acf of the process X_i decays rapidly and vanishes for all time lags $k \geq 5$. The deviation between the estimates Q_m for the original and random shuffled time series in Figure 9.3 indicating strong SRC effects together with the small sampling error and the fast convergence of the ultimate approximation by (A) yields the Gumbel return level Q_m as the estimate with lowest statistical uncertainty.

9. Improved estimation of return levels by incorporating outside knowledge

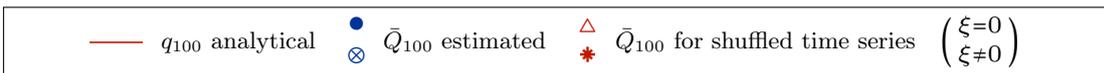
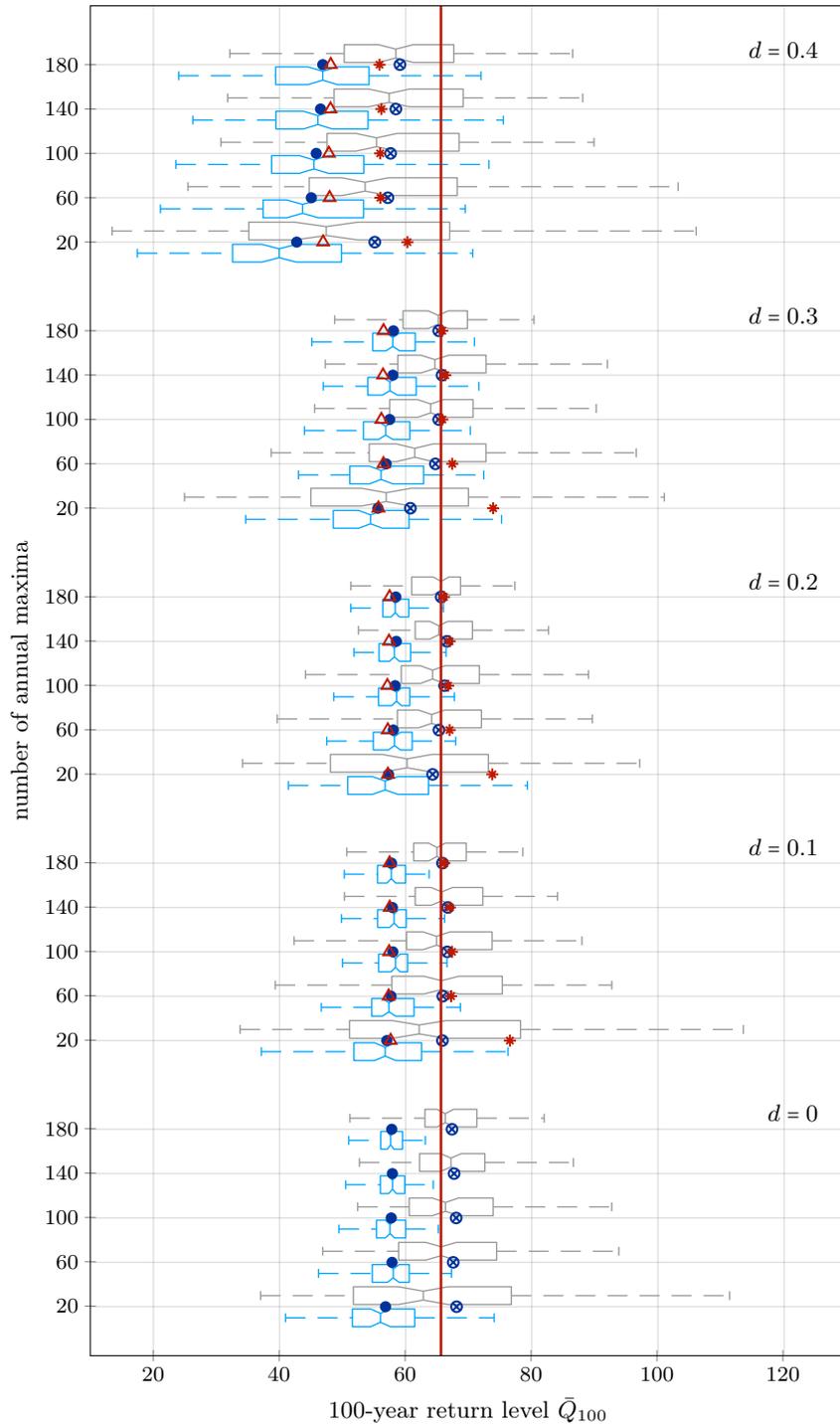
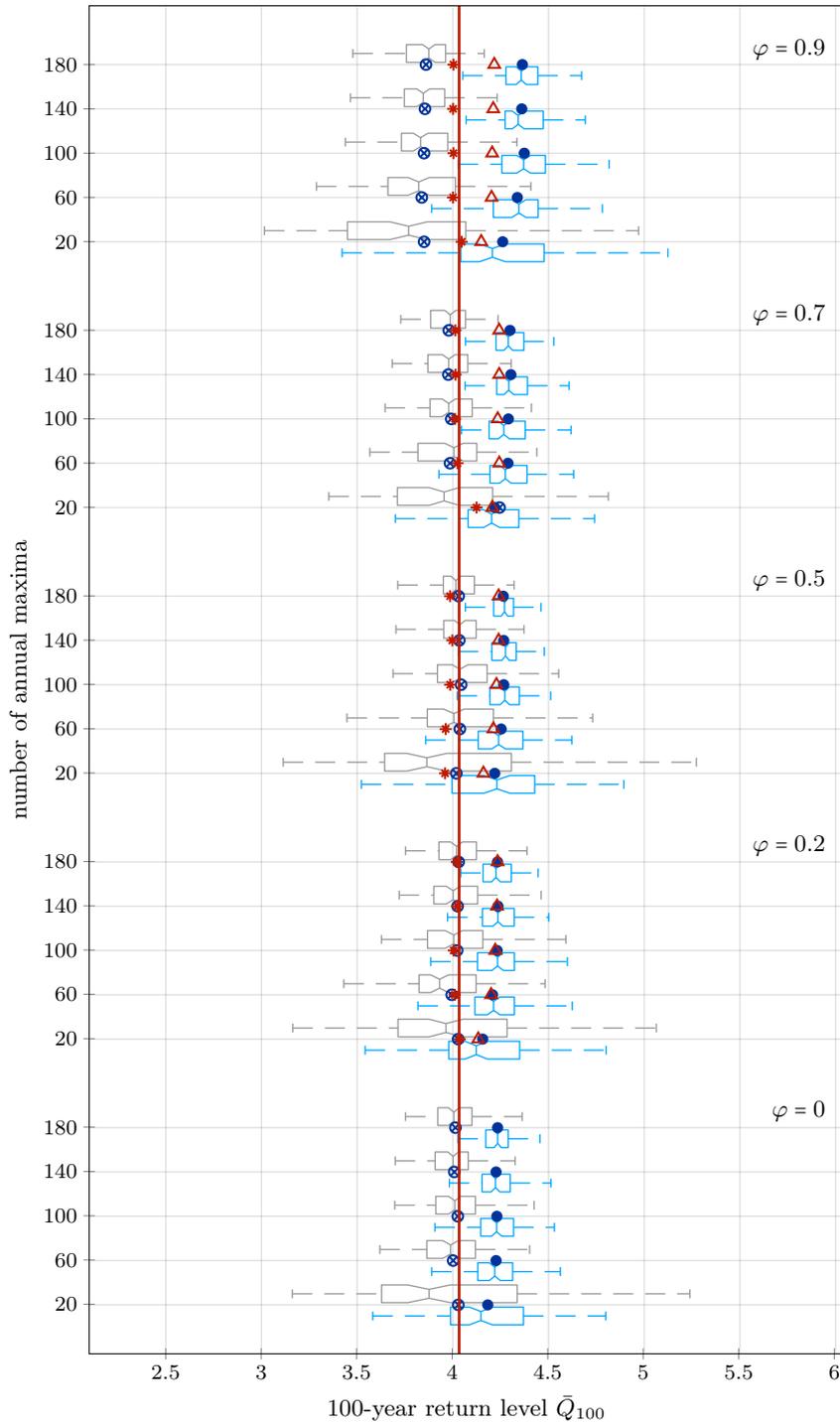


Fig. 9.1 GEV and Gumbel estimates of the 100-year return level $Q_{100} \approx 65.6$ of tGp-transformed ARFIMA(0, d , 0) time series with location $\nu = 0$, scale $\sigma = 1$ and exponent $\eta = 3$ for different LRC parameters d and different numbers of annual maxima included. The estimate \bar{Q}_{100} depicts the mean of the estimates of 100 sample time series visualized by the box plots. The boxes show their median and the 25% and 75% quantiles, the whiskers extend to the adjacent value each.



— q_{100} analytical
● \bar{Q}_{100} estimated
△ \bar{Q}_{100} for shuffled time series $\begin{pmatrix} \xi=0 \\ \xi \neq 0 \end{pmatrix}$

Fig. 9.2 GEV and Gumbel estimates of the 100-year return level Q_{100} of standard Gaussian AR(1) time series for different AR(1) parameters φ and different numbers of annual maxima included. The 100-year return level by the marginal distribution reads $q_{100} \approx 4.03$. The estimate \bar{Q}_{100} depicts the mean of the estimates of 100 sample time series visualized by the box plots. The boxes show their median and the 25% and 75% quantiles, the whiskers extend to the adjacent value each.

9. Improved estimation of return levels by incorporating outside knowledge

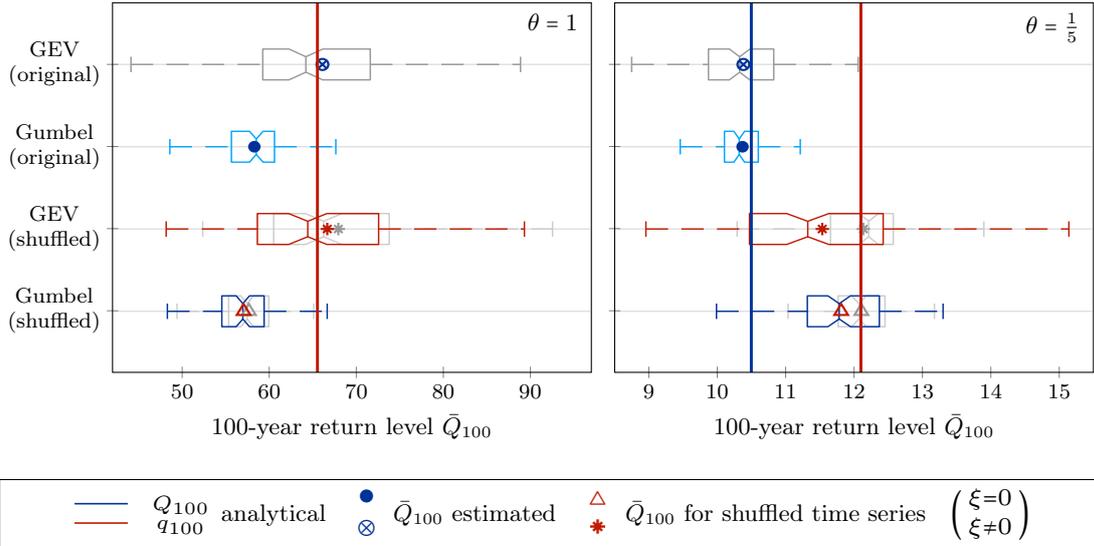


Fig. 9.3 Impact of SRC on Gumbel and GEV estimates of the return level Q_m (9.4) for the tGp-transformed ARFIMA(0, d , 0) process from Figure 9.1 with exponent $\eta = 3$ and LRC parameter $d = 0.2$ (left) and the process X_i from Example 8.11 with extremal index $\theta = \frac{1}{5}$ (right). The depicted estimates are for block size $n = 365$ and time series with $N = 36,500$ samples or equivalently $m = 100$ years. Note that the left panel reproduces the estimates located in the center of Figure 9.1.

The central conclusions for an assessment of the reliability and statistical uncertainty of return level estimates for stationary processes by EVT are

- Outside knowledge about the marginal distribution improves decisions about the validity of a Gumbel estimation based on the domain of attraction of the marginal distribution or if applicable on the rate of the convergence to the Gumbel limit.
- Outside knowledge about the long-range dependence structure among individual measurements improves the validity of return level estimates subject to effective sample sizes.
- Outside knowledge about the short-range dependence structure improves the reliability of the return level given by the marginal distribution.

The results in Section 9.1 and Section 9.2 remain valid for any return level other than the visualized 100-year return level as the convergence to the limit EVD in the generalized extreme value theorem Theorem 7.9 is uniform (cp. Appx. C). The ML estimate \bar{Q}_{100} yields results with lowest statistical uncertainty for the event that occurs on average once a year, which is represented by the location of the bulk of the corresponding EVD. The longer the considered return period is the higher gets the statistical uncertainty. In Section 9.2.2 we apply the comparison of return levels to the data sets of precipitation measurements discussed in Chapter 6 for the estimation of 100-year return levels of extreme precipitation events. Section 9.3 establishes an alternative approach to the estimation of return levels based on specifically chosen norming constants.

9.2.2. Application to daily precipitation amounts

Section 6.3 provides a detailed validation of tGp-transformed ARFIMA processes for modeling time series of mid-latitude daily precipitation amounts. The marginal distribution of such an estimated tGp model provides the return level q_m and allows for a comparison of this estimate to those of the return level Q_m by Gumbel and GEV estimates.

Figure 9.4 shows a comparison of the 100-year return levels Q_{100} and q_{100} estimated for the 20 recorded data sets introduced in Chapter 6 (cp. Tab. E.1). The relative deviation between the GEV estimates \bar{Q}_{100} and the estimates \bar{q}_{100} by the marginal distribution defined by $\frac{Q_{100}-\bar{q}_{100}}{Q_{100}}$ is less than $\pm 10\%$ and less than ± 10 mm for all data sets except for one (Schwerin with $\approx 18\%$) and two (Schwerin with ≈ 20 mm, Valencia with ≈ 26 mm), respectively. The analogously defined relative deviation between the GEV and Gumbel estimates of the return level Q_{100} , however, ranges between $\approx -3\%$ and $\approx 30\%$ with more than 10% difference for about half of the data sets (Fig. 9.4, left panel).

Applying the comparison described in Section 9.2 by incorporating outside knowledge about the individual measurements additionally to the annual maxima amounts only yields validity of the GEV estimates \bar{Q}_{100} due to high coincidence with the estimates \bar{q}_{100} by the marginal distribution. All the empirical time series exhibit both weak seasonality and weak LRC, so that neither violation of the assumption of stationarity nor a large sampling error crucially confound the estimates of return levels by EVT (cp. Sect. 6.1.1, the regression values in Tab. E.3 and the LRC parameters d in Tab. E.1). SRC in the time series are low to moderate with AR parameters φ in the range of 0.2 to 0.45 (cp. the AR parameters φ in Tab. E.1). Section 9.4 further discusses the impact of SRC.

The tGp distribution is in the Gumbel domain of attraction. As recapped in Chapter 8 the slow $\frac{1}{\ln n}$ -rate of the convergence of the corresponding maxima distribution to its limit Gumbel shape as the block size $n \rightarrow \infty$ improves to $\frac{1}{(\ln n)^2}$ using a penultimate approximation by a GEV distribution. We find that the larger the estimated shape parameter η of the tGp marginal distribution is, in other words, the heavier the tail of the marginal distribution is, the larger is the relative deviation between the Gumbel and GEV estimates (Fig. 9.4, right panel). Section 9.3 provides an explanation of the tendency of Gumbel estimates for maxima of tGp samples towards an underestimation of return levels and uses the observation of less deviation for exponents η close to two for an alternative approach to the estimation of return levels.

The higher appropriateness of GEV estimates for maxima of stationary tGp processes comes along with increased statistical uncertainty due to one additional parameter for the stochastic description of extremes. Aiming at smaller confidence intervals, the estimate \bar{q}_{100} by the marginal distribution provides an adequate alternative estimator given low SRC and LRC by Section 9.1. Using a Bayesian approach, for three example data sets Figure 9.5 shows confidence intervals for the estimates \bar{Q}_{100} by Gumbel and GEV estimation and \bar{q}_{100} by the marginal distribution, respectively, based on 100 synthetic sample time series obtained from an adjusted tGp model.

9. Improved estimation of return levels by incorporating outside knowledge

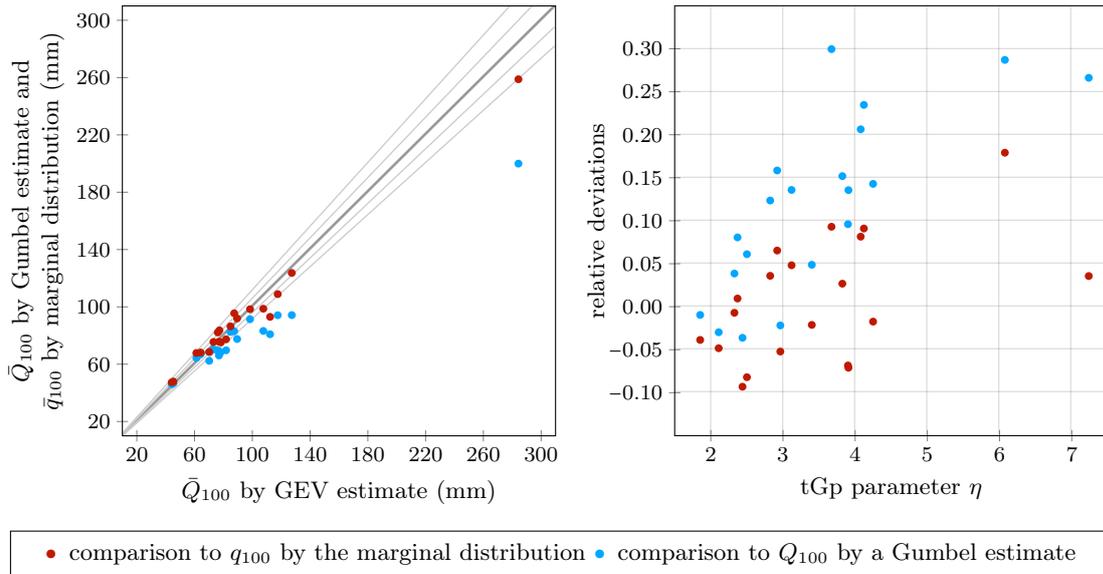


Fig. 9.4 Comparison of the Gumbel and GEV estimates \bar{Q}_{100} and the estimate \bar{q}_{100} by the marginal distribution (see list (9.4)) of the 100-year return level for the 20 data sets of daily precipitation records in Table E.2. The inner and outer straight lines next to the equality line mark 5% and 10% relative deviation (left panel). Table E.1 collects the shape parameters η of the tGp model adjusted to the empirical time series each (used in right panel).

The three depicted examples in Figure 9.5 show decreased statistical uncertainty for the 100-year return level each by smaller confidence intervals, or more precisely credible intervals in the Bayesian context, for the estimates \bar{q}_{100} compared to GEV estimates \bar{Q}_{100} . For the “Fichtelberg” data set (a) the mean GEV estimate for 100 shuffled versions of the original time series clearly falls below the estimate for the original time series. This observation is possibly due to prominent SRC. Section 9.4 gives an outlook on such effects. Nonetheless, either estimate for the stations (a) to (c) lies within one standard deviation of the mean estimate \bar{q}_{100} . Note that the time series (c) does not contain measurements taken at a single station but daily records accumulated over several locations in the region of central England. The thus lower variance among the daily measurements results in smaller variation of return level estimates accordingly.

As a result, knowledge about the marginal distribution of the predominantly stationary daily measurements well described by a tGp distribution yields higher reliability of GEV than of Gumbel estimates for return levels of mid-latitude precipitation extremes. Given weak SRC and LRC additionally, estimating return levels by the marginal distribution allows for decreased statistical uncertainty. Note that these estimates depend on the method used for adjusting a model distribution to the daily data (cp. Rem. 9.2). The confidence intervals in Figure 9.5 base on the tail oriented parameter estimation introduced in Section 6.2.1 but other techniques such as entropy based parameter estimations [140] are possible. Section 9.3 aims at further improvements by making use of knowledge about the marginal distribution and the corresponding Gumbel limit at once.

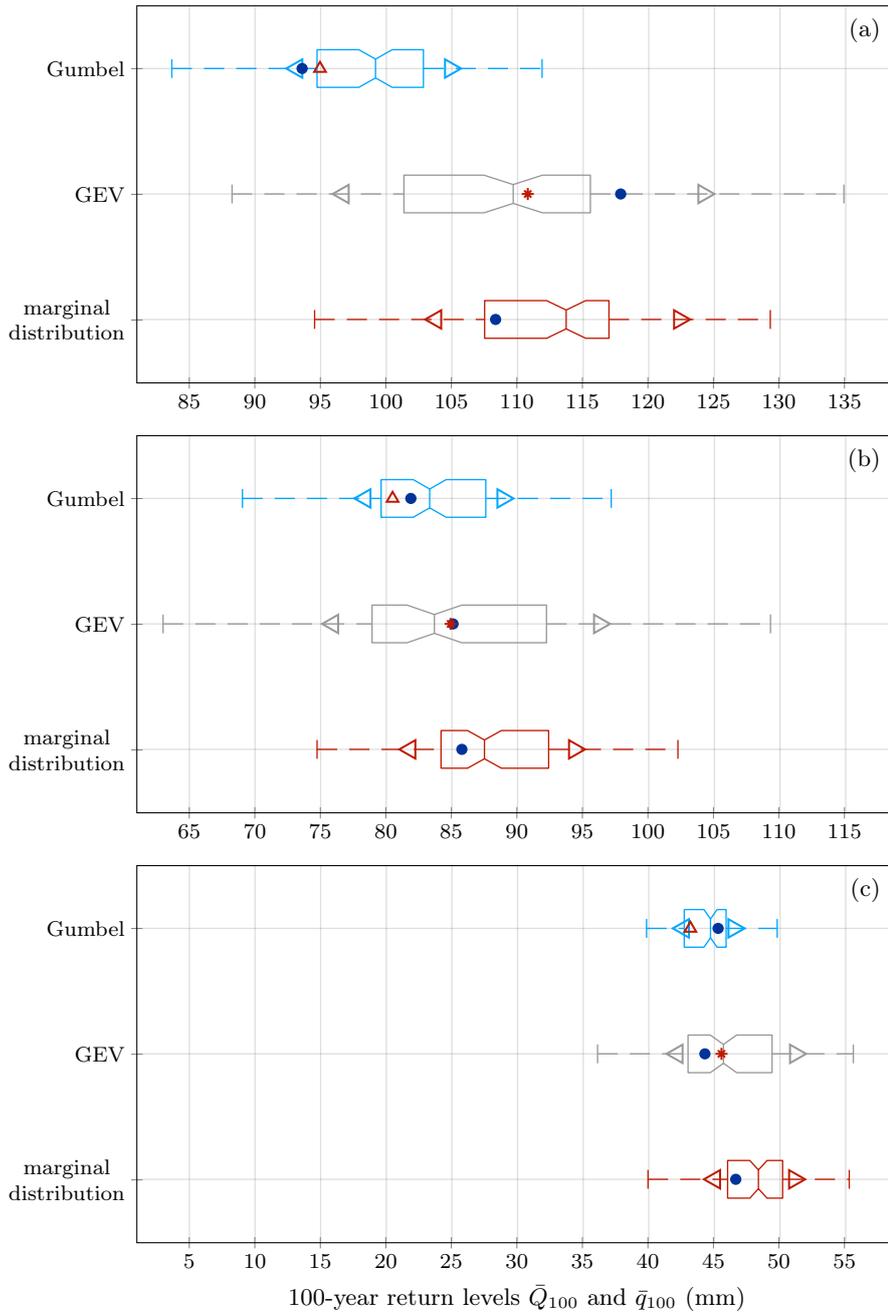


Fig. 9.5 Comparison of the Gumbel and GEV estimates \bar{Q}_{100} and the estimate \bar{q}_{100} by the marginal distribution (see list (9.4)) of the 100-year return level for the three empirical precipitation data sets (a) to (c) in Table 6.1. The estimates for the shuffled time series are the mean of the same estimates for 100 shuffled versions of the original empirical time series. The box plots comprise the estimates \bar{Q}_{100} and \bar{q}_{100} for 100 model time series for each data set according to Section 6.2. The boxes show the median and the 25% and 75% quantiles each, the whiskers extend to the adjacent value, and the arrowheads include one standard deviation from the box mean.

9.3. Estimation of return levels by norming constants in the Gumbel domain

It is well known and visualized in Sections 7.3 and 8.2 that the convergence of the distribution of properly normed extremes to its limit GEV shape in Theorem 7.9 is particularly slow for some distributions in the Gumbel domain of attraction. For the Gaussian and tGp distribution this rate of convergence as the block size $n \in \mathbb{N}$ increases asymptotically reads $\frac{1}{\ln n}$ and even $\frac{1}{\sqrt{\ln n}}$ for the lognormal distribution (cp. Ex. 8.5). Due to this slow convergence, adjusting a Gumbel distribution to empirical maxima of blocks of finite length possibly yields inappropriate estimates of return levels (cp. Sect. 7.4).

Using a penultimate approximation by adjusting a GEV distribution instead of the ultimate approximation by a Gumbel distribution (cp. Sect. 7.3.3) provides one possibility of approaching the above effect. As an example, the rates of the convergence for maxima of Gaussian or tGp samples then accelerates to $\frac{1}{(\ln n)^2}$ (cp. Ex. 8.7). Section 9.2 visualizes that the mean estimation of return levels indeed improves but with increased statistical uncertainty due to an additional parameter needed for a GEV distribution and in case of low dependencies among the individual samples, estimating return levels by the marginal distribution instead of EVT yields decreased reliability.

For some distributions, however, we observe reliable estimates of return levels by the Gumbel distribution despite a known slow convergence. In Figure 9.4 we find highest coincidence between the Gumbel and GEV estimates of the 100-year return level of extreme precipitation events for those empirical data sets for which the estimated shape parameter of the marginal tGp distribution is closest to $\eta \approx 2$. Hall [68] derived a faster optimal rate of $\frac{1}{(\ln n)^2}$ for the ultimate approximation of the maxima of the squared absolute values of Gaussian samples (see Ex. 8.3) and Castro proved $\frac{1}{(\ln n)^2}$ for the Gamma distribution [31]. Figure 9.6 shows estimates of the 100-year return level of synthetic iid tGp samples with shape parameter $\eta = 2.1$ (left panel) and Gamma (cp. Tab. B.1) distributed samples (right panel). Contrary to the still slow convergence of the corresponding maxima distributions we find high coincidence between the ultimate and penultimate estimates for 100 sample time series in either case and thus increased reliability by Section 9.2. The shape of the marginal distribution each provides an explanation of this effect. The two distributions considered in Figure 9.6 have in common a nearly exponentially decreasing tail of their pdfs. The tGp pdf decays at rate $y^{\frac{1-\eta}{\eta}} e^{-\frac{y^{2/\eta}}{2\sigma^2}}$ by its asymptotics (6.5) and the pdf of the Gamma distribution at rate $y^{\alpha-1} e^{-\beta y}$ by Table B.1 as $y \rightarrow \infty$.

9.3. Estimation of return levels by norming constants in the Gumbel domain

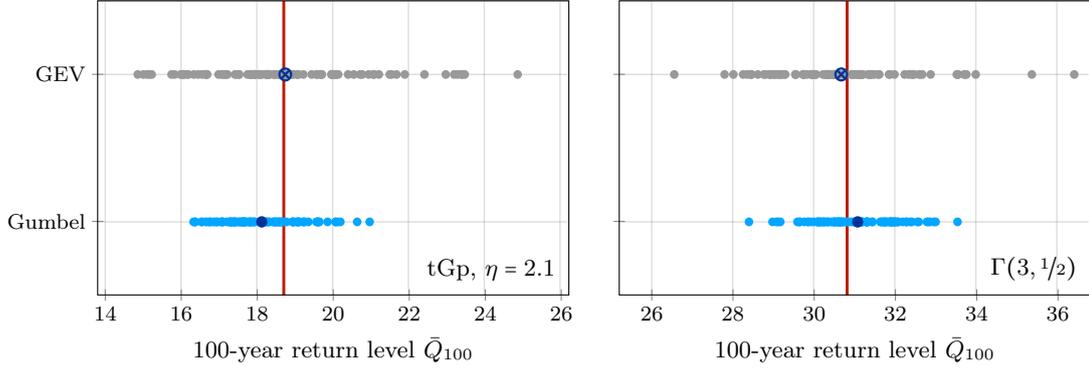


Fig. 9.6 Mean (dark) and individual (light) GEV and Gumbel estimates of the 100-year return level for 100 synthetic sample time series of length $N = 36,500$ of an iid process with tGp marginal distribution with location $\nu = 0$, scale $\sigma = 1$ and shape $\eta = 2.1$ (left) and Gamma marginal distribution with shape $\alpha = 3$ and scale $\beta = \frac{1}{2}$. The straight lines marks the return level $Q_{100} \approx 18.7$ and $Q_{100} \approx 30.8$, respectively, analytically known by equation (9.1).

This observation of improved Gumbel estimates in the presence of a tail of the marginal pdf close to the exponential tail of the Gumbel pdf rises the idea of further decreasing statistical uncertainty of return level estimates by incorporating both the marginal distribution and the knowledge about an asymptotic Gumbel shape of the maxima distribution in Section 9.3.1. Section 9.3.2 provides corresponding results for synthetic time series and Section 9.3.4 applies the method to empirical precipitation measurements. An outlook in Section 9.4 is dedicated to generalizations of the idea to processes with strong SRC.

Remark 9.3: From a theoretical perspective, inverse sampling is capable of improving return level estimates for extremes of samples with a slow convergence of the corresponding maxima distribution to the limit Gumbel shape. The return level Q_m^X with return period $\frac{1}{m}$ of maxima of blocks of size $n \in \mathbb{N}$ of samples with cdf F_X are given by

$$Q_m^X = F_X^{-1} (F_Y(Q_m^Y)) \quad (9.5)$$

for a distribution with cdf F_Y with return level Q_m^Y with the same return period for the corresponding maxima distribution due to the monotonicity of the quantile function F_Y^{-1} . Using a distribution with a faster convergence in the extremal types theorem Theorem 7.9 then yields both return levels Q_m^X and Q_m^Y with improved reliability.

For practical purposes, however, the cdf F_X and its quantile function are only given empirically. The calculation of the return level Q_m^X by equation (9.5) requires an estimation of the empirical cdf up to the probability $1 - \frac{1}{nm}$. By equation (9.3) this approach resembles the estimation of return levels by the marginal distribution. A transformation of empirical data to a distribution with improved convergence properties in EVT is thus only a loop way to the direct estimation by the marginal distribution.

9.3.1. Error reduction for return levels by norming constants

The norming constants describe the relation between the marginal distribution and the limit GEV distribution in Theorem 7.3. Fisher and Tippett [52] derived the universal limit shape along with the three types of EVDs by calculating the mode and scale of the distribution of the maximum M_n (7.1) of $n \in \mathbb{N}$ samples with marginal cdf F given by $P(M_n \leq x) = F(x)^n$ for values $x \in \mathbb{R}$ in the domain of the cdf F . After that norming constants have typically based on the idea (7.20) and as described in Table 7.3. Norming constants are thus given by choosing thresholds $u_n = \frac{x}{a_n} + b_n$ such that

$$F(u_n) \approx 1 - \frac{1}{n} \quad (9.6)$$

for large numbers n of samples representing the empirical maximum of n samples. In particular, for $x = 0$ the centering constant satisfies $F(b_n) = 1 - \frac{1}{n}$ and approximates the position of the mode of the maximum distribution in case of a Gumbel limit distribution. For the Gaussian cdf Φ , the approximation (7.53) with $1 - \Phi(x) = \frac{\phi(x)}{x} + \mathcal{O}\left(\frac{1}{x^3}\right)$ as $x \rightarrow \infty$ yields the centering constant β_n implicitly by

$$\frac{\phi(\alpha_n)}{\alpha_n} = \frac{1}{n} \quad (9.7)$$

together with the scaling constant $\beta_n = \alpha_n$ (cp. Ex. 7.7). The maximum distribution normed by constants chosen by equation (9.6) approximates the limit Gumbel distribution best in its mode located at $x = 0$ and with probability $G(0) = \frac{1}{e}$ for cdfs G of all three types of EVDs. Note that $F(b_n)^n = \left(1 - \frac{1}{n}\right)^n \rightarrow \frac{1}{e}$ as the sample size $n \rightarrow \infty$. Figure 7.4 shows this convergence of the pdf of the normed maximum $\alpha_n(M_n - \beta_n)$ to the Gumbel pdf for the maximum of Gaussian variates.

Most probability mass is located in the neighborhood of the mode of a distribution by definition. A ML adjustment of a Gumbel distribution to empirical maxima thus gives parameters close to the norming constants obtained by assumption (9.6) for distributions in the Gumbel domain of attraction disregarding either tails. The mode of Type-II and Type-III GEV distributions is not located precisely at $x = 0$, so that GEV parameters estimated by ML do not reflect the assumption (9.6) for distributions in their domain of attraction. The above observation of improved return level estimates due to similarity between the tail of the pdf of the normed maximum distribution and the limit exponential tail of the Gumbel pdf indicates decreased estimation errors due to slow convergence when choosing alternative norming constants with decreased deviation of the tails rather than the bulks of the two pdfs. The knowledge about the limit GEV distribution allows for defining norming constants by assuming

$$F\left(\frac{Q}{a_n} + b_n\right)^n \stackrel{!}{=} G(Q) \quad \text{and} \quad (9.8)$$

$$\frac{d}{dQ} F\left(\frac{Q}{a_n} + b_n\right)^n \stackrel{!}{=} G'(Q) \quad (9.9)$$

9.3. Estimation of return levels by norming constants in the Gumbel domain

and thus equating both the cdf and pdf of the maximum distribution for finite block size n to those of the GEV limit in a certain quantile $Q \in \mathbb{R}$ in the domain of the cdf F . Setting $Q = 0$ reflects the approach (9.6) cdfs in the Gumbel domain of attraction. We have $G(0) = G'(0) = \frac{1}{e}$ for all three types of EVDs (cp. Fig. 7.3).

Aiming at agreement of the two distributions in the neighborhood of a return level Q_m of annual extremes with return period $m \in \mathbb{N}$ other than $m = \frac{e}{e-1}$ like for the approach (9.6), we consider the return level $Q_m^G := G^{-1}\left(1 - \frac{1}{m}\right)$. Solving equations (9.8) and (9.9) then yields norming constants such that the deviation between the normed maximum distribution and the limit GEV distribution vanishes at the return level Q_m .

Let $Q \in \mathbb{R}$ be a return level of the Gumbel distribution with return period $m \in \mathbb{N}$ such that with $p = 1 - \frac{1}{m}$ we have $Q = \Lambda^{-1}(p) = -\ln(-\ln p)$ and denote by $u_n = \frac{Q}{a_n} + b_n$ corresponding thresholds. Then for the tGp with shape parameter η using equations (9.8) and (9.9), we obtain

$$p^{\frac{1}{n}} = \Lambda(Q) \stackrel{!}{=} \Phi(u_n^{1/\eta}) \quad \text{and thus} \quad \frac{\phi(u_n^{1/\eta})}{u_n^{1/\eta}} \stackrel{!}{=} -\frac{\ln p}{n} \quad \text{by (7.53)} \quad (9.10)$$

$$-p \ln p = \Lambda'(Q) \stackrel{!}{=} \frac{n u_n^{1/\eta-1}}{\eta a_n} \Phi(u_n^{1/\eta})^{n-1} \phi(u_n^{1/\eta}). \quad (9.11)$$

The assumptions (9.10) together with $\Phi(x) \rightarrow 1$ as $x \rightarrow \infty$ transform the assumptions (9.11) to $u_n = (\eta a_n)^{\frac{\eta}{2-\eta}}$ so that the definition of the thresholds u_n implies

$$\frac{\phi\left((\eta a_n)^{\frac{1}{2-\eta}}\right)}{(\eta a_n)^{\frac{1}{2-\eta}}} \stackrel{!}{=} -\frac{\ln p}{n} \quad \text{and} \quad b_n = (\eta a_n)^{\frac{\eta}{2-\eta}} - \frac{Q}{a_n} \quad (9.12)$$

as an implicit approach to norming constants with the desired properties (9.8) and (9.9). The constants (9.12) with shape $\eta = 1$ provide norming constants for maxima of Gaussian samples in Lemma 9.4 as a special case. Choosing additionally $p = \frac{1}{e}$ and $Q = 0$ reproduces the constants (9.7). Note the similarity between the scaling constants α_n and a_n given for the Gaussian distribution by assumption (9.7) and (9.12), respectively. The centering constants β_n and b_n differ by the shift $\frac{Q}{a_n}$ of the return level Q scaled accordingly. As a remark, Zarfaty et al. [188] mention replacing the right hand side of the condition (9.7) by $\frac{p}{n}$ as a promising approach to improving estimates of the uniform error of the $\frac{1}{\ln n}$ -convergence in the extremal types theorem Theorem 7.3 for maxima of Gaussians. Lemma 9.4 and Corollary 9.5 prove the validity of the norming constants (9.12) for properly norming the maximum of Gaussian and tGp samples, respectively, by applying Khintchine's theorem Theorem 7.4.

9. Improved estimation of return levels by incorporating outside knowledge

Lemma 9.4: *Using the shape $\eta = 1$, the norming constants obtained from equations (9.8) and (9.9) for quantiles $Q = -\ln(-\ln p)$ of probability $p \in (0, 1)$ are valid norming constants for the maximum distribution of standard Gaussian iid variates to retain convergence in the extremal types theorem Theorem 7.3.*

Proof. Let α_n, β_n be the norming constants (9.7) for the maximum distribution of standard Gaussian iid variates and denote by $\tilde{\alpha}_n, \tilde{\beta}_n$ the norming constants obtained from equations (9.8) and (9.9) and satisfying $\frac{\phi(\tilde{\alpha}_n)}{\tilde{\alpha}_n} = \frac{c}{n}$ and $\tilde{\beta}_n = \tilde{\alpha}_n - \frac{Q}{\tilde{\alpha}_n}$ with the numerator $c := -\ln p$ and the quantile $Q = -\ln c$ by approximation (9.12). We derive the asymptotic behavior of the scaling constant $\tilde{\alpha}_n$ analogously to Hall's method [67] for the asymptotics (7.58) of the scaling constant α_n (9.7) in the case of $c = 1$.

The assumption $n^2 = 2\pi c^2 \tilde{\alpha}_n^2 e^{\tilde{\alpha}_n^2} \sim e^{\tilde{\alpha}_n^2}$ for large block sizes $n \in \mathbb{N}$ implies $\tilde{\alpha}_n^2 \sim 2 \ln n + u$ with $u \ll \ln n$. Then analogously to the derivation of the asymptotis (7.58) we obtain

$$\begin{aligned} \tilde{\alpha}_n^2 &\sim \alpha_n^2 - 2 \ln c = \alpha_n^2 + 2Q \\ \tilde{\alpha}_n &\sim \alpha_n + Q/\alpha_n + \mathcal{O}(1/\alpha_n^3) \end{aligned} \quad (n \rightarrow \infty). \quad (9.13)$$

Khintchine's theorem Theorem 7.4 states that any pair of norming constants satisfying the asymptotic relations (7.13) are valid. For the norming constants $\tilde{\alpha}_n, \tilde{\beta}_n$ using $\alpha_n = \beta_n$ the approximations (9.13) imply

$$\begin{aligned} \frac{\alpha_n}{\tilde{\alpha}_n} &= \frac{\alpha_n}{\alpha_n(1 + \mathcal{O}(1/\alpha_n^2))} \rightarrow 1 \\ (\tilde{\beta}_n - \beta_n)\alpha_n &= (\tilde{\alpha}_n - Q/\tilde{\alpha}_n - \alpha_n)\alpha_n = Q \left(1 - \frac{1}{1 + \mathcal{O}(1/\alpha_n^2)} \right) \rightarrow 0 \end{aligned} \quad (n \rightarrow \infty), \quad (9.14)$$

so that by Khintchine's theorem the norming constants $\tilde{\alpha}_n, \tilde{\beta}_n$ are valid. \square

Corollary 9.5: *The norming constants obtained from equations (9.8) and (9.9) for quantiles $Q = -\ln(-\ln p)$ of probability $p \in (0, 1)$ are valid norming constants for the maximum distribution of iid tGp variates with shape $\eta \in \mathbb{R}_{>0}$, location $\nu = 0$ and scale $\sigma = 1$ to retain convergence in the extremal types theorem Theorem 7.3.*

Proof. Denote by \tilde{a}_n, \tilde{b}_n the norming constants obtained from equations (9.8) and (9.9). Let $\alpha_n = \beta_n$ be the norming constants (9.7) for the standard Gaussian distribution and let a_n, b_n be the norming constants for the tGp distribution with shape parameter $\eta \in \mathbb{R}_{>0}$, location $\nu = 0$ and scale $\sigma = 1$ deduced as $a_n = \frac{\alpha_n^{2-\eta}}{\eta}$ and $b_n = \beta_n^\eta$ in equations (8.12) of Example 8.3. Due to equations (9.12) and definition (9.7), using the notion of Lemma 9.4, the norming constants satisfy

$$\tilde{a}_n = \frac{1}{\eta} \tilde{\alpha}_n^{2-\eta} \quad \text{and} \quad \tilde{b}_n = a_n^\eta - \frac{Q}{\tilde{a}_n}. \quad (9.15)$$

Then arguments analogous to Lemma 9.4 yield $\frac{a_n}{\tilde{a}_n} \rightarrow 1$ and $(\tilde{b}_n - b_n)a_n \rightarrow 0$ and thus the validity of the norming constants \tilde{a}_n, \tilde{b}_n for the convergence in Theorem 7.3. \square

9.3. Estimation of return levels by norming constants in the Gumbel domain

Remark 9.6: Given valid norming constants α_n, β_n for the maximum distribution of random variates X , the norming constants $a_n = \frac{\alpha_n}{\sigma}$ and $b_n = \sigma\beta_n + \nu$ are valid for properly norming the maximum of the affine linear transformation $\sigma X + \nu$ to obtain convergence in the extremal types theorem Theorem 7.3 (cp. Sect. 8.1). Applying this relation to a tGp distribution with location and scale parameters other than zero and unity, respectively, yields corresponding norming constants by the reasoning of Corollary 9.5.

Aiming at a best possible visualization of the approximation in a prescribed quantile in Figure 9.7, we abstain from considering the asymptotic solutions (9.12) of equations (9.8) and (9.9) for large block sizes n but determine norming constants for the maxima of tGp samples directly by denoting $q_n := \Phi^{-1}(p^{1/n})$ as

$$\begin{aligned}\tilde{a}_n &= -\frac{n}{\eta\sigma p^{1/n} \ln p} \phi(q_n) (\sigma q_n + \nu)^{1-\eta} \quad \text{and} \\ \tilde{b}_n &= (\sigma q_n + \nu)^\eta - \frac{Q}{a_n}.\end{aligned}\tag{9.16}$$

Figure 9.7 shows the normed pdfs of the maximum distribution for iid standard Gaussian and tGp variates using the norming constants (9.16) with coincidence with the limit Gumbel pdf and cdf in the return levels with probability $p = \frac{1}{e}$ and $p = 1 - \frac{1}{100}$ (left panel) and $p = 0.6$ (right panel). Such a choice possibly produces huge deviations in return levels other than the one equalized, however, these errors do not affect the estimation of return levels close to this return level in Section 9.3.2.

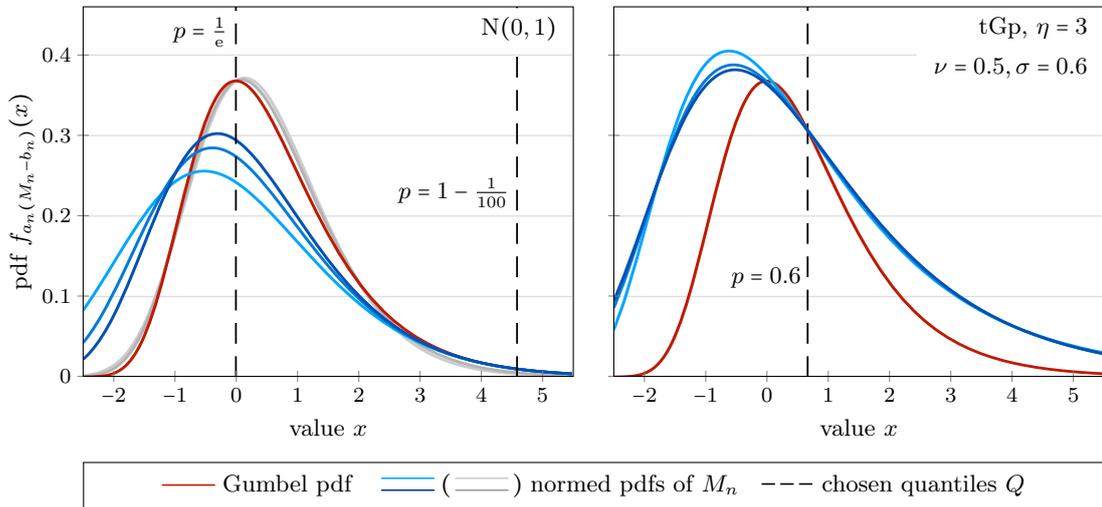


Fig. 9.7 Evolution of the pdfs $f_{a_n(M_n - b_n)}$ (7.6) of the normed maximum $a_n(M_n - b_n)$ of iid standard Gaussian (left) and tGp variates with shape $\eta = 3$, location $\nu = 0.5$ and scale $\sigma = 0.6$ (right), respectively, using the norming constants (9.16) for block sizes $n \in \{50, 500, 5000\}$ with darkening lines as n increases. The left panel depicts the normed pdfs with coincidence in the return level $Q = 0$ with $p = \Lambda(0) = \frac{1}{e}$ and $Q \approx 4.6$ with $p = \Lambda(Q) = 1 - \frac{1}{100}$, the right panel uses the return level $Q \approx 0.7$ with $p = \Lambda(Q) = 0.6$.

9.3.2. Statistical inference by norming constants

Let F be a cdf in the domain of attraction of the Gumbel distribution. Then given valid norming constants a_n, b_n , the Gumbel distribution asymptotically well describes the distribution of the normed maximum $a_n(M_n - b_n)$ by the extremal types theorem Theorem 7.3 as $P(a_n(M_n - b_n) \leq x) \sim \Lambda(x)$ as the block size $n \rightarrow \infty$. Using the norming constants established in Section 9.3.1 for a return level Q of the Gumbel distribution with probability $p = \Lambda(Q)$ yields the coincidence $F\left(\frac{Q}{a_n} + b_n\right)^n = \Lambda(Q)$. By the continuity of the Gumbel cdf we find the similarity

$$F(q)^n \approx \Lambda(a_n(q - b_n)) = p \quad (9.17)$$

for return levels q in the neighborhood of the return level Q . Choosing the probability $p = 1 - \frac{1}{m}$ of a return level Q_m of the Gumbel distribution with return period $m \in \mathbb{N}$ therefore yields the return level \tilde{Q}_m of the maximum M_n with the same return period m as

$$\tilde{Q}_m = \frac{Q_m}{a_n} + b_n. \quad (9.18)$$

The return level $Q_m = \Lambda^{-1}(p)$ for a desired return period m is analytically known by the Gumbel cdf Λ . The estimation of return levels \tilde{Q}_m of a stationary process with marginal cdf F by equation (9.18) requires the estimation of the norming constants a_n, b_n such that deviations are low between the Gumbel pdf and the pdf of the normed maximum. Note that in typical applications of EVT the norming constants are captured by ML estimates of GEV parameters. The *modified Gumbel estimate* above directly uses the norming constants defined by equations (9.8) and (9.9) instead.

9.3.3. Application to synthetic data

Figure 9.8 shows the 100-year return levels estimated by the modified Gumbel approach for extremes of Gaussian and of tGp iid time series. Aiming at an application to precipitation data in Section 9.3.4, we consider both time series with continuously generated samples (top panels) and with samples Y_t discretized by $\frac{\lfloor 10Y_t \rfloor}{10}$ with a precision of 0.1 (bottom panels) (cp. Sect. 1.4). This discretization is in accordance with mapping all model values less than 0.1 mm to zero precipitation in Section 6.3.1.

The Gumbel and GEV estimates and their statistical uncertainty do not differ for continuous and discretized model data by a comparison of the bottom and top panels in Figure 9.8. Adjusting a model distribution to synthetic data obtained from this distribution by ML yields clearly smaller confidence intervals for both estimates by the marginal distribution and the modified Gumbel approach (top panels). Section 6.3.1 and Section 6.3.7, however, elaborate the inappropriateness of a ML adjustment of the tGp model to daily precipitation amounts when focussing upon proper representation of extreme events.

9.3. Estimation of return levels by norming constants in the Gumbel domain

The bottom panels show application-oriented confidence intervals obtained from the tail-oriented estimation of the marginal distribution by synthetic time series with precision 0.1 for the estimation of the 100-year return level by both this marginal distribution directly and the modified Gumbel estimation based on estimating the norming constants (9.18). Note that the choice of the probability $p = 1 - \frac{1}{m}$ for the modified Gumbel approach reproduces the mean result and uncertainty of the estimate by the marginal distribution. This observation is due to the definition (9.10) for the norming constants used for the former method and the coincidence of the estimates \bar{Q}_m and \bar{q}_m by $(1 - \frac{1}{m})^{1/n} \approx 1 - \frac{1}{nm}$ as discussed in Section 9.2. If the modified Gumbel estimate for the probability $p = \frac{1}{e}$ reproduces the original Gumbel estimate depends on the specific choice of the approximation of the norming constants (9.16). The deviation between the normed pdf of the maximum distribution and the limit Gumbel pdf vanishes in the quantile $Q = 0$ in this case but the error in the return level $Q = \Lambda^{-1}(p)$ depends on the uniform error between the two distributions (see Fig. 9.7 and cp. Sect. 7.3).

We obtain decreased uncertainty of the estimate of the return level Q_m by choosing a probability $\frac{1}{e} < p < 1 - \frac{1}{m} =: p_m$ close to but not equal to the upper bound p_m , which represents the estimation by the marginal distribution directly. For lower probabilities the estimation of the marginal distribution is required only up to return levels with probability $p < p_m$. The tail-oriented estimation then generates the smaller confidence intervals the smaller the chosen probability p (Fig. 9.8, bottom panels).

The choice of the tail-oriented parameter estimation for the marginal distribution and confidence intervals obtained by Bayesian approach and 100 sample time series provide a proof of concept for reducing statistical uncertainty of return level estimates for extreme events by combining outside knowledge about the marginal distribution and the known Gumbel limit. The slight underestimation of the analytically known return level Q_{100} for the time series in Figure 9.8 is due to normalization of the empirical cdf by $N + 1$, which yields the model cdf asymptotically only (cp. Rem. 6.1). Section 9.4 provides an outlook on more sophisticated and general application of the modified Gumbel approach.

9.3.4. Application to daily precipitation amounts

Figure 9.9 extends Figure 9.5 by the modified Gumbel estimate of the 100-year return level of annual maximal daily precipitation amounts for the three empirical data sets listed in Table 6.1. The panels add on the results of the Gumbel and GEV estimates \bar{Q}_{100} , the estimates $q_{\bar{100}}$ by the marginal distribution. For all stations both estimates \bar{Q}_{100} and $q_{\bar{100}}$ are within one standard deviation of the mean modified Gumbel estimate with coincidence (9.8) and (9.9) assumed for the return level Q with $p = \Lambda(Q) = 0.95$. We thus find decreased statistical uncertainty of the modified Gumbel estimate for the 100-year return level compared to estimates by a GEV or the marginal distribution measured by the standard deviation obtained from modified Gumbel estimates for 100 model time series each as established in Section 9.3.1. Mind that the data set (a) exhibits strong SRC (cp. Sect. 9.2.2), which explains the visible deviations between the estimates.

9. Improved estimation of return levels by incorporating outside knowledge

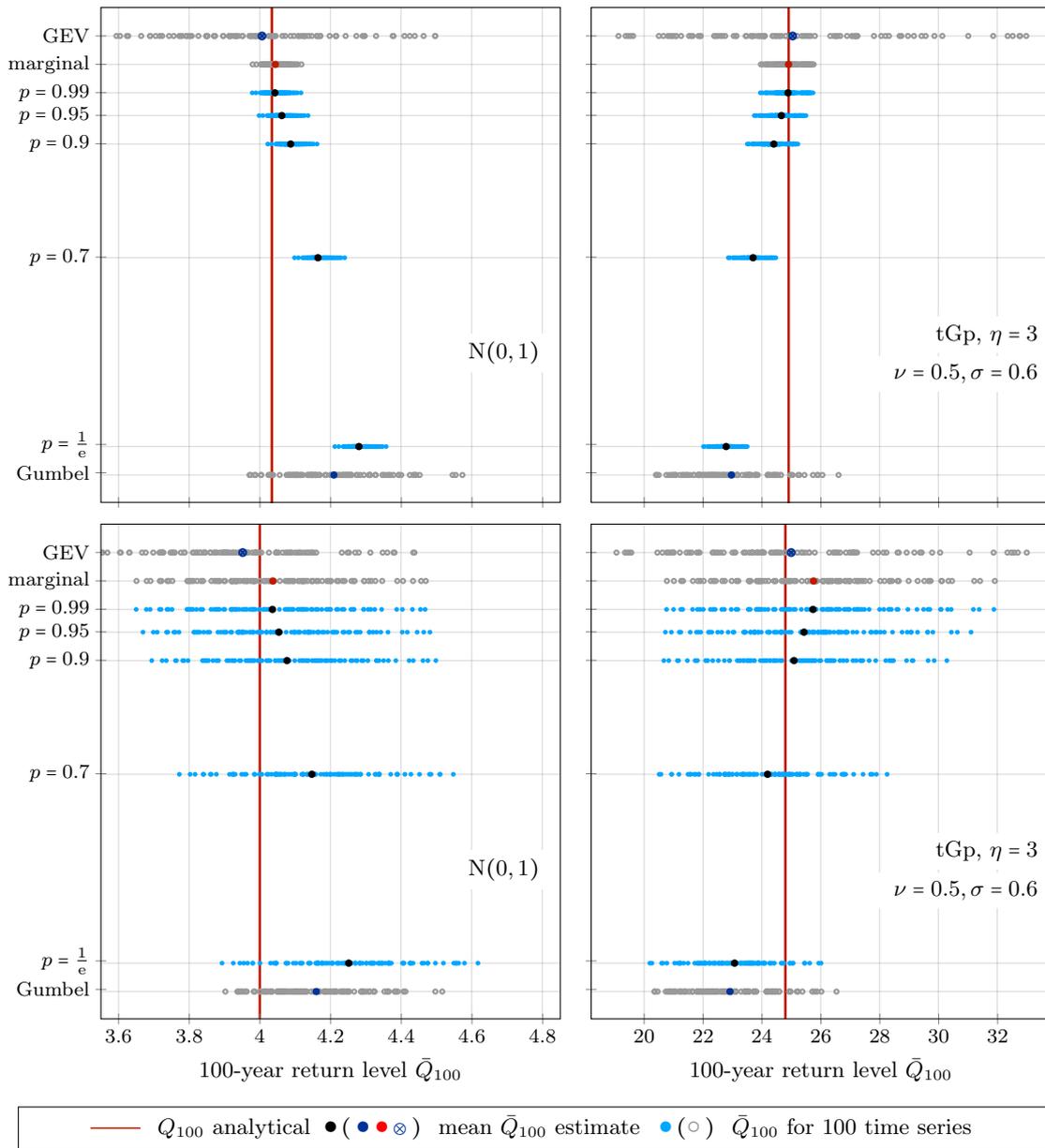


Fig. 9.8 Estimates of the 100-year return level Q_{100} of maxima of blocks of size $n = 365$ of standard Gaussian (left) and tGp samples with shape $\eta = 3$, location $\nu = 0.5$ and scale $\sigma = 0.6$ (right) by the four methods: Gumbel and GEV estimate, estimate by the marginal distribution (see list (9.4)) and modified Gumbel estimate. The modified Gumbel estimate is depicted for different choices of the probability p , so that the Gumbel return level $Q = \Lambda^{-1}(p)$ coincides with the maxima distribution for the finite block size n in that return level. The upper panels show the results for 100 synthetic sample time series with the marginal distribution estimated by ML, for the lower panels the sample time series are discretized and the marginal distribution is estimated by the tail-oriented approach of Sect. 6.2.1. In Section 9.3.4 we apply the lower method for the generation of confidence intervals for return levels of precipitation amounts, which obey the precision 0.1 mm.

9.3. Estimation of return levels by norming constants in the Gumbel domain

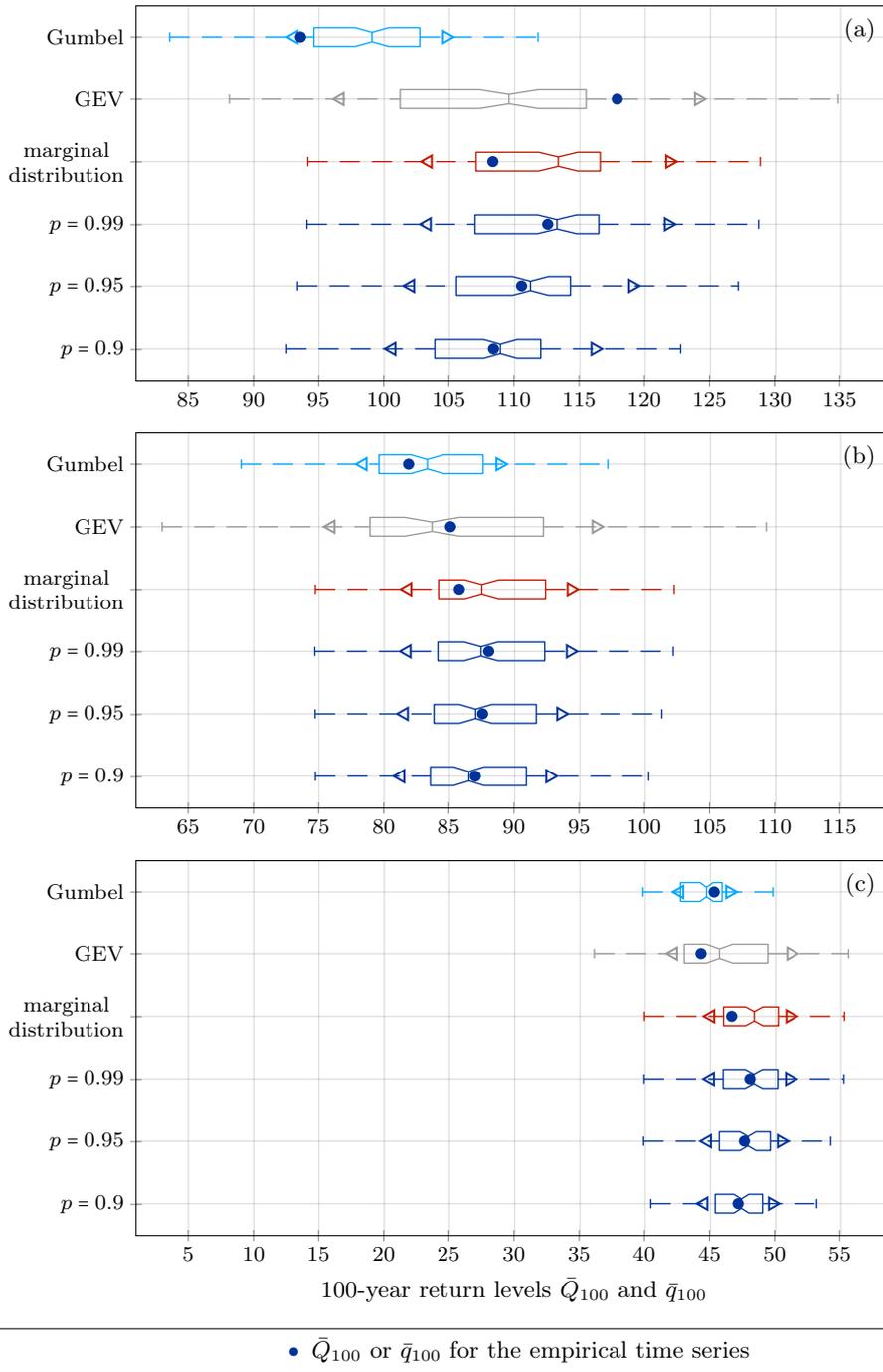


Fig. 9.9 Estimates of the 100-year return level Q_{100} of annual maxima for the three empirical precipitation data sets (a) to (c) in Table 6.1 by the four methods: Gumbel and GEV estimates, the estimate by the marginal distribution (see list (9.4)) and the modified Gumbel estimate. The modified Gumbel estimate is depicted for the choices $p \in \{0.9, 0.95, 0.99\}$ of the probability p , so that the Gumbel return level $Q = \Lambda^{-1}(p)$ coincides with the maxima distribution for the finite block size n in that return level. The box plots comprise the estimates for 100 model time series each for each data set according to Section 6.2. The boxes show the median and the 25% and 75% quantiles each, the whiskers extend to the adjacent value, and the arrowheads include one standard deviation from the box mean.

9.4. Closing remarks and outlook

The return levels of stationary processes are subject to the marginal distribution and the dependence structure of the process. Outside knowledge about both properties allows for assessing the validity of return levels Q_m obtained from EVT by a comparison to the corresponding return level q_m of the marginal distribution (cp. Sect. 9.2).

In Figure 9.4 we find the largest deviation between the GEV estimate \bar{Q}_{100} and the estimate \bar{q}_{100} for the data set “Valencia”, which obeys both the largest marginal variance and strongest SRC among the 20 considered data sets in Table E.1 and Table E.2. Section 9.2 elaborates the impact of strong SRC on the resulting maxima distribution even for processes with asymptotic unit extremal index. A detailed investigation of the finite-block-size effects on the mean cluster size and the relation to the heaviness of the marginal distribution along with a systematic analysis of the influence and distinction of strong long- and strong short-term dependence remains an open question. The Figures 7.16, 9.1 and 9.2 apply the straightforward approach to this question by considering the return level estimates for shuffled versions of the original time series with destroyed dependencies.

Along with such a study above a comparison of the estimates by the marginal distribution in Section 9.2 and the modified Gumbel estimates established in Section 9.3 to the POT approach provides questions for future research. In case of availability of data beyond annual maxima only, the “point-process” or POT approach is the method of choice for discussing extremes of iid processes [48, p. 340]. Considering not only block maxima but all threshold excesses is capable of improving both the sampling error and the error of the approximation (7.10). By Theorem 7.28 the parameters of the GPD adjusted to threshold exceedances, however, are directly related to the parameters of the GEV distribution adjusted to corresponding block maxima, so that statistical uncertainty obeys the same influencing factors as described in Section 7.4. In particular, for processes with a slow convergence to a limit GEV or Gumbel shape like Gaussian or tGp processes, the choice of a threshold is potentially delicate and introduces further uncertainty (cp. Ex. 7.29), which is even more relevant in the presence of strong SRC.

The modified Gumbel estimate assumes weak dependence among the individual samples. Note that weak dependence refers to both regimes short- and long-term dependence since strong LRC crucially confound statistical estimates anyways and strong SRC like in Example 8.11 possibly change the distribution of finite-block-size maxima significantly by an effective extremal index other than unity. A potential generalization of the modified Gumbel approach to stationary processes with strong short-range dependence might include an estimation of an effective finite-block-size extremal index by the mean cluster size [131] and applying an affine linear transformation to the norming constants defined by equations (9.8) and (9.9). Such a method accounted for the possible shift of the maximum distribution by an extremal index other than unity (cp. Fig. 9.3). The same opportunity of a generalization applies to the estimation of return levels by the marginal distribution and by the POT method.

Our estimation of the norming constants for the modified Gumbel approach applies the tail-oriented parameter estimation described in Section 6.2.1, which highly depends on the precision of the considered data as visualized in Figure 9.8. An advantage of the tail-oriented estimation is the direct access to a comparison of statistical uncertainty by the required accuracy of marginal distribution for large events as described in Section 9.3.2. Such a comparison is not as straightforward for example when using norming constants defined by the Lamber W function (cp. Tab. 7.4) and satisfying the conditions (9.8) and (9.9) even though such constants are expected to provide a lower uniform error for large events [188]. Nonetheless, alternative methods for the estimation of the norming constants might further reduce the size of confidence intervals for return level estimates and potentially allow for more detailed insights into appropriate choices of return levels in the neighborhood of the desired return level.

The modified Gumbel approach is further not limited to Gaussian or tGp processes. The conditions (9.8) and (9.9) readily transfer to other distribution with a differentiable cdf such as the lognormal, exponential or Pareto distribution. The approach thus also generalizes to distributions in the domain of attraction of Type-II or -III EVDs. A future project could study whether also in these situations the modified Gumbel approach significantly improves statistical uncertainty since the convergence to the limit GEV distribution is typically not as slow as for the prominent Gaussian and meta-Gaussian processes (cp. Sects. 7.3 and 8.2). Obtaining a proper shape parameter for the limit GEV distribution by the Hill estimator (cp. Rem. 7.26) might additionally reduce statistical unertainty in the general GEV setting for stationary processes with weak dependence.

10. Conclusion

The present thesis deals with non-Gaussian stationary stochastic processes with special focus on risk assessment and statistical uncertainty due to the presence of LRC. My contribution is the extension and application of existing concepts from pure and applied mathematics to the research field of theoretical physics. I consider the following topics.

- (i) Data models for non-Gaussian long-range correlated processes
- (ii) The theory of effective sample sizes for stationary processes
- (iii) Extreme value theory for stationary and particularly meta-Gaussian processes
- (iv) Application of the concepts (i) to (iii) to empirical precipitation records

My major contribution to the topics above is twofold. First, I show that meta-Gaussian data models obtained from nonlinear and memoryless transformations of Gaussian AR-FIMA processes provide a flexible approach to the modeling of non-Gaussian LRC data. Second, I specify the influence of LRC on statistical inference by a detailed interpretation of existing theory about stationary stochastic processes for non-Gaussian correlated data and provide mathematical methods for assessing and improving the reliability of statistical estimates for general stationary and, in particular, meta-Gaussian processes.

The main outcome of topic (i) is a parametric alternative to the generation of synthetic non-Gaussian LRC time series with prescribed correlations. By establishing the effective sample size as a finite-sample property of the sample mean of stationary processes in topic (ii), I quantify the impact of both SRC and LRC on the statistics of non-Gaussian data. Regarding topic (iii), I obtain theoretical results for the extreme value theory of meta-Gaussian processes in terms of rates of convergence in the generalized extremal types theorem and corresponding norming constants. Using these results, I formulate a procedure for assessing the reliability of return level estimates obtained from EVT by a comparison to estimates obtained from the marginal distribution directly. Beyond that, I establish a modified Gumbel method for the estimation of return levels based on a specific choice of norming constants depending on the marginal distribution.

Applying the concepts of topic (i) in topic (iv), I establish a step-by-step modeling procedure for daily precipitation amounts. I also elaborate that meta-Gaussian AR-FIMA models are capable of reproducing key statistical quantities of empirical non-Gaussian LRC precipitation records by a detailed statistical assessment. In addition, I apply the effective sample size of topic (ii) to obtain distributions and confidence intervals for the daily mean and annual total precipitation amounts. Using both methods introduced in topic (iii), I determine reliable estimates with decreased statistical uncertainty of the 100-year return level of extreme precipitation events.

Section 10.1 lists my major and minor contributions to the for topics (i) to (iv) considered in the present thesis. Section 10.2 mentions open questions and provides impulses for future research.

10.1. List of central contributions

As a basis for my results and methods the present thesis includes required existing knowledge from pure and applied mathematics, theoretical physics and climate science. I draw relations between previous and contemporary results about stochastic processes, extreme value theory and precipitation modeling and extend them. The collections below list my central contributions to the topics (i) to (iv) beyond the main results above.

(i) Data models for non-Gaussian long-range correlated processes

- The Hermite approach to the acf of meta-Gaussian processes is usually reported for transformations of Gaussian zero-mean and unit-variance processes. In general, the acf of the transformed process, however, depends nonlinearly on the original standard deviation. Aiming at a flexible and direct application to empirical data, I reproduce the Hermite approach by applying Mehler's formula to zero-mean Gaussian processes with arbitrary variance (cp. Sect. 3.1).
- I provide a theoretical derivation of properties of the acf of meta-Gaussian processes, which are occasionally mentioned in application-oriented literature but not proven explicitly, such as the convex relation to the original acf (cd. Sect. 3.2).
- I contribute a new theoretical example by the exact calculation of the Hermite expansion and a closed-form representation of the acf of the pointwise absolute value of Gaussian processes (cp. Sect. 3.3).
- As an example of an alternative approach to non-Gaussian LRC models I elaborate summed squares of consecutive members of a Gaussian process. This transformation is not memoryless and yields a non-Gaussian LRC model with approximately exponential marginal distribution (cp. 3.4.3).
- For the numerical detection of LRC by the wavelet transform based on the original Haar wavelet with only one vanishing moment I define generalized Haar wavelets with an increased capability of detrending and prove their vanishing moments of any desired number (cp. 4.2.3).

(ii) The theory of effective sample sizes for stationary processes

- As a generalization of the previously known asymptotic time scale of SRC processes based on the variance of the sample mean, I introduce and determine effective sample sizes as an interpretable, yet asymptotically infinite, finite-sample property of stationary LRC processes (cp. Chap. 5).

- By citing a theorem about the distribution of sums of stationary processes, I contribute to ongoing discussions about the validity of the Gaussian distribution, which is appropriate in case of iid processes by the CLT. If the transformation generating a meta-Gaussian process obeys an Hermite rank larger than unity, then the sum over the members of such a process is not Gaussian but given by the marginal distribution of a corresponding Hermite processes (cp. Sect. 5.1.1).
- Establishing the concept of a finite-size decorrelation time, I determine finite-sample and asymptotic confidence intervals for the sample mean of general stationary and, in particular, ARFIMA processes. These results generalize the previously known asymptotic distribution of the sample mean for AR(1) processes (cp. Sect. 5.1.2).
- The power-law sample-size dependence of the asymptotic scaling of the variance of the sample mean of Gaussian ARFIMA samples has been known before [119]. I determine this relation precisely by calculating the corresponding proportionality factor in closed form (cp. Sect. 5.1.2).
- I provide an explanation of the origin of the observation that for AR(1) processes asymptotically the ratio of the given and effective size approaches the value two [120] for AR parameters close to unity. The symmetry of the acf regarding time lags yields an affine linear relation with slope two between the decorrelation time and the correlation time by definition (cp. Sect. 5.1.3).
- Applying the Hermite approach to the calculation of the (finite-size) decorrelation times, I specify the theoretical results above about finite-size and asymptotic effective sample sizes to meta-Gaussian processes (cp. Sect. 5.2).
- Linking the above results about effective sample sizes to large deviation theory, I clarify previously unknown constants in closed-form formulae for large deviation probabilities of Gaussian LRC processes (cp. Sect. 5.3.3).

(iii) Extreme value theory for stationary and meta-Gaussian processes

- As a basis for an application of norming constants, which are typically estimated GEV parameters adjusted to empirical block maxima, I reproduce in detail previously known results from EVT for both iid and stationary processes with particular focus on rates of convergence and the norming constants in the extremal types theorem (cp. Chap. 7).
- By numerical experiments I visualize the impact of a slow convergence in the extremal types theorem on return level estimates by the block maximum or POT approach, in particular, for Gaussian data (cp. Sect. 7.4).

10. Conclusion

- I find a significant impact of long-range dependencies on block maxima estimates of return levels only in the presence of strong LRC by numerical experiments with synthetic ARFIMA and meta-Gaussian ARFIMA samples (cp. Sect. 7.4.4 and Sect. 9.2.1)
- I formulate second order conditions on the transformation generating a meta-Gaussian process, such that the norming constants obtained by a Taylor expansion of the transformation are valid to retain convergence in the extremal types theorem and I derive the corresponding rate of convergence. I apply these results to obtain norming constants for the maxima of tGp samples with arbitrary location and scale (cp. Sect. 8.1).
- I derive a theoretical relation between return levels obtained from the block maxima distribution and from the marginal distribution. In case of iid processes, these two conceptions coincide up to negligible deviations and I show that the marginal distribution provides proper return level estimates especially in case of a slow convergence in the extremal types theorem. Based on this result, I formulate a procedure for evaluating return level estimates obtained from EVT by incorporating additional information about the marginal distribution and both SRC and LRC given by individual samples beyond block maxima only (cp. Sect. 9.2).
- I show that for some distributions in the Gumbel domain of attraction return level estimates by EVT exhibit a high accuracy despite a slow convergence of the corresponding maximum distribution to its limit shape due to an already high agreement of its tail and the limit Gumbel tail. Based on this observation, I formulate a modified Gumbel approach as an alternative method for the estimation of return levels with decreased uncertainty in the presence of individual samples. I define norming constants, such that the tail of the normed maximum distribution coincides with the limit Gumbel tail in a prescribed return level. Generalizing Hall's method [69], I derive the validity of these alternative norming constants in the extremal types theorem (cp. Sect. 9.3).

A central conclusion of my work on effective sample sizes and EVT for stationary processes is a noteworthy difference between the impact of SRC and LRC on inference about sample means and return levels. LRC qualitatively change the asymptotic sample-size dependence of the standard deviation of the sample mean to a power-law decay slower than the square-root rate. In contrast, power-law decaying LRC do not change the asymptotic rate of convergence and limit distribution in the extremal types theorem, whereas SRC are capable of qualitatively changing the location of the limit maximum distribution by a shift due to clustering of extremes. For finite-sample statistics strong SRC or strong LRC nonetheless affect sample mean and return level estimates, respectively, by introducing an effective sample size or an effective extremal index.

Estimating the SRC and LRC in empirical data allows for an evaluation of the uncertainty accompanying sample mean and return level estimates. These conclusions are crucial for my assessment and modeling of daily precipitation amounts in topic (iv).

(iv) Application of the concepts (i) to (iii) to empirical precipitation records

- I formulate an easily applicable and parametric data model for stationary LRC precipitation measurements by applying a tGp transformation to an ARFIMA process (cp. Chap. 6) and elaborate the validity of this model for records from 20 empirical mid-latitude European data sets.
- Reasoning the model design, I assess the stationarity of the empirical data by a comparison of the variance and the annual cycle amongst the individual daily records, and the presence and significance of LRC in the data by regression values describing the linearity of the asymptotic strength of fluctuations measured by DFA and a comparison of the obtained Hurst exponents to randomly shuffled versions of the original data sets with destroyed correlations (cp. Sect. 6.1).
- I provide a step-by-step modeling procedure for the estimation of the five model parameters with focus on extreme events. For the estimation of the location, scale and shape of the tGp transformation I apply a tail-oriented approach by least-square adjusting the model marginal survival function to the empirical survival function in semi-logarithmic scaling. Such an approach discriminates very small and emphasizes large precipitation amounts. I estimate the LRC parameter of the underlying ARFIMA model by DFA(3) directly from the data since the acf of the tGp model obeys the same asymptotic power-law decay due to unit Hermite rank. For the estimation of the AR parameter for highly non-Gaussian data I formulate an alternative method based on conditional probabilities of the occurrence of a prescribed amount of precipitation on two consecutive days (cp. Sect. 6.2).
- I validate my model by a detailed statistical analysis concerning key statistical quantities, namely daily and annual amounts, SRC and LRC, wet and dry spells and annual maxima. A main result is that an ML or an MM estimate of the tGp marginal distribution crucially underestimates extreme precipitation events but represent mean daily and annual totals precisely. The tail-oriented parameter estimation, however, captures the occurrence of extremes events with deviations for small and mean amounts below the measurement precision of 0.1 mm or within one standard deviation (cp. Sect. 6.3).

10. Conclusion

- Applying my estimation of effective sample sizes from topic (ii) to the estimated LRC tGp models, for the 20 data sets I derive effective samples sizes between about 12% to 50% along with confidence intervals for the daily mean and annual total amount (cp. Sect. 6.3.5 and Sect. 6.3.6).
- Applying my procedure from topic (iii) for the evaluation of return level estimates by extreme value theory yields that the tGp distribution obeys slow convergence of Gumbel estimates in the extremal types theorem and GEV estimates are to be preferred. All data sets show moderate to weak SRC and LRC, respectively, such that direct return level estimates by the marginal distribution are valid. Applying my modified Gumbel approach additionally, I further decrease the statistical uncertainty of these estimates of the 100-year return level (cp. Sect. 9.2.2 and Sect. 9.3.4).

10.2. Open questions and impulses for future research

- I draw a relation between the Hermite approach and the recent inverse of idea of Papalexiou [138], who maps the acf of a non-Gaussian to the acf of a Gaussian process. The interpretation of the parameters estimated for this relation as effective Hermite ranks helps choosing an appropriate transformation for the meta-Gaussian approach. Future research could address this idea in detail and with application to other nonlinear transformations of Gaussian processes than the tGp transformation (cp. Sect. 3.4.1).
- I visualize the sensitivity of different estimators of LRC obtained from numerical experiments with nonstationary toy models. The maximal Hurst exponent the estimators yield seemingly depends on the prescribed order of detrending. Rescaled-range statistics are known to obey a unit Hurst exponent at maximum. The slopes by DFA(2) and DFA(3) yields for a cubic additive trend differ from unity and from each other. A precise assessment of these maximal slope requires future research (cp. Sect. 4.3).
- I deduce an explicit formula for the asymptotic large deviation probabilities of the sample mean of Gaussian-LRC processes by applying my results on effective sample size. A direct next step of research could be an application of the same idea to large deviations of meta-Gaussian processes.
- My modified Gumbel approach to the estimation of return levels bases on the tail-oriented estimation of the marginal distribution and shows the capability of reducing statistical uncertainty by the corresponding norming constants. Incorporating more sophisticated methods for modeling the marginal distribution might further improve return level estimates by the modified Gumbel approach (cp. Sect. 9.3).

10.2. Open questions and impulses for future research

- Similar to effective sample sizes prominent SRC could be interpreted as an effective extremal index and thus induce an effective block size as a finite sample property for EVT. Asymptotically such an effective block size tends to the original block for processes with unit extremal index such as ARFIMA processes (cp. Sect. 8.3.1).
- Section 8.3.2 summarizes existing theory on large deviation probabilities for the maxima of iid processes with focus on the expected maximum value and the rate function for large deviations. Both influencing factors depend on the marginal distribution. Possible generalizations are incorporating meta-Gaussian iid or LRC processes. As for power-law decaying autocorrelations the asymptotic rate of convergence in the extremal types theorem coincides with the iid setting, also the results for large deviations of the maximum are expected to agree with the iid results but require further research (cp. Sect. 8.3.2)
- My data model for daily precipitation amounts is appropriate for empirical records with only weak nonstationarities. For the application of this approach to data sets recorded in regions with a strong annual seasonality in the occurrence and daily amount of precipitation nonlinear transformations of LRC processes with a cyclicity provides a generalization of transforming stationary ARFIMA processes.

Appendix

A. Mehler's formula

Mehler's formula [123] provides an Hermite polynomial expansion of the density of a bivariate normal distribution. This formula is the basis of the relation (Thm. 3.2) between the acf of a stationary Gaussian process and the acf of an \mathcal{L}^2 -transformation of this process as described in Section 3.1.

Let $(X, Y) \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma\right)$ be a bivariate Gaussian vector with (two-dimensional) zero mean, unit marginal variance each and autocovariance matrix $\Sigma := \begin{pmatrix} 1 & \varrho \\ \varrho & 1 \end{pmatrix}$ with the correlation $\varrho = \text{Corr}(X, Y)$. Then for values $x, y \in \mathbb{R}$ the density $f_{(X,Y)}$ of the joint distribution of the variates X and Y reads

$$\begin{aligned} f_{(X,Y)}(x, y) &= \frac{1}{2\pi\sqrt{\det(\Sigma)}} e^{-\frac{x^2+y^2-2\varrho xy}{2(1-\varrho^2)}} \\ &= \phi(x)\phi(y) \sum_{j=0}^{\infty} \frac{\varrho^j}{j!} H_j(x)H_j(y) \end{aligned} \quad (\text{A.1})$$

$$= \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} \sum_{j=0}^{\infty} \frac{\varrho^j}{j!} H_j(x)H_j(y), \quad (\text{A.2})$$

where H_j denotes the j -th Hermite polynomial for indices $j \in \mathbb{N}_0$ (cp. Sect. 3.1). The expression (A.1) goes back to Kibble [93]. Using equation (A.2) and the orthogonality (3.7) of the Hermite polynomials, a basic but lengthy calculation yields

$$\mathbb{E}[H_i(X)H_j(Y)] = \varrho^j \langle H_i, H_j \rangle = j! \varrho^j \delta_{ij} \quad (\text{A.3})$$

for indices $i, j \in \mathbb{N}_0$. For arbitrary variances $\text{Var}(\sigma X) = \text{Var}(\sigma Y) = \sigma^2 \in \mathbb{R}_{>0}$ consider the bivariate Gaussian distribution $\mathcal{N}\left((0, 0), \sigma^2 \Sigma\right)$. Then the bivariate joint density $f_{(\sigma X, \sigma Y)}$ of the scaled variates σX and σY satisfies

$$f_{(\sigma X, \sigma Y)}(x, y) = \frac{1}{\sigma^2} f_{(X,Y)}\left(\frac{x}{\sigma}, \frac{y}{\sigma}\right). \quad (\text{A.4})$$

In the setting of the generalized Hermite polynomials $H_j^{\sigma^2}$, equation (A.3) translates to

$$\mathbb{E}\left[H_i^{\sigma^2}(\sigma X)H_j^{\sigma^2}(\sigma Y)\right] = \sigma^{i+j} \mathbb{E}[H_i(X)H_j(Y)] = \sigma^{2j} j! \varrho^j \delta_{ij} \quad (\text{A.5})$$

for indices $i, j \in \mathbb{N}_0$.

B. Specific distributions

distribution	parameter	support	cdf F (and pdf f)
Uniform, $U(a, b)$	$a, b \in \mathbb{R}$ (limits) $a > b$	$a \leq x \leq b$	$F(x) = \frac{x-b}{b-a}$ $f(x) = \frac{1}{b-a}$
Exponential, $\text{Exp}(\lambda)$	$\lambda \in \mathbb{R}_{>0}$ (scale)	$x \in \mathbb{R}_{\geq 0}$	$F(x) = 1 - e^{-\lambda x}$ $f(x) = \lambda e^{-\lambda x}$
Normal, $N(\mu, \sigma^2)$	$\mu \in \mathbb{R}$ (location) $\sigma \in \mathbb{R}_{>0}$ (standard deviation)	$x \in \mathbb{R}$	$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$ $\Phi(x) := \frac{1}{2}\left(1 + \text{erf}\left(\frac{x}{\sqrt{2}}\right)\right)$ $f(x) = \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)$ $\phi(x) := \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$
Chi-square ¹ , $\chi^2(k)$	$k \in \mathbb{N}_{>0}$ (degrees of freedom)	$x \in \mathbb{R}_{>0}$, if $k = 1$ $x \in \mathbb{R}_{\geq 0}$, if $k > 1$	$F(x) = 1 - \frac{\gamma\left(\frac{k}{2}, \frac{x}{2}\right)}{\Gamma\left(\frac{k}{2}\right)}$ $f(x) = \frac{x^{\frac{k}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)}$
Pareto, $\text{Par}(x_{\min}, \alpha)$	$x_{\min} \in \mathbb{R}_{>0}$ (scale) $\alpha \in \mathbb{R}_{>0}$ (shape)	$x \in \mathbb{R}_{\geq x_{\min}}$	$F(x) = 1 - \left(\frac{x_{\min}}{x}\right)^\alpha$ $f(x) = \frac{\alpha x_{\min}^\alpha}{x^{\alpha+1}}$
Generalized extreme value distribution, $\text{GEV}(\mu, \sigma, \xi)$	$\mu \in \mathbb{R}$ (location) $\sigma \in \mathbb{R}_{>0}$ (scale) $\xi \in \mathbb{R}$ (shape)	$x \in \mathbb{R}$: $1 + \xi\left(\frac{x-\mu}{\sigma}\right) > 0$	$F(x) = e^{-(1 + \xi\left(\frac{x-\mu}{\sigma}\right))^{-1/\xi}}$
Gamma $\Gamma(\alpha, \beta)$	$\beta \in \mathbb{R}_{\geq 0}$ (scale) $\alpha \in \mathbb{R}_{>0}$ (shape)	$x \in \mathbb{R}_{>0}$	$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$

Table B.1 Collection of frequently used distributions.

¹In the cdf of the χ^2 -distribution the two-argument function γ denotes the lower incomplete gamma function. For $k = 1$ degree of freedom the cdf and pdf are given in equation (3.28).

C. Convergence of distribution functions

Let $F_n, F : \mathbb{R} \rightarrow [0, 1]$ be cdfs.

Definition (Pointwise convergence): A series $(F_n)_{n \in \mathbb{N}}$ converges pointwise to a limit cdf F , denoted by $F_n \xrightarrow{\text{p.w.}} F$, if for all $x \in \mathbb{R}$

$$\lim_{n \rightarrow \infty} F_n(x) = F(x).$$

Definition (Convergence in distribution): A series $(F_n)_{n \in \mathbb{N}}$ converges in distribution to a limit cdf F , denoted by $F_n \xrightarrow{d} F$, if the cdfs F_n converge pointwise to the limit cdf F in all the continuity points $x \in \mathbb{R}$ of F .

Definition (Uniform distribution): A series $(F_n)_{n \in \mathbb{N}}$ converges uniformly to a limit cdf F if the cdfs F_n converge to F uniformly on \mathbb{R} , in other words, if for all $x \in \mathbb{R}$

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = 0. \quad (\text{C.1})$$

Lemma C.1: Let $F_n, F : \mathbb{R} \rightarrow [0, 1], n \in \mathbb{N}$, be cdfs such that F is continuous and $F_n \xrightarrow{\text{p.w.}} F$ as $n \rightarrow \infty$. Then $F_n \xrightarrow{d} F$ uniformly on \mathbb{R} as $n \rightarrow \infty$.

Proof. Based on the pointwise convergence of the series $(F_n)_{n \in \mathbb{N}}$ to F , uniform convergence is implied by the continuity and asymptotics of the limit cdf F and the monotony of both all F_n and F as follows.

Let $\varepsilon > 0$. By the continuity and asymptotics of F we can choose a finite $(m+1)$ -tuple $-\infty =: x_0 < x_1 < \dots < x_{m-1} < x_m := \infty, m \in \mathbb{N}$, such that

$$|F(x_{i-1}) - F(x_i)| < \frac{\varepsilon}{5} \quad (\text{C.2})$$

for all $i = 1, \dots, m$. Note that, in particular, the convergence of F as $x \rightarrow \pm\infty$ allows for choices of x_1 and x_{m-1} such that $|F(x_0) - F(x_1)| = |0 - F(x_1)| < \varepsilon/5$ and $|F(x_{m-1}) - F(x_m)| = |F(x_{m-1}) - 1| < \varepsilon/5$. Hence, for all $n \in \mathbb{N}$ we obtain

$$\begin{aligned} & \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \\ & \leq \max_{i=1, \dots, m} \sup_{x \in [x_{i-1}, x_i]} |F_n(x) - F_n(x_i)| + |F_n(x_i) - F(x_i)| + |F(x_i) - F(x)| \\ & \leq \max_{i=1, \dots, m} |F_n(x_{i-1}) - F_n(x_i)| + |F_n(x_i) - F(x_i)| + |F(x_{i-1}) - F(x_i)| \end{aligned} \quad (\text{C.3})$$

Appendix

by the monotony of F_n and F . Let $n_0 \in \mathbb{N}$ be such that by the pointwise convergence of the cdfs F_n

$$|F_n(x_i) - F(x_i)| < \frac{\varepsilon}{5} \quad (\text{C.4})$$

for all $i = 1, \dots, m$ and $n \geq n_0$. Such an n_0 exists since the number $m + 1$ of points x_i is finite. We can further estimate the first summand in (C.3) by

$$\begin{aligned} & |F_n(x_{i-1}) - F_n(x_i)| \\ & \leq |F_n(x_{i-1}) - F(x_{i-1})| + |F(x_{i-1}) - F(x_i)| + |F_n(x_i) - F(x_i)|. \end{aligned} \quad (\text{C.5})$$

Applying (C.2) and (C.4) to (C.3) and (C.5), for all $n \geq n_0$ we conclude

$$\begin{aligned} & \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \\ & \leq \max_{i=1, \dots, k} |F_n(x_{i-1}) - F(x_{i-1})| + 2|F_n(x_i) - F(x_i)| + 2|F(x_{i-1}) - F(x_i)| < \varepsilon \end{aligned}$$

and obtain convergence in distribution or $F_n \xrightarrow{d} F$, respectively. \square

D. Weak dependence structures of stochastic processes

The two extreme scenarios of dependence in a stochastic process are the setting of full independence and of a completely dependent series of just the same random variable repeated. In between these two limit cases there exist various characteristics describing the dependence structure [56, Chap. 3]. The collection below contains *short-memory* characterisations that do not change the shape of the limit distribution both of the CLT (cp. Rem. 5.1 in Sect. 5.1) and in the EVT (cp. Thm. 7.10 in Sect. 7.2.1). Let

$$\alpha(k) := \sup \left\{ |\mathrm{P}(A \cap B) - \mathrm{P}(A)\mathrm{P}(B)| \mid A \in \mathcal{F}_0^-, B \in \mathcal{F}_k^+ \right\},$$

where \mathcal{F}_0^- and \mathcal{F}_k^+ are the σ -algebras generated by the “past information” $X_s, s \in \mathbb{Z}_{\leq k}$, and the “future information” $X_s, s \in \mathbb{Z}_{>k}$, respectively [56].

Definition (α -mixing, strong mixing): A stationary process X_t is α -mixing or *strong mixing* if $\alpha(k) \rightarrow 0$ as $k \rightarrow \infty$.

The conditions $D(u_n)$ and $D'(u_n)$ (Sect. 7.2.1) are a generalization of strong mixing. Leadbetter et al. [101, Sect. 3.2] introduced these weaker requirements as a convenient condition on the dependence structure of extreme events in EVT.

Definition (m -dependence): A stationary process X_t obeys *m -dependence* if there exists an integer $m \in \mathbb{N}$ such that $\alpha(k) = 0$ for all time lags $k \in \mathbb{Z}_{>m}$.

Definition (Markov property): A discrete-time stochastic process X_t possesses the *Markov property* if for all times $t \in \mathbb{N}_0$

$$\mathrm{P}(X_t = x_t \mid X_{t-1} = x_{t-1}, \dots, X_0 = x_0) = \mathrm{P}(X_t = x_t \mid X_{t-1} = x_{t-1}). \quad (\text{D.1})$$

The Markov property of continuous-time processes is defined by means of the filtration adapted to the process. Markov chains are discrete-time Markov processes. Basic calculations imply an exponentially decaying acf for such processes by raising their diagonalized transition matrix to the power of increasing time lags. Note that the eigenvalues of stochastic matrices are positive and bounded from above by one. An example of a continuous-time Markov process is the Brownian motion.

E. Estimated model parameters

Table E.1 provides the estimated parameter values of the LRC tGp model for measurement data from 20 different mid-latitude European locations (cp. Sect. 6.3) together with effective sample sizes for these time series (cp. Sect. 6.3.5). Table E.2 compares the statistics of the empirical data and the adjusted models for all 20 data sets. Table E.3 collects the estimated Hurst exponents along with their validation by regression values for all 20 data sets.

data set	country	period	N	N_{eff}	η	ν	σ	d	φ
Aachen	GER	1946 - 2010	23741	6121	2.923	0.759	0.906	0.063	0.301
Cottbus	GER	1947 - 2018	26298	7206	7.25	0.886	0.262	0.081	0.221
Fichtelberg	GER	1916 - 2018	37621	5074	4.083	1.048	0.521	0.096	0.284
Greifswald	GER	1978 - 2018	14975	3953	3.906	0.747	0.58	0.081	0.212
Hamburg	GER	1936 - 2018	30316	4575	3.404	0.844	0.671	0.095	0.278
Jena	GER	1946 - 2018	26663	7470	4.255	0.764	0.527	0.076	0.205
Karlsruhe	GER	1876 - 2007	48212	9250	4.127	0.885	0.534	0.083	0.249
Magdeburg	GER	1947 - 2018	26298	5980	3.913	0.646	0.607	0.092	0.196
Schwerin	GER	1947 - 2018	26298	6235	6.085	0.927	0.292	0.079	0.235
Trier	GER	1947 - 1998	18993	3179	2.441	0.544	1.254	0.094	0.281
Central England	GBR	1931 - 2018	32142	4472	2.967	0.843	0.696	0.099	0.222
Southeast England	GBR	1931 - 2018	32142	3054	1.854	0.279	1.913	0.119	0.247
Bordeaux	FRA	1946 - 2018	26663	2779	2.329	0.429	1.57	0.111	0.356
Luxembourg	LUX	1947 - 2018	26298	3974	2.111	0.356	1.735	0.094	0.3
Valencia	ESP	1938 - 2018	29585	12598	3.678	-0.401	1.222	0.052	0.439
Zaragoza	ESP	1941 - 2018	28489	12496	2.503	-1.009	1.779	0.047	0.356
Sarajevo	BIH	1901 - 2018	43099	16700	2.372	0.341	1.626	0.04	0.289
Vaernes	NOR	1946 - 2018	26663	2408	2.826	0.846	0.893	0.121	0.283
Malmö	SWE	1945 - 2018	27028	7504	3.828	0.793	0.569	0.073	0.189
Bollerup	SWE	1945 - 2018	27028	5820	3.123	0.681	0.817	0.087	0.2

Table E.1 Estimated model parameters for 20 mid-latitude European locations.

data set	country	mean		variance		P(data = 0)	P(model ≤ 0.1)	P(data ≤ 4)	P(model ≤ 4)
		data	model	data	model				
Aachen	GER	2.249	2.28	20.881	20.373	0.465	0.369	0.817	0.825
Cottbus	GER	1.567	1.674	16.982	16.562	0.54	0.273	0.881	0.892
Fichtelberg	GER	3.14	3.349	36.003	33.939	0.411	0.179	0.759	0.753
Greifswald	GER	1.645	1.689	14.379	14.07	0.523	0.37	0.869	0.879
Hamburg	GER	2.071	2.159	18.013	17.094	0.469	0.308	0.832	0.837
Jena	GER	1.639	1.703	15.707	15.358	0.515	0.365	0.875	0.881
Karlsruhe	GER	2.176	2.322	23.463	22.064	0.52	0.279	0.828	0.832
Magdeburg	GER	1.401	1.424	12.572	12.419	0.535	0.44	0.894	0.9
Schwerin	GER	1.737	1.905	15.904	14.312	0.508	0.203	0.86	0.87
Trier	GER	2.089	2.096	18.875	18.529	0.536	0.451	0.829	0.835
Central England	GBR	1.783	1.805	9.595	9.295	0.301	0.291	0.856	0.86
Southeast England	GBR	2.026	1.969	14.193	14.407	0.362	0.502	0.83	0.831
Bordeaux	FRA	2.506	2.538	30.48	29.924	0.532	0.486	0.813	0.811
Luxembourg	LUX	2.32	2.303	22.626	22.524	0.505	0.495	0.815	0.818
Valencia	ESP	1.236	1.246	42.509	44.495	0.812	0.778	0.935	0.936
Zaragoza	ESP	0.897	0.884	12.794	13.051	0.787	0.786	0.938	0.939
Sarajevo	BIH	2.532	2.604	36.073	35.102	0.571	0.509	0.822	0.814
Vaernes	NOR	2.327	2.394	19.481	18.824	0.441	0.326	0.805	0.811
Malmø	SWE	1.68	1.778	14.375	13.532	0.537	0.333	0.863	0.871
Bollerup	SWE	1.81	1.881	16.831	16.036	0.586	0.402	0.851	0.859

Table E.2 Comparison between the statistics of the empirical data sets in Table E.1 and the adjusted models.

data set	country	DFA	R/s	wavelet	linear regression values each			
Aachen	GER	0.563	0.575	0.586	0.9988	0.9986	0.9977	0.9977
Cottbus	GER	0.581	0.597	0.591	0.9989	0.9988	0.9983	0.9983
Fichtelberg	GER	0.596	0.614	0.589	0.9989	0.9993	0.998	0.998
Greifswald	GER	0.581	0.578	0.597	0.9977	0.998	0.9967	0.9967
Hamburg	GER	0.595	0.594	0.614	0.9983	0.9974	0.9979	0.9979
Jena	GER	0.576	0.594	0.592	0.9982	0.9987	0.9981	0.9981
Karlsruhe	GER	0.583	0.615	0.567	0.9994	0.9995	0.9944	0.9944
Magdeburg	GER	0.592	0.61	0.603	0.9995	0.9996	0.9977	0.9977
Schwerin	GER	0.579	0.583	0.56	0.9987	0.9981	0.9928	0.9928
Trier	GER	0.594	0.618	0.597	0.9988	0.9995	0.9962	0.9962
Central England	GBR	0.599	0.603	0.624	0.9984	0.9983	0.997	0.997
Southeast England	GBR	0.62	0.615	0.637	0.998	0.9973	0.9964	0.9964
Bordeaux	FRA	0.611	0.633	0.582	0.9973	0.999	0.9944	0.9944
Luxembourg	LUX	0.594	0.616	0.593	0.9987	0.9994	0.9968	0.9968
Valencia	ESP	0.552	0.571	0.554	0.9975	0.9985	0.9952	0.9952
Zaragoza	ESP	0.547	0.566	0.528	0.9986	0.9986	0.9951	0.9951
Sarajevo	BIH	0.54	0.57	0.542	0.9974	0.9982	0.9981	0.9981
Vaernes	NOR	0.621	0.619	0.667	0.9974	0.9974	0.9974	0.9974
Malmø	SWE	0.573	0.582	0.593	0.9976	0.9976	0.996	0.996
Bollernp	SWE	0.587	0.593	0.603	0.9975	0.9977	0.9969	0.9969

Table E.3 Estimated Hurst exponents for the empirical data sets in Table E.1 obtained by DFA(3), R/s and wavelet analysis, respectively. Regression values close to one assess the linearity of the growth of the fluctuation function, rescaled range and wavelet coefficients in double-logarithmic scaling.

F. Influence of tails in extreme value theory

Let $M_n = \{X_1, \dots, X_n\}$ be the maximum of $n \in \mathbb{N}$ iid random variables X_i , $i = 1, \dots, n$, with common cdf F . The shape of the distribution of M_n is dominated by the tail behavior of the cdf F (cp. Lem. 7.6). Lemma F.1 provides further insight into how the speed of the decay of the survival function $1 - F$ influences the distribution of extremes (cp. [101, Thm. 1.7.13]).

Lemma F.1 (Thm. 3.1.3 in [48]): *Let $x_m \leq \infty$ be the right endpoint (7.4) of a cdf F and let $\tau \in \mathbb{R}_{>0}$. A sequence $(u_n)_{n \in \mathbb{N}}$ of exceedances $u_n \in \mathbb{R}$ such that*

$$n(1 - F(u_n)) \rightarrow \tau \quad (n \rightarrow \infty) \tag{F.1}$$

exists if and only if

$$\lim_{x \nearrow x_m} \frac{1 - F(x)}{1 - F(x_-)} = 1 \quad \text{and} \quad F(x_{m-}) = 1, \tag{F.2}$$

where $F(x_-) := \lim_{t \nearrow x} F(t)$ denotes the left or lower limit of F in $x \in \mathbb{R}$.

Lemma F.1 applies, in particular, to discrete distributions with infinite upper endpoint. Prominent examples of such distributions, which violate limit condition (F.2), are the Poisson and the geometric distribution. In either case, the decay of the survival function is too fast, in the sense that the attained limit (F.1) is less than one (Exs. 3.1.4 and 3.1.5 in [48]). In other words, the jump heights of F decay too slowly to yield a non-degenerate limit distribution of M_n as $n \rightarrow \infty$.

GEV(μ, σ, ξ)	characterization	examples
$\xi > 0$	$\alpha = \frac{1}{\xi}$ $1 - F(x) = x^{-\alpha}L(x), x \in \mathbb{R}$ $L : \mathbb{R} \rightarrow \mathbb{R}$ slowly varying von Mises (sufficient) condition: $\lim_{x \rightarrow \infty} \frac{xf(x)}{1 - F(x)} = \alpha$	Pareto $\text{Par}(x_{\min}, \alpha)$, Cauchy, Burr, α -stable with $\alpha < 2$
$\xi < 0$	$\alpha = \frac{1}{\xi}, x_m < \infty$ $1 - F(x_m - \frac{1}{x}) = x^{-\alpha}L(x), x \in \mathbb{R}$ $L : \mathbb{R} \rightarrow \mathbb{R}$ slowly varying von Mises (sufficient) condition: $\lim_{x \nearrow x_m} \frac{(x_m - x)f(x)}{1 - F(x)} = \alpha$	uniform $U(a, b)$, power-law tail with finite upper endpoint Beta distribution
$\xi = 0$	$x_m \leq \infty, \exists z < x < x_m :$ $1 - F(x) = c(x)e^{-\int_z^x \frac{g(t)}{h(t)} dt}, x \in \mathbb{R}$ $c(x) \rightarrow c > 0, g(x) \rightarrow 1, h'(x) \rightarrow 0 (x \nearrow x_m)$ and moderate further conditions (necessary and sufficient) condition: $\lim_{x \nearrow x_m} \frac{1 - F(x + f\tilde{h}(x))}{1 - F(x)} = e^{-t}, t \in \mathbb{R}$ with, e.g., $\tilde{h}(x) = \int_x^{x_m} \frac{1 - F(t)}{1 - F(x)} dt, x < x_m$	Gaussian, Lognormal Weibull, Gamma, exponential, exponential tail with finite upper endpoint

Table F.1 Characterization of domains of attractions of EVDs and example distributions

G. Mean residual life plot

The application of EVT by the POT approach requires the choice of an appropriate threshold for the selection of points for the statistical analysis (cp. Sect. 7.4). See [37] for a more detailed description of the methodology of threshold selection below.

Let $n \in \mathbb{N}$ denote a number of iid samples of a random variable X with a cdf in the domain of attraction of a $\text{GEV}(\mu, \sigma, \xi)$ distribution with $\mu, \xi \in \mathbb{R}$ and $\sigma \in \mathbb{R}_{>0}$. Then by Theorem 7.28 approximately for large numbers n and large thresholds $u \in \mathbb{R}$, the distribution of the excesses $X - u$ conditional on $X > u$ follow a $\text{GPD}(0, \sigma_u, \xi)$ with $\sigma_u := \sigma + \xi(u - \mu)$. The mean of such a GPD reads $\frac{\sigma_u}{1-\xi}$ if $\xi < 1$, and is infinite otherwise. Therefore, for the mean size of the threshold exceedances we find

$$\mathbb{E}[X - u \mid X > u] = \frac{\sigma + \xi(u - \mu)}{1 - \xi}. \quad (\text{G.1})$$

Relation (G.1) reveals a linear dependency of the mean size of excesses on the threshold u in the asymptotic regime of the GPD. Testing for a linear change of the mean conditional exceedances under altering thresholds helps locate a valid regime of thresholds. An appropriate estimator for these means is the sample mean of the empirical excesses. Let $N_u \in \mathbb{N}$ denote the number of empirical values above a threshold u . A plot of the points

$$\left(u, \frac{1}{N_u} \sum_{i=1}^{N_u} (X_{(i)} - u) \right) \quad (\text{G.2})$$

is called *mean residual life plot* and visualizes the shape of the dependency of the mean excesses from the height of the threshold. Any threshold from a linear region in such a plot forms a suitable choice for the POT approach. For a visualization of this procedure see Example 7.29.

Elementary arguments yield exact formulae for the mean excess $m_u := \mathbb{E}[X - u \mid X > u]$ of an exponentially (with rate parameter $\lambda \in \mathbb{R}_{>0}$), Gaussian or Pareto (with location $x_{\min} = 1$ and shape parameter $\alpha \in \mathbb{R}_{>1}$) distributed random variable X as

$$\text{Exp}(\lambda) : \quad m_u = \frac{1}{\lambda^2} \quad (\text{G.3})$$

$$\text{N}(0, 1) : \quad m_u = \frac{\phi(u)}{1 - \Phi(u)} - u \quad (\text{G.4})$$

$$\text{Par}(1, \alpha) : \quad m_u = \frac{u}{\alpha - 1}. \quad (\text{G.5})$$

Abbreviations

General abbreviations

appx.	appendix
chap.	chapter
cor.	corollary
cp.	compare (to a given reference)
e.g.	for example (exempli gratia (lat.))
equ.	equation (plural: equs.)
ex.	example (plural: exs.)
i.e.	that is (id est (lat.))
lem.	lemma
p.	page
prop.	proposition
rem.	remark (plural: rems.)
sec.	section
thm.	theorem

Specific abbreviations

acf	autocorrelation function (plural: acfs)
cdf	cumulative distribution function (plural: cdfs)
cfp	conditional follow-up probability
ecdf	empirical (cumulative) distribution function
EVD	extreme value distribution (plural: EVDs)
EVT	extreme value theory
fBm	fractional Brownian motion
fGn	fractional Gaussian noise
GEV	generalized extreme value (distribution)
iid	independent and identically distributed
LRC	long-range correlations
ML	maximum likelihood
MM	method of moments
pdf	probability density function (plural: pdfs)
sf	survival function
SRC	short-range correlations
tGp	truncated-Gaussian-power (distribution)

Symbols

General symbols

\mathbb{N}	set of all natural numbers excluding the number zero
\mathbb{N}_0	set of all natural numbers including the number zero
\mathbb{R}	set of all real numbers
$\mathbb{R}_{\geq 0}$	set of all non-negative real numbers
$\mathbb{R}_{> 0}$	set of all positive real numbers
\ln	natural logarithm
χ_A	indicator function of a set A
Φ	cdf of the standard Gaussian distribution
ϕ	pdf of the standard Gaussian distribution
$\mathcal{L}^2(\mathbb{R}, \mu)$	space of all with respect to the measure μ square-integrable functions on $\mathbb{R} \rightarrow \mathbb{R}$
$\stackrel{d}{=}$	equality in distribution
\xrightarrow{d}	convergence in distribution
\implies	weak convergence in $D[0, 1]$:= $\{[0, 1] \rightarrow \mathbb{R} \mid \text{right-cont. on } [0, 1), \text{ finite limits on } (0, 1]\}$
$\text{dom}(f)$	domain of a function f
$P(X \leq x)$	probability of an event $\{X \leq x\} = \{\omega \in \Omega \mid X(\omega) \leq x\}$
$X \sim P$	A random variable X follows a distribution P
$X \sim F$	A random variable X follows a distribution with cdf F
$f(x) \sim g(x)$ $(x \rightarrow \infty)$	<i>Asymptotic equivalence</i> : The functions f and g are called asymptotically equivalent if $\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = 1$.
$f(x) \propto g(x)$ $(x \rightarrow \infty)$	<i>Asymptotic proportionality</i> : The functions f and g are called asymptotically proportional if $\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = c \in (0, \infty)$.

Specific symbols

$\varrho, (\varrho_X)$	autocorrelation function (of a stochastic process X_t)	12,227
$\tau_D(N)$	finite-size decorrelation time of $N \in \mathbb{N}$ samples	73
τ_D	decorrelation time	77
N_{eff}	effective sample size of $N \in \mathbb{N}$ samples of a stationary process	74
$\Delta_n(x)$	local error in Theorem 7.3 at a position $x \in \mathbb{R}$, $n \in \mathbb{N}$	145
Δ_n	global error in the extremal types theorem, $n \in \mathbb{N}$	145
$\Delta'_n, \Delta''_n, \Delta'''_n$	specific errors affecting the global error Δ_n , $n \in \mathbb{N}$	146
Q_m	$\in \mathbb{R}$, return level of block maxima with return period $m \in \mathbb{N}$	159
q_m	$\in \mathbb{R}$, m -year return level of a distribution of daily data, $m \in \mathbb{N}$	159

Bibliography

- [1] Abadir, K.M., Talmain, G.: Autocovariance functions of series and of their transforms. *Journal of Econometrics* **124**(2), 227–252 (2005). DOI: [10.1016/j.jeconom.2004.02.015](https://doi.org/10.1016/j.jeconom.2004.02.015)
- [2] Abramowitz, M., Stegun, I.A.: *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*. 55 (1965). URL: http://people.math.sfu.ca/~cbm/aands/abramowitz_and_stegun.pdf
- [3] Abry, P., Flandrin, P., Taqqu, M.S., Veitch, D.: Self similarity and long-range dependence through the wavelet lens. In: Gubner, J.A. (ed.) *Theory and Applications of Longrange Dependence*, pp. 591–614. Cambridge University Press, Cambridge (2003). DOI: [10.1017/CBO9780511813610.017](https://doi.org/10.1017/CBO9780511813610.017)
- [4] Abry, P., Veitch, D.: Wavelet analysis of long-range-dependent traffic. *IEEE Transactions on Information Theory* **44**(1), 2–15 (1998). DOI: [10.1109/18.650984](https://doi.org/10.1109/18.650984)
- [5] Aghion, E., Meyer, P.G., Adlakha, V., Kantz, H., Bassler, K.E.: Moses, Noah and Joseph effects in Lévy walks. *New Journal of Physics* **23**(2), 023002 (2021). DOI: [10.1088/1367-2630/abd43c](https://doi.org/10.1088/1367-2630/abd43c)
- [6] Ailliot, P., Thompson, C., Thomson, P.: Space-time modelling of precipitation by using a hidden Markov model and censored Gaussian distributions. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **58**(3), 405–426 (2009). DOI: [10.1111/j.1467-9876.2008.00654.x](https://doi.org/10.1111/j.1467-9876.2008.00654.x)
- [7] Alodah, A., Seidou, O.: The adequacy of stochastically generated climate time series for water resources systems risk and performance assessment. *Stochastic Environmental Research and Risk Assessment* **33**(1), 253–269 (2019). DOI: [10.1007/s00477-018-1613-2](https://doi.org/10.1007/s00477-018-1613-2)
- [8] Altmann, E.G., Kantz, H.: Recurrence time analysis, long-term correlations, and extreme events. *Physical Review E* **71**(5), 056106 (2005). DOI: [10.1103/PhysRevE.71.056106](https://doi.org/10.1103/PhysRevE.71.056106)
- [9] Ayache, A., Taqqu, M.S.: Rate Optimality of Wavelet Series Approximations of Fractional Brownian Motion. *Journal of Fourier Analysis and Applications* **9**(5), 451–471 (2003). DOI: [10.1007/s00041-003-0022-0](https://doi.org/10.1007/s00041-003-0022-0)
- [10] Bai, S., Taqqu, M.s.: Sensitivity of the Hermite rank. *Stochastic Processes and their Applications* **129**(3), 822–840 (2019). DOI: [10.1016/j.spa.2018.03.020](https://doi.org/10.1016/j.spa.2018.03.020)
- [11] Baillie, R.T.: Long memory processes and fractional integration in econometrics. *Journal of Econometrics* **73**(1), 5–59 (1996). DOI: [10.1016/0304-4076\(95\)01732-1](https://doi.org/10.1016/0304-4076(95)01732-1)

- [12] Balkema, A.A., de Haan, L.: Residual Life Time at Great Age. *The Annals of Probability* **2**(5), 792–804 (1974). URL: <http://projecteuclid.org/euclid.aop/1176996548>
- [13] Bardet, J.M., Kammoun, I.: Asymptotic Properties of the Detrended Fluctuation Analysis of Long-Range-Dependent Processes. *IEEE Transactions on Information Theory* **54**(5), 2041–2052 (2008). DOI: [10.1109/TIT.2008.920328](https://doi.org/10.1109/TIT.2008.920328)
- [14] Bárdossy, A., Pegram, G.G.S.: Copula based multisite model for daily precipitation simulation. *Hydrology and Earth System Sciences* **13**(12), 2299–2314 (2009). DOI: [10.5194/hess-13-2299-2009](https://doi.org/10.5194/hess-13-2299-2009)
- [15] Bardossy, A., Plate, E.J.: Space-time model for daily rainfall using atmospheric circulation patterns. *Water Resources Research* **28**(5), 1247–1259 (1992). DOI: [10.1029/91WR02589](https://doi.org/10.1029/91WR02589)
- [16] Barunik, J., Kristoufek, L.: On Hurst exponent estimation under heavy-tailed distributions. *Physica A: Statistical Mechanics and its Applications* **389**(18), 3844–3855 (2010). DOI: [10.1016/j.physa.2010.05.025](https://doi.org/10.1016/j.physa.2010.05.025)
- [17] Bashan, A., Bartsch, R., Kantelhardt, J.W., Havlin, S.: Comparison of detrending methods for fluctuation analysis. *Physica A: Statistical Mechanics and its Applications* **387**(21), 5080–5090 (2008). DOI: [10.1016/j.physa.2008.04.023](https://doi.org/10.1016/j.physa.2008.04.023)
- [18] Baxevani, A., Lennartsson, J.: A spatiotemporal precipitation generator based on a censored latent Gaussian field. *Water Resources Research* **51**(6), 4338–4358 (2015). DOI: [10.1002/2014WR016455](https://doi.org/10.1002/2014WR016455)
- [19] Beirlant, J., Caeiro, F., Ivette Gomes, M.: An overview and open research topics in statistics of univariate extremes. *Revstat Statistical Journal* **10**(1), 1–31 (2012). URL: <https://ine.pt/revstat/pdf/rs120101.pdf>
- [20] Bennett, B., Thyer, M., Leonard, M., Lambert, M., Bates, B.: A comprehensive and systematic evaluation framework for a parsimonious daily rainfall field model. *Journal of Hydrology* **556**, 1123–1138 (2018). DOI: [10.1016/j.jhydrol.2016.12.043](https://doi.org/10.1016/j.jhydrol.2016.12.043)
- [21] Beran, J.: *Statistical Methods for Data with Long-Range Dependence*. *Statistical Science* **7**(4), 404–416 (1992). DOI: [10.1214/ss/1177011122](https://doi.org/10.1214/ss/1177011122)
- [22] Beran, J., Feng, Y., Ghosh, S., Kulik, R.: *Long-Memory Processes*. Springer Berlin Heidelberg, Berlin, Heidelberg (2013). DOI: [10.1007/978-3-642-35512-7](https://doi.org/10.1007/978-3-642-35512-7)
- [23] Berman, S.M.: Limit Theorems for the Maximum Term in Stationary Sequences. *The Annals of Mathematical Statistics* **35**(2), 502–516 (1964). DOI: [10.1214/aoms/1177703551](https://doi.org/10.1214/aoms/1177703551)

- [24] Borak, S., Härdle, W.K., López Cabrera, B.: *Statistics of Financial Markets*. Springer Berlin Heidelberg, Berlin, Heidelberg (2010). DOI: [10.1007/978-3-642-11134-1](https://doi.org/10.1007/978-3-642-11134-1)
- [25] Botev, Z.I.: The normal law under linear restrictions: simulation and estimation via minimax tilting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79**(1), 125–148 (2017). DOI: [10.1111/rssb.12162](https://doi.org/10.1111/rssb.12162)
- [26] Box, G.E.P., Cox, D.R.: An Analysis of Transformations. *Journal of the Royal Statistical Society: Series B (Methodological)* **26**(2), 211–243 (1964). DOI: [10.1111/j.2517-6161.1964.tb00553.x](https://doi.org/10.1111/j.2517-6161.1964.tb00553.x)
- [27] Box, G.E.P., Jenkins, G.M., Reinsel, G.C.: *Time Series Analysis, Wiley Series in Probability and Statistics*, vol. 20. Wiley (2008). DOI: [10.1002/9781118619193](https://doi.org/10.1002/9781118619193)
- [28] Brockwell, P.J., Davis, R.A.: *Time Series: Theory and Methods*, 2 edn. Springer Series in Statistics. Springer New York, New York (1991). DOI: [10.1007/978-1-4419-0320-4](https://doi.org/10.1007/978-1-4419-0320-4)
- [29] Bunde, A., Eichner, J.F., Kantelhardt, J.W., Havlin, S.: Long-Term Memory: A Natural Mechanism for the Clustering of Extreme Events and Anomalous Residual Times in Climate Records. *Physical Review Letters* **94**(4), 048701 (2005). DOI: [10.1103/PhysRevLett.94.048701](https://doi.org/10.1103/PhysRevLett.94.048701)
- [30] Bărbulescu, A., Serban, C., Maftai, C.: Evaluation of Hurst exponent for precipitation time series. *Proceedings of the 14th WSEAS international conference on Computers II*, 590–595 (2010). URL: <http://www.wseas.us/e-library/conferences/2010/Corfu/COMPUTERS/COMPUTERS2-31.pdf>
- [31] Canto E Castro, L.: Uniform rates of convergence in extreme-value theory. Normal and gamma models. *Annales scientifiques de l’Université de Clermont-Ferrand 2. Série Probabilités et applications* **90**(6), 25–41 (1987). URL: http://www.numdam.org/article/ASCFPA_1987__90_6_25_0.pdf
- [32] Charras-Garrido, M., Lezaud, P.: Extreme Value Analysis: an Introduction. *Journal de la Société Française de Statistique* **154**(2) (2013). URL: <https://hal-enac.archives-ouvertes.fr/hal-00917995/>
- [33] Chen, L., Bassler, K.E., McCauley, J.L., Gunaratne, G.H.: Anomalous scaling of stochastic processes and the Moses effect. *Physical Review E* **95**(4), 042141 (2017). DOI: [10.1103/PhysRevE.95.042141](https://doi.org/10.1103/PhysRevE.95.042141)

- [34] Chowdhury, A.F.M.K., Lockart, N., Willgoose, G., Kuczera, G., Kiem, A.S., Parana Manage, N.: Development and evaluation of a stochastic daily rainfall model with long-term variability. *Hydrology and Earth System Sciences* **21**(12), 6541–6558 (2017). DOI: [10.5194/hess-21-6541-2017](https://doi.org/10.5194/hess-21-6541-2017)
- [35] Cohen, J.P.: Convergence rates for the ultimate and pentultimate approximations in extreme-value theory. *Advances in Applied Probability* **14**(4), 833–854 (1982). DOI: [10.2307/1427026](https://doi.org/10.2307/1427026)
- [36] Cohen, J.P.: The penultimate form of approximation to normal extremes. *Advances in Applied Probability* **14**(2), 324–339 (1982). DOI: [10.2307/1426524](https://doi.org/10.2307/1426524)
- [37] Coles, S.: *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. Springer London, London (2001). DOI: [10.1007/978-1-4471-3675-0](https://doi.org/10.1007/978-1-4471-3675-0)
- [38] Corless, R.M., Gonnet, G.H., Hare, D.E.G., Jeffrey, D.J., Knuth, D.E.: On the LambertW function. *Advances in Computational Mathematics* **5**(1), 329–359 (1996). DOI: [10.1007/BF02124750](https://doi.org/10.1007/BF02124750)
- [39] Crouse, M.S., Baraniuk, R.G.: Fast , Exact Synthesis of Gaussian and nonGaussian Long-Range-Dependent Processes (1999). URL: <https://scholarship.rice.edu/handle/1911/19819>
- [40] David, F.N., Cramer, H.: *Mathematical Methods of Statistics*. *Biometrika* **34**(3/4), 374 (1947). DOI: [10.2307/2332454](https://doi.org/10.2307/2332454)
- [41] Dekkers, A.L.M., Einmahl, J.H.J., De Haan, L.: A Moment Estimator for the Index of an Extreme-Value Distribution. *The Annals of Statistics* **17**(4), 1833–1855 (1989). URL: <http://www.jstor.org/stable/2241667>
- [42] Deng, W., Barkai, E.: Ergodic properties of fractional Brownian-Langevin motion. *Physical Review E* **79**(1), 011112 (2009). DOI: [10.1103/PhysRevE.79.011112](https://doi.org/10.1103/PhysRevE.79.011112)
- [43] Deutscher Wetterdienst (DWD): (2018). URL: <https://www.dwd.de/DE/klimaumwelt/cdc/cdc{ }node.html>. URL: <https://www.dwd.de/DE/klimaumwelt/cdc/cdc{ }node.html>. Accessed: 2019-09-26
- [44] Dittmann, I., Granger, C.W.J.: Properties of nonlinear transformations of fractionally integrated processes. *Journal of Econometrics* **110**(2), 113–133 (2002). DOI: [10.1016/S0304-4076\(02\)00089-1](https://doi.org/10.1016/S0304-4076(02)00089-1)
- [45] Du, L., Chen, S.: Asymptotic properties for distributions and densities of extremes from generalized gamma distribution. *Journal of the Korean Statistical Society* **45**(2), 188–198 (2016). DOI: [10.1016/j.jkss.2015.09.005](https://doi.org/10.1016/j.jkss.2015.09.005)

- [46] Eichner, J.F., Kantelhardt, J.W., Bunde, A., Havlin, S.: Extreme value statistics in records with long-term persistence. *Physical Review E* **73**(1), 016130 (2006). DOI: [10.1103/PhysRevE.73.016130](https://doi.org/10.1103/PhysRevE.73.016130)
- [47] Eichner, J.F., Kantelhardt, J.W., Bunde, A., Havlin, S.: The Statistics of Return Intervals, Maxima, and Centennial Events Under the Influence of Long-Term Correlations. In: Kropp, J., Schellnhuber, H.J. (eds.) *In Extremis*, pp. 2–43. Springer Berlin Heidelberg, Berlin, Heidelberg (2011). DOI: [10.1007/978-3-642-14863-7_1](https://doi.org/10.1007/978-3-642-14863-7_1)
- [48] Embrechts, P., Klüppelberg, C., Mikosch, T.: *Modelling Extremal Events*, vol. 97. Springer Berlin Heidelberg, Berlin, Heidelberg (1997). DOI: [10.1007/978-3-642-33483-2](https://doi.org/10.1007/978-3-642-33483-2)
- [49] European Climate Assessment and Dataset (ECA&D): <https://www.ecad.eu/dailydata/predefinedseries.php> (2018). Accessed: 2019-09-11
- [50] Feller, W.: The Asymptotic Distribution of the Range of Sums of Independent Random Variables. *The Annals of Mathematical Statistics* **22**(3), 427–432 (1951). DOI: [10.1214/aoms/1177729589](https://doi.org/10.1214/aoms/1177729589)
- [51] Feng, Y., Zhou, C.: Forecasting financial market activity using a semiparametric fractionally integrated Log-ACD. *International Journal of Forecasting* **31**(2), 349–363 (2015). DOI: [10.1016/j.ijforecast.2014.09.001](https://doi.org/10.1016/j.ijforecast.2014.09.001)
- [52] Fisher, R.A., Tippett, L.H.C.: Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society* **24**(2), 180–190 (1928). DOI: [10.1017/S0305004100015681](https://doi.org/10.1017/S0305004100015681)
- [53] Gallager, R.G.: *Discrete stochastic processes*. Springer, New York (1995). DOI: [10.1007/978-1-4615-2329-1](https://doi.org/10.1007/978-1-4615-2329-1)
- [54] Gasull, A., Jolis, M., Utzet, F.: On the norming constants for normal maxima. *Journal of Mathematical Analysis and Applications* **422**(1), 376–396 (2015). DOI: [10.1016/j.jmaa.2014.08.025](https://doi.org/10.1016/j.jmaa.2014.08.025)
- [55] Geisel, T., Thomae, S.: Anomalous Diffusion in Intermittent Chaotic Systems. *Physical Review Letters* **52**(22), 1936–1939 (1984). DOI: [10.1103/PhysRevLett.52.1936](https://doi.org/10.1103/PhysRevLett.52.1936)
- [56] Giraitis, L., Koul, H.L., Surgailis, D.: *Large Sample Inference for Long Memory Processes*. Imperial College Press (2012). DOI: [10.1142/p591](https://doi.org/10.1142/p591)

- [57] Giraitis, L., Surgailis, D.: Central limit theorem for the empirical process of a linear sequence with long memory. *Journal of Statistical Planning and Inference* **80**(1-2), 81–93 (1999). DOI: [10.1016/S0378-3758\(98\)00243-2](https://doi.org/10.1016/S0378-3758(98)00243-2)
- [58] Gnedenko, B.: Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire. *The Annals of Mathematics* **44**(3), 423 (1943). DOI: [10.2307/1968974](https://doi.org/10.2307/1968974)
- [59] Gomes, M.I., Haan, L.D.: Approximation by Penultimate Extreme Value Distributions. *Extremes* **2**(1), 71–85 (1999). DOI: [10.1023/A:1009920327187](https://doi.org/10.1023/A:1009920327187)
- [60] Granger, C.W.J., Joyeux, R.: An Introduction to Long-Memory Time Series Models and Fractional Differencing. *Journal of Time Series Analysis* **1**(1), 15–29 (1980). DOI: [10.1111/j.1467-9892.1980.tb00297.x](https://doi.org/10.1111/j.1467-9892.1980.tb00297.x)
- [61] Graves, T., Gramacy, R., Watkins, N., Franzke, C.: A Brief History of Long Memory: Hurst, Mandelbrot and the Road to ARFIMA, 1951 – 1980. *Entropy* **19**(9), 437 (2017). DOI: [10.3390/e19090437](https://doi.org/10.3390/e19090437)
- [62] Grossmann, S., Fujisaka, H.: Diffusion in discrete nonlinear dynamical systems. *Physical Review A* **26**(3), 1779–1782 (1982). DOI: [10.1103/PhysRevA.26.1779](https://doi.org/10.1103/PhysRevA.26.1779)
- [63] Guillot, G.: Approximation of Sahelian rainfall fields with meta-Gaussian random functions, Part 1: model definition and methodology. *Stochastic Environmental Research and Risk Assessment (SERRA)* **13**(1-2), 100–112 (1999). DOI: [10.1007/s004770050034](https://doi.org/10.1007/s004770050034)
- [64] Guillot, G., Lebel, T.: Approximation of Sahelian rainfall fields with meta-Gaussian random functions, Part 2: parameter estimation and comparison to data. *Stochastic Environmental Research and Risk Assessment (SERRA)* **13**(1-2), 113–130 (1999). DOI: [10.1007/s004770050035](https://doi.org/10.1007/s004770050035)
- [65] de Haan, L., Ferreira, A.: *Extreme Value Theory*. Springer Series in Operations Research and Financial Engineering. Springer New York, New York, NY (2006). DOI: [10.1007/0-387-34471-3](https://doi.org/10.1007/0-387-34471-3)
- [66] de Haan, L., Resnick, S.I.: Second-order regular variation and rates of convergence in extreme-value theory. *The Annals of Probability* **24**(1), 97–124 (1996). DOI: [10.1214/aop/1042644709](https://doi.org/10.1214/aop/1042644709)
- [67] Hall, P.: On the rate of convergence of normal extremes. *Journal of Applied Probability* **16**(2), 433–439 (1979). DOI: [10.2307/3212912](https://doi.org/10.2307/3212912)
- [68] Hall, P.: Estimating probabilities for normal extremes. *Advances in Applied Probability* **12**(2), 491–500 (1980). DOI: [10.2307/1426608](https://doi.org/10.2307/1426608)

- [69] Hall, W.J., Wellner, J.A.: The rate of convergence in law of the maximum of an exponential sample. *Statistica Neerlandica* **33**(3), 151–154 (1979). DOI: [10.1111/j.1467-9574.1979.tb00671.x](https://doi.org/10.1111/j.1467-9574.1979.tb00671.x)
- [70] Hamed, K.H.: Improved finite-sample Hurst exponent estimates using rescaled range analysis. *Water Resources Research* **43**(4), 1–9 (2007). DOI: [10.1029/2006WR005111](https://doi.org/10.1029/2006WR005111)
- [71] Hardstone, R., Poil, S.S., Schiavone, G., Jansen, R., Nikulin, V.V., Mansvelder, H.D., Linkenkaer-Hansen, K.: Detrended Fluctuation Analysis: A Scale-Free View on Neuronal Oscillations. *Frontiers in Physiology* **3**(November), 1–13 (2012). DOI: [10.3389/fphys.2012.00450](https://doi.org/10.3389/fphys.2012.00450)
- [72] Haslett, J., Raftery, A.E.: Space-Time Modelling with Long-Memory Dependence: Assessing Ireland’s Wind Power Resource. *Applied Statistics* **38**(1), 1 (1989). DOI: [10.2307/2347679](https://doi.org/10.2307/2347679)
- [73] Helgason, H., Pipiras, V., Abry, P.: Synthesis of multivariate stationary series with prescribed marginal distributions and covariance using circulant matrix embedding. *Signal Processing* **91**(8), 1741–1758 (2011). DOI: [10.1016/j.sigpro.2011.01.020](https://doi.org/10.1016/j.sigpro.2011.01.020)
- [74] Hernandez, A., Guenni, L., Sanso, B.: Extreme limit distribution of truncated models for daily rainfall Aracelis. *Environmetrics* **20**, 962–980 (2009). DOI: [10.1002/env.967](https://doi.org/10.1002/env.967)
- [75] Hill, B.M.: A Simple General Approach to Inference About the Tail of a Distribution. *The Annals of Statistics* **3**(5), 1163–1174 (1975). URL: <http://www.jstor.org/stable/2958370>
- [76] Höll, M., Kantz, H.: The relationship between the detrended fluctuation analysis and the autocorrelation function of a signal. *The European Physical Journal B* **88**(12), 327 (2015). DOI: [10.1140/epjb/e2015-60721-1](https://doi.org/10.1140/epjb/e2015-60721-1)
- [77] Höll, M., Kantz, H., Zhou, Y.: Detrended fluctuation analysis and the difference between external drifts and intrinsic diffusionlike nonstationarity. *Physical Review E* **94**(4), 042201 (2016). DOI: [10.1103/PhysRevE.94.042201](https://doi.org/10.1103/PhysRevE.94.042201)
- [78] Hosking, J.R.M.: Fractional differencing. *Biometrika* **68**(1), 165–176 (1981). DOI: [10.1093/biomet/68.1.165](https://doi.org/10.1093/biomet/68.1.165)
- [79] Hosking, J.R.M.: Modeling persistence in hydrological time series using fractional differencing. *Water Resources Research* **20**(12), 1898–1908 (1984). DOI: [10.1029/WR020i012p01898](https://doi.org/10.1029/WR020i012p01898)

- [80] Hosseini, A., Hosseini, R., Zare-Mehrjerdi, Y., Abooi, M.H.: Capturing the time-dependence in the precipitation process for weather risk assessment. *Stochastic Environmental Research and Risk Assessment* **31**(3), 609–627 (2017). DOI: [10.1007/s00477-016-1285-8](https://doi.org/10.1007/s00477-016-1285-8)
- [81] Hu, K., Ivanov, P.C., Chen, Z., Carpena, P., Eugene Stanley, H.: Effect of trends on detrended fluctuation analysis. *Physical Review E* **64**(1), 011114 (2001). DOI: [10.1103/PhysRevE.64.011114](https://doi.org/10.1103/PhysRevE.64.011114)
- [82] Hurst, H.E.: Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers* **116**(1), 770—799 (1951). DOI: [10.1061/TACEAT.0006518](https://doi.org/10.1061/TACEAT.0006518)
- [83] Hurst, H.E.: The Problem of Long-Term Storage in Reservoirs. *International Association of Scientific Hydrology. Bulletin* **1**(3), 13–27 (1956). DOI: [10.1080/02626665609493644](https://doi.org/10.1080/02626665609493644)
- [84] Ibragimov, I.A., Rozanov, Y.A.: *Gaussian Random Processes*, vol. 1. Springer, New York (1978). DOI: [10.1016/0377-2217\(81\)90062-X](https://doi.org/10.1016/0377-2217(81)90062-X)
- [85] Janicki, R., McElroy, T.S.: Hermite expansion and estimation of monotonic transformations of Gaussian data. *Journal of Nonparametric Statistics* **28**(1), 207–234 (2016). DOI: [10.1080/10485252.2016.1139880](https://doi.org/10.1080/10485252.2016.1139880)
- [86] Kai, S., Chun-qiong, L., Nan-shan, A., Xiao-hong, Z.: Using three methods to investigate time-scaling properties in air pollution indexes time series. *Nonlinear Analysis: Real World Applications* **9**(2), 693–707 (2008). DOI: [10.1016/j.nonrwa.2007.06.003](https://doi.org/10.1016/j.nonrwa.2007.06.003)
- [87] Kantelhardt, J.W., Koscielny-Bunde, E., Rego, H.H.A., Havlin, S., Bunde, A.: Detecting long-range correlations with detrended fluctuation analysis. *Physica A: Statistical Mechanics and its Applications* **295**(3-4), 441–454 (2001). DOI: [10.1016/S0378-4371\(01\)00144-3](https://doi.org/10.1016/S0378-4371(01)00144-3)
- [88] Kantelhardt, J.W., Koscielny-Bunde, E., Rybski, D., Braun, P., Bunde, A., Havlin, S.: Long-term persistence and multifractality of precipitation and river runoff records. *Journal of Geophysical Research* **111**(D1), D01106 (2006). DOI: [10.1029/2005JD005881](https://doi.org/10.1029/2005JD005881)
- [89] Kantelhardt, J.W., Rybski, D., Zschiegner, S.A., Braun, P., Koscielny-Bunde, E., Livina, V., Havlin, S., Bunde, A.: Multifractality of river runoff and precipitation: comparison of fluctuation analysis and wavelet methods. *Physica A: Statistical Mechanics and its Applications* **330**(1-2), 240–245 (2003). DOI: [10.1016/j.physa.2003.08.019](https://doi.org/10.1016/j.physa.2003.08.019)

- [90] Kasdin, N.: Discrete simulation of colored noise and stochastic processes and $1/f^\alpha$ power law noise generation. *Proceedings of the IEEE* **83**(5), 802–827 (1995). DOI: [10.1109/5.381848](https://doi.org/10.1109/5.381848)
- [91] Kavasseri, R.G., Seetharaman, K.: Day-ahead wind speed forecasting using f-ARIMA models. *Renewable Energy* **34**(5), 1388–1393 (2009). DOI: [10.1016/j.renene.2008.09.006](https://doi.org/10.1016/j.renene.2008.09.006)
- [92] Kelly, K.S., Krzysztofowicz, R.: A bivariate meta-Gaussian density for use in hydrology. *Stochastic Hydrology and Hydraulics* **11**(1), 17–31 (1997). DOI: [10.1007/BF02428423](https://doi.org/10.1007/BF02428423)
- [93] Kibble, W.F.: An extension of a theorem of Mehler’s on Hermite polynomials. *Mathematical Proceedings of the Cambridge Philosophical Society* **41**(01), 12–15 (1945). DOI: [10.1017/S0305004100022313](https://doi.org/10.1017/S0305004100022313)
- [94] Klafter, J., Sokolov, I.M.: *First Steps in Random Walks*. Oxford University Press (2011). DOI: [10.1093/acprof:oso/9780199234868.001.0001](https://doi.org/10.1093/acprof:oso/9780199234868.001.0001)
- [95] Klein Tank, A.M.G., Wijngaard, J.B., Können, G.P., Böhm, R., Demarée, G., Gocheva, A., Mileta, M., Pashiardis, S., Hejkrlik, L., Kern-Hansen, C., Heino, R., Bessemoulin, P., Müller-Westermeier, G., Tzanakou, M., Szalai, S., Pálsdóttir, T., Fitzgerald, D., Rubin, S., Capaldo, M., Maugeri, M., Leitass, A., Bukantis, A., Aberfeld, R., van Engelen, A.F.V., Forland, E., Mielus, M., Coelho, F., Mares, C., Razuvaev, V., Nieplova, E., Cegnar, T., Antonio López, J., Dahlström, B., Moberg, A., Kirchhofer, W., Ceylan, A., Pachaliuk, O., Alexander, L.V., Petrovic, P.: Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *International Journal of Climatology* **22**(12), 1441–1453 (2002). DOI: [10.1002/joc.773](https://doi.org/10.1002/joc.773)
- [96] Koutsoyiannis, D.: On the appropriateness of the gumbel distribution in modelling extreme rainfall. In: Brath, A., Montanari, A., Toth, E. (eds.) *Hydrological Risk: recent advances in peak river flow modelling, prediction and real-time forecasting. Assessment of the impacts of land-use and climate changes*, pp. 303–319. Bologna (2003). DOI: [10.13140/RG.2.1.3811.6080](https://doi.org/10.13140/RG.2.1.3811.6080)
- [97] Kriřtoufek, L.: Rescaled Range Analysis and Detrended Fluctuation Analysis: Finite Sample Properties and Confidence Intervals. *Czech Economic Review* **4**(3), 315–329 (2010). URL: http://ideas.repec.org/a/fau/aucoz/au2010_315.html
- [98] Kriřtoufek, L.: How are rescaled range analyses affected by different memory and distributional properties? A Monte Carlo study. *Physica A: Statistical Mechanics and its Applications* **391**(17), 4252–4260 (2012). DOI: [10.1016/j.physa.2012.04.005](https://doi.org/10.1016/j.physa.2012.04.005)

- [99] Leadbetter, M.: On a basis for ‘Peaks over Threshold’ modeling. *Statistics & Probability Letters* **12**(4), 357–362 (1991). DOI: [10.1016/0167-7152\(91\)90107-3](https://doi.org/10.1016/0167-7152(91)90107-3)
- [100] Leadbetter, M.R.: On extreme values in stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **28**(4), 289–303 (1974). DOI: [10.1007/BF00532947](https://doi.org/10.1007/BF00532947)
- [101] Leadbetter, M.R., Lindgren, G., Rootzén, H.: *Extremes and Related Properties of Random Sequences and Processes*. Springer Series in Statistics. Springer New York, New York, NY (1983). DOI: [10.1007/978-1-4612-5449-2](https://doi.org/10.1007/978-1-4612-5449-2)
- [102] Leadbetter, M.R., Rootzén, H.: *Extremal Theory for Stochastic Processes*. The Annals of Probability **16**(2), 431–478 (1988). DOI: [10.1214/aop/1176991767](https://doi.org/10.1214/aop/1176991767)
- [103] Leland, W.E., Taquq, M.S., Willinger, W., Wilson, D.V.: On the self-similar nature of Ethernet traffic. *ACM SIGCOMM Computer Communication Review* **23**(4), 183–193 (1993). DOI: [10.1145/167954.166255](https://doi.org/10.1145/167954.166255)
- [104] Li, Z., Brissette, F.P., Chen, J.: Finding the most appropriate precipitation probability distribution for stochastic weather generation and hydrological modelling in Nordic watersheds. *Hydrological Processes* **27**(25), 3718–3729 (2013). DOI: [10.1002/hyp.9499](https://doi.org/10.1002/hyp.9499)
- [105] Liao, X., Peng, Z.: Convergence rates of limit distribution of maxima of lognormal samples. *Journal of Mathematical Analysis and Applications* **395**(2), 643–653 (2012). DOI: [10.1016/j.jmaa.2012.05.077](https://doi.org/10.1016/j.jmaa.2012.05.077)
- [106] Liu, Y., Zhang, W., Shao, Y., Zhang, K.: A comparison of four precipitation distribution models used in daily stochastic models. *Advances in Atmospheric Sciences* **28**(4), 809–820 (2011). DOI: [10.1007/s00376-010-9180-6](https://doi.org/10.1007/s00376-010-9180-6)
- [107] Løvstletten, O.: Consistency of detrended fluctuation analysis. *Physical Review E* **96**(1), 012141 (2017). DOI: [10.1103/PhysRevE.96.012141](https://doi.org/10.1103/PhysRevE.96.012141)
- [108] Lucarini, V., Faranda, D., Freitas, A.C.M., Freitas, J.M., Holland, M., Kuna, T., Nicol, M., Todd, M., Vaienti, S.: *Extremes and Recurrence in Dynamical Systems*. John Wiley & Sons, Inc, Hoboken, NJ, USA (2016). DOI: [10.1002/9781118632321](https://doi.org/10.1002/9781118632321)
- [109] Makkonen, L.: Bringing Closure to the Plotting Position Controversy. *Communications in Statistics - Theory and Methods* **37**(3), 460–467 (2008). DOI: [10.1080/03610920701653094](https://doi.org/10.1080/03610920701653094)

- [110] Mandelbrot, B.B.: Statistical methodology for nonperiodic cycles: from the covariance to r/s analysis. In: Berg, S.V. (ed.) *Annals of Economic and Social Measurement*, vol. 1, pp. 259–290. National Bureau of Economic Research, Inc (1972). URL: <https://www.nber.org/system/files/chapters/c9433/c9433.pdf>
- [111] Mandelbrot, B.B.: Limit Theorem on the Self-Normalized Range for Weakly and Strongly Dependent Processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **31**, 271–285 (1975). DOI: [10.1007/BF00534968](https://doi.org/10.1007/BF00534968)
- [112] Mandelbrot, B.B., Van Ness, J.W.: Fractional Brownian Motions, Fractional Noises and Applications. *SIAM Review* **10**(4), 422–437 (1968). DOI: [10.1137/1010093](https://doi.org/10.1137/1010093)
- [113] Mandelbrot, B.B., Wallis, J.R.: Noah, Joseph, and Operational Hydrology. *Water Resources Research* **4**(5), 909–918 (1968). DOI: [10.1029/WR004i005p00909](https://doi.org/10.1029/WR004i005p00909)
- [114] Mandelbrot, B.B., Wallis, J.R.: Computer Experiments with Fractional Gaussian Noises: Part 2, Rescaled Ranges and Spectra. *Water Resources Research* **5**(1), 242–259 (1969). DOI: [10.1029/WR005i001p00242](https://doi.org/10.1029/WR005i001p00242)
- [115] Mandelbrot, B.B., Wallis, J.R.: Computer Experiments with Fractional Gaussian Noises: Part 3, Mathematical Appendix. *Water Resources Research* **5**(1), 260–267 (1969). DOI: [10.1029/WR005i001p00260](https://doi.org/10.1029/WR005i001p00260)
- [116] Mandelbrot, B.B., Wallis, J.R.: Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence. *Water Resources Research* **5**(5), 967–988 (1969). DOI: [10.1029/WR005i005p00967](https://doi.org/10.1029/WR005i005p00967)
- [117] Manneville, P., Pomeau, Y.: Different ways to turbulence in dissipative dynamical systems. *Physica D: Nonlinear Phenomena* **1**(2), 219–226 (1980). DOI: [10.1016/0167-2789\(80\)90013-5](https://doi.org/10.1016/0167-2789(80)90013-5)
- [118] Mason, D.M., Schuenemeyer, J.H.: A Modified Kolmogorov-Smirnov Test Sensitive to Tail Alternatives. *The Annals of Statistics* **11**(3), 933–946 (1983). DOI: [10.1214/aos/1176346259](https://doi.org/10.1214/aos/1176346259)
- [119] Massah, M., Kantz, H.: Confidence intervals for time averages in the presence of long-range correlations, a case study on Earth surface temperature anomalies. *Geophysical Research Letters* **43**(17), 9243–9249 (2016). DOI: [10.1002/2016GL069555](https://doi.org/10.1002/2016GL069555)
- [120] Massah, M., Nicol, M., Kantz, H.: Large-deviation probabilities for correlated Gaussian processes and intermittent dynamical systems. *Physical Review E* **97**(5), 052147 (2018). DOI: [10.1103/PhysRevE.97.052147](https://doi.org/10.1103/PhysRevE.97.052147)

- [121] Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J., Maycock, T., Waterfield, T., Yelekçi, O., Yu, R., Zhou, B. (eds.): IPCC, 2021: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press (2021). URL: <https://www.ipcc.ch/report/ar6/wg1/>. (in press)
- [122] Matsoukas, C., Islam, S., Rodriguez-Iturbe, I.: Detrended fluctuation analysis of rainfall and streamflow time series. *Journal of Geophysical Research: Atmospheres* **105**(D23), 29165–29172 (2000). DOI: [10.1029/2000JD900419](https://doi.org/10.1029/2000JD900419)
- [123] Mehler, F.G.: Ueber die Entwicklung einer Function von beliebig vielen Variablen nach Laplaceschen Functionen höherer Ordnung. *Journal fuer die reine und angewandte mathematik* **66**, 161–176 (1866). DOI: [10.1515/crll.1866.66.161](https://doi.org/10.1515/crll.1866.66.161)
- [124] Met Office Hadley Centre: <https://www.metoffice.gov.uk/hadobs/> (2018). Accessed: 2018-09-26
- [125] Meyer, P., Kantz, H.: Infinite invariant densities due to intermittency in a nonlinear oscillator. *Physical Review E* **96**(2), 022217 (2017). DOI: [10.1103/PhysRevE.96.022217](https://doi.org/10.1103/PhysRevE.96.022217)
- [126] Meyer, P.G., Anvari, M., Kantz, H.: Identifying characteristic time scales in power grid frequency fluctuations with DFA. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **30**(1), 013130 (2020). DOI: [10.1063/1.5123778](https://doi.org/10.1063/1.5123778)
- [127] Meyer, P.G., Kantz, H.: A simple decomposition of European temperature variability capturing the variance from days to a decade. *Climate Dynamics* (0123456789) (2019). DOI: [10.1007/s00382-019-04965-0](https://doi.org/10.1007/s00382-019-04965-0)
- [128] Meyer, P.G., Kantz, H.: Inferring characteristic timescales from the effect of autoregressive dynamics on detrended fluctuation analysis. *New Journal of Physics* **21**(3), 033022 (2019). DOI: [10.1088/1367-2630/ab0a8a](https://doi.org/10.1088/1367-2630/ab0a8a)
- [129] Mielniczuk, J., Wojdyło, P.: Estimation of Hurst exponent revisited. *Computational Statistics & Data Analysis* **51**(9), 4510–4525 (2007). DOI: [10.1016/j.csda.2006.07.033](https://doi.org/10.1016/j.csda.2006.07.033)
- [130] Mohammed, S.E.A.: Stochastic Differential Systems With Memory: Theory, Examples and Applications. In: Decreusefond, L., Øksendal, B., Gjerde, J., Üstünel, A.S. (eds.) *Stochastic Analysis and Related Topics VI*, vol. 53, pp. 1–77. Birkhäuser Boston, Boston, MA (1998). DOI: [10.1007/978-1-4612-2022-0_1](https://doi.org/10.1007/978-1-4612-2022-0_1)

- [131] Moloney, N.R., Faranda, D., Sato, Y.: An overview of the extremal index. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **29**(2), 022101 (2019). DOI: [10.1063/1.5079656](https://doi.org/10.1063/1.5079656)
- [132] Montanari, A., Rosso, R., Taqqu, M.S.: Some long-run properties of rainfall records in Italy. *Journal of Geophysical Research: Atmospheres* **101**(D23), 29431–29438 (1996). DOI: [10.1029/96JD02512](https://doi.org/10.1029/96JD02512)
- [133] Movahed, M.S., Jafari, G.R., Ghasemi, F., Rahvar, S., Tabar, M.R.R.: Multifractal detrended fluctuation analysis of sunspot time series. *Journal of Statistical Mechanics: Theory and Experiment* **2006**(02), P02003–P02003 (2006). DOI: [10.1088/1742-5468/2006/02/P02003](https://doi.org/10.1088/1742-5468/2006/02/P02003)
- [134] Müller, P.: Climex - extreme value analysis of climatological time series. <https://gitlab.com/theGreatWhiteShark/climex> (2018). open source software, web application access via <http://climex.pks.mpg.de/climex/>
- [135] Nerantzaki, S.D., Papalexiou, S.M.: Tails of extremes: Advancing a graphical method and harnessing big data to assess precipitation extremes. *Advances in Water Resources* **134**, 103448 (2019). DOI: [10.1016/j.advwatres.2019.103448](https://doi.org/10.1016/j.advwatres.2019.103448)
- [136] Oriani, F., Mehrotra, R., Mariethoz, G., Straubhaar, J., Sharma, A., Renard, P.: Simulating rainfall time-series: how to account for statistical variability at multiple scales? *Stochastic Environmental Research and Risk Assessment* **32**(2), 321–340 (2018). DOI: [10.1007/s00477-017-1414-z](https://doi.org/10.1007/s00477-017-1414-z)
- [137] Palma, W.: *Long-Memory Time Series*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, USA (2007). DOI: [10.1002/9780470131466](https://doi.org/10.1002/9780470131466)
- [138] Papalexiou, S.M.: Unified theory for stochastic modelling of hydroclimatic processes: Preserving marginal distributions, correlation structures, and intermittency. *Advances in Water Resources* **115**, 234–252 (2018). DOI: [10.1016/j.advwatres.2018.02.013](https://doi.org/10.1016/j.advwatres.2018.02.013)
- [139] Papalexiou, S.M.: Unified theory for stochastic modelling of hydroclimatic processes: Preserving marginal distributions, correlation structures, and intermittency. *Advances in Water Resources* **115**, 234–252 (2018). DOI: [10.1016/j.advwatres.2018.02.013](https://doi.org/10.1016/j.advwatres.2018.02.013)
- [140] Papalexiou, S.M., Koutsoyiannis, D.: Entropy based derivation of probability distributions : A case study to daily rainfall. *Advances in Water Resources* **45**, 51–57 (2012). DOI: [10.1016/j.advwatres.2011.11.007](https://doi.org/10.1016/j.advwatres.2011.11.007)

- [141] Papalexiou, S.M., Koutsoyiannis, D.: Battle of extreme value distributions: A global survey on extreme daily rainfall. *Water Resources Research* **49**(1), 187–201 (2013). DOI: [10.1029/2012WR012557](https://doi.org/10.1029/2012WR012557)
- [142] Papalexiou, S.M., Koutsoyiannis, D.: A global survey on the seasonal variation of the marginal distribution of daily precipitation. *Advances in Water Resources* **94**, 131–145 (2016). DOI: [10.1016/j.advwatres.2016.05.005](https://doi.org/10.1016/j.advwatres.2016.05.005)
- [143] Papalexiou, S.M., Koutsoyiannis, D., Makropoulos, C.: How extreme is extreme? An assessment of daily rainfall distribution tails. *Hydrology and Earth System Sciences* **17**(2), 851–862 (2013). DOI: [10.5194/hess-17-851-2013](https://doi.org/10.5194/hess-17-851-2013)
- [144] Papalexiou, S.M., Markonis, Y., Lombardo, F., AghaKouchak, A., Fofoula-Georgiou, E.: Precise Temporal Disaggregation Preserving Marginals and Correlations (DiPMaC) for Stationary and Nonstationary Processes. *Water Resources Research* **54**(10), 7435–7458 (2018). DOI: [10.1029/2018WR022726](https://doi.org/10.1029/2018WR022726)
- [145] Papalexiou, S.M., Serinaldi, F.: Random Fields Simplified: Preserving Marginal Distributions, Correlations, and Intermittency, With Applications From Rainfall to Humidity. *Water Resources Research* **56**(2) (2020). DOI: [10.1029/2019WR026331](https://doi.org/10.1029/2019WR026331)
- [146] Pelletier, J.D., Turcotte, D.L.: Long-range persistence in climatological and hydrological time series: analysis, modeling and application to drought hazard assessment. *Journal of Hydrology* **203**(1-4), 198–208 (1997). DOI: [10.1016/S0022-1694\(97\)00102-9](https://doi.org/10.1016/S0022-1694(97)00102-9)
- [147] Peng, C.K., Buldyrev, S.V., Havlin, S., Simons, M., Stanley, H.E., Goldberger, A.L.: Mosaic organization of DNA nucleotides. *Physical Review E* **49**(2), 1685–1689 (1994). DOI: [10.1103/PhysRevE.49.1685](https://doi.org/10.1103/PhysRevE.49.1685)
- [148] Percival, D.B.: Three Curious Properties of the Sample Variance and Autocovariance for Stationary Processes with Unknown Mean. *The American Statistician* **47**(4), 274–276 (1993). DOI: [10.2307/2685286](https://doi.org/10.2307/2685286)
- [149] Peters, O., Hertlein, C., Christensen, K.: A Complexity View of Rainfall. *Physical Review Letters* **88**(1), 018701 (2001). DOI: [10.1103/PhysRevLett.88.018701](https://doi.org/10.1103/PhysRevLett.88.018701)
- [150] Pickands III, J.: Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics* **3**(1), 119–131 (1975). DOI: [10.1214/aos/1176343003](https://doi.org/10.1214/aos/1176343003)
- [151] Pipiras, V., Taqqu, M.S.: Long-Range Dependence and Self-Similarity. Cambridge University Press, Cambridge (2017). DOI: [10.1017/CBO9781139600347](https://doi.org/10.1017/CBO9781139600347)

- [152] Polotzek, K., Sato, Y., Klages, R., Moloney, N.: Anomalous diffusion in random dynamical systems. In: Summer School, London Mathematical Laboratory (2016). URL: https://lml.org.uk/wp-content/uploads/2016/08/2016_SumSchoolPresentation_Polotzek.pdf
- [153] Resnick, S.I.: Extreme Values, Regular Variation and Point Processes. Springer Series in Operations Research and Financial Engineering. Springer New York, New York, NY (1987). DOI: [10.1007/978-0-387-75953-1](https://doi.org/10.1007/978-0-387-75953-1)
- [154] Richardson, C.W.: Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research* **17**(1), 182–190 (1981). DOI: [10.1029/WR017i001p00182](https://doi.org/10.1029/WR017i001p00182)
- [155] Rootzén, H.: The rate of convergence of extremes of stationary normal sequences. *Advances in Applied Probability* **15**(1), 54–80 (1983). DOI: [10.2307/1426982](https://doi.org/10.2307/1426982)
- [156] Rootzén, H.: Attainable rates of convergence of maxima. *Statistics & Probability Letters* **2**(4), 219–221 (1984). DOI: [10.1016/0167-7152\(84\)90019-1](https://doi.org/10.1016/0167-7152(84)90019-1)
- [157] Rust, H.W.: The effect of long-range dependence on modelling extremes with the generalised extreme value distribution. *The European Physical Journal Special Topics* **174**(1), 91–97 (2009). DOI: [10.1140/epjst/e2009-01092-8](https://doi.org/10.1140/epjst/e2009-01092-8)
- [158] Rust, H.W., Kallache, M., Schellnhuber, H.J., Kropp, J.P.: Confidence Intervals for Flood Return Level Estimates Assuming Long-Range Dependence. In: Kropp, J., Schellnhuber, H.J. (eds.) *In Extremis*, pp. 60–88. Springer Berlin Heidelberg, Berlin, Heidelberg (2011). DOI: [10.1007/978-3-642-14863-7_3](https://doi.org/10.1007/978-3-642-14863-7_3)
- [159] Rybski, D., Bunde, A., Havlin, S., Kantelhardt, J.W., Koscielny-Bunde, E.: Detrended Fluctuation Studies of Long-Term Persistence and Multifractality of Precipitation and River Runoff Records. In: Kropp, J., Schellnhuber, H.J. (eds.) *In Extremis*, pp. 216–248. Springer Berlin Heidelberg, Berlin, Heidelberg (2011). DOI: [10.1007/978-3-642-14863-7_11](https://doi.org/10.1007/978-3-642-14863-7_11)
- [160] Sakia, R.M.: The Box-Cox Transformation Technique: A Review. *The Statistician* **41**(2), 169 (1992). DOI: [10.2307/2348250](https://doi.org/10.2307/2348250)
- [161] Samorodnitsky, G.: Stochastic Processes and Long Range Dependence. Springer Series in Operations Research and Financial Engineering. Springer International Publishing, Cham (2016). DOI: [10.1007/978-3-319-45575-4](https://doi.org/10.1007/978-3-319-45575-4)
- [162] Sánchez Granero, M., Trinidad Segovia, J., García Pérez, J.: Some comments on Hurst exponent and the long memory processes on capital markets. *Physica A: Statistical Mechanics and its Applications* **387**(22), 5543–5551 (2008). DOI: [10.1016/j.physa.2008.05.053](https://doi.org/10.1016/j.physa.2008.05.053)

- [163] Sang, H., Sang, Y.: Memory properties of transformations of linear processes. *Statistical Inference for Stochastic Processes* **20**(1), 79–103 (2017). DOI: [10.1007/s11203-016-9134-4](https://doi.org/10.1007/s11203-016-9134-4)
- [164] Sanso, B., Guenni, L.: A stochastic model for tropical rainfall at a single location. *Journal of Hydrology* **214**(1-4), 64–73 (1999). DOI: [10.1016/S0022-1694\(98\)00241-8](https://doi.org/10.1016/S0022-1694(98)00241-8)
- [165] Sato, Y., Klages, R.: Anomalous Diffusion in Random Dynamical Systems. *Physical Review Letters* **122**(17), 174101 (2019). DOI: [10.1103/PhysRevLett.122.174101](https://doi.org/10.1103/PhysRevLett.122.174101)
- [166] Scherrer, A., Larrieu, N., Owezarski, P., Borgnat, P., Abry, P.: Non-Gaussian and Long Memory Statistical Characterizations for Internet Traffic with Anomalies. *IEEE Transactions on Dependable and Secure Computing* **4**(1), 56–70 (2007). DOI: [10.1109/TDSC.2007.12](https://doi.org/10.1109/TDSC.2007.12)
- [167] Serinaldi, F.: Use and misuse of some Hurst parameter estimators applied to stationary and non-stationary financial time series. *Physica A: Statistical Mechanics and its Applications* **389**(14), 2770–2781 (2010). DOI: [10.1016/j.physa.2010.02.044](https://doi.org/10.1016/j.physa.2010.02.044)
- [168] Serinaldi, F., Kilsby, C.G.: Rainfall extremes: Toward reconciliation after the battle of distributions. *Water Resources Research* **50**(1), 336–352 (2014). DOI: [10.1002/2013WR014211](https://doi.org/10.1002/2013WR014211)
- [169] Serinaldi, F., Lombardo, F.: BetaBit: A fast generator of autocorrelated binary processes for geophysical research. *EPL (Europhysics Letters)* **118**(3), 30007 (2017). DOI: [10.1209/0295-5075/118/30007](https://doi.org/10.1209/0295-5075/118/30007)
- [170] Shelton, D.P.: Long-range orientation correlation in water. *The Journal of Chemical Physics* **141**(22), 224506 (2014). DOI: [10.1063/1.4903541](https://doi.org/10.1063/1.4903541)
- [171] Sigrist, F., Künsch, H.R., Stahel, W.A.: A dynamic nonstationary spatio-temporal model for short term prediction of precipitation. *The Annals of Applied Statistics* **6**(4), 1452–1477 (2012). DOI: [10.1214/12-AOAS564](https://doi.org/10.1214/12-AOAS564)
- [172] Sikora, G., Höll, M., Gajda, J., Kantz, H., Chechkin, A., Wyłomańska, A.: Probabilistic properties of detrended fluctuation analysis for Gaussian processes. *Physical Review E* **101**(3), 032114 (2020). DOI: [10.1103/PhysRevE.101.032114](https://doi.org/10.1103/PhysRevE.101.032114)
- [173] Smith, R.L.: Uniform rates of convergence in extreme-value theory. *Advances in Applied Probability* **14**(3), 600–622 (1982). DOI: [10.2307/1426676](https://doi.org/10.2307/1426676)
- [174] von Storch, H., Zwiers, F.W.: *Statistical Analysis in Climate Research*. Cambridge University Press (1984). DOI: [10.1017/CBO9780511612336](https://doi.org/10.1017/CBO9780511612336)

- [175] Taqqu, M.S.: The Rosenblatt Process. In: Davis, R.A., Lii, K.S., Politis, D.N. (eds.) *Selected Works of Murray Rosenblatt*, pp. 29–45. Springer New York, New York, NY (2011). DOI: [10.1007/978-1-4419-8339-8_6](https://doi.org/10.1007/978-1-4419-8339-8_6)
- [176] Taqqu, M.S., Teverovsky, V., Willinger, W.: Estimators for Long-Range Dependence: an Empirical Study. *Fractals* **03**(04), 785–798 (1995). DOI: [10.1142/S0218348X95000692](https://doi.org/10.1142/S0218348X95000692)
- [177] Thiébaux, H.J., Zwiers, F.W.: The Interpretation and Estimation of Effective Sample Size. *Journal of Climate and Applied Meteorology* **23**(5), 800–811 (1984). DOI: [10.1175/1520-0450\(1984\)023<0800:TIAEOE>2.0.CO;2](https://doi.org/10.1175/1520-0450(1984)023<0800:TIAEOE>2.0.CO;2)
- [178] Tough, R.J.A., Ward, K.D.: The correlation properties of gamma and other non-Gaussian processes generated by memoryless nonlinear transformation. *Journal of Physics D: Applied Physics* **32**(23), 3075–3084 (1999). DOI: [10.1088/0022-3727/32/23/314](https://doi.org/10.1088/0022-3727/32/23/314)
- [179] Trenberth, K.E.: Some Effects of Finite Sample Size and Persistence on Meteorological Statistics Part I Autocorrelations. *Monthly Weather Review* **112**(12), 2359–2368 (1984). DOI: [10.1175/1520-0493\(1984\)112<2359:SEOFSS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1984)112<2359:SEOFSS>2.0.CO;2)
- [180] Tschernig, R.: *Wechselkurse, Unsicherheit und Long Memory*. Studies in Contemporary Economics. Physica-Verlag HD, Heidelberg (1994). DOI: [10.1007/978-3-642-95912-7](https://doi.org/10.1007/978-3-642-95912-7)
- [181] Veillette, M.S., Taqqu, M.S.: Properties and numerical evaluation of the Rosenblatt distribution. *Bernoulli* **19**(3), 982–1005 (2013). DOI: [10.3150/12-BEJ421](https://doi.org/10.3150/12-BEJ421)
- [182] Vivo, P.: Large deviations of the maximum of independent and identically distributed random variables. *European Journal of Physics* **36**(5), 055037 (2015). DOI: [10.1088/0143-0807/36/5/055037](https://doi.org/10.1088/0143-0807/36/5/055037)
- [183] Weron, R.: Measuring long-range dependence in electricity prices. In: *Empirical Science of Financial Fluctuations*, pp. 110–119. Springer Japan, Tokyo (2002). DOI: [10.1007/978-4-431-66993-7_12](https://doi.org/10.1007/978-4-431-66993-7_12)
- [184] Wilks, D.S., Wilby, R.L.: The weather generation game: a review of stochastic weather models. *Progress in Physical Geography* **23**(3), 329–357 (1999). DOI: [10.1191/030913399666525256](https://doi.org/10.1191/030913399666525256)
- [185] Wilson, P.S., Toumi, R.: A fundamental probability distribution for heavy rainfall. *Geophysical Research Letters* **32**(14), 1–4 (2005). DOI: [10.1029/2005GL022465](https://doi.org/10.1029/2005GL022465)

- [186] Yalcin, G.C., Rabassa, P., Beck, C.: Extreme event statistics of daily rainfall: dynamical systems approach. *Journal of Physics A: Mathematical and Theoretical* **49**(15), 154001 (2016). DOI: [10.1088/1751-8113/49/15/154001](https://doi.org/10.1088/1751-8113/49/15/154001)
- [187] Zarfaty, L., Barkai, E., Kessler, D.A.: Accurately approximating extreme value statistics (2020). URL: <http://arxiv.org/abs/2006.13677>
- [188] Zarfaty, L., Barkai, E., Kessler, D.A.: Accurately approximating extreme value statistics. *Journal of Physics A: Mathematical and Theoretical* **54**(31), 315205 (2021). DOI: [10.1088/1751-8121/abf767](https://doi.org/10.1088/1751-8121/abf767)
- [189] Zumofen, G., Klafter, J.: Scale-invariant motion in intermittent chaotic systems. *Physical Review E* **47**(2), 851–863 (1993). DOI: [10.1103/PhysRevE.47.851](https://doi.org/10.1103/PhysRevE.47.851)

Confirmation

I herewith declare that I have produced this paper without the prohibited assistance of third parties and without making use of aids other than those specified; notions taken over directly or indirectly from other sources have been identified as such. This paper has not previously been presented in identical or similar form to any other German or foreign examination board.

Date

Signature

Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Datum

Unterschrift