Beyond popular distributions in hydroclimatology and stochastic simulation reproducing the behaviour of extremes

Simon Michael Papalexiou
sm.papalexiou@usak.ca
Global Institute for Water Security
University of Saskatchewan, Canada

A Taste of (Summer School) Extremes in Water Science

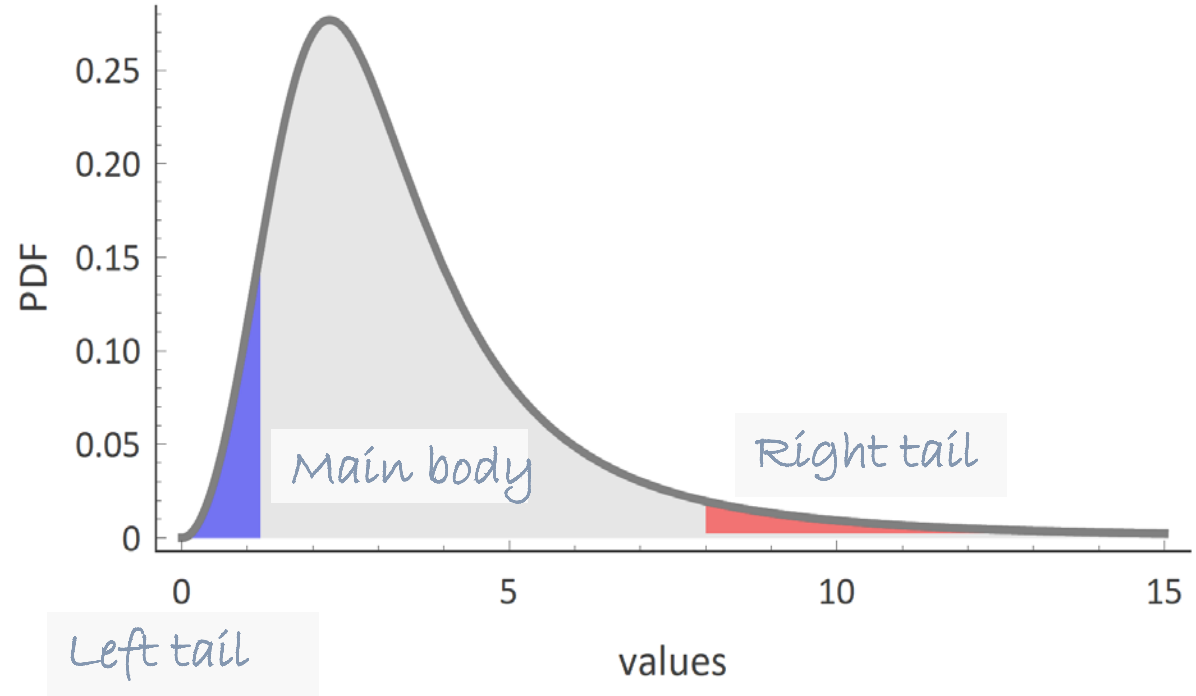
Online even | 4-7pm, 13 July 2020

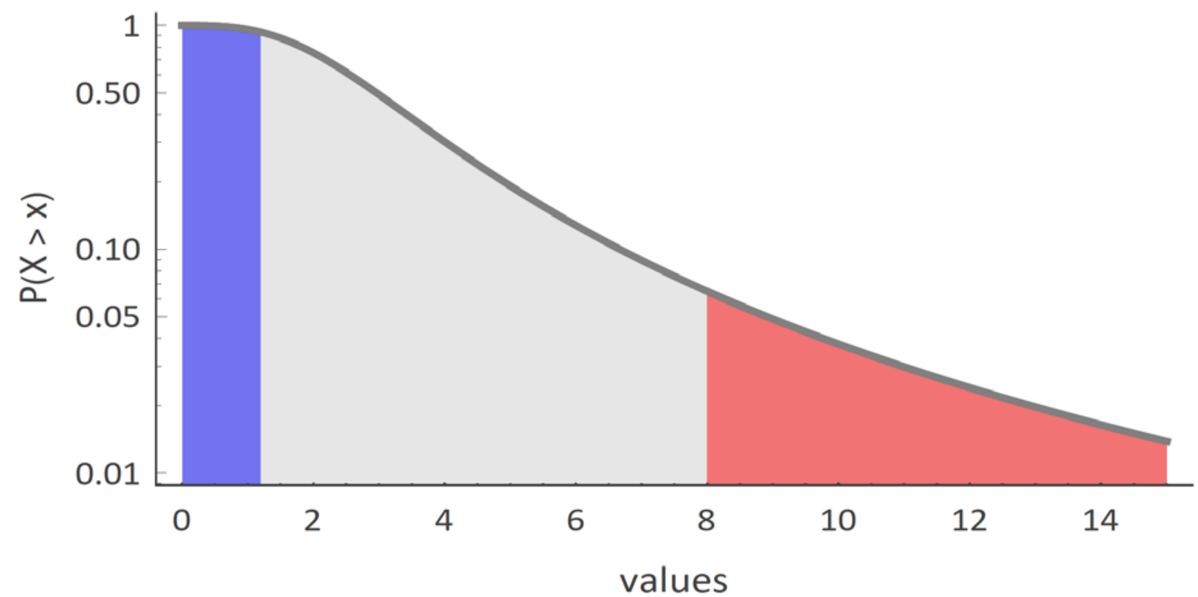
Organized by:
WARREDOC
Università per Stranieri di Perugia

Scientific Commitee:
Roberto Deidda, Salvatore Grimaldi, Marco Marani, Elena Volpi
Organizing Commitee:
Fabio Castelli, Aldo Fiori, Fernando Nardi, Valerio Leonardo Noto, Maria Cristina Rulli

Part I
Probability Distributions

Basic Anatomy I





How the tail approaches to zero defines the behavior of extremes

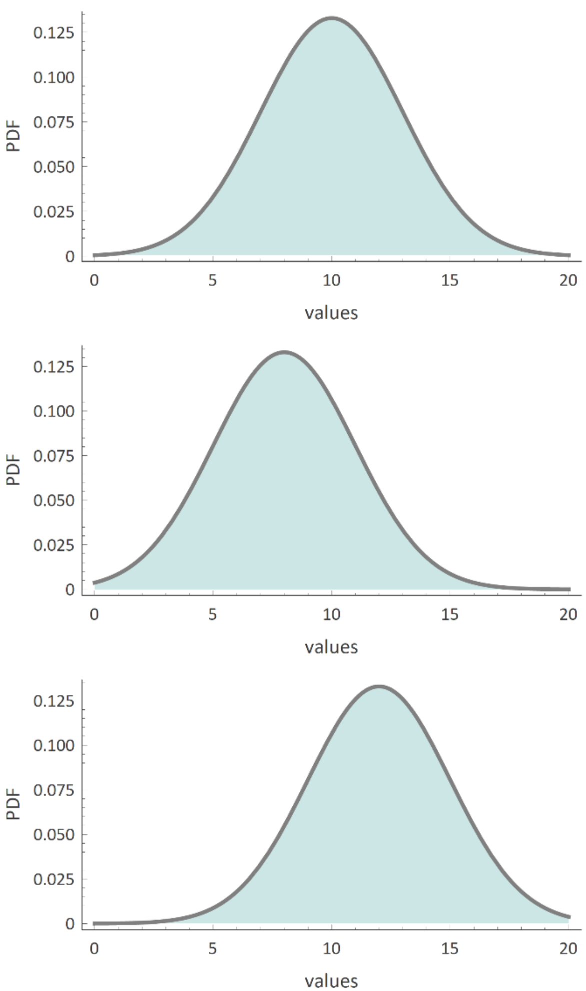
For example it could approach as

- $P(X > x) \sim x^{-1/\gamma}$
- $P(X > x) \sim \exp(-x^{\gamma})$

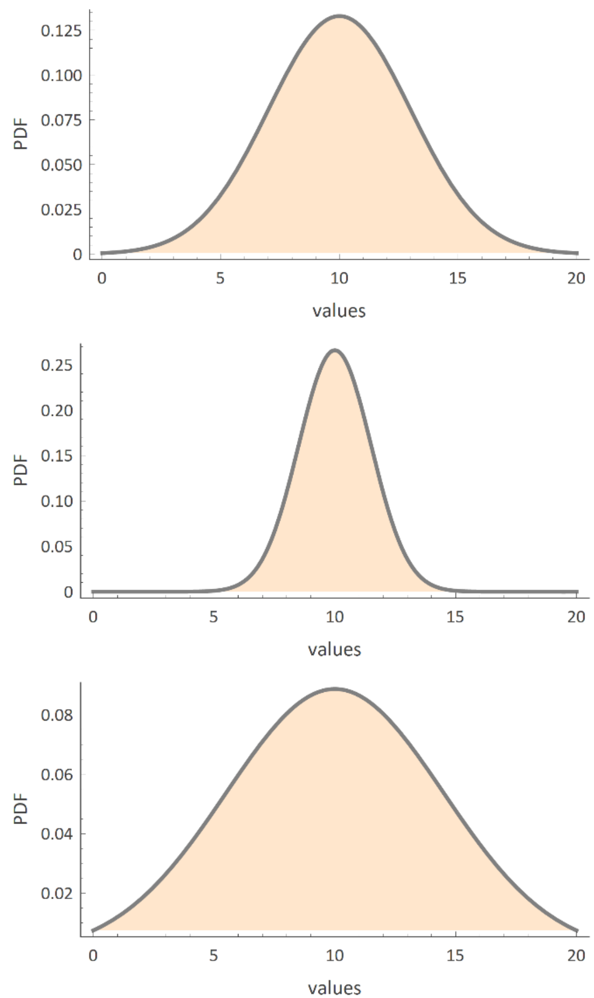
- Classifying tails in general is a difficult task. There are too many classes.

Basic Anatomy II

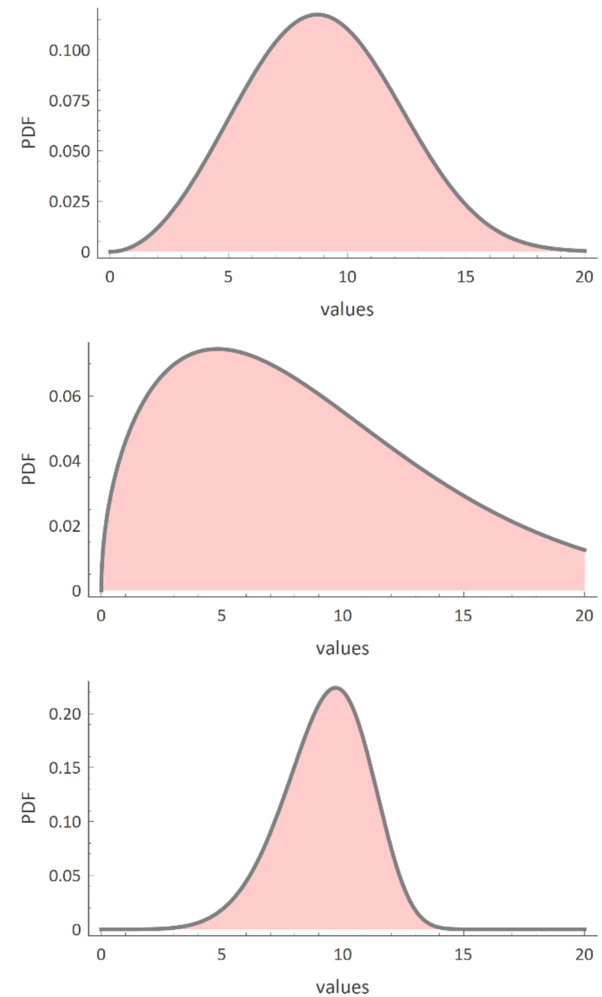
Location a



Scale β

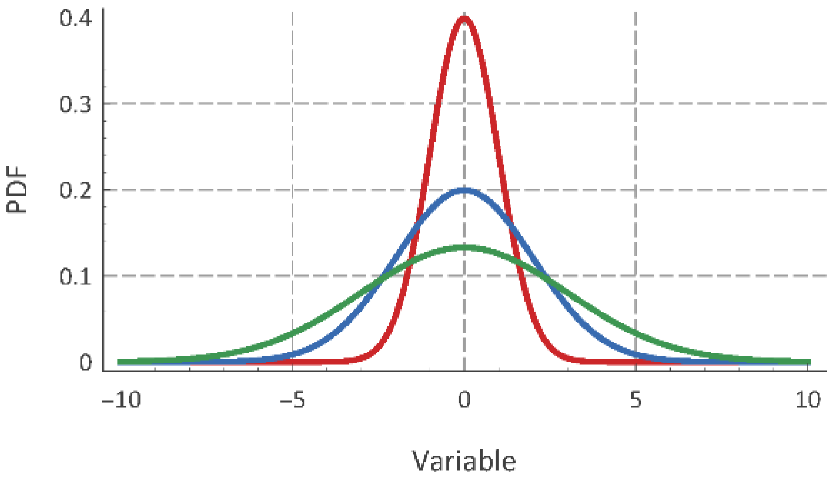


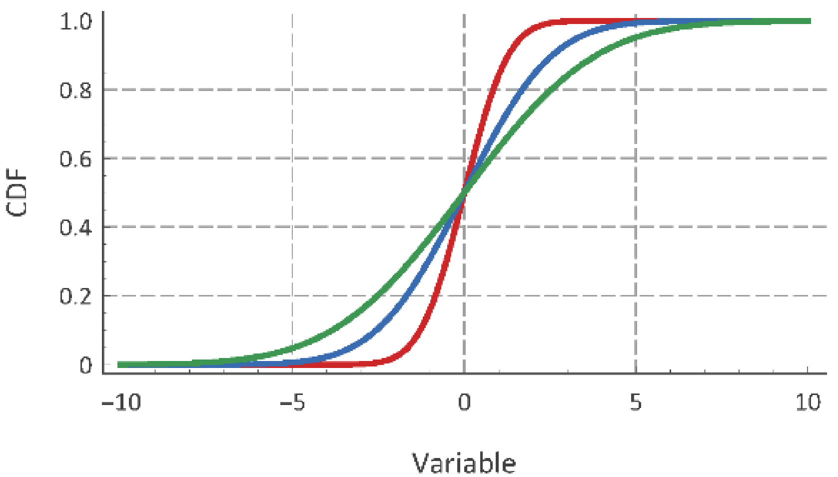
Shape γ



Normal distribution

Type	Normal distribution equations
Range	$-\infty < x < \infty$
pdf	$f_{\mathcal{N}}(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$
cdf	$F_{\mathcal{N}}(x) = \frac{1}{2}\mathrm{erfc}\left(\frac{\mu - x}{\sqrt{2}\sigma}\right)$
Quantile	$Q_{\mathcal{N}}(u) = \mu - \sqrt{2}\sigma\mathrm{erfc}^{-1}(2u)$
Mean	$\mu_{\mathcal{N}} = \mu$
SD	$\sigma_{\mathcal{N}} = \sigma$
Skew	$\mathrm{CS}_{\mathcal{N}} = 0$
Kurtosis	$\mathrm{CK}_{\mathcal{N}} = 3$

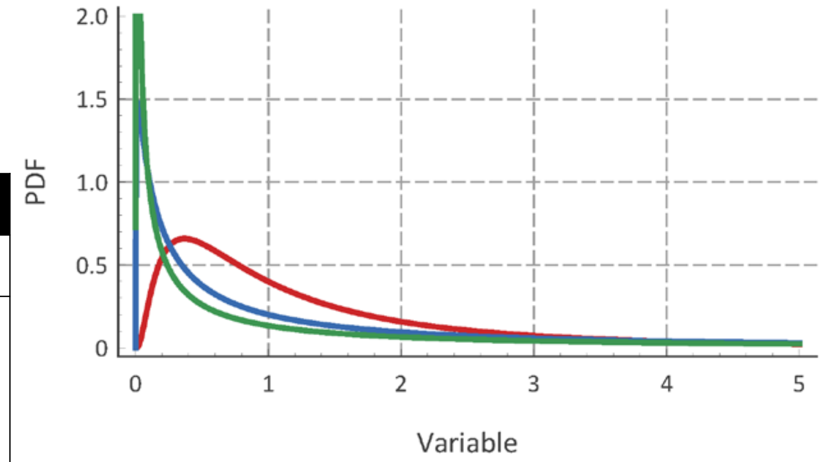


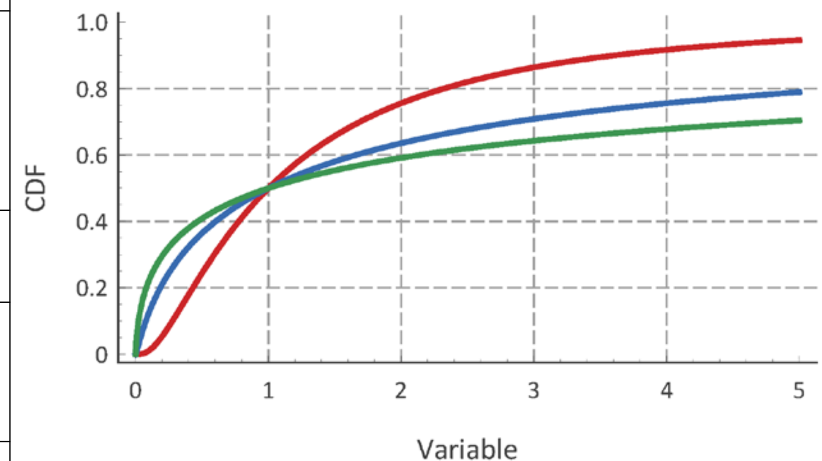


- The most famous distribution in general.
- It has two parameter: the location μ and scale σ.
- It is symmetric and thus not ideal for hydroclimatic variables.
- **No tail control**

Log-Normal distribution

Type	Lognormal distribution equations
Range	$x \geq 0$
pdf	$f_{\mathcal{LN}}(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)$
cdf	$F_{\mathcal{LN}}(x) = \frac{1}{2} \operatorname{erfc}\left(\frac{\mu - \ln x}{\sqrt{2}\sigma}\right)$
Quantile	$Q_{\mathcal{LN}}(u) = \exp\left(\mu - \sqrt{2}\sigma \operatorname{erfc}^{-1}(2u)\right)$
Mean	$\mu_{\mathcal{LN}} = \exp\left(\mu + \frac{\sigma^2}{2}\right)$
SD	$\sigma_{\mathcal{LN}} = \sqrt{(\exp \sigma^2 - 1)\exp(2\mu + \sigma^2)}$
Skew	$\mathrm{CS}_{\mathcal{LN}} = \sqrt{\exp \sigma^2 - 1}\left(\exp \sigma^2 + 2\right)$
Kurtosis	$\mathrm{CK}_{\mathcal{LN}} = 3\exp(2\sigma^2) + 2\exp(3\sigma^2) + \exp(4\sigma^2) - 3$

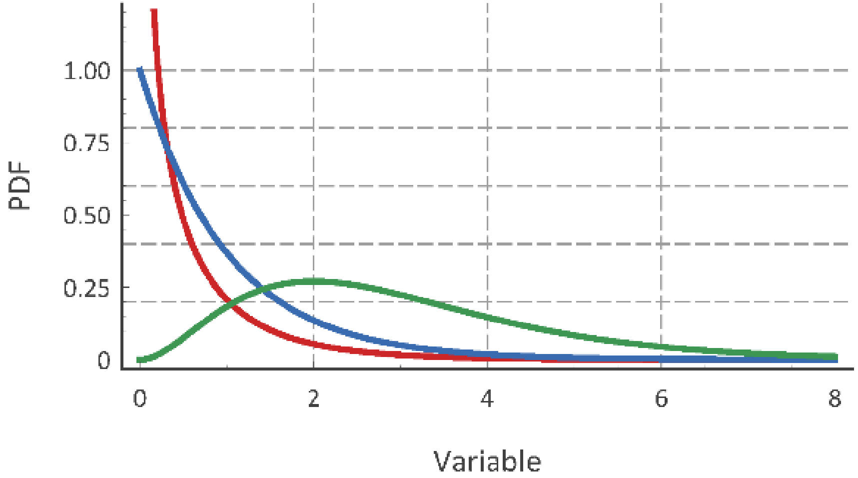


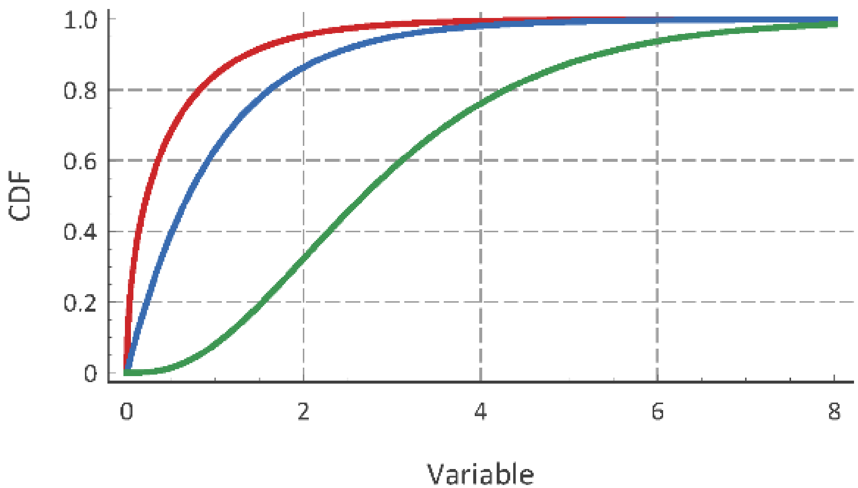


- A popular distribution in hydrology.
- Emerges as a transformation from the normal, i.e., $Y = \exp X$
- It has two parameter: the scale μ and shape σ.
- **One shape controlling both tails.**

Gamma distribution

Type	Gamma distribution equations
Range	$x \geq 0$
pdf	$f_{\mathcal{G}}(x) = \frac{\beta^{-\gamma}}{\Gamma(\gamma)} x^{-1+\gamma} \exp\left(-\frac{x}{\beta}\right)$
cdf	$F_{\mathcal{G}}(x) = \int_0^x f_{\mathcal{G}}(t)\,\mathrm{d}t$
Quantile	$Q_{\mathcal{G}}(u) = F_{\mathcal{G}}^{-1}(u)$
Mean	$\mu_{\mathcal{G}} = \beta\gamma$
SD	$\sigma_{\mathcal{G}} = \beta\sqrt{\gamma}$
Skew	$\mathrm{CS}_{\mathcal{G}} = 2/\sqrt{\gamma}$
Kurtosis	$\mathrm{CK}_{\mathcal{G}} = 3 + 6/\gamma$





- It has two parameters: the scale β and shape γ
- Commonly used to model precipitation.
- **The shape control mainly the left tail as the exponential part dominates quickly**

So, what's the problem?

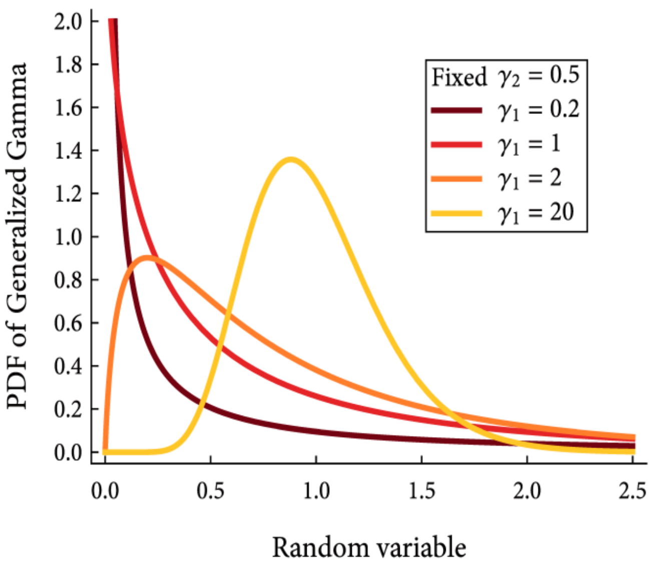
- We can add a location parameter to any distribution, e.g., three-parameter Gamma, Lognormal, Pareto, etc.
- This will increase flexibility but then the lower bound will **not be zero** and it could be **even negative**.
- One shape parameter either will control both tails or mainly one of them.
- *Could we capture the behavior of extremes in hydroclimatic variables with limited flexibility as well as the behavior of the left tail, and the shape of main body*?

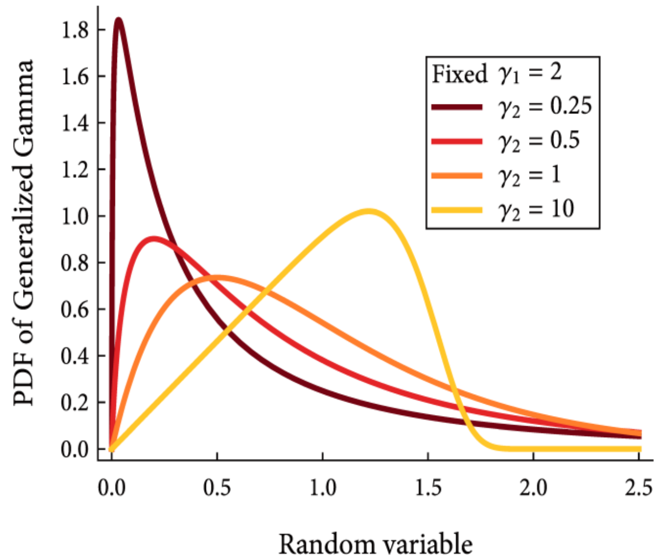
Some more cool distributions

- Most RV's in nature are defined in the positive axis
- For consistency at least we should choose distribution function defined in $(0, \infty)$
- Flexible distribution should have a scale parameters β and shape parameters γ_1 and γ_2, e.g.,
- $f_{\mathcal{GG}}(x; \beta, \gamma_1, \gamma_2) = \frac{\gamma_2}{\beta\, \Gamma(\gamma_1/\gamma_2)} \left(\frac{x}{\beta}\right)^{\gamma_1 - 1} \exp\left(-\left(\frac{x}{\beta}\right)^{\gamma_2}\right)$
- $F_{\mathcal{Br}\mathrm{XII}}(x; \beta, \gamma_1, \gamma_2) = 1 - \left(1 + \gamma_2 \left(\frac{x}{\beta}\right)^{\gamma_1}\right)^{-\frac{1}{\gamma_1 \gamma_2}}$
- $F_{\mathcal{Br}\mathrm{III}}(x; \beta, \gamma_1, \gamma_2) = \left(1 + \frac{1}{\gamma_1} \left(\frac{x}{\beta}\right)^{-\frac{1}{\gamma_2}}\right)^{-\gamma_1 \gamma_2}$

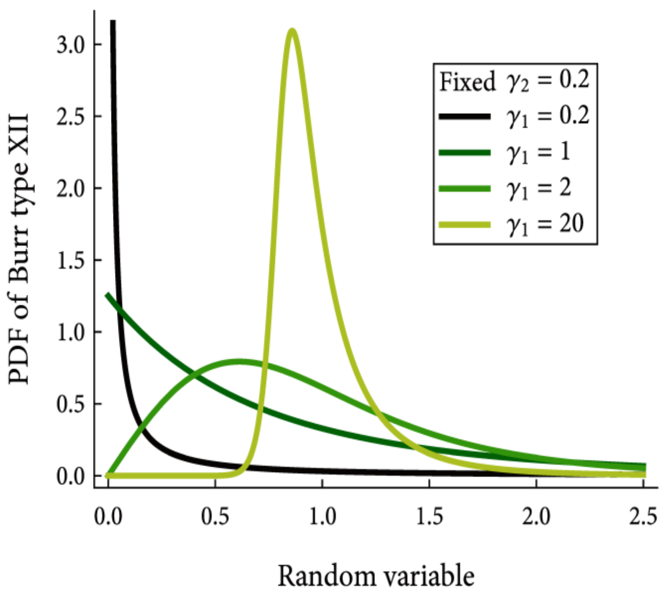
More details in Papalexiou and Koutsoyiannis, 2013 and Papalexiou, 2018

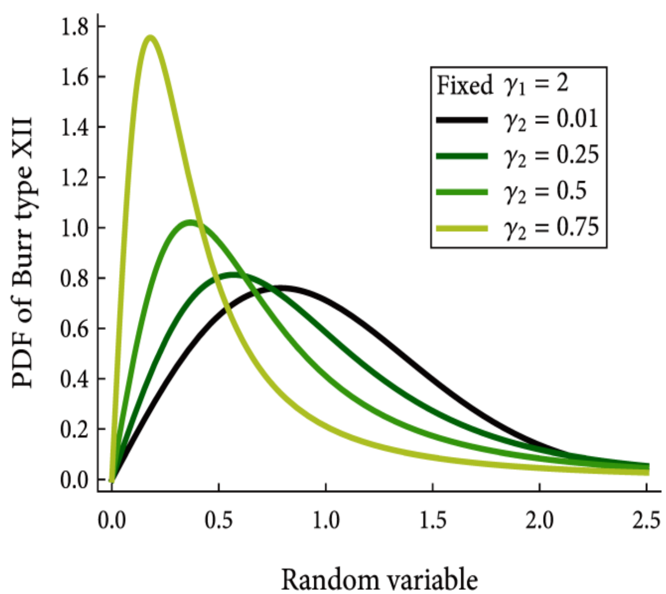
Shape examples





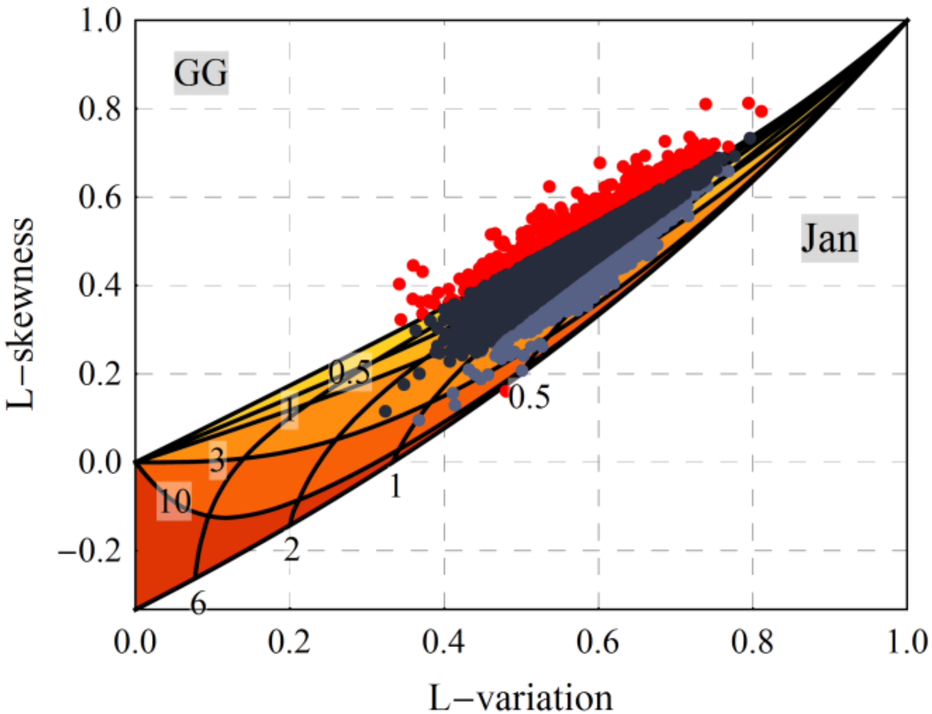
Probability density functions of the Generalized Gamma distribution for various shape parameter values. The values of scale parameter *β* were chosen so that mean value of each distribution equals 1.

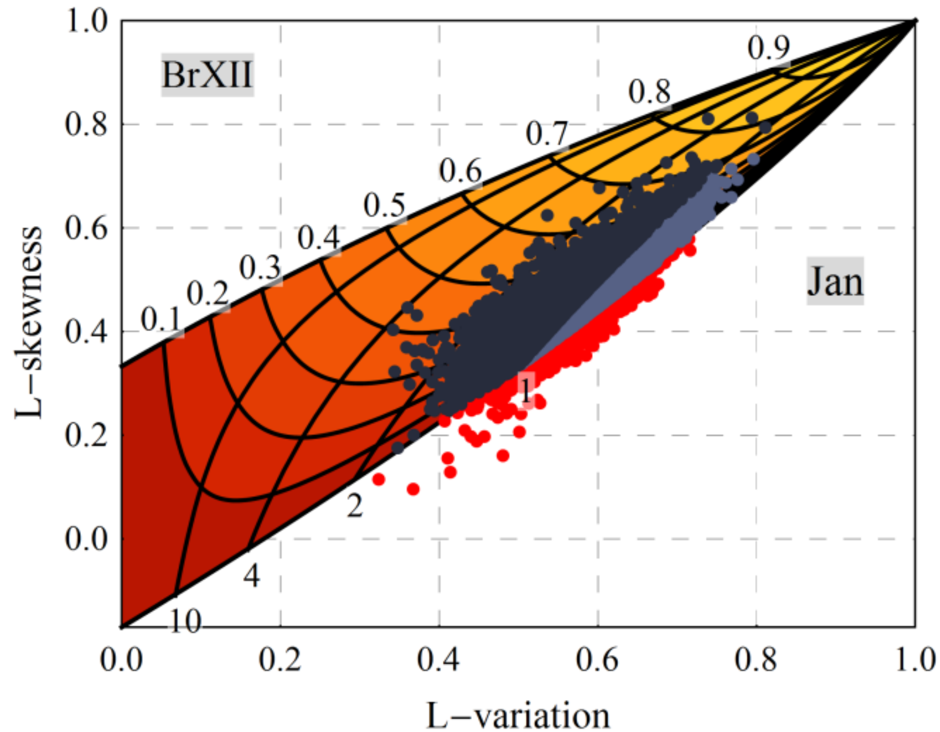




Probability density functions of the Burr type XII distribution for various shape parameter values. The values of scale parameter *β* were chosen so that mean value of each distribution equals 1.

Global analysis on daily precipitation

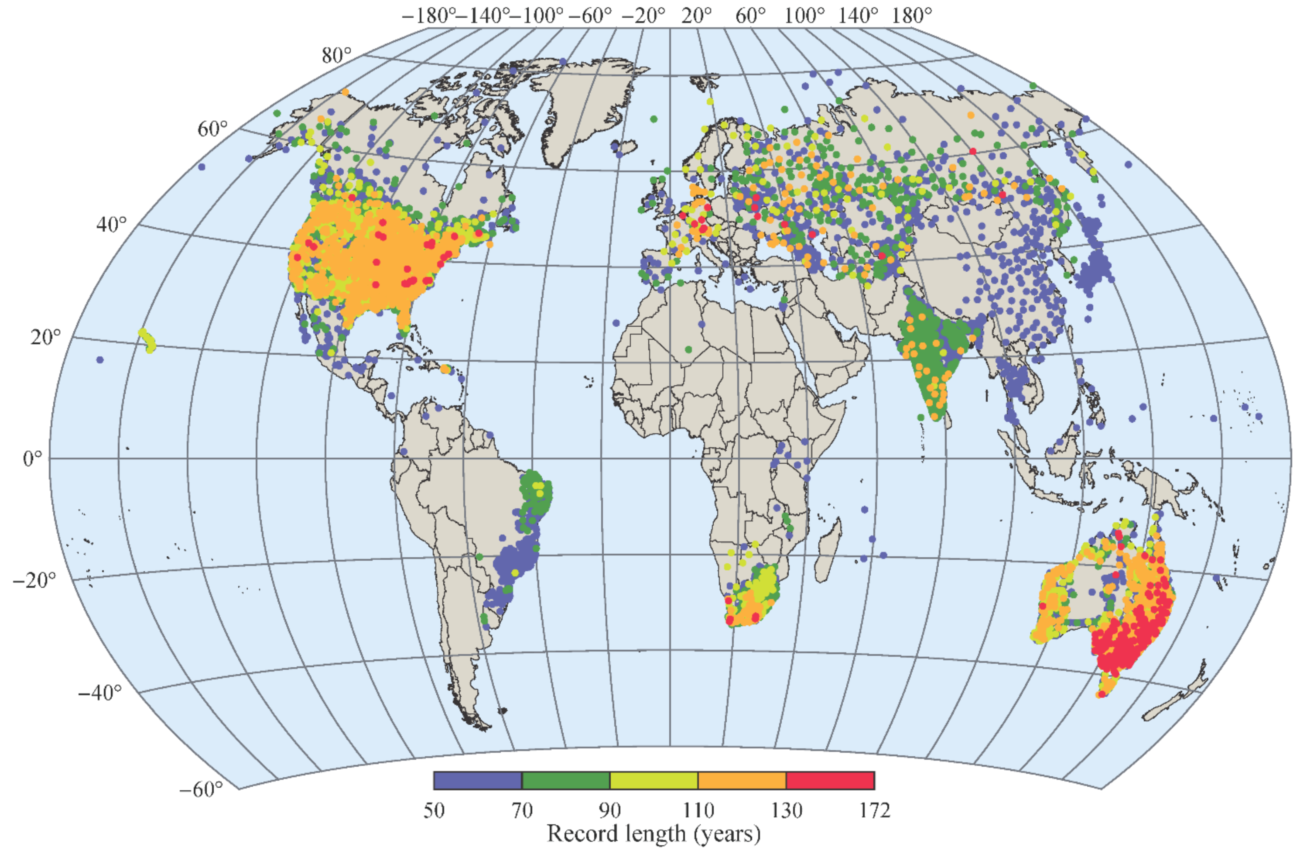




Observed L-points for the month of January of the 14 157 daily rainfall records studied in comparison to the theoretical L-areas of (a) the BrXII distribution and (b) the GG distribution. Red-colored L-points lie outside the L-area; dark-colored indicate a Bell-shaped distribution; light-colored indicate a J-shaped distribution.

A global survey on the seasonal variation of the marginal distribution of daily precipitation.
https://doi.org/10.1016/j.advwatres.2016.05.005

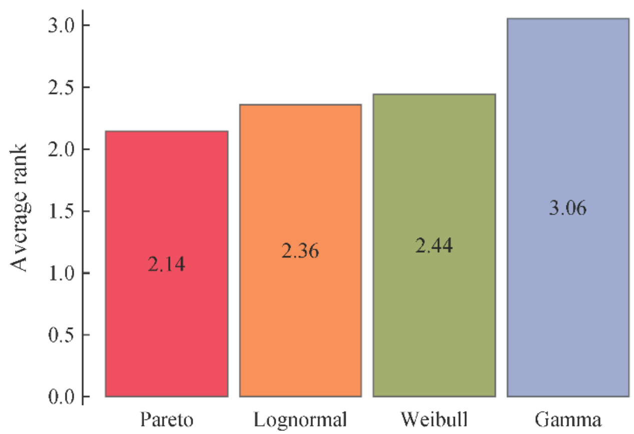
Focusing on tails



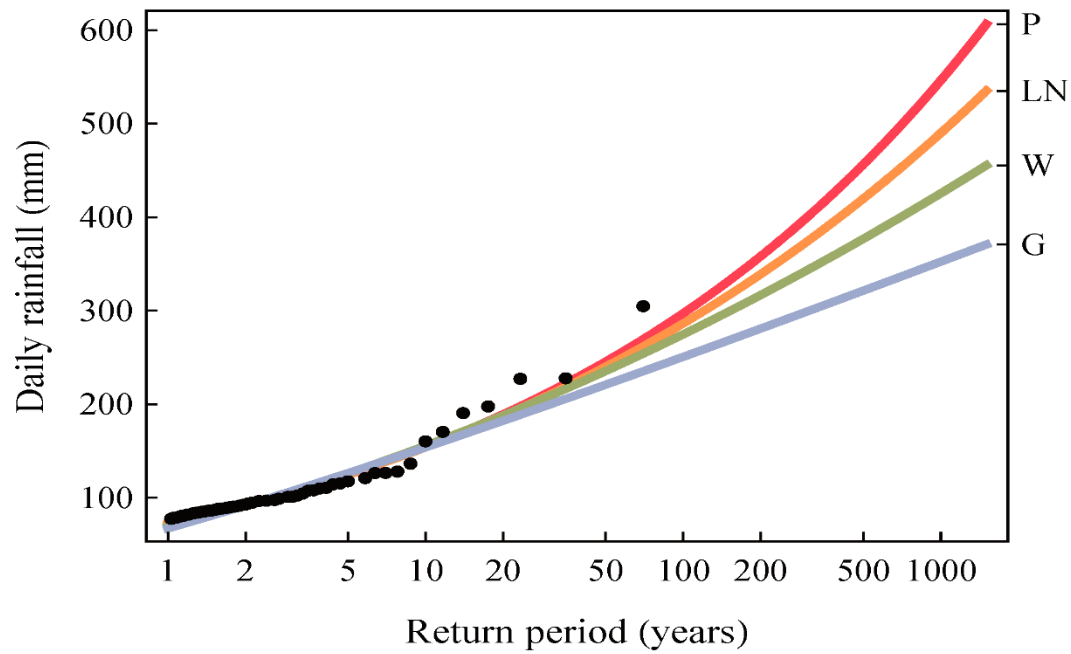
Locations of the stations studied (a total of 15 137 daily rainfall records with time series length greater than 50 years). Note that there are overlaps with points corresponding to high record lengths shadowing (being plotted in front of) points of lower record lengths.

Papalexiou et al. (2013). *How extreme is extreme? An assessment of daily rainfall distribution tails*. *Hydrol. Earth Syst. Sci.*, *17*(2), 851–862. https://doi.org/10.5194/hess-17-851-2013

Comparing popular tails

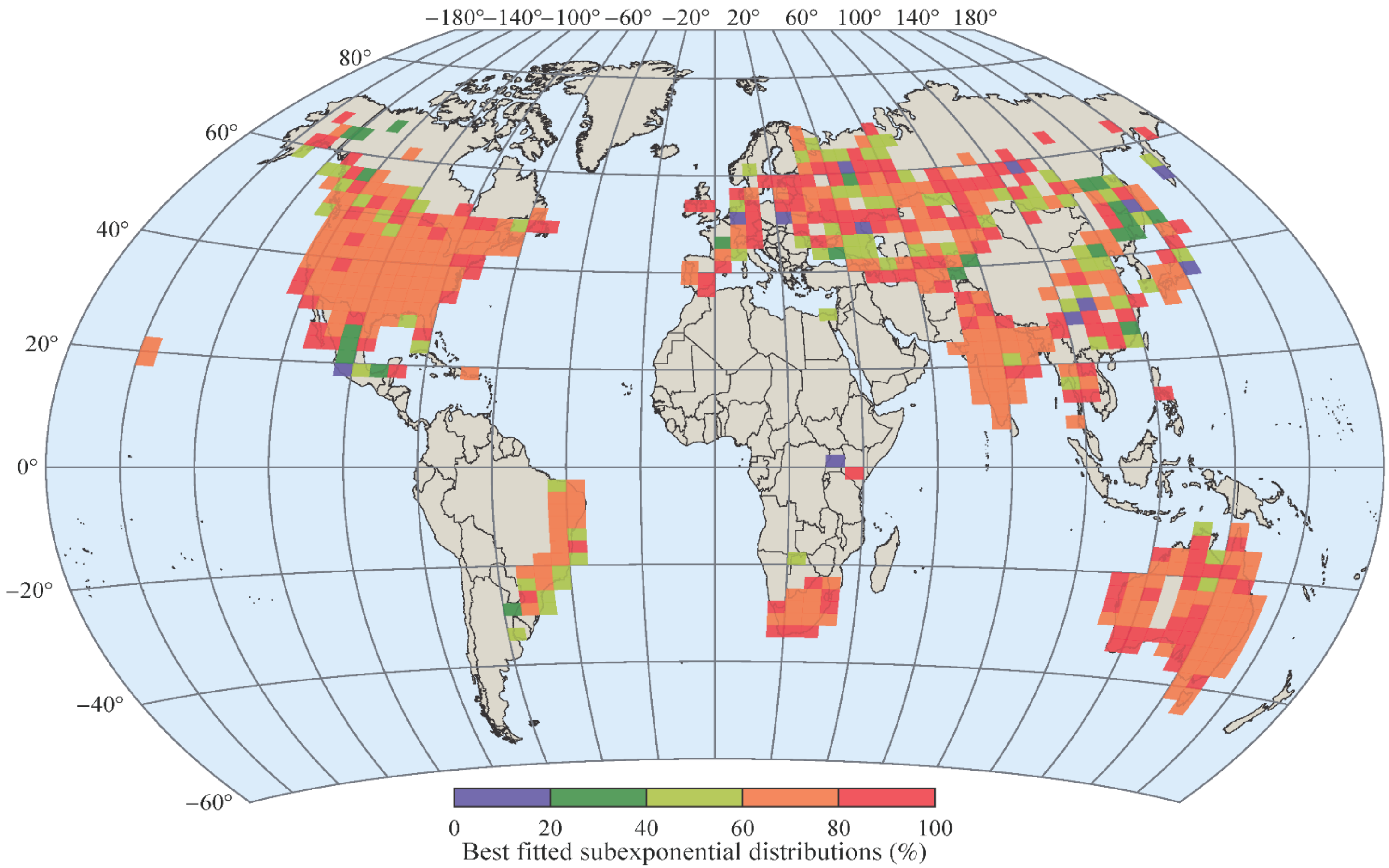


Mean ranks of the tails for all records. The best-fitted tail is ranked as 1 while the worst-fitted as 4. A lower average rank indicates a better performance.



Four different distribution tails fitted to an empirical tail (P, LN, W and G stands for the Pareto, the Lognormal, the Weibull and the Gamma distribution). A wrong choice may lead to severely underestimated or overestimated rainfall for large return periods.

Heavy tails prevail



Geographical variation of the percentage of best fitted subexponential tails in cells defined by latitude difference $\Delta\varphi = 2.5°$ and longitude difference $\Delta\lambda = 5°$. In total, in **72.6%** of the 15 029 records analysed, the subexponential tails were the best fitted.

Battle of Extreme Value Distributions

If a random variable (RV) X follows the distribution $F_X(x)$ then the distribution function of the maximum of n independent and identically distributed RV's, i.e., $Y_n = \max(X_1,\dots,X_n)$ will be,

$$G_{Y_n}(x) = \left(F_X(x)\right)^n \tag{1}$$

Now, if $n \to \infty$ there are three limiting laws, the type I or Gumbel (G), the type II or Fréchet (F) and the type III or reversed Weibull (RW) with distribution functions respectively given by

$$G_{\mathrm{G}}(x) = \exp\left(-\exp\left(-\frac{x-\alpha}{\beta}\right)\right) \qquad x \in \mathbb{R} \tag{2}$$

$$G_{\mathrm{F}}(x) = \exp\left(-\left(\frac{x-\alpha}{\beta}\right)^{-\gamma}\right) \qquad x \geq \alpha \tag{3}$$

$$G_{\mathrm{RW}}(x) = \exp\left(-\left(-\frac{x-\alpha}{\beta}\right)^{\gamma}\right) \qquad x \leq \alpha \tag{4}$$

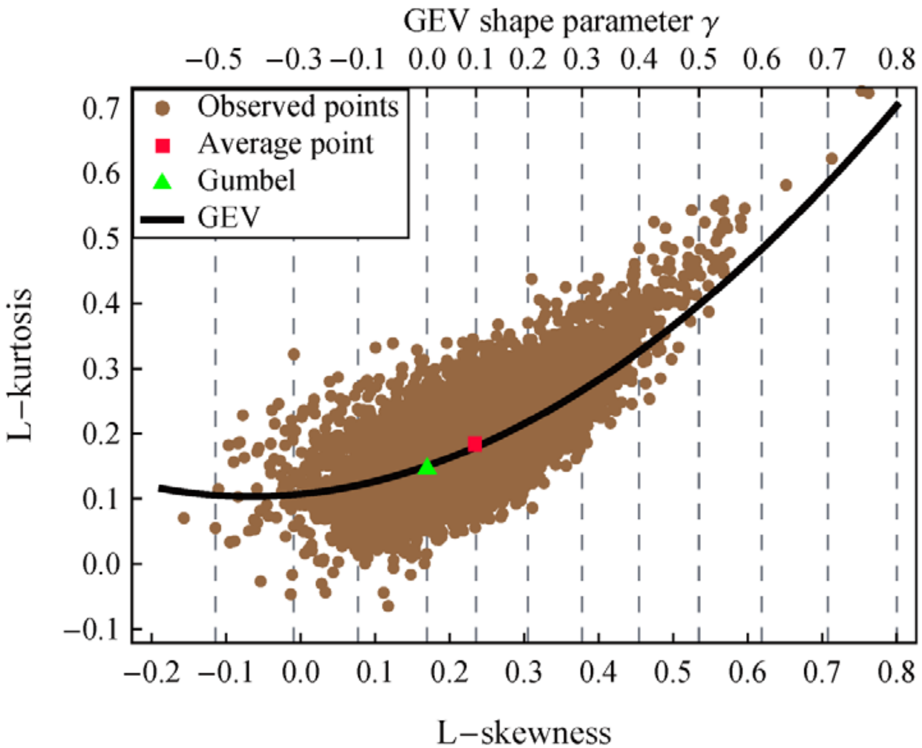
These distributions comprise a location parameter $\alpha \in \mathbb{R}$ and a scale parameter $\beta > 0$, with the Fréchet and the reversed Weibull distributions having the additional shape parameter $\gamma > 0$.

These three distributions can be unified into a single expression known as the Generalized Extreme Value (GEV) distribution (also known as the Fisher-Tippet) with distribution function given by

$$G_{\mathrm{GEV}}(x) = \exp\left(-\left(1+\gamma\frac{x-\alpha}{\beta}\right)^{-1/\gamma}\right) \qquad 1+\gamma\frac{x-\alpha}{\beta} \geq 0 \tag{5}$$

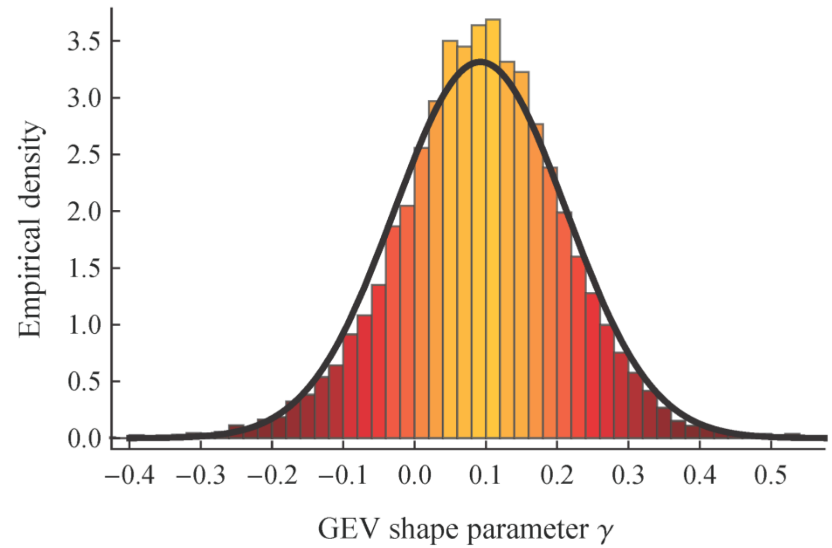
This simple reparameterization exploits the limiting definition of the exponential function so that the Gumbel distribution can also emerges for $\gamma \to 0$.

L-moments fitting



L-kurtosis *vs.* L-skewness plot the 15 137 observed points. Interestedly, only 20% of points lies on the left of the Gumbel distribution, corresponding thus to a GEV distribution with $\gamma < 0$ (reversed Weibull law), while 80% of points lies on the right corresponding to a GEV distribution with $\gamma > 0$ (Fréchet law). The average point lies almost exactly on the GEV line and corresponds to $\gamma \approx 0.09$. The figure may not reveal the percentage of points that could be described by a Gumbel distribution, yet, it offers a clear indication that the Fréchet law prevails.

The figure depicts the empirical distribution of the GEV shape parameter as well as a fitted normal distribution with mean 0.093 and standard deviation 0.12.
It is worth noting the large variation of the estimated GEV shape parameter as it ranges from −0.59 to 0.76 with mean value 0.093; the 90% empirical confidence interval is much smaller, i.e., from −0.11 to 0.28 .

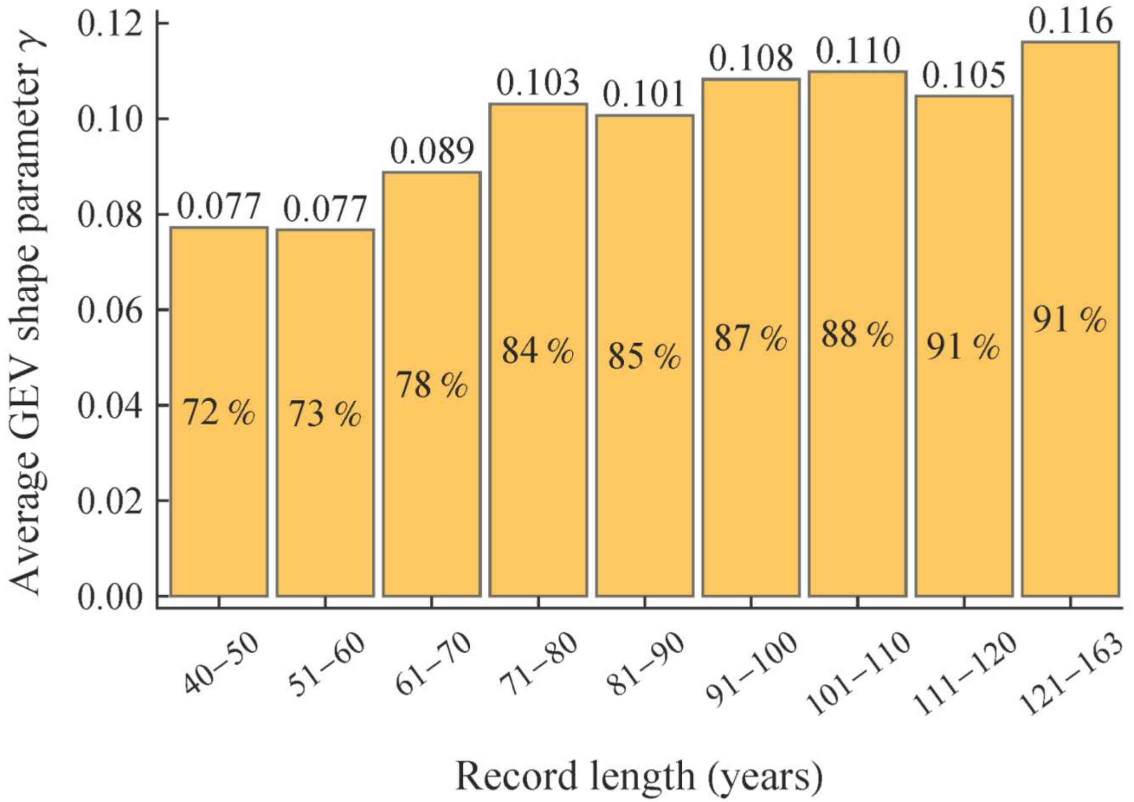


More in “Battle of Extreme Value Distributions”

It’s the record length that makes the difference

Larger samples offer more accurate estimates; in this direction we study the estimated GEV shape parameter in relationship with the record length, as the records studied here vary in length from 40 to 163 years.

We grouped the 15 137 estimated shape parameter values into nine groups based on the length of the record that were estimated; and second, we estimated various statistics for each group.



Clearly, the figure indicates an upward “trend” over record length both in the mean shape parameter value and in the percentage of records having positive shape parameter.

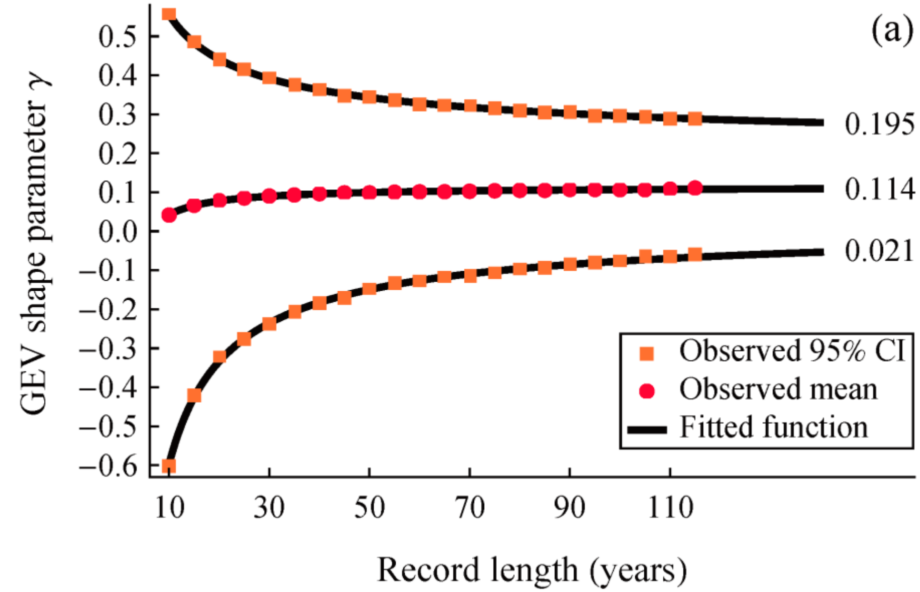
The figure depicts the mean value of the GEV shape parameter for various ranges of record length. While the number in the boxes indicates the percentage of records with positive shape parameter value.

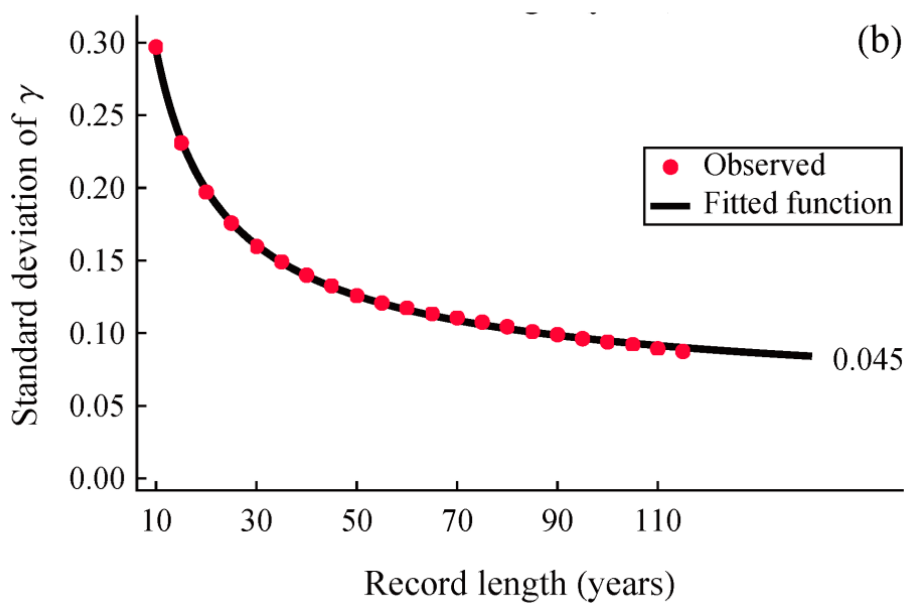
Bounded GEV is probably an artefact

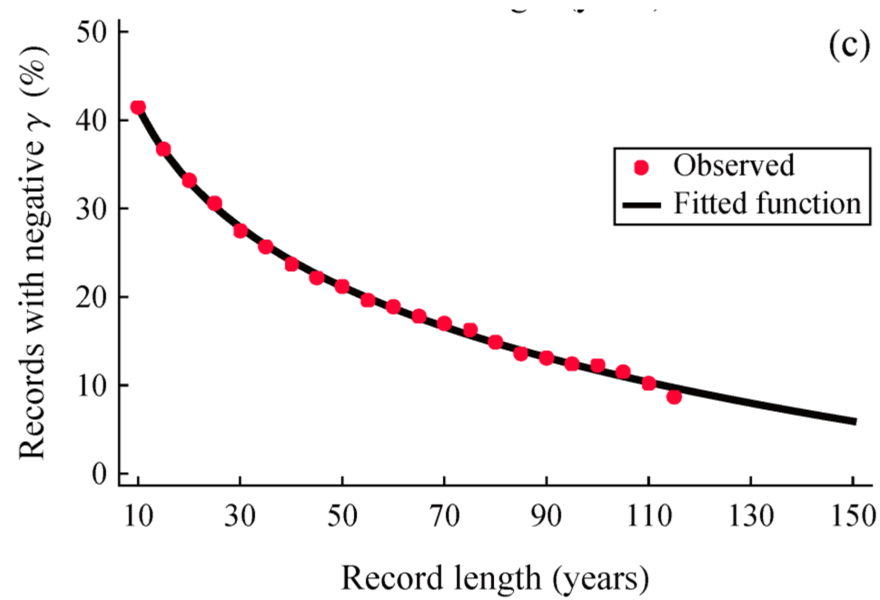
(a) Mean, Q_5 and Q_{95} observed points *vs.* record length, and the estimated asymptotic values of the fitted curves.

(b) Standard deviation vs. record length,

(c) Percentage of record with negative shape parameter vs. record length.

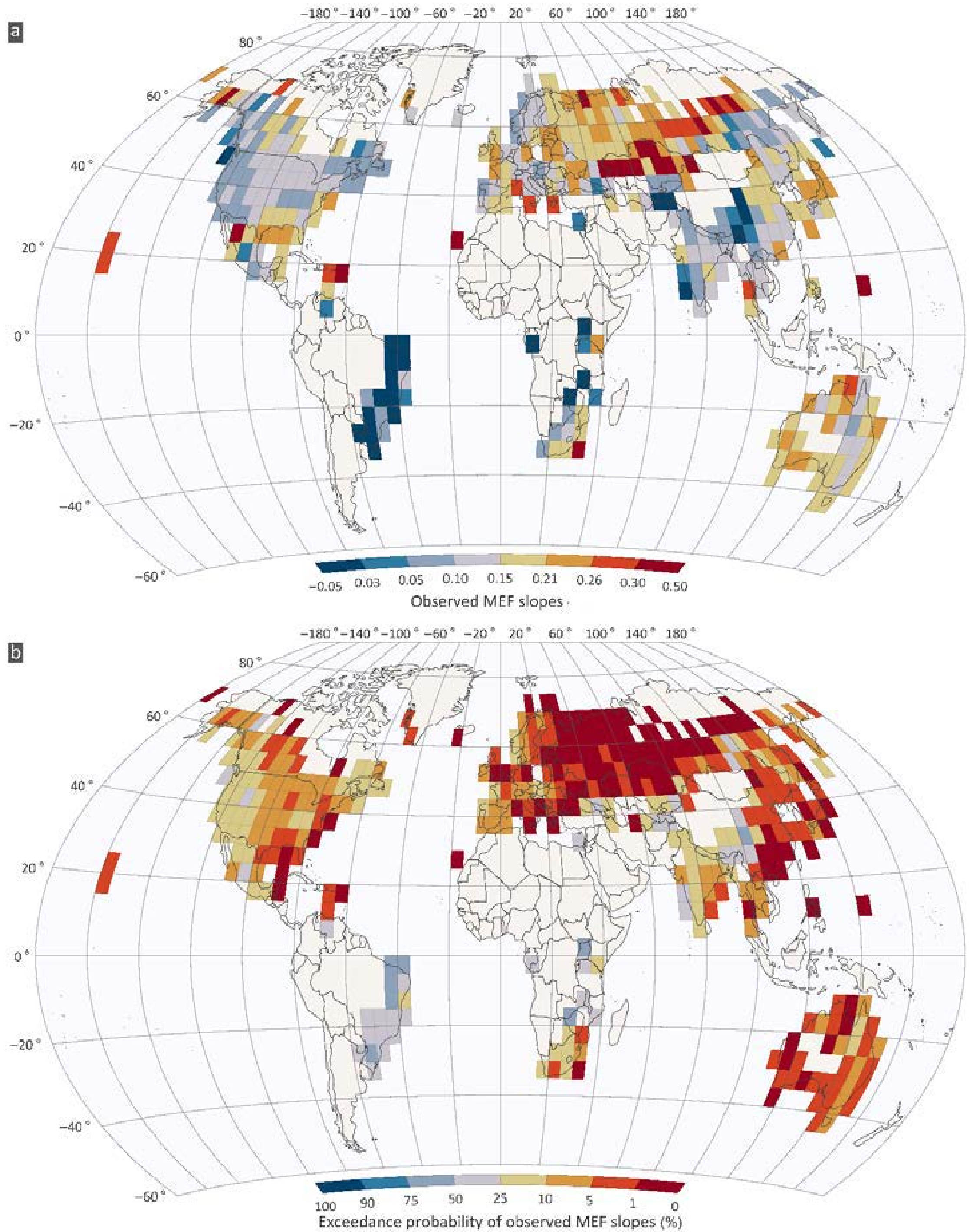






Recent Evidence

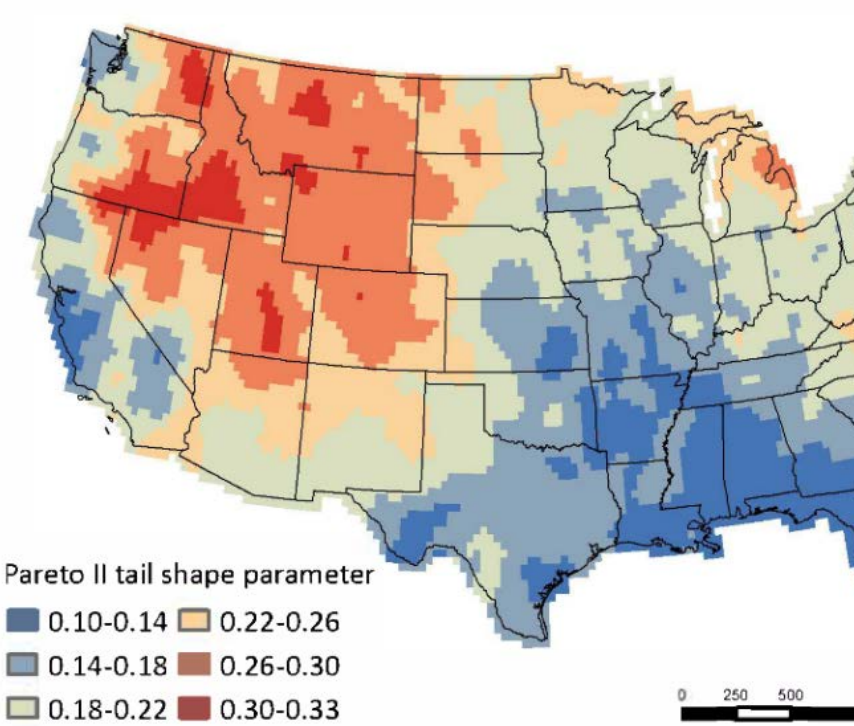
(a) Spatial distribution of the absolute values of the empirical MEF slopes for the 21,348 records of daily precipitation,
(b) spatial distribution of the corresponding exceedance probability of the observed MEF slopes (%); grid boxes show mean values.

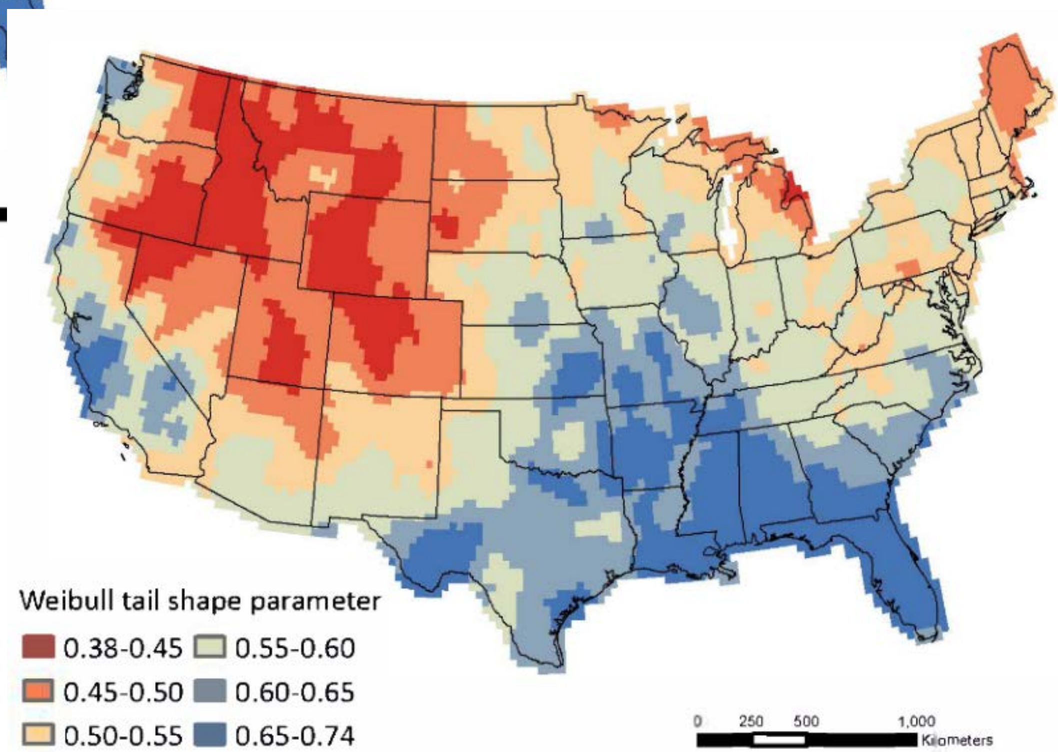


Nerantzaki, S., Papalexiou, S.M., *Tails of Extremes: Advancing a Graphical Method and Harnessing Big Data to Assess Precipitation Extremes*, *Advances in Water Resources* (2019), https://doi.org/10.1016/j.advwatres.2019.103448

Hourly records

Spatial variation of Pareto II (PII) and Weibull (W) tail indices.

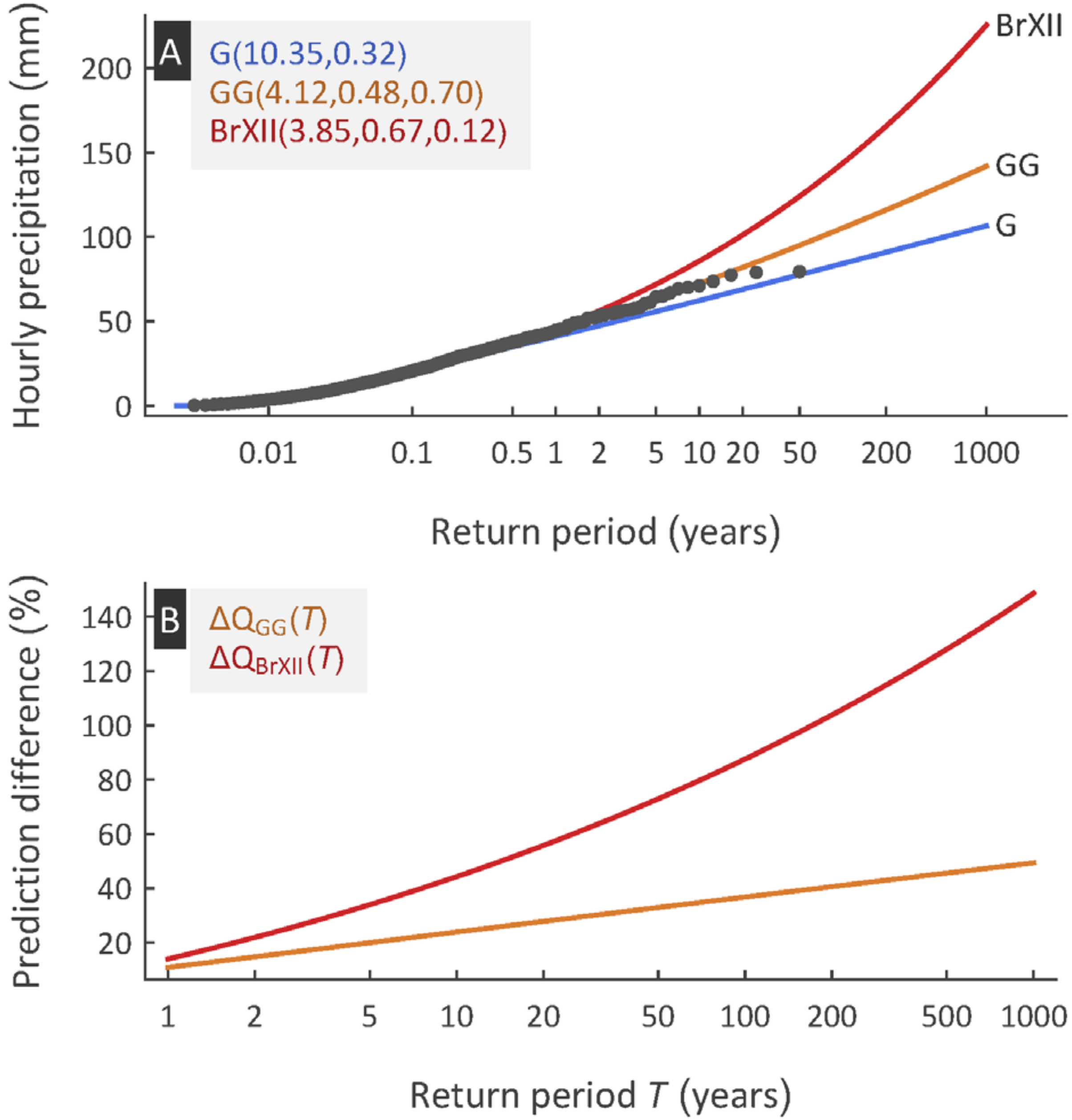




Papalexiou, S. M., AghaKouchak, A., & Foufoula-Georgiou, E. (2018). *A Diagnostic Framework for Understanding Climatology of Tails of Hourly Precipitation Extremes in the United States*. *Water Resources Research*. https://doi.org/10.1029/2018WR022732

Missing the tail is bad news

(A) Fitted Gamma (G), Burr type XII (BrXII) and Generalized Gamma (GG) distributions for an hourly precipitation record of Lake Charles regional Airport in Louisiana. Tail indices of BrXII and GG distributions were fixed from the maps of Fig. 6. (B) Prediction difference of the GG and the BrXII distributions compared to a G distribution for a typical hourly precipitation record in USA.



Papalexiou, S. M., AghaKouchak, A., & Foufoula-Georgiou, E. (2018). *A Diagnostic Framework for Understanding Climatology o Tails of Hourly Precipitation Extremes in the United States*. *Water Resources Research*. https://doi.org/10.1029/2018WR022732

Maximum likelihood

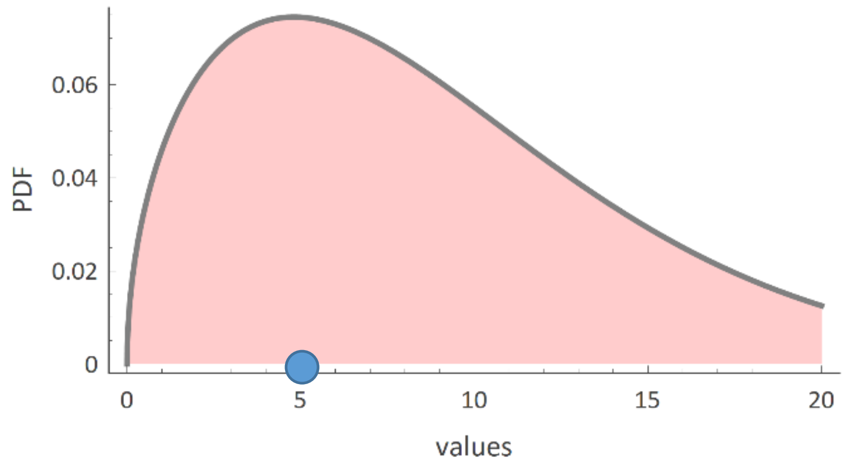
A generic estimation method is the maximum likelihood (ML). The likelihood function is defined by:

$$\mathcal{L}(\boldsymbol{\theta}) = \frac{1}{n}\prod_{i=1}^{n} f_X(x_i;\boldsymbol{\theta})$$

But in practice we use the log-likelihood function

$$\log\mathcal{L}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \log f_X(x_i;\boldsymbol{\theta})$$

as sums are simpler to handle and the two expressions are equivalent in terms of the parameter values that maximize them. The maximum likelihood estimates (MLE) are given by $\boldsymbol{\theta} = \text{argmax}\,\log\mathcal{L}(\boldsymbol{\theta})$



Least square fitting

We can estimate the parameters of any distribution by using generic optimization methods like minimizing the Square Error of an objective function (OF). We can construct several OF for fitting distributions:

I	$\boldsymbol{\theta} = \mathrm{argmin} \sum_{i=1}^{n} \left(x_i - Q_X(u_i; \boldsymbol{\theta})\right)^2$
II	$\boldsymbol{\theta} = \mathrm{argmin} \sum_{i=1}^{n} \left(\frac{x_i}{Q_X(u_i; \boldsymbol{\theta})} - 1\right)^2$
III	$\boldsymbol{\theta} = \mathrm{argmin} \sum_{i=1}^{n} \left(F_X(x_i; \boldsymbol{\theta}) - F_N(x_i)\right)^2$
IV	$\boldsymbol{\theta} = \mathrm{argmin} \sum_{i=1}^{n} \left(\frac{F_X(x_i; \boldsymbol{\theta})}{F_N(x_i)} - 1\right)^2$

Where $u_i = F_N(x_i) = r(x_i)/(n+1)$

Moments and L-moments

- Generally, If a distribution has *n* parameters, we need to form a system of *n* equations and solve for the unknown parameters.
- This is accomplished by equating the theoretical moments (or L-moments) with the corresponding sample ones. i.e., $\mu_q = \hat{\mu}_q$ for $q = 1,2, \dots n$
- For example, if a distribution has two parameters then we can create a system of equations by using the expressions of the theoretical mean and standard deviation and the corresponding ones, i.e., $\{\mu = \hat{\mu}, \sigma = \hat{\sigma}\}$

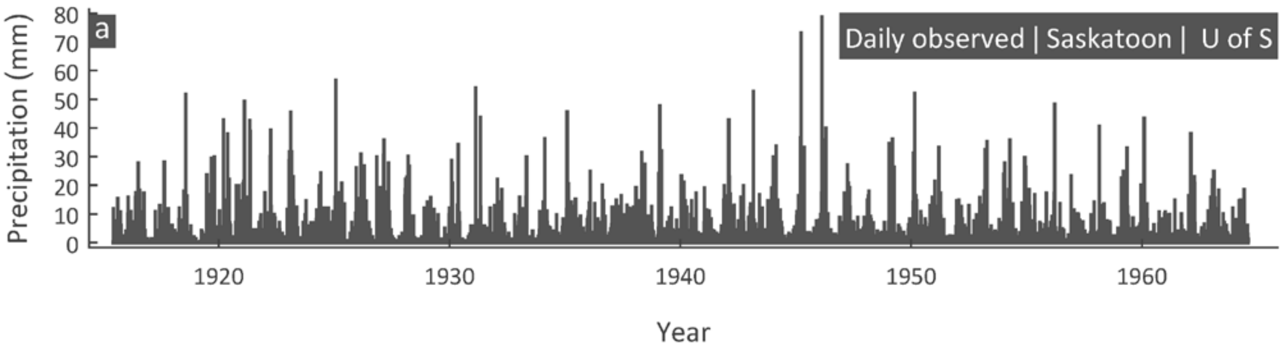
Caution!

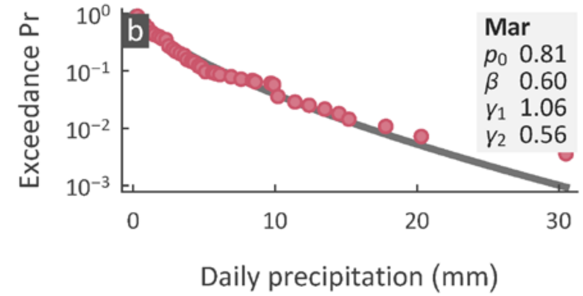
- Distributions with two shape parameters offer great flexibility
- But it is not always trivial to fit
- It easy to get tail behavior that are unrealistic, e.g., infinite variance
- Regional methods and different fitting methods should be explored
- It has become very common to use software packages to fit tenths of distributions to select the best…

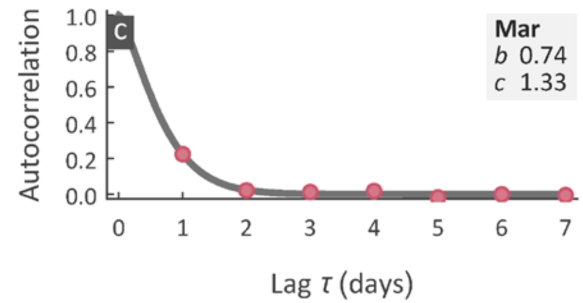
Please don't do that…

Part II
Stochastic Modelling

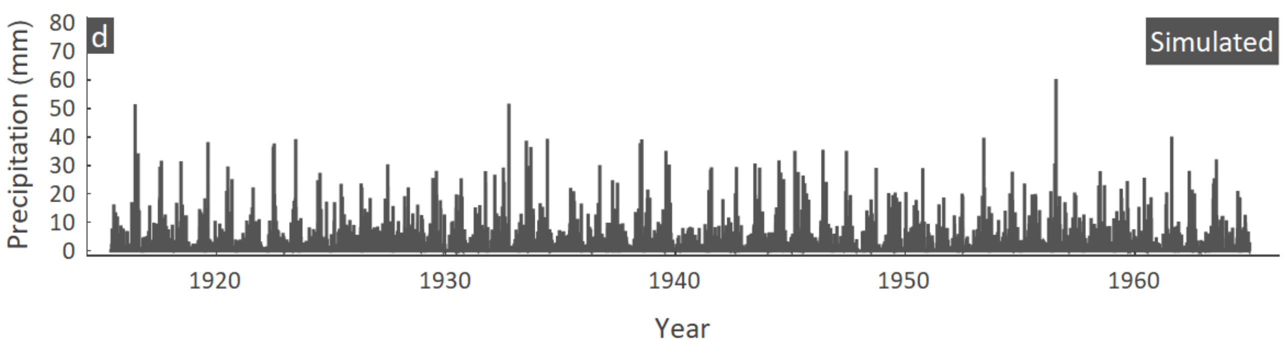
Reality as observed…

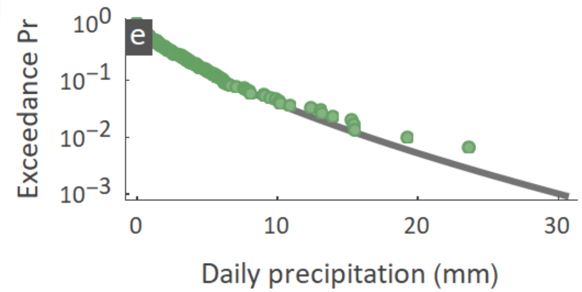


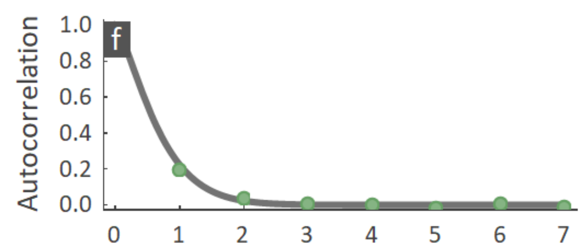




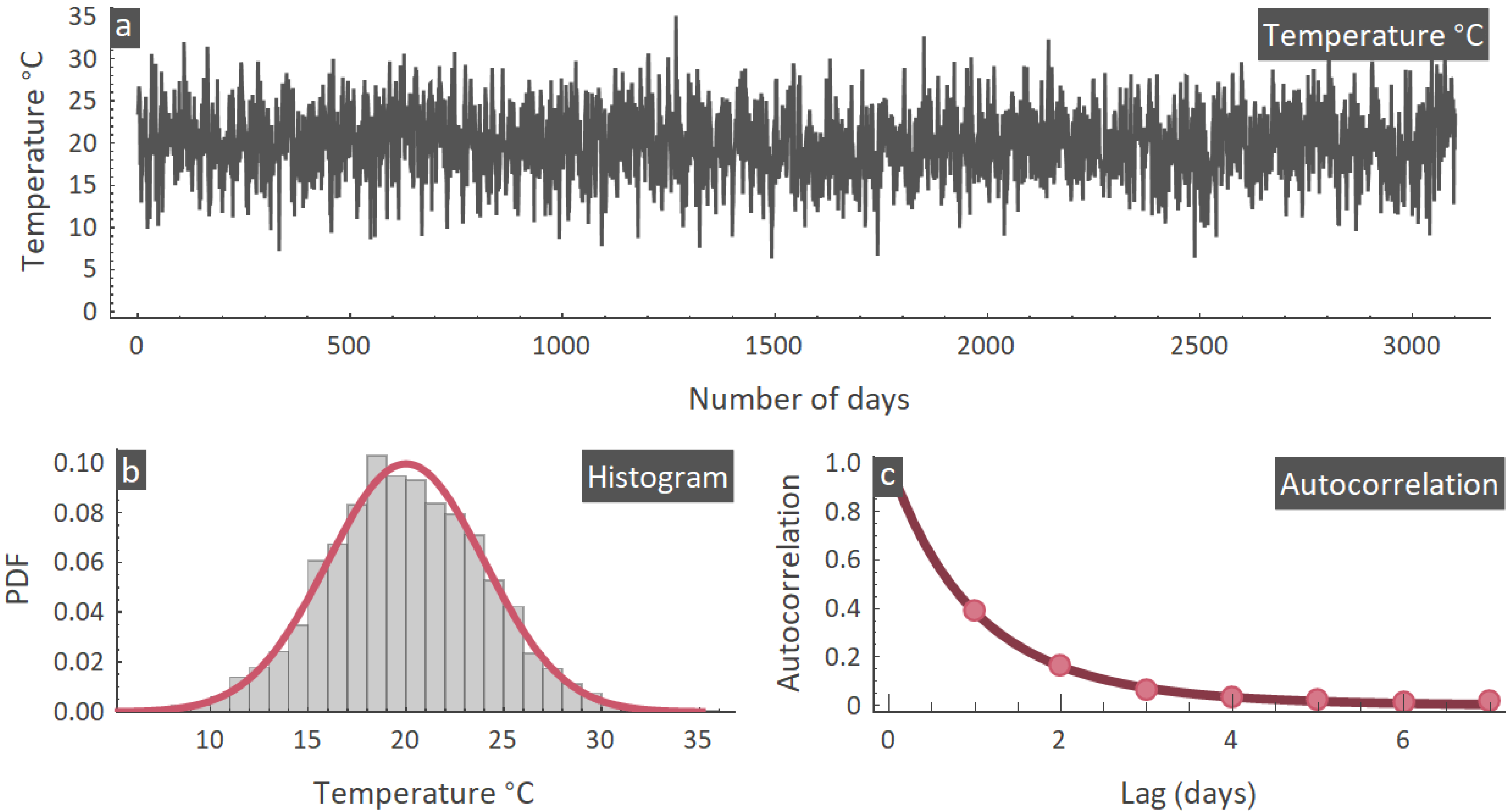
… "virtual reality" of stochastic modelling… if done right it'll reproduce the behavior or extremes





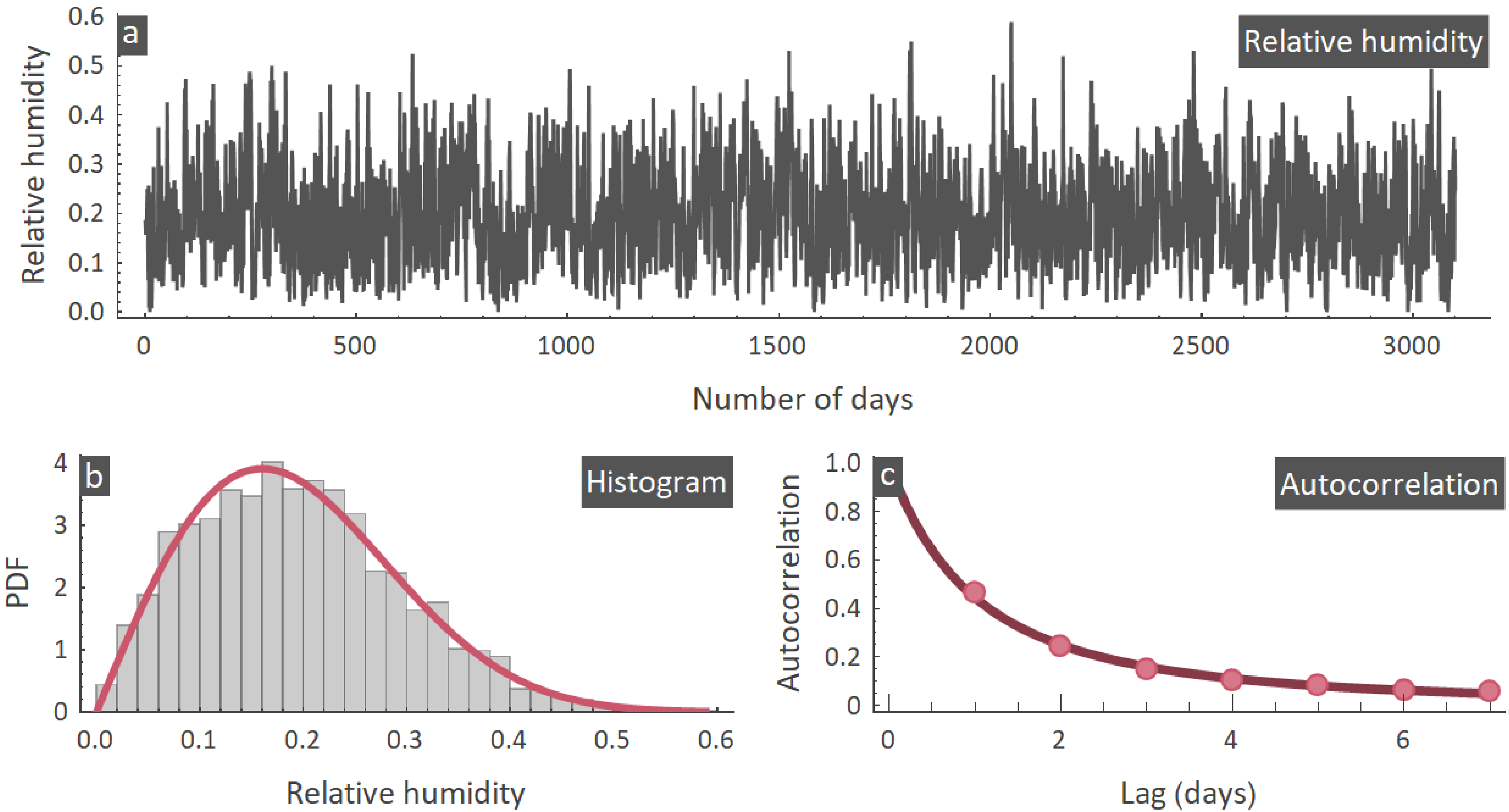


Daily temperature | changing autocorrelation



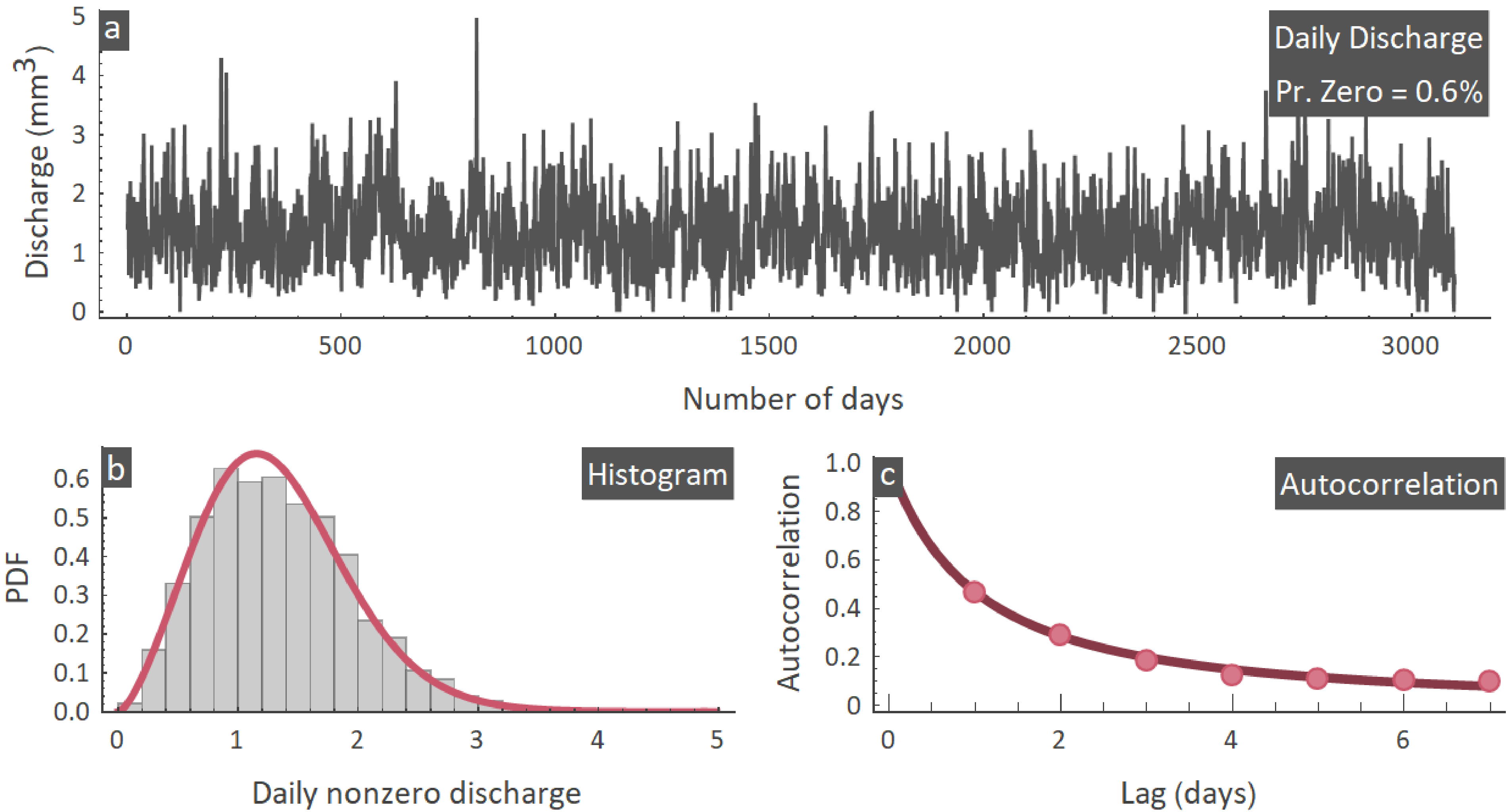
The distribution law remains the same yet the time series profoundly changes as the ACS becomes more intense.

Relative humidity | changing shape



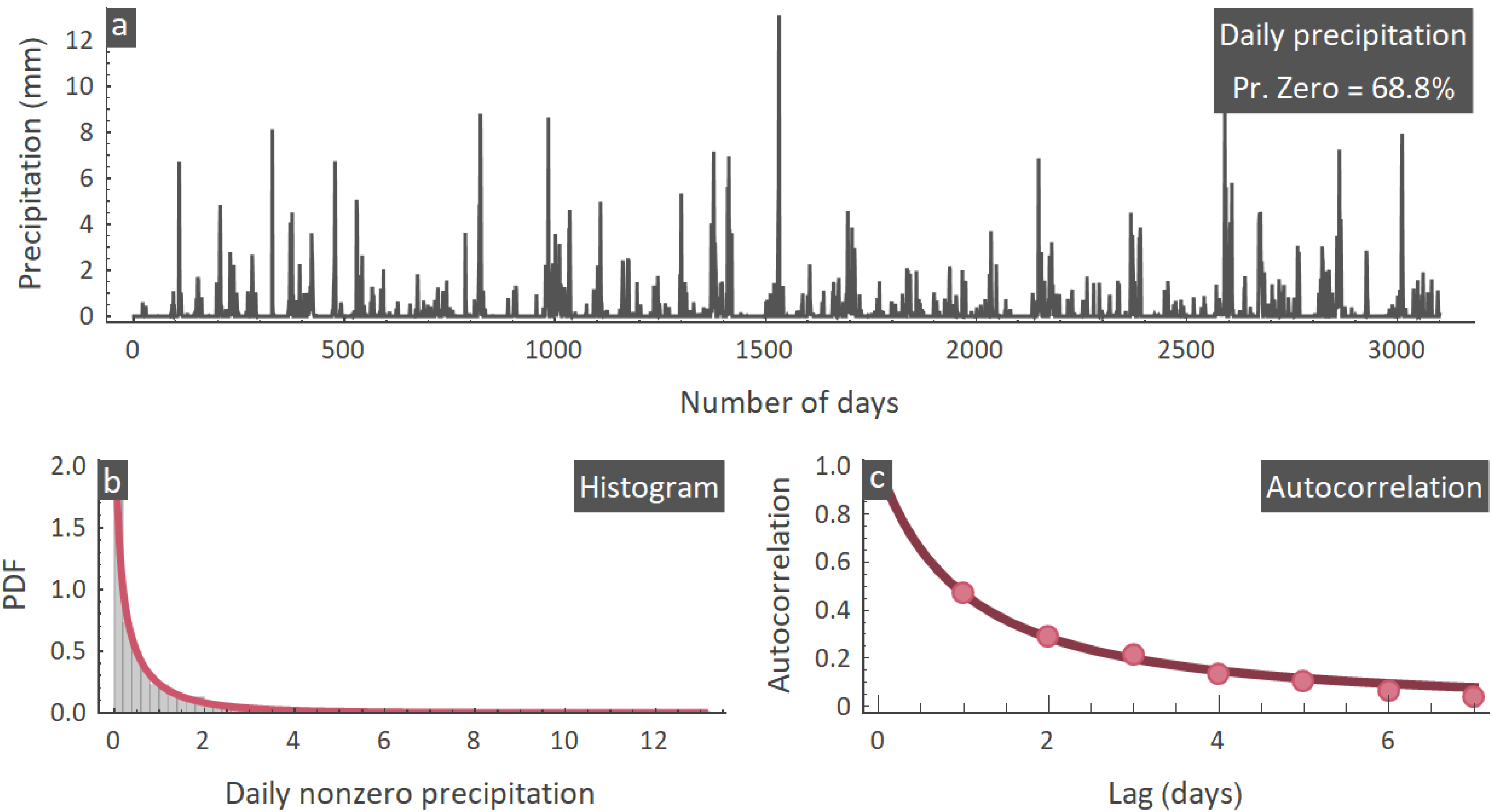
See how changes in the shape of distribution affect the time series!

Daily discharge | changing tail



Natural phenomena with heavy distribution tails can kill!

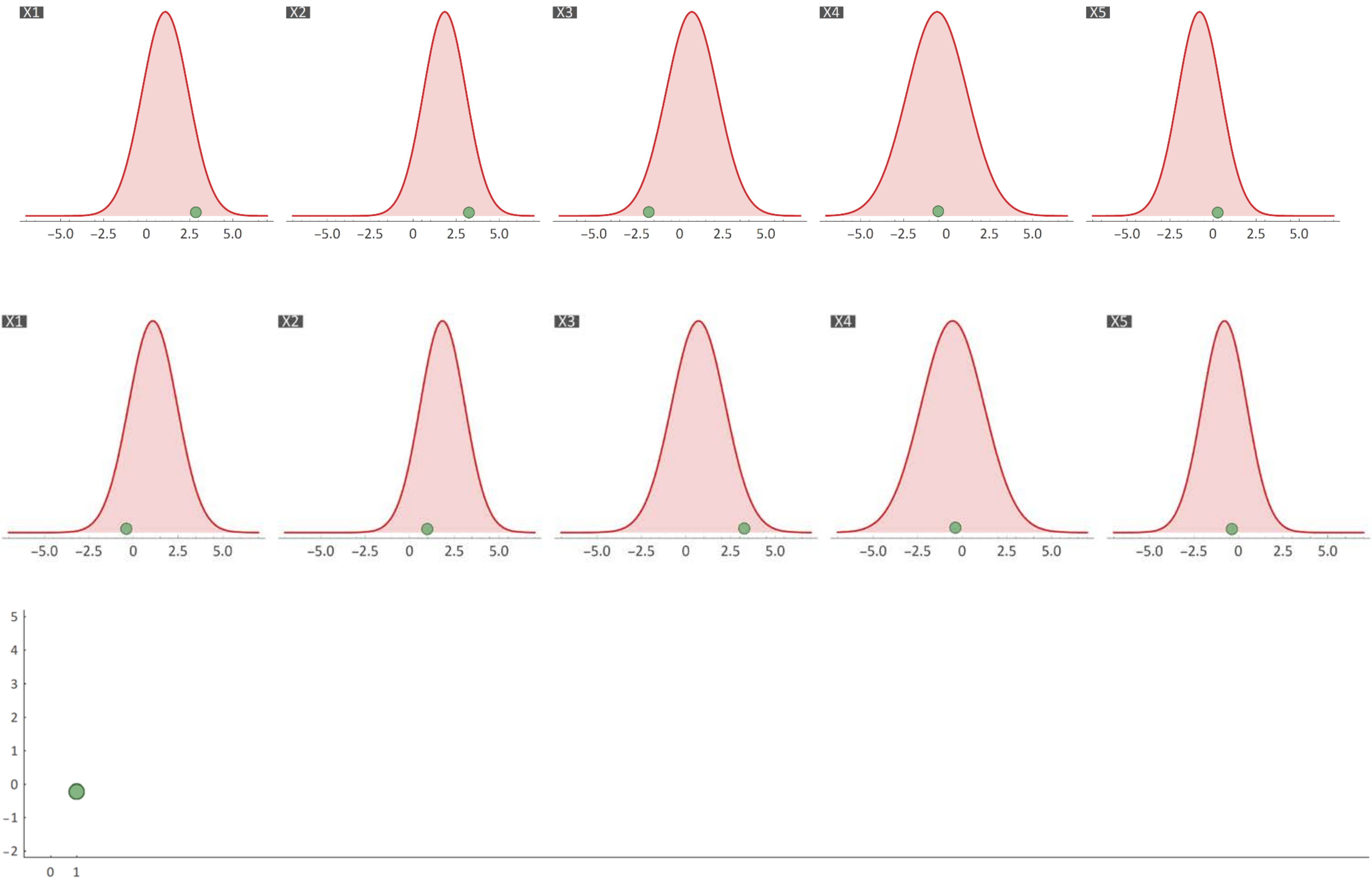
Daily precipitation | changing shape



From a RV to a random process

- A random process is just a sequence of random variables commonly denoted as $\{X(t)|t \in T\}$ where T is an indexed set.
- We can assume that the time series we observe or record in nature are the outcome or the realization of a random process.
- Stochastic process ≠ Time series

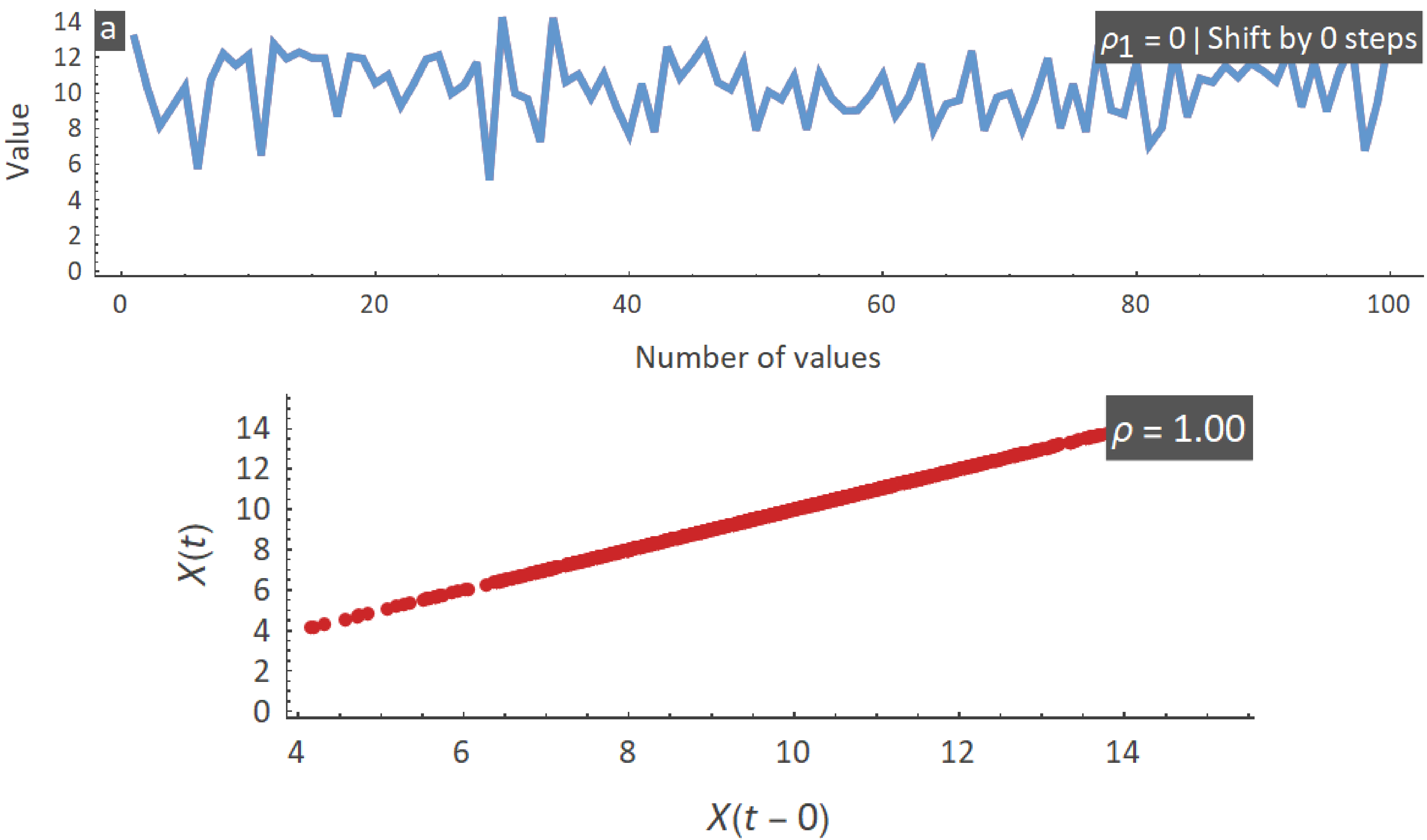
So what is a random process?



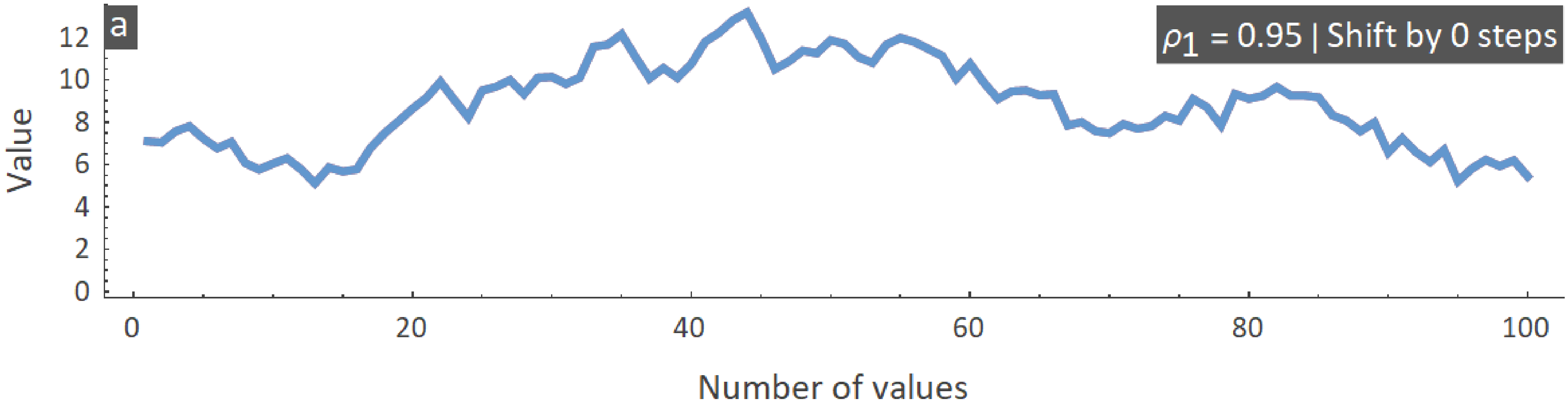
Things not to Forget…

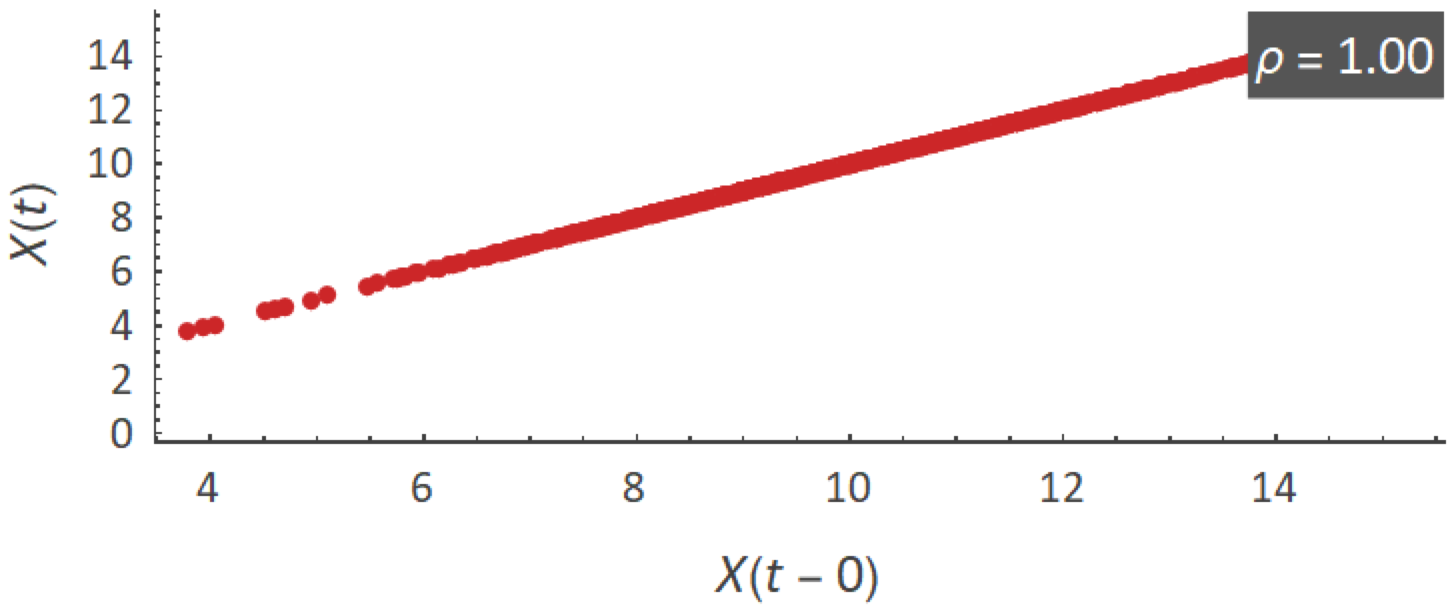
- Nature likes to connect things
- The random variables $\{X_1, X_2, X_3, \ldots\}$ that make the random process are typically connected with each other.
- Waldo Tobler's First Law of Geography, stating that "near things are more related than distant things"
- But we can also say that: near in time "things" are more related than distance in time "things"

Not correlated

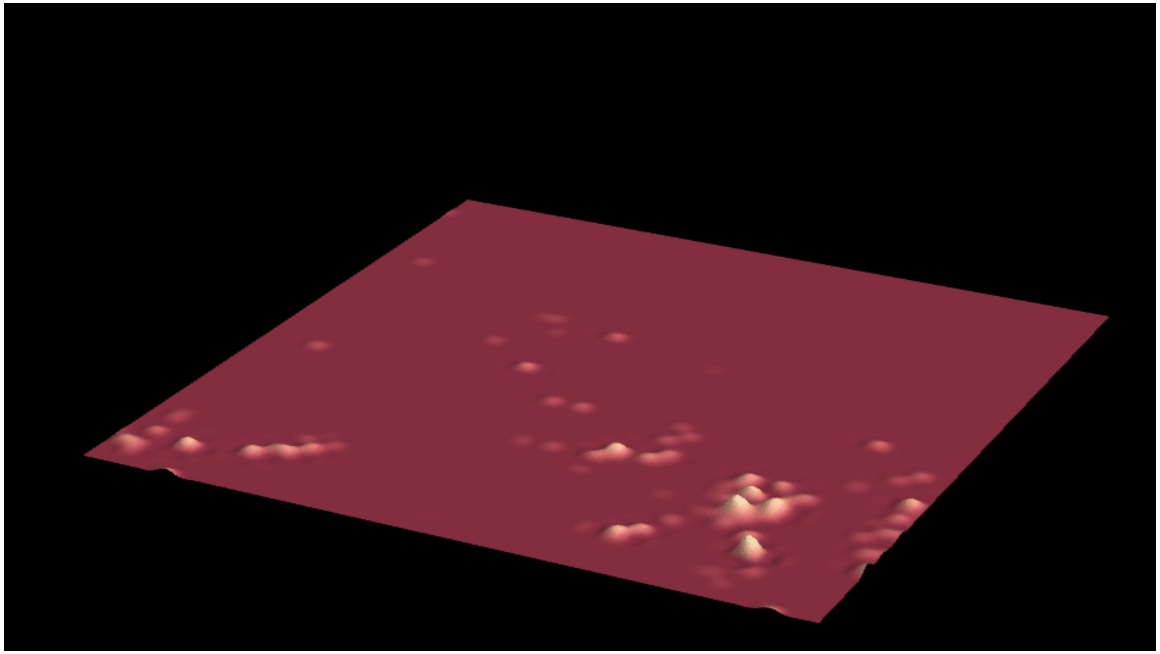


Temporal Correlation matters for extremes





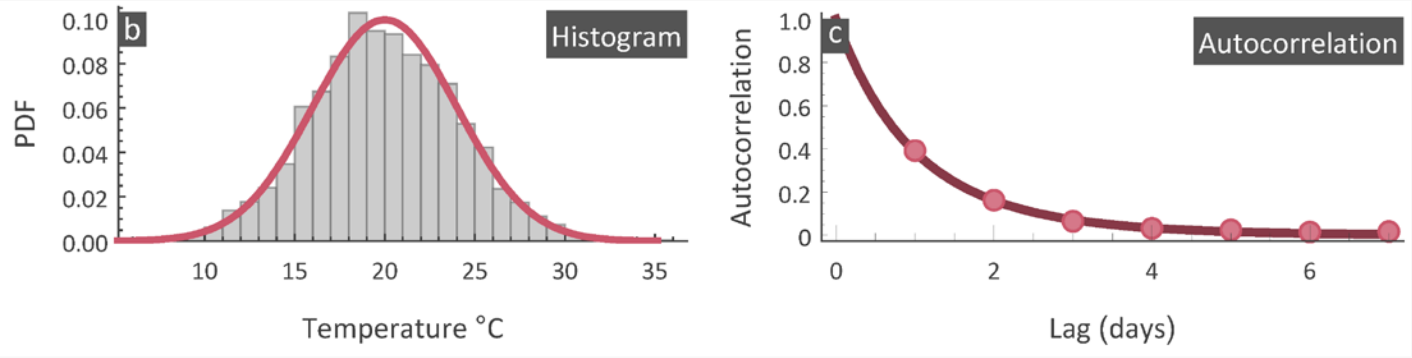
Spatial Correlation matters for extremes



Make a stochastic process simple

In most cases we could approximate well a process by two major component

- Its marginal distribution (for stationary processes)
- the autocorrelation structure quantified commonly by the correlation coefficient (a measure of linear dependence).



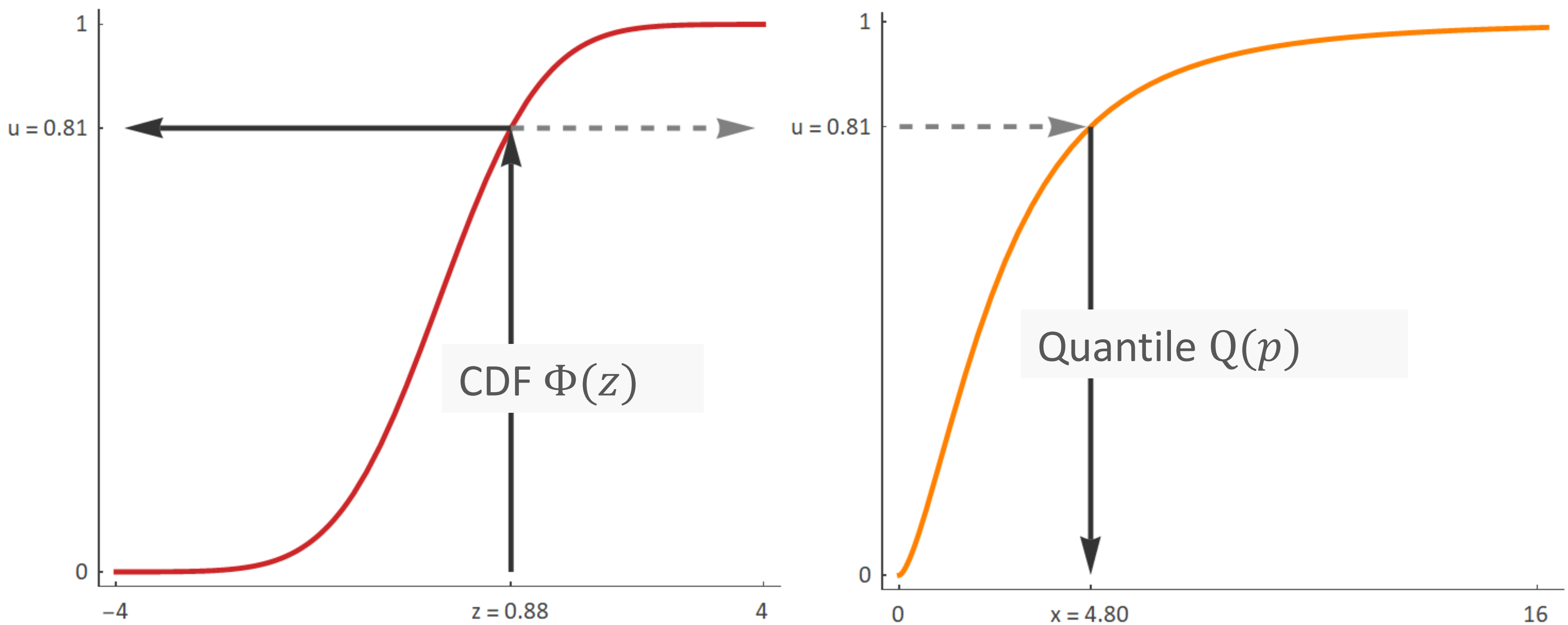
- The lag-τ autocorrelation of a stochastic process is defined as

$$\rho_X(\tau) = \frac{E\Big(\big(X(t) - \mu_{X(t)}\big)\big(X(t-\tau) - \mu_{X(t-\tau)}\big)\Big)}{\sigma_{X(t)}\sigma_{X(t-\tau)}} = \frac{E\big(X(t)X(t-\tau)\big) - \mu_X^2}{\sigma_X^2}$$

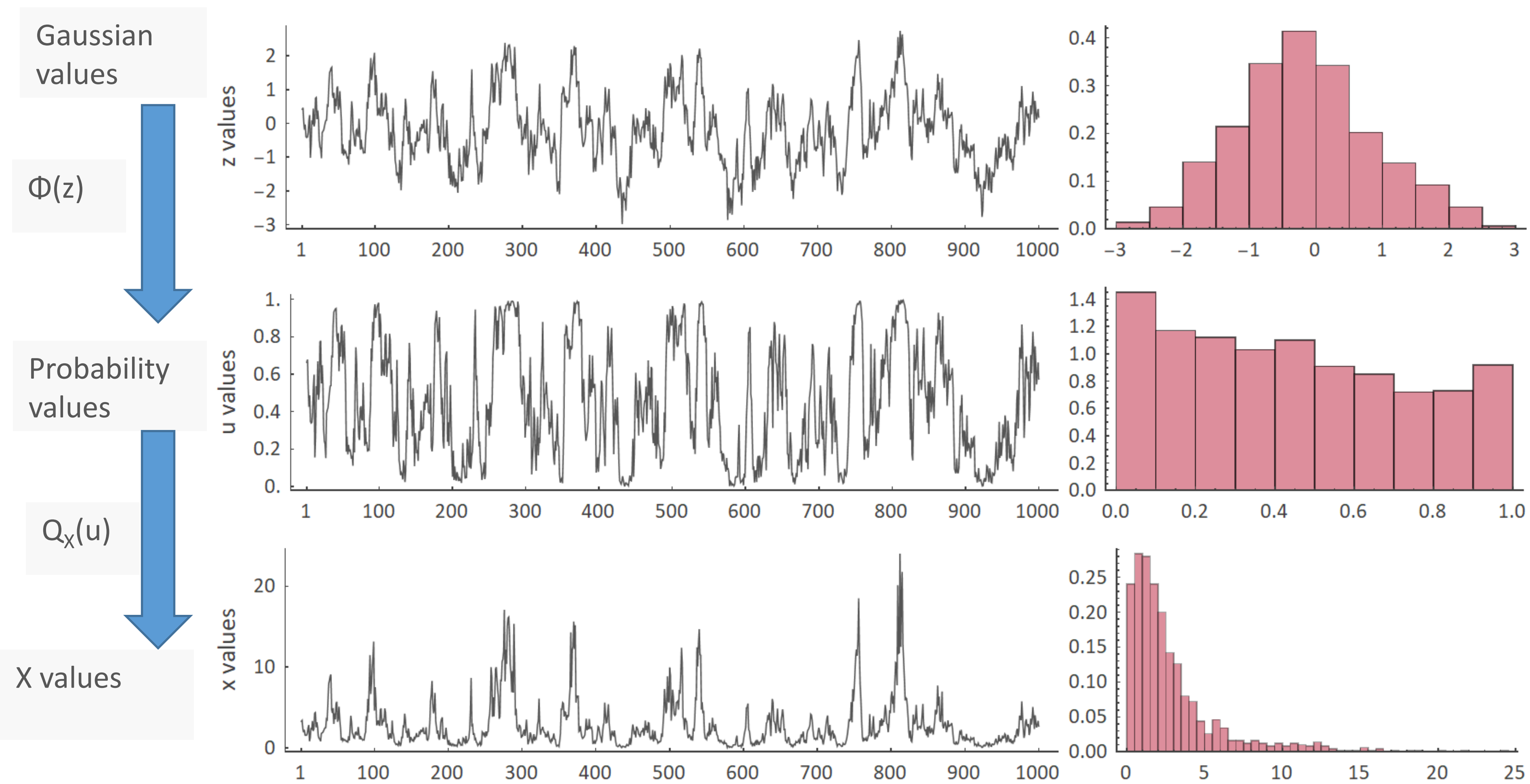
$$E\big(X(t)X(t-\tau)\big) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x(t)x(t-\tau)f_{X(t)X(t-\tau)}\big(x(t), x(t-\tau)\big)\,\mathrm{d}x(t)\mathrm{d}x(t-\tau)$$

From one distribution to another

- From a value z using the CDF we can find the corresponding probability u
- And from u using the quantile function of another distribution we can find the corresponding value x



Transforming time series



Correlation transformation I

Basic idea: Find the ACS of the parent-Gaussian process

From Z to X we use: $X(t) = g\big(Z(t)\big) = Q_X\left(\Phi_Z\big(Z(t)\big)\right)$

From X to Z we use: $Z(t) = g^{-1}\big(X(t)\big) = Q_Z\left(F_X\big(X(t)\big)\right)$

We can estimate the bivariate $f_{X(t)X(\tau)}\big(x(t), x(\tau)\big)$ by transforming the the bivariate Normal Distribution $\varphi_{Z(t)Z(\tau)}\big(z(t), z(\tau); \rho_Z(\tau)\big)$ which relates two Gaussian RV's that are correlated by $\rho_Z(\tau)$

$$f_{X(t)X(\tau)}\big(x(t), x(\tau)\big) = \varphi_{Z(t)Z(\tau)}\left(g^{-1}\big(x(t)\big), g^{-1}\big(x(\tau)\big); \rho_Z(\tau)\right)\mathbf{J}\big(x(t), x(\tau)\big)$$

$$\mathbf{J}\big((x(t), (x(\tau)\big) = \partial g^{-1}(x(t))/\partial x(t)\partial g^{-1}(x(\tau))/\partial x(\tau)$$

$$\mathcal{C}\big(\boldsymbol{\theta}_X, \rho_Z(\tau)\big) = E\big(X(t)X(\tau)\big) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x(t)x(\tau)f_{X(t)X(\tau)}\big(x(t), x(\tau); \rho_Z(\tau)\big)\,\mathrm{d}x(t)\mathrm{d}x(\tau)$$

$$\rho_X(\tau) = \mathcal{R}\big(\boldsymbol{\theta}_X, \rho_Z(\tau)\big) := \frac{\mathcal{C}\big(\boldsymbol{\theta}_X, \rho_Z(\tau)\big) - \mu_X^2}{\sigma_X^2}$$

Correlation transformation II

Basic idea: Find the ACS of the parent-Gaussian process

From Z to X we use the Transformation: $X(t) = Q_X\left(\Phi_Z(Z(t))\right)$

The bivariate Normal Distribution $\varphi_{Z(t)Z(\tau)}\left(z(t), z(\tau); \rho_Z(\tau)\right)$ relates two Gaussian RV's that are correlated by $\rho_Z(\tau)$. Using the mean value theorem of the Transformation of a RV we can create a link between $\rho_X(\tau)$ and $\rho_Z(\tau)$.

$$\mathcal{C}\left(\boldsymbol{\theta}_X, \rho_Z(\tau)\right) \coloneqq E\left(X(t)X(\tau)\right) = E\left(Q_X\left(\Phi_Z(Z(t))\right) Q_X\left(\Phi_Z(Z(\tau))\right)\right)$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} Q_X\left(\Phi_Z(z(t)\right) Q_X\left(\Phi_Z(z(\tau)\right) \varphi_{Z(t)Z(\tau)}\left(z(t), z(\tau); \rho_Z(\tau)\right) \mathrm{d}z(t)\mathrm{d}z(\tau)$$

$$\rho_X(\tau) = \mathcal{R}\left(\boldsymbol{\theta}_X, \rho_Z(\tau)\right) \coloneqq \frac{\mathcal{C}\left(\boldsymbol{\theta}_X, \rho_Z(\tau)\right) - \mu_X^2}{\sigma_X^2}$$

Intermittency

- Several natural processes like precipitation at fine temporal scales (e.g., at daily or subdaily scales), discharge of small streams, or even wind, are intermittent processes.
- This means that their marginal distribution is of mixed-type.

$$F_X(x) = (1 - p_0) F_{X|X>0}(x) + p_0 \quad x \geq 0 \tag{1}$$

$$f_X(x) = \begin{cases} p_0 & x = 0 \\ (1 - p_0) f_{X|X>0}(x) & x > 0 \end{cases} \tag{2}$$

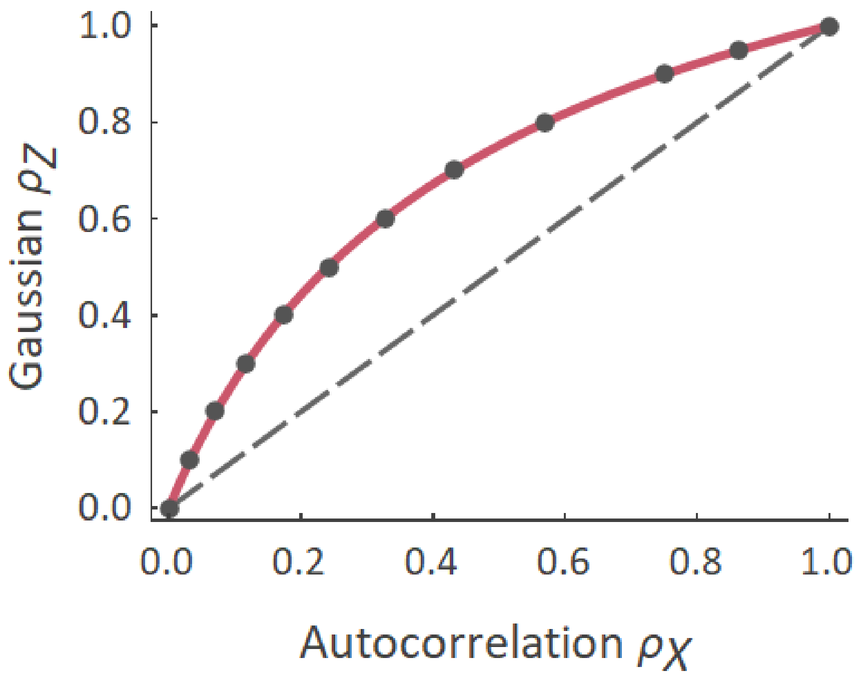
$$x_u = Q_X(u) = \begin{cases} 0 & 0 \leq u \leq p_0 \\ Q_{X|X>0}\left(\dfrac{u - p_0}{1 - p_0}\right) & p_0 < u \leq 1 \end{cases} \tag{3}$$

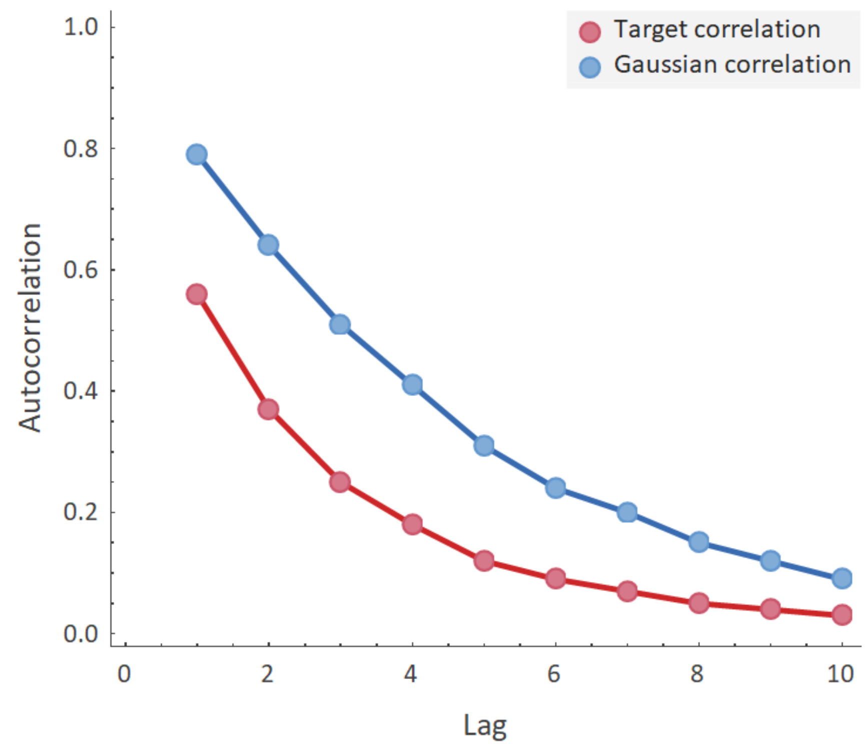
$$\mu_X = (1 - p_0)\mu_{X|X>0} \tag{4}$$

$$\sigma_X^2 = (1 - p_0)\sigma_{X|X>0}^2 + p_0(1 - p_0)\mu_{X|X>0}^2 \tag{5}$$

ACS transformation

So how can we find corresponding Gaussian ASC?

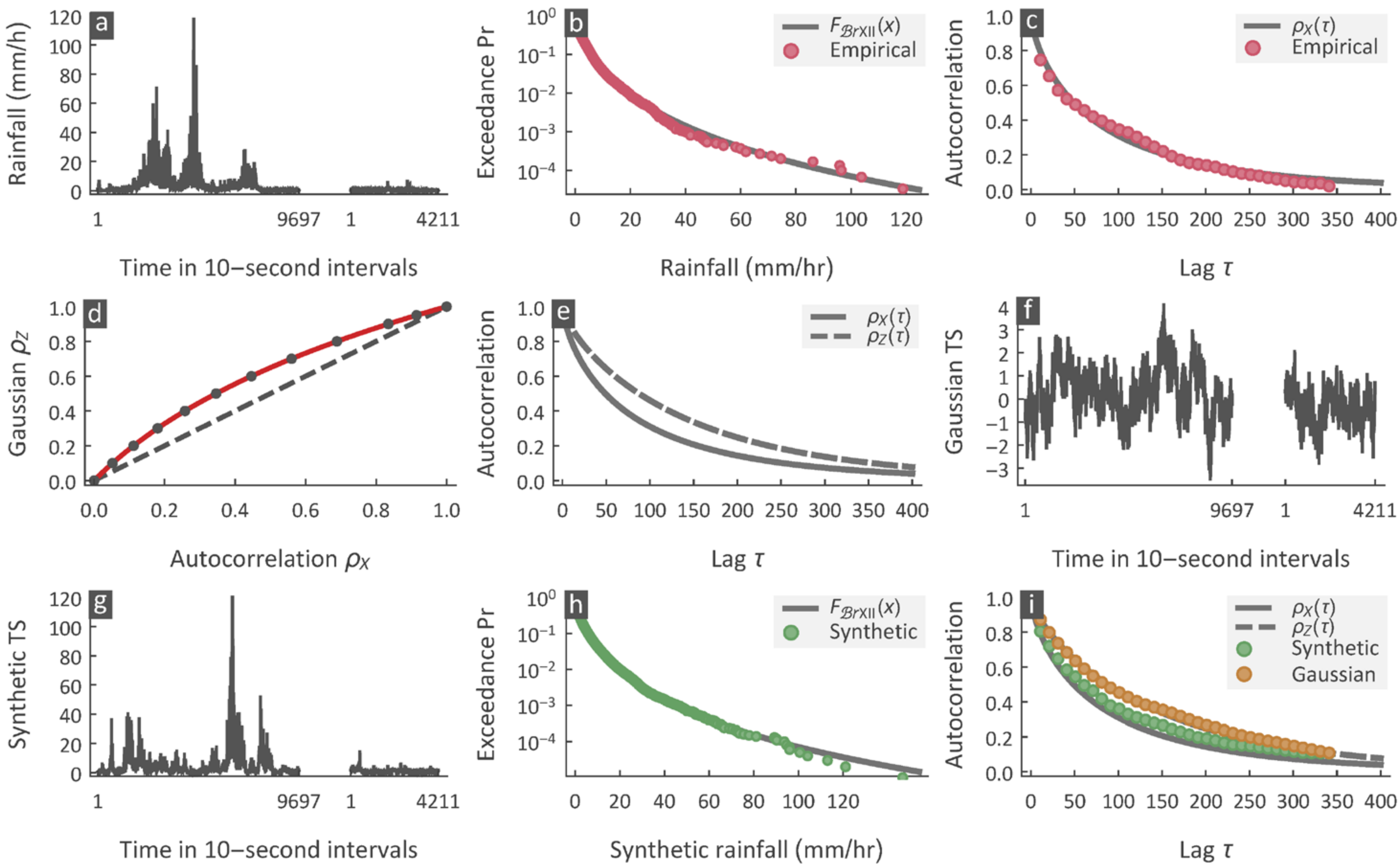




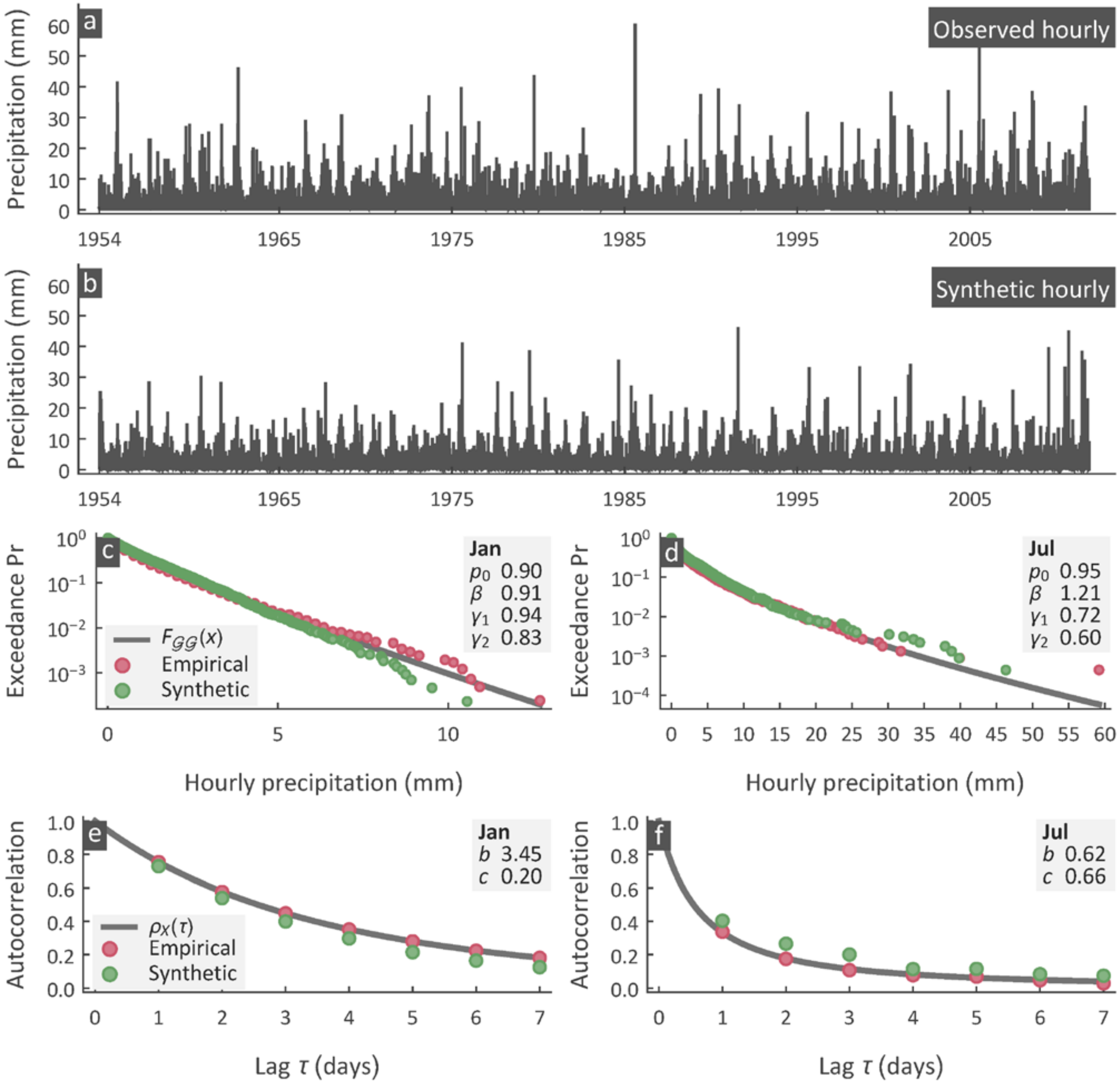
Target	0.56	0.37	0.25	0.18	0.12	0.09	0.07	0.05	0.04	0.03
Gaussian	0.79	0.64	0.51	0.41	0.31	0.24	0.2	0.15	0.12	0.09

Instead of values we could use functions!

Unified framework | Graphical summary



Hourly rainfall simulation

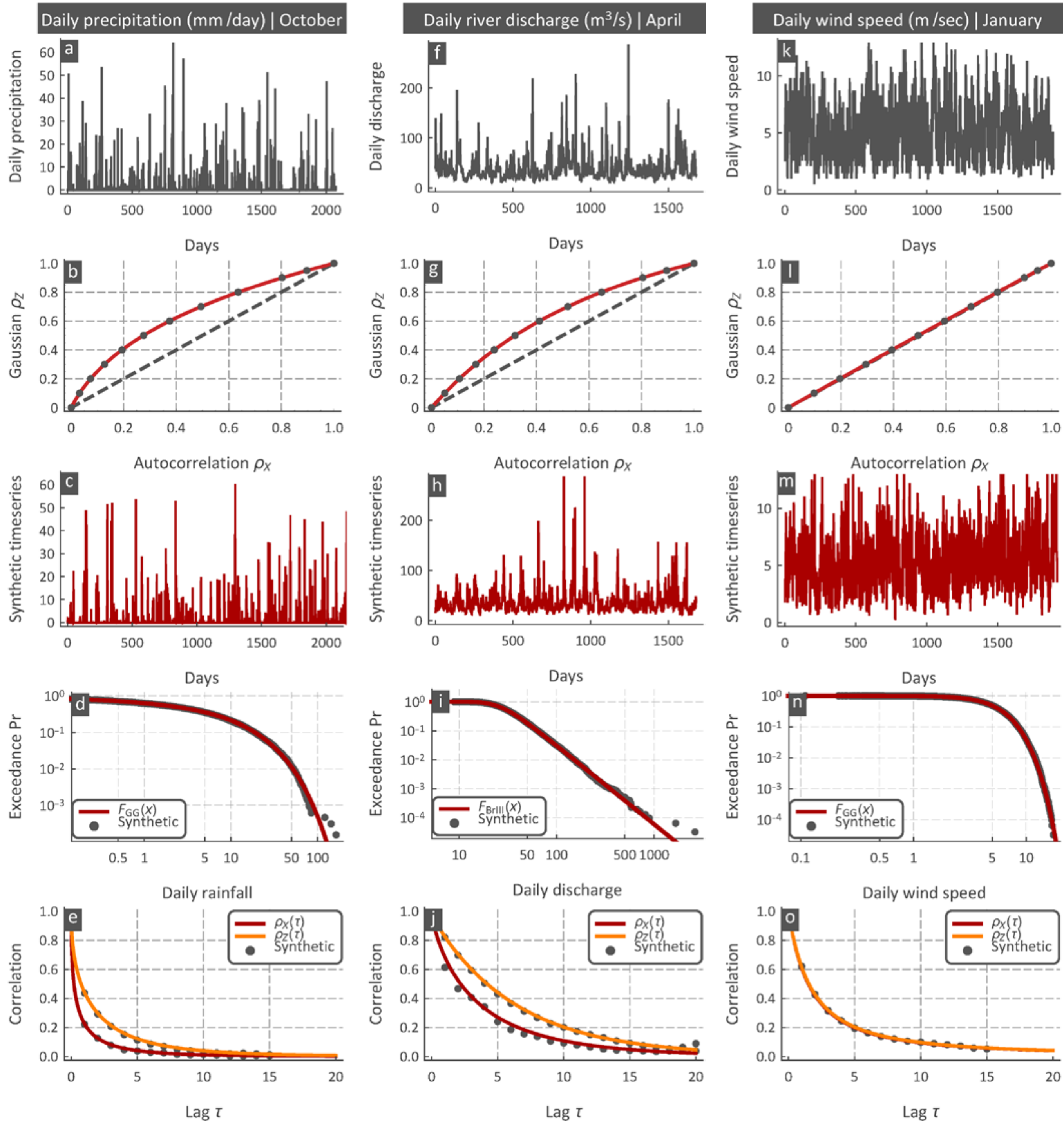


Real World Examples I

Daily precipitation
Mixed-type marginal
p_0: 78%
PDF: $\mathcal{GG}(16.5, 0.39, 0.97)$
ACS: $\rho_{\mathrm{W}}(\tau; 0.43, 0.48)$

Daily river discharge
Continuous marginal
p_0: 0%
PDF: $\mathcal{Br}\mathrm{III}(40.5, 12.6, 0.37)$
ACS: $\rho_{\mathrm{W}}(\tau; 3.5, 0.79)$

Daily wind speed
Continuous marginal
p_0: 0%
PDF: $\mathcal{GG}(4.4, 2.66, 1.76)$
ACS: $\rho_{\mathrm{PII}}(\tau; 1.7, 0.68)$

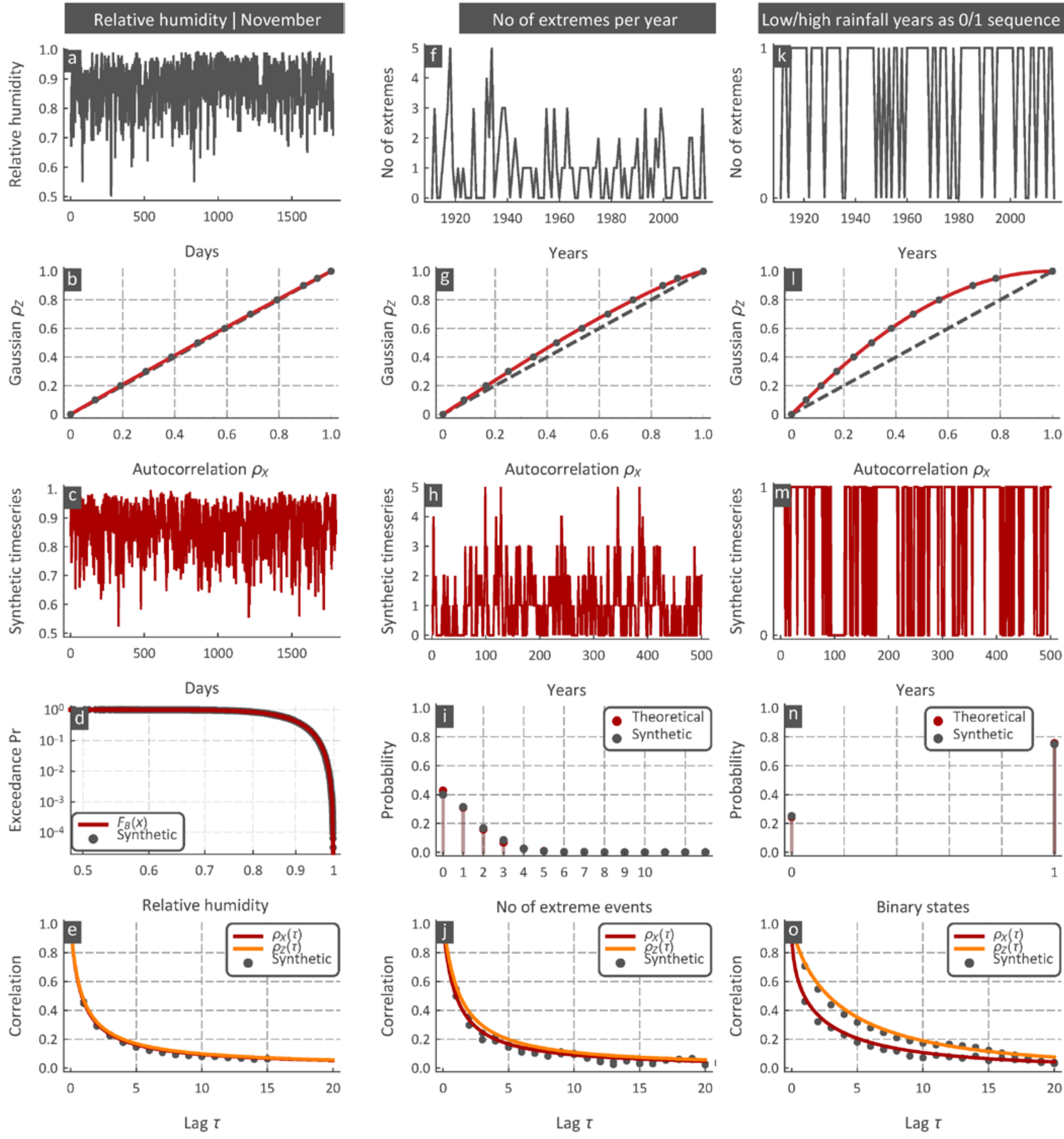


Real World Examples II

Relative humidity
Bounded in [0,1] marginal
p_0: 0%
PDF: $\mathcal{B}(16.1, 2.3)$
ACS: $\rho_{\text{PII}}(\tau; 0.80, 1, 16)$

Extremes per year
Discrete marginal
p_0: 0%
PMF: $\mathcal{PA}(0.85, 0.15)$
ACS: $\rho_{\text{PII}}(\tau; 1, 1)$

Low/high rainfall years
Binary
PMF: $p_0 = 0.25, p_1 = 0.75$
ACS: $\rho_{\text{W}}(\tau; 2, 0.5)$



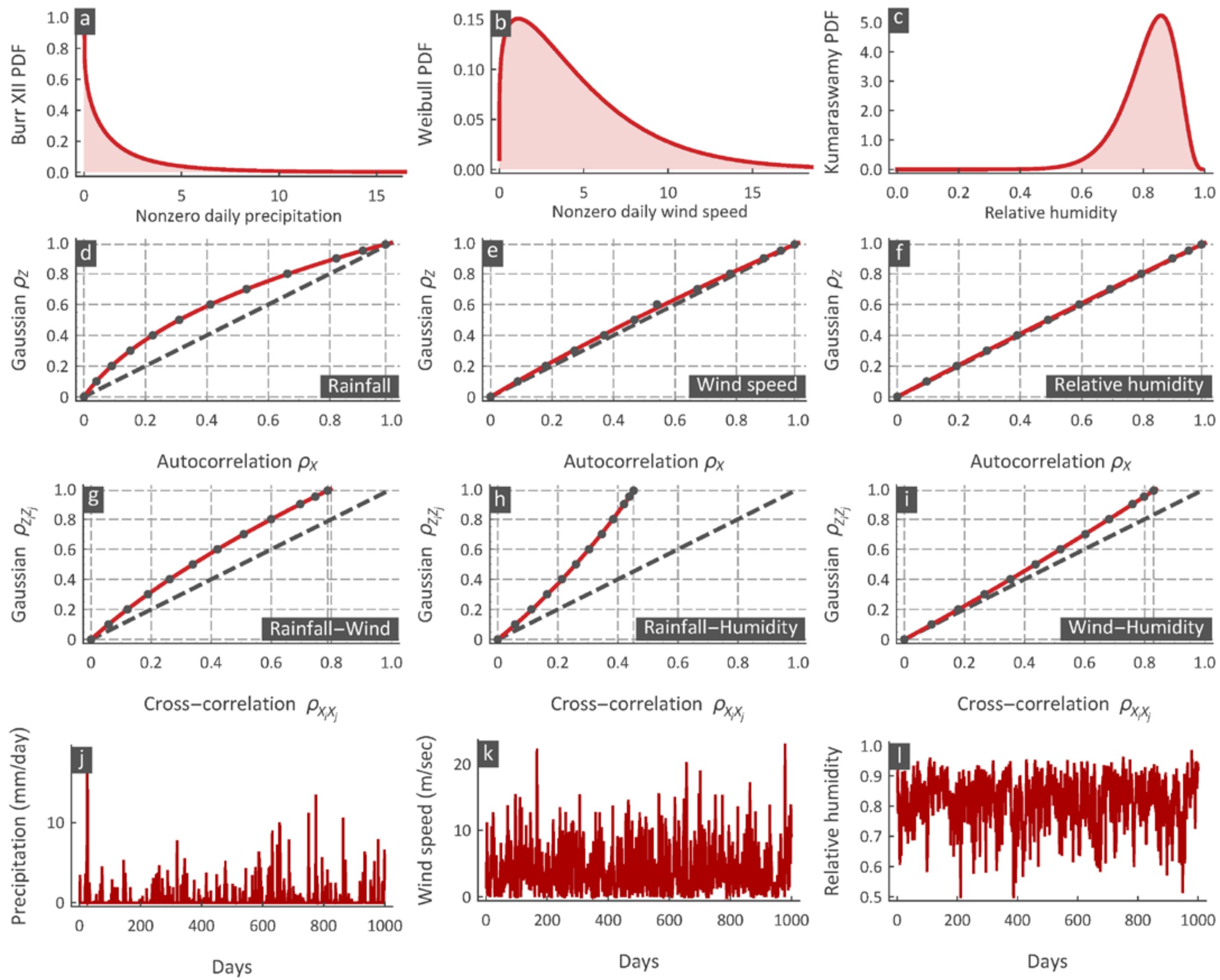
A multivariate case

Simulation of daily precipitation, wind speed and relative humidity.

The lag-0 and lag-1 correlation matrices are:

$$\boldsymbol{K}_X(0) = \begin{pmatrix} 1 & 0.50 & 0.35 \\ 0.50 & 1 & 0.60 \\ 0.35 & 0.60 & 1 \end{pmatrix}$$

$$\boldsymbol{K}_X(1) = \begin{pmatrix} 0.30 & 0.25 & 0.15 \\ 0.10 & 0.40 & 0.35 \\ 0.12 & 0.30 & 0.50 \end{pmatrix}$$



Details in...

Advances in Water Resources 115 (2018) 234–252

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

ELSEVIER

Unified theory for stochastic modelling of hydroclimatic processes: Preserving marginal distributions, correlation structures, and intermittency

Check for updates

Simon Michael Papalexiou

Department of Civil and Environmental Engineering, University of California, Irvine, CA, USA

A R T I C L E   I N F O

Article history:
Received 25 October 2017
Revised 14 February 2018
Accepted 14 February 2018
Available online 15 February 2018

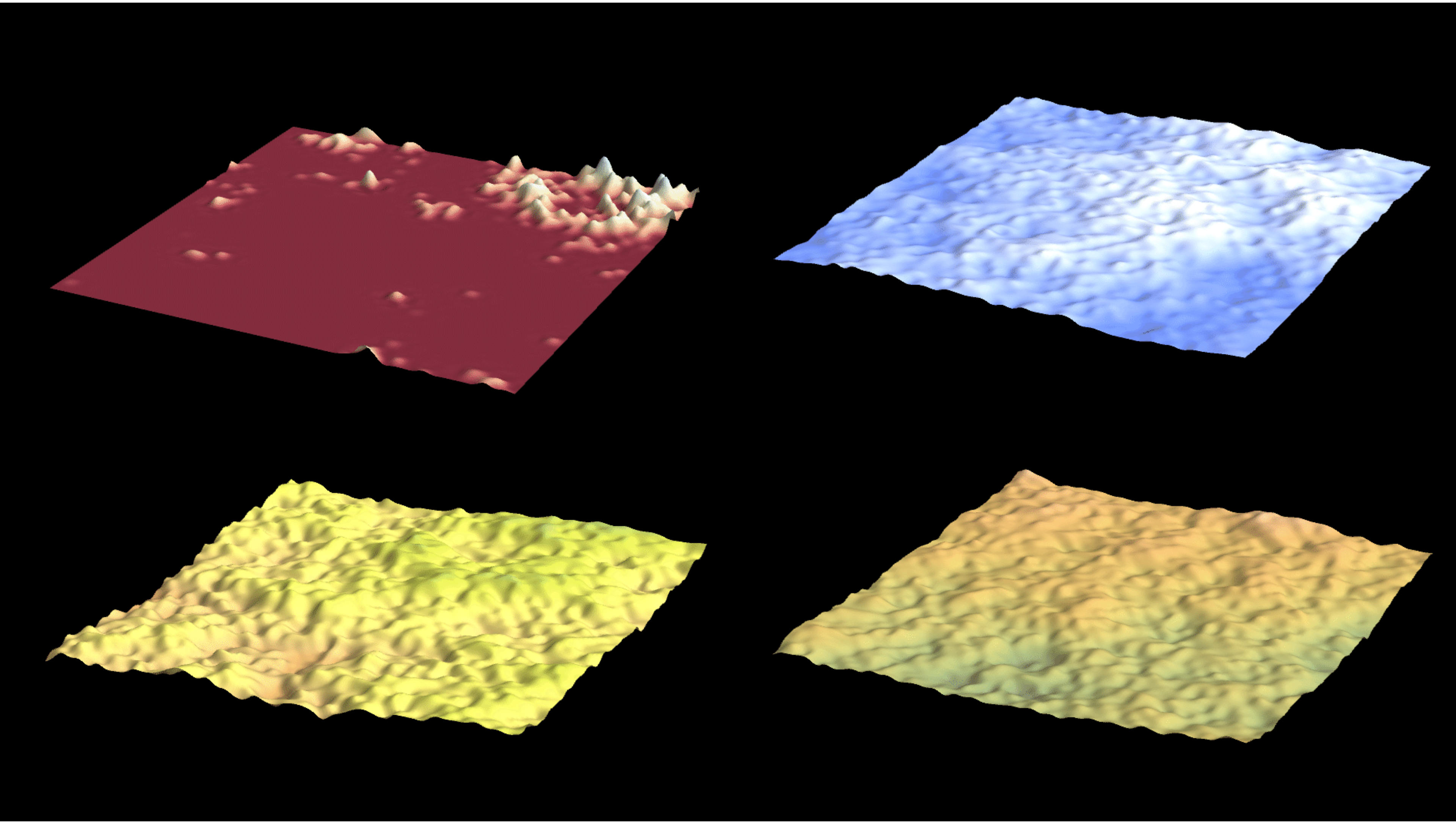
Keywords:
Stochastic modelling
Weather generator
Parent-Gaussian framework

A B S T R A C T

Hydroclimatic processes come in all “shapes and sizes”. They are characterized by different spatiotemporal correlation structures and probability distributions that can be continuous, mixed-type, discrete or even binary. Simulating such processes by reproducing precisely their marginal distribution and linear correlation structure, including features like intermittency, can greatly improve hydrological analysis and design. Traditionally, modelling schemes are case specific and typically attempt to preserve few statistical moments providing inadequate and potentially risky distribution approximations. Here, a single framework is proposed that unifies, extends, and improves a general-purpose modelling strategy, based on the assumption that any process can emerge by transforming a specific “parent” Gaussian process. A novel mathematical representation of this scheme, introducing parametric correlation transformation

https://doi.org/10.1016/j.advwatres.2018.02.013

Random Fields



Details in…

Water Resources Research

RESEARCH ARTICLE
10.1029/2019WR026331

Random Fields Simplified: Preserving Marginal Distributions, Correlations, and Intermittency, With Applications From Rainfall to Humidity

Simon Michael Papalexiou[1,2,3] **and Francesco Serinaldi**[4,5]

[1]Department of Civil, Geological and Environmental Engineering, University of Saskatchewan, Saskatoon, Saskatchewan, Canada, [2]Global Institute for Water Security, Saskatoon, Saskatchewan, Canada, [3]Faculty of Environmental Sciences, Czech University of Life Sciences Prague, [4]School of Engineering, Newcastle University, Newcastle, Upon Tyne, UK, [5]Willis Research Network, London, UK

Key Points:
- Versatile and easy modeling and simulation of spatiotemporal random fields
- Generated fields reproduce any continuous, discrete, and mixed-type marginal distribution, any correlation structure, and intermittency
- New spatiotemporal correlation structures based on copulas and survival functions

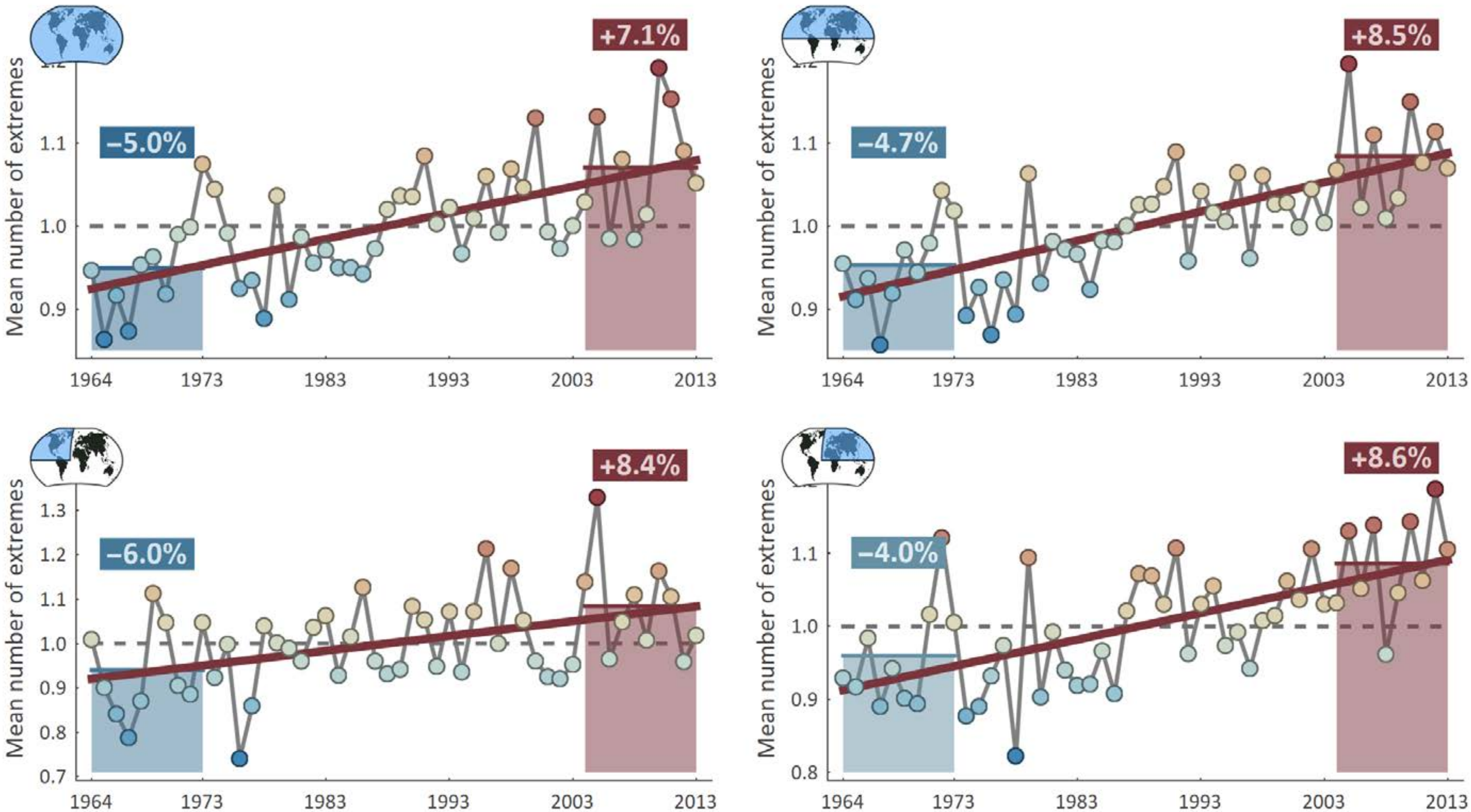
Correspondence to:
S. M. Papalexiou,
sm.papalexiou@usask.ca

Citation:
Papalexiou, S. M., & Serinaldi, F. (2020). Random fields simplified: Preserving marginal distributions, correlations, and intermittency, with applications from rainfall to humidity. *Water Resources Research*, *56*, e2019WR026331. https://doi.org/10.1029/2019WR026331

Abstract Nature manifests itself in space and time. The spatiotemporal complexity of processes such as precipitation, temperature, and wind, does not allow purely deterministic modeling. Spatiotemporal random fields have a long history in modeling such processes, and yet a single unified framework offering the flexibility to simulate processes that may differ profoundly does not exist. Here we introduce a blueprint to efficiently simulate spatiotemporal random fields that preserve any marginal distribution, any valid spatiotemporal correlation structure, and intermittency. We suggest a set of parsimonious yet flexible marginal distributions and provide a rule of thumb for their selection. We propose a new and unified approach to construct flexible spatiotemporal correlation structures by combining copulas and survival functions. The versatility of our framework is demonstrated by simulating conceptual cases of intermittent precipitation, double-bounded relative humidity, and temperature maxima fields. As a real-word case we simulate daily precipitation fields. In all cases, we reproduce the desired properties. In an era characterized by advances in remote sensing and increasing availability of spatiotemporal data, we deem that this unified approach offers a valuable and easy-to-apply tool for modeling complex spatiotemporal processes.

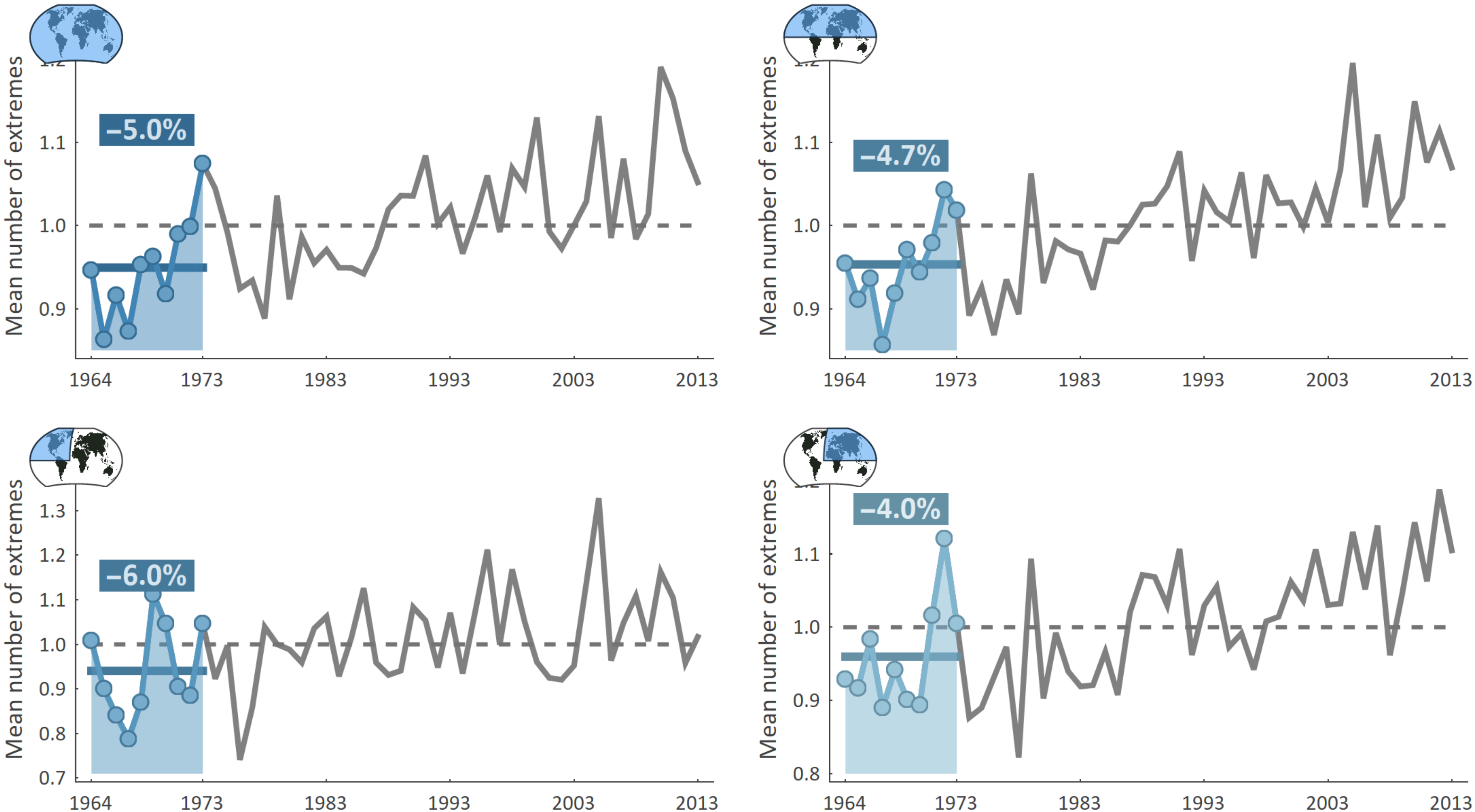
https://doi.org/10.1029/2019WR026331

A quick comment on changes

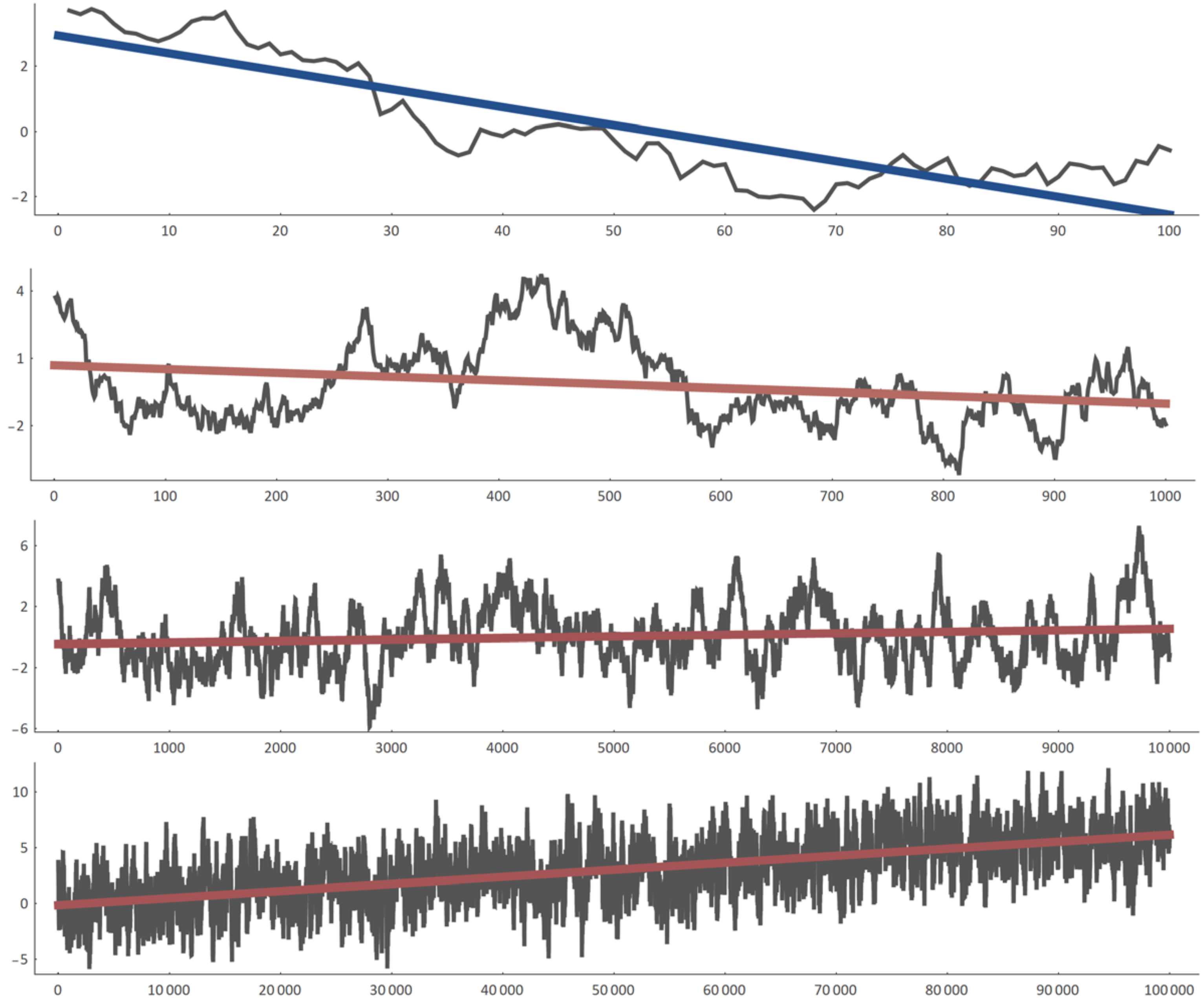


Papalexiou, S. M., & Montanari, A. (2019). Global and Regional Increase of Precipitation Extremes under Global Warming. *Water Resources Research*. https://doi.org/2018WR024067

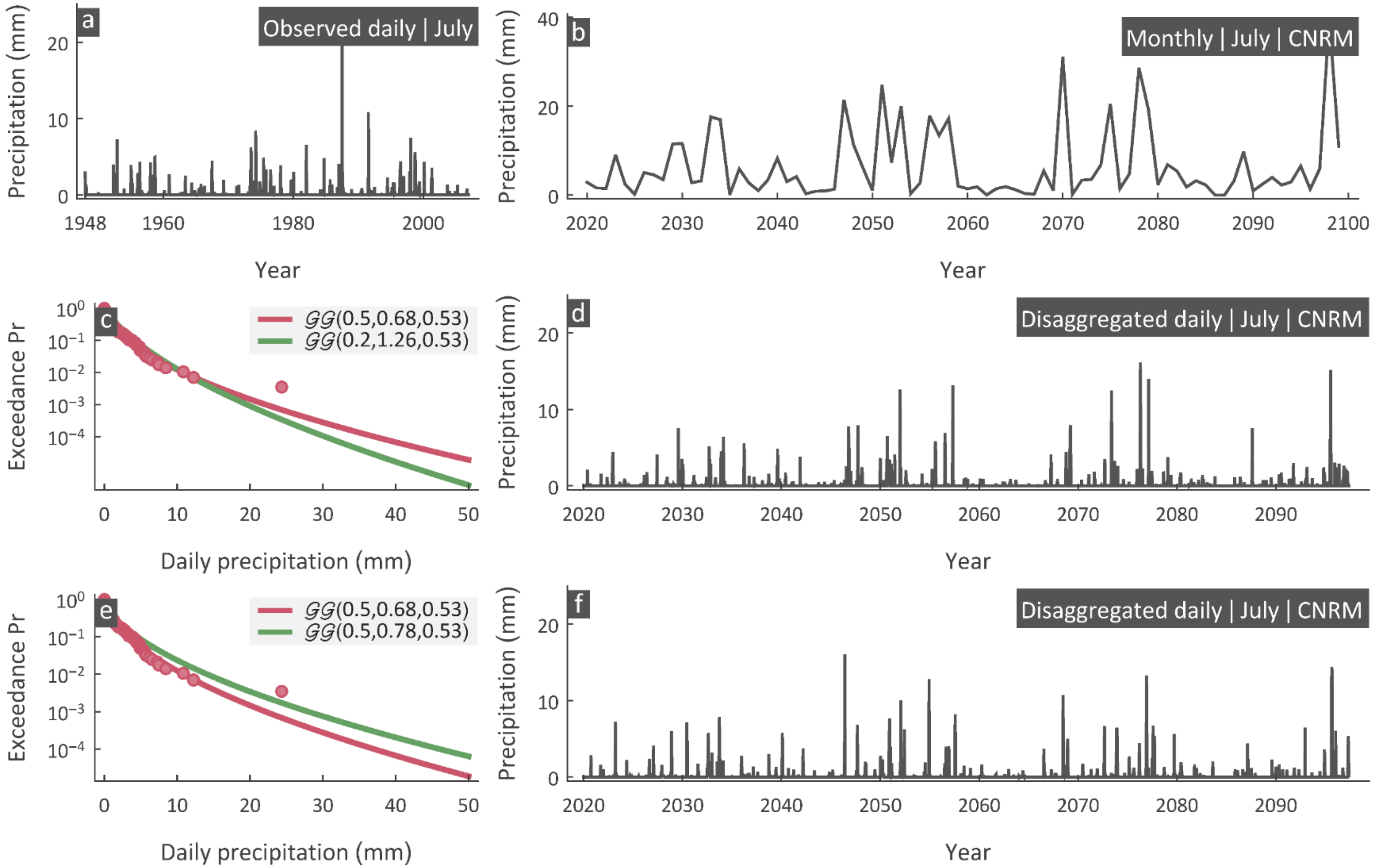
Changes per decade



Quiz: find the Future



Non-stationary simulation



(c) 40% linear increase in the mean value of nonzero precipitation
(e) assuming 40% linear increase in the mean value and 30% in the standard deviation

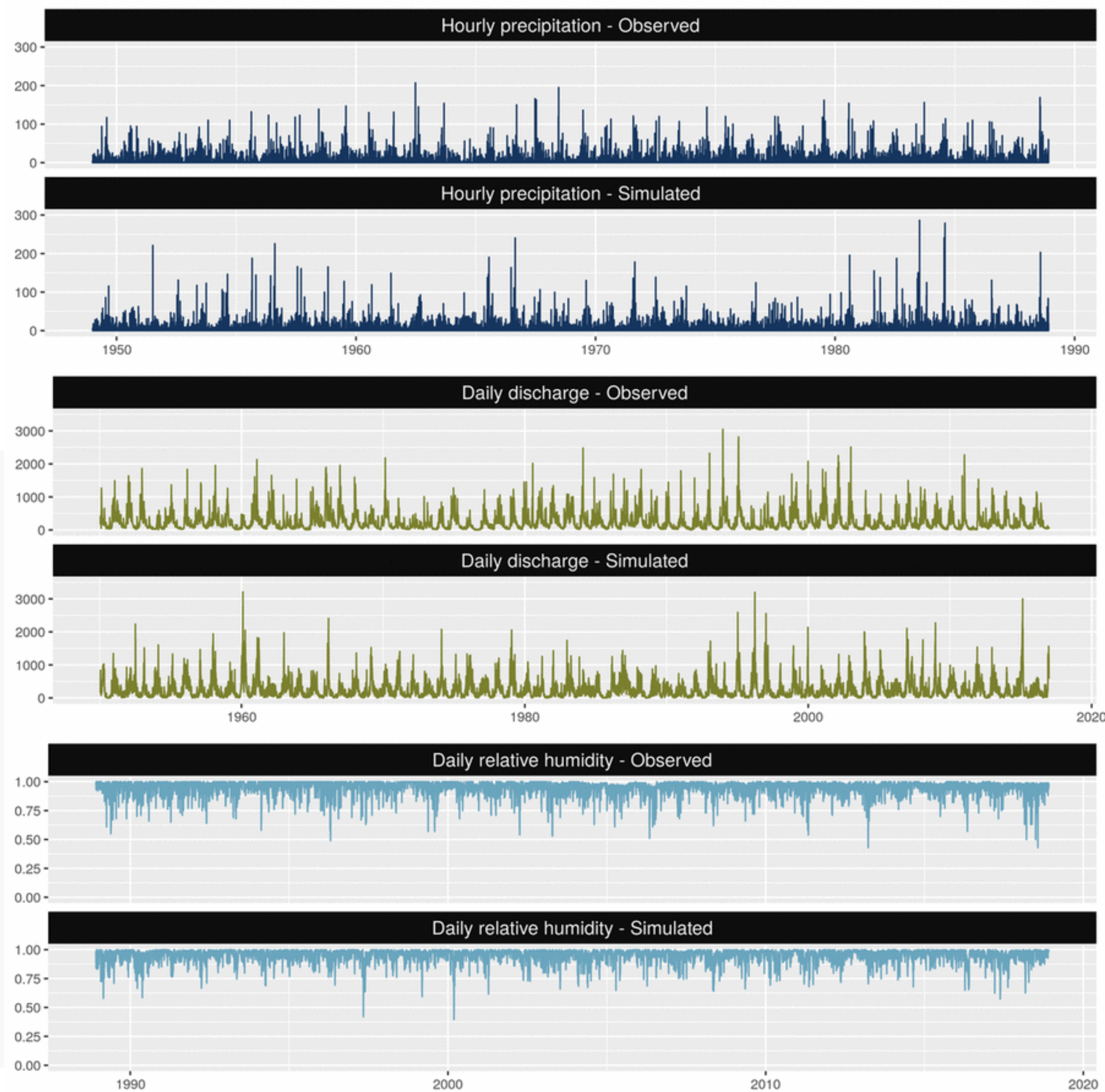
From Papalexiou et al. (2018). Precise Temporal Disaggregation Preserving Marginals and Correlations (DiPMaC) for Stationary and Nonstationary Processes. *Water Resources Research*. https://doi.org/10.1029/2018WR022726

CoSMoS-R… Makes it Easy!

…just use our R-package!

CoSMoS: Complete Stochastic Modeling Solution

Just search for CoSMoS in CRAN and install it like any other R-package.

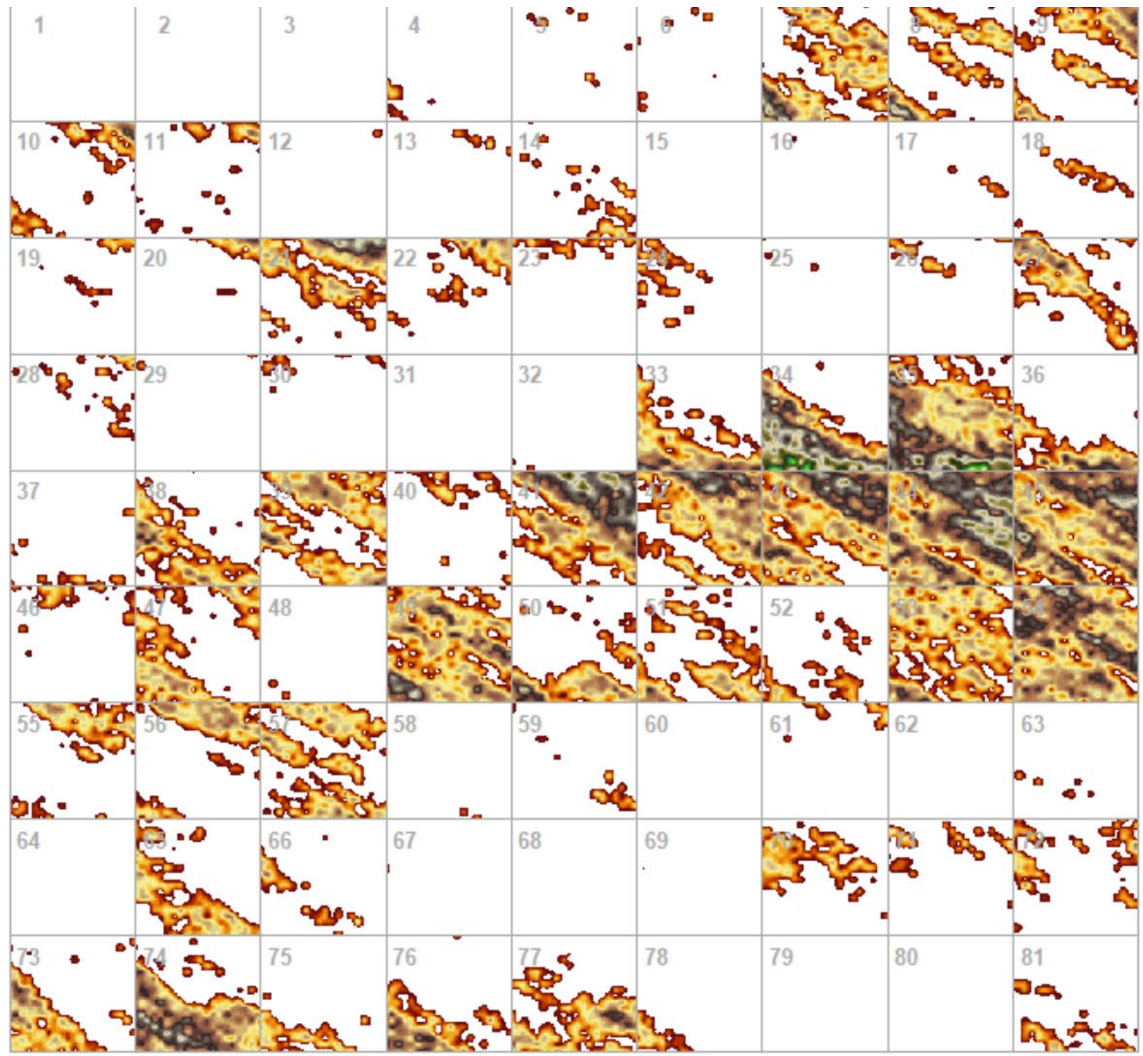


CoSMoS-R… Makes it Easy!

…just use our R-package!

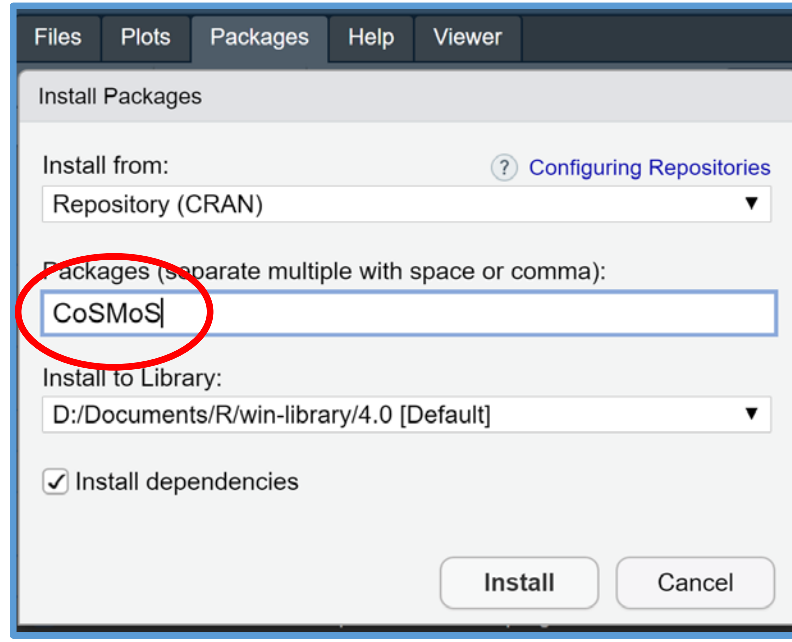
CoSMoS: Complete Stochastic Modeling Solution

Just search for CoSMoS in CRAN and install it like any other R-package.



Install CoSMoS-R…

CoSMoS initially conceived back in 2009.
It's a long story… more details here.



To install

1. Open RStudio
2. Type CoSMoS
3. Click Install

Alternatively, just copy-paste the following text in Rstudio:

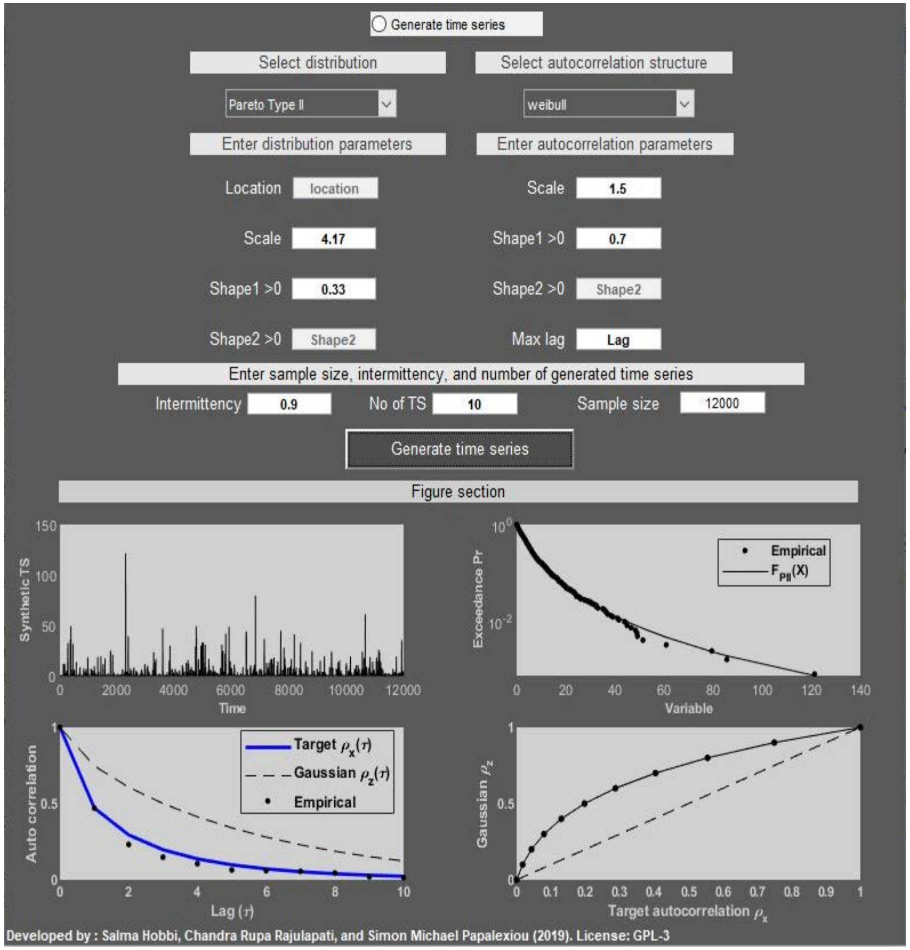
```
if (!require('devtools')) {install.packages('devtools'); library(devtools)}
install_github('TycheLab/CoSMoS', upgrade = 'never', build_vignettes = TRUE)
library(CoSMoS)
?`CoSMoS-package`
```

To check documentation with many examples copy-paste and run this:

```
RShowDoc("vignette", type = c("html"), "CoSMoS")
```

Or just click the link:
https://cran.r-project.org/web/packages/CoSMoS/vignettes/vignette.html

There’s a CoSMoS-MATLAB



CoSMoS MATLAB available at:

- MathWorks
- GitHub
  https://github.com/SMPLab

- Super simple GUI!
- Select the probability distribution and autocorrelation structure from the drop-down lists.
- Enter the parameters of selected distribution and autocorrelation structure.
- Enter the intermittency value (as probability zero), sample size (time series length), and number of time series you wish to generate.
- Click the “Generate time series” button and that’s it!

So, targets for next year…

1. Have fun - *really important!*
2. Explore various distribution focusing on their tail behavior
3. Try many fitting methods and learn when to use the right one
4. Learn how to generate time series and random fields to mimic reality
5. Explore spatiotemporal dynamics of extremes
6. Do a large-scale research study and write a paper

…and most importantly HAVE FUN!

Keep Calm
And
Generate Time Series
Preserving Extremes!

Thank you!