Hydrological risk assessment:
Return period and probability of failure

E. Volpi

elena.volpi@uniroma3.it

Dept. of Engineering, Roma Tre University, Rome, Italy

Objectives

The return period concept was first introduced by **Fuller (1914)** – who pioneered statistical flood frequency analysis in USA – to **quantify the rareness of hydrologic events** (e.g. floods, droughts, etc.) that might cause huge damages to the society and the environment.

Despite well-established in the literature, the return period concept has recently attracted renewed interest stimulated by the need of **efficiently dealing with complex processes in a *changing* environment**.

This lecture presents the concept of return period and the related probability of failure by making use of a *general mathematical framework* with the aim to:

(i) help for a better understanding of the return period formulation that is commonly adopted in practical engineering applications;

(ii) be applied under more general conditions, i.e. by relaxing the hypotheses of stationarity and independence usually (sometimes implicitly) assumed in practical problems to derive simple analytic formulations;

(iii) exploit all the information provided by rich (?) data sets.

➢ Fuller, W. (1914), Flood flows. Trans. Am. Soc. Civ. Eng., 77, 564-617

Outline

1. Basics of return period and probability of failure: general definitions and possible issues
2. General mathematical framework
3. Stationary and independent processes
4. Non-stationary or time-dependent processes
5. Remarks on return period estimation
6. Summary and conclusions

1.1 Basics of return period

The return period is a probabilistic concept used to measure and communicate the random occurrence of geophysical events that may produce huge economic, social and environmental damages.

Under some regularity conditions (stationarity and independence), once the random variable (Z) quantifying the event of interest is identified, with $P_Z(z) = \Pr\{Z \leq z\}$ its probability distribution function (cdf), the return period of a possible dangerous value, $T(z)$ (years), is

$$T(z) \sim \frac{1}{1 - P_Z(z)} \tag{1}$$

The threshold value that corresponds to T is generally named as the T-year event

$$z_T = P_Z^{-1}\left(1 - \frac{1}{T}\right)$$

- The probability that the T-year event is exceeded is $1/T$ in every year, i.e. the T-year event occurs on average every T years.

- Although time is generally more understandable than probability to a general audience, the rationale behind return period was often misunderstood to mean that one T-year event should occur exactly every T year **(Stedinger et al., 1993)**.

- Stedinger, J. R., Vogel, R. M., & Foufoula-Georgiou, E. (1993). Frequency analysis of extreme events. In D. Maidment (Ed.), *Handbook of hydrology* (chap. 18). New York: McGraw-Hill

1.2 Basics of probability of failure

The classical approach of risk management is that of designing hydraulic structures to control up hazardous events to a predefined T-year event.

The level of protection (i.e. selected T value) is determined by broad considerations of risk conditions:

- the expected damage that occurs or will be exceeded with a certain probability in a certain time period, e.g. one year (Vogel and Castellarin, 2017).

The probability that the T-year event is exceeded is specified period of time (e.g. the design life l) that is **the probability of failure** is

$$R(z,l) = 1 - \left(1 - \frac{1}{T(z)}\right)^{l} \qquad (2)$$

- the probability of failure due to the T-year event reaches the non-negligible value of about 63% after T years (i.e., for $l = T$) for large events (e.g., $T > 10$ years).

➢ Vogel, R., & Castellarin, A. (2017). Risk, Reliability, and Return Periods and Hydrologic Design. In V. Singh (Ed.), Handbook of applied hydrology (chap. 78). New York:McGraw-Hill.

1.3 Return period estimation

Although the concept and application of return period are well established in the literature, estimating T and the T-year event is still challenging (especially in the *extrapolation range*).

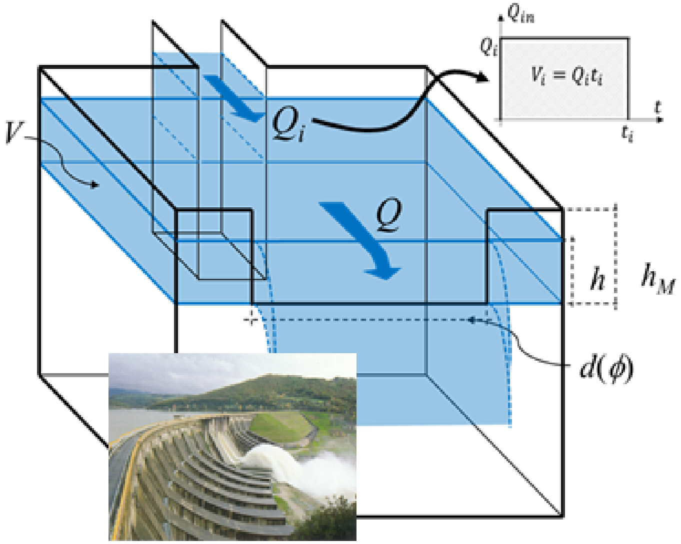
General estimation procedure

1. Identify the mechanism(s) of system failure that determine the random event of interest, and subsequently the random variable describing it (Z) (Schumann, 2017):
   i. **single random variable** describing the hydrological load and, as a consequence, the state of the system (univariate case);
   ii. joint behavior of **several random variables** (whose pairwise correlation is in general not negligible) that are all relevant to risk analyses (Brunner et al., 2016)
2. Collect (or derive from observations) time series of the random variable of interest.
3. Fit a probability distribution function (a parametric or non-parametric model) to the underlying variable and determine $T(z)$ and/or the T-year event (Coles, 2002; Gaume, 2018)

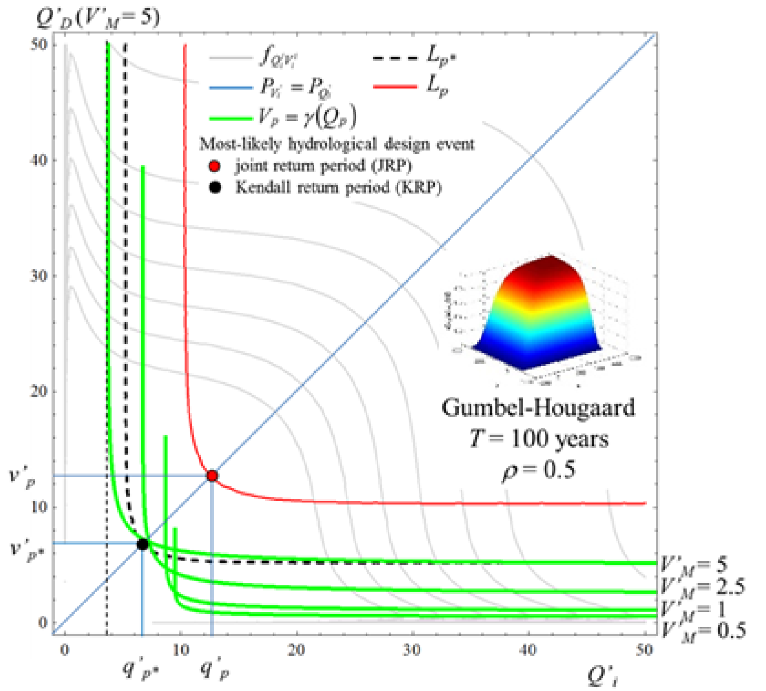
- Schumann, A. (2017). Flood safety versus remaining risks - Options and limitations of probabilistic concepts in flood management. Water Resources Management, 31 (10), 3131-3145, doi: 10.1007/s11269-017-1700-z
- Brunner, M. I., Seibert, J., & Favre, A.-C. (2016). Bivariate return periods and their importance for flood peak and volume estimation. Wiley Interdisciplinary Reviews: Water, 3 (6), 819-832, doi: 10.1002/wat2.1173
- Coles, S. (2001). An introduction to statistical modeling of extreme values. London: Springer.
- Gaume, E. (2018). Flood frequency analysis: The Bayesian choice. Wiley Interdisciplinary Reviews: Water, 5 (4), e1290. doi: 10.1002/wat2.1290

Return period in a multivariate framework

When the hydrological load acting on the system is characterized by two (or more) random variables those can be combined to define different types of event (e.g., joint occurrence: AND event etc.).



(Volpi and Fiori, 2014)



In a multivariate framework:

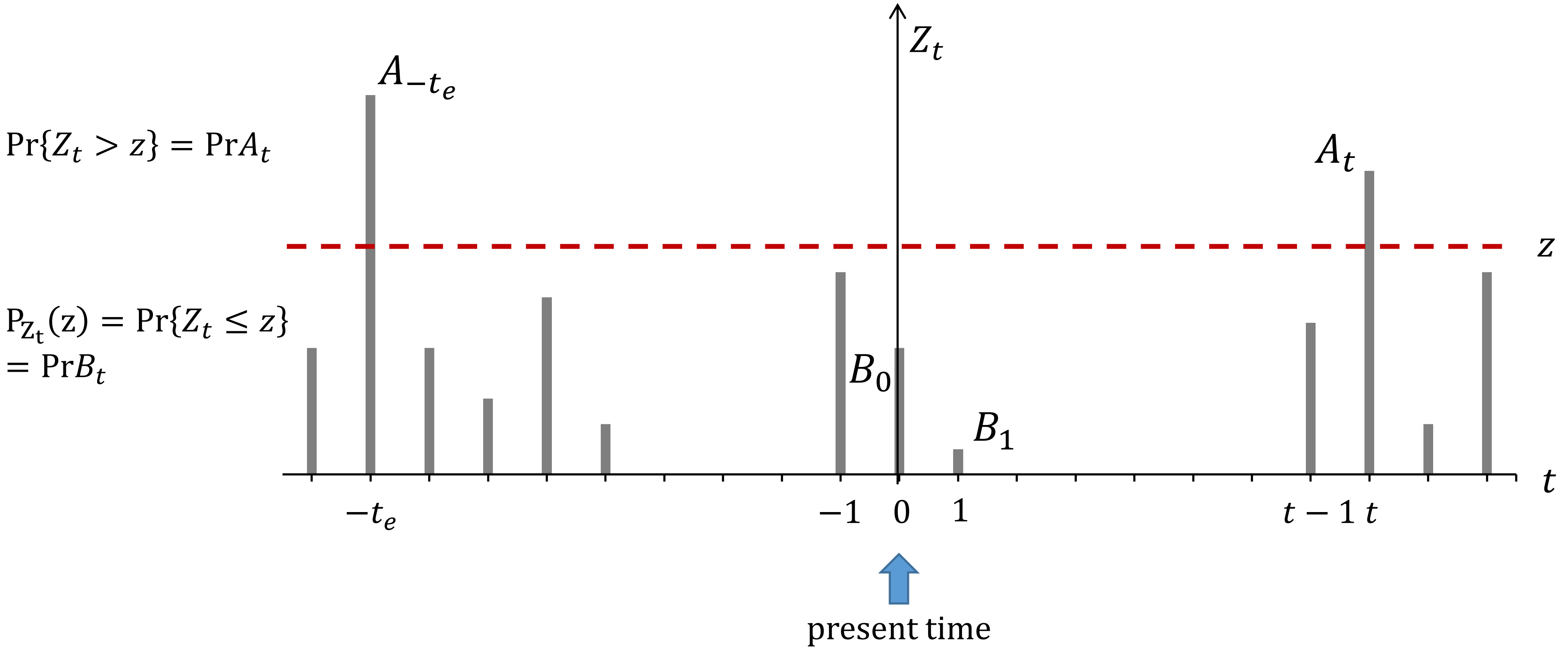
- the return period of system failure does not generally correspond to that of the hydrological load (e.g. of the AND event);
- system failure should be quantified in probabilistic terms by determining the proper combination of the random variables describing the interaction between the hydrological load and the system.

➢ Volpi, E., & Fiori, A. (2014). Hydraulic structures subject to bivariate hydrological loads: Return period, design, and risk assessment. Water Resources Research(50), 885-897, doi: 10.1002/2013WR014214

2.1 General mathematical framework

Let's consider a discrete-time stochastic process Z_t that is obtained by sampling a natural process evolving in continuous time at constant time intervals $\Delta\tau$ (Koutsoyiannis, 2016)

We are interested in the occurrence of events that might cause the failure of a structure or system, e.g. $A_t = \{Z_t > z\}$ at time t.



➢ Koutsoyiannis, D. (2016). Generic and parsimonious stochastic modelling for hydrology and beyond. Hydrological Sciences Journal, 61 (2), 225-244, doi: 10.1080/02626667.2015.1016950

2.2 General mathematical framework: Definitions

The **return period** is the time interval quantifying the average occurrence of the dangerous event A; it is a means of expressing the exceedance probability of A in terms of time units $\Delta\tau$, typically years

$$\frac{T}{\Delta\tau} = \mathrm{E}[X] = \sum_{x=1}^{\infty} x\, f_X(x)$$

where $f_X(x) = \Pr\{X = x\}$ is its probability mass function.

The **probability of failure** $R(l)$ is defined as the probability that a dangerous event occur in the period of time l

$$R(z, l) = \Pr\{X \leq l\} = \sum_{x=1}^{l} f_X(x)$$

X may be defined as (e.g., Fernández and Salas, 1999):

(i) the **waiting time** to the next event;

(ii) the **interrarrival time** between successive events.

➢ Fernández, B., & Salas, J. D. (1999). Return period and risk of hydrologic events. I: mathematical formulation. Journal of Hydrologic Engineering, 4 (4), 297-307, doi:10.1061/(ASCE)1084-0699(1999)4:4(297)

2.3.1 General mathematical framework: waiting time, f_W

Probability mass function, $f_X(x)$

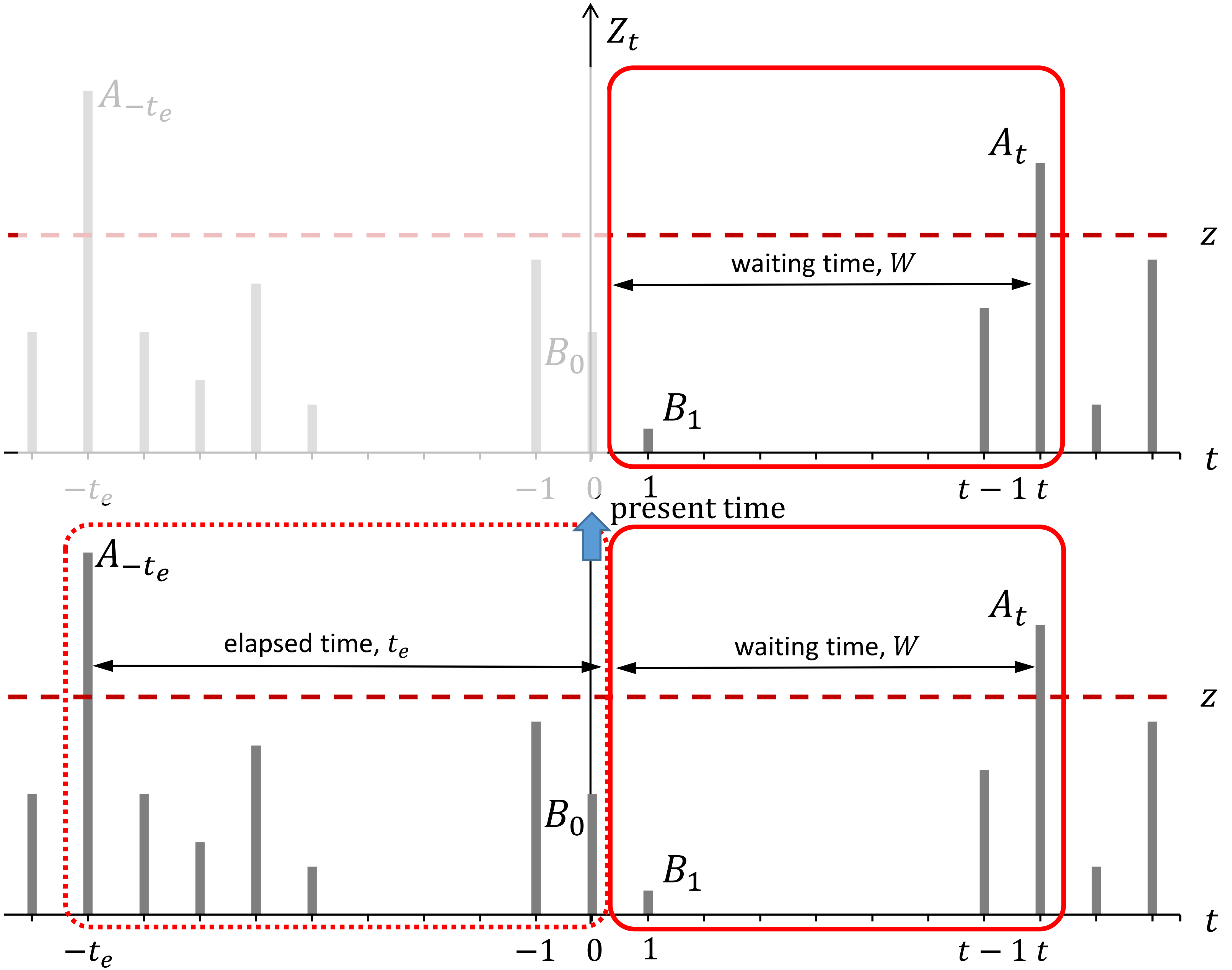
Unconditional waiting time, W

$f_W(t) = \Pr(B_1, B_2, .. A_t)$

Conditional waiting time, $W|t_e$

$f_{W|t_e}(t) =$

$\Pr\left(B_1, B_2, .. A_t | A_{-t_e}, .. B_0\right)$



2.3.2 General mathematical framework: interarrival time, f_N

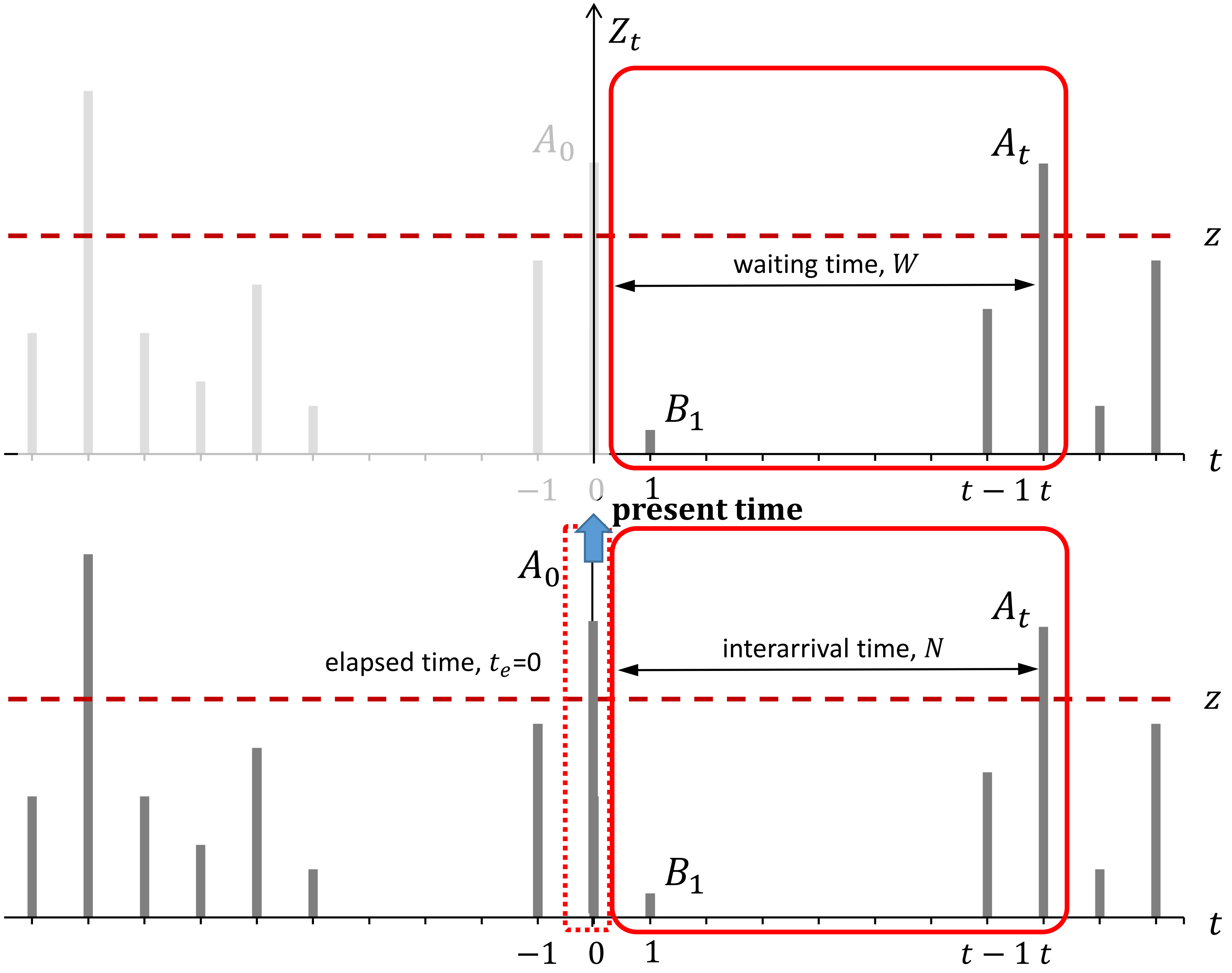
Probability mass function, $f_X(x)$

Unconditional waiting time, W

$f_W(t) = \Pr(B_1, B_2, .. A_t)$

Interarrival time, $N = W|(t_e = 0)$

$f_N(t) = \Pr(B_1, B_2, .. A_t | A_0)$



3.1 Stationary and independent processes: return period T

Under the common hypotheses that:

- events arise from a **stationary distribution**, e.g. $\Pr B_1 = \Pr B = \Pr\{Z \leq z\} = P_Z\ (z)$
- are **independent** of one another, i.e. $\Pr(B_1, B_2, .. A_t) = \Pr B_1 \Pr B_2 \ldots \Pr A_t = (\Pr B)^{t-1} \Pr A$

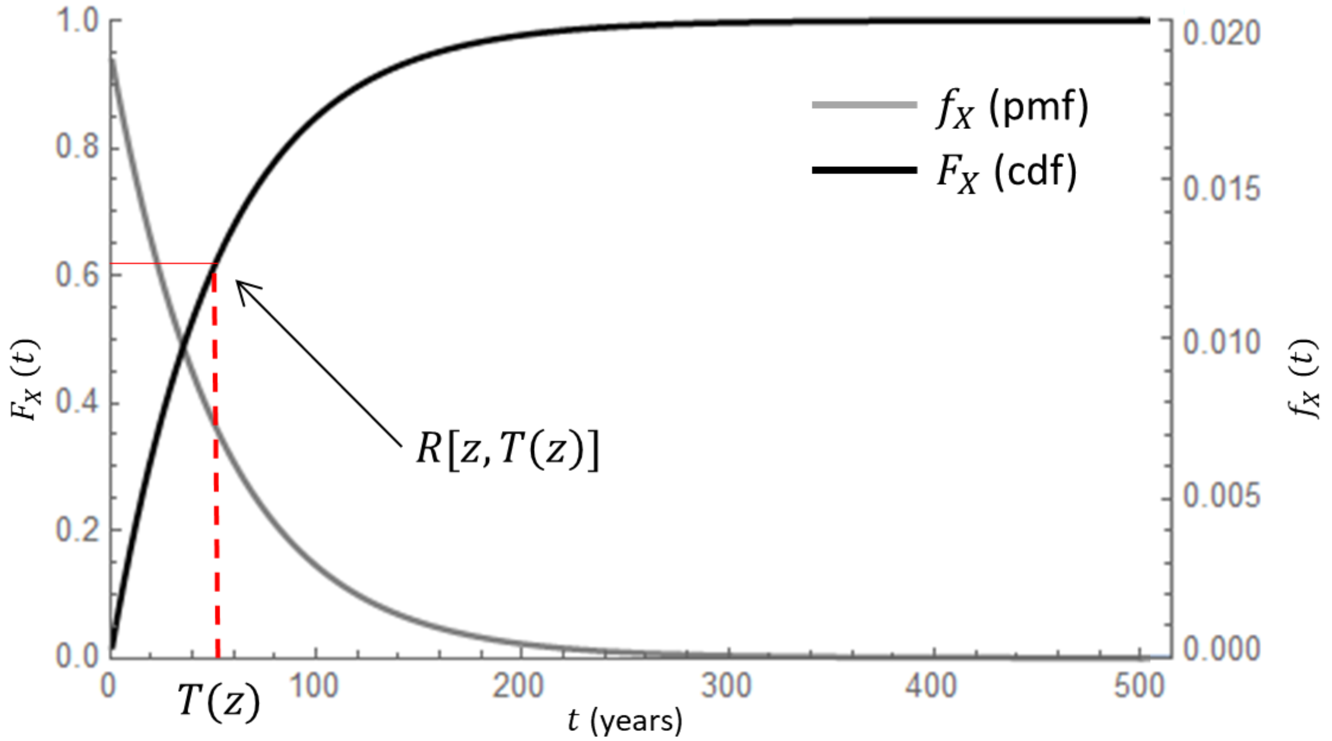
what has happened in the past does not influence future realizations of A; hence W $(W|t_e)$ and N follow the same **geometric distribution**

$$f_X(t) = [P_Z(z)]^{t-1}[1 - P_Z(z)]$$

$$F_X(t) = \sum_{x=1}^{t} f_X(x) = 1 - P_Z(z)^t$$

with mean:

$$\boxed{T(z) = \frac{\Delta\tau}{1 - P_Z(z)}} \qquad (1)$$



3.2 Stationary and independent processes: probability of failure R

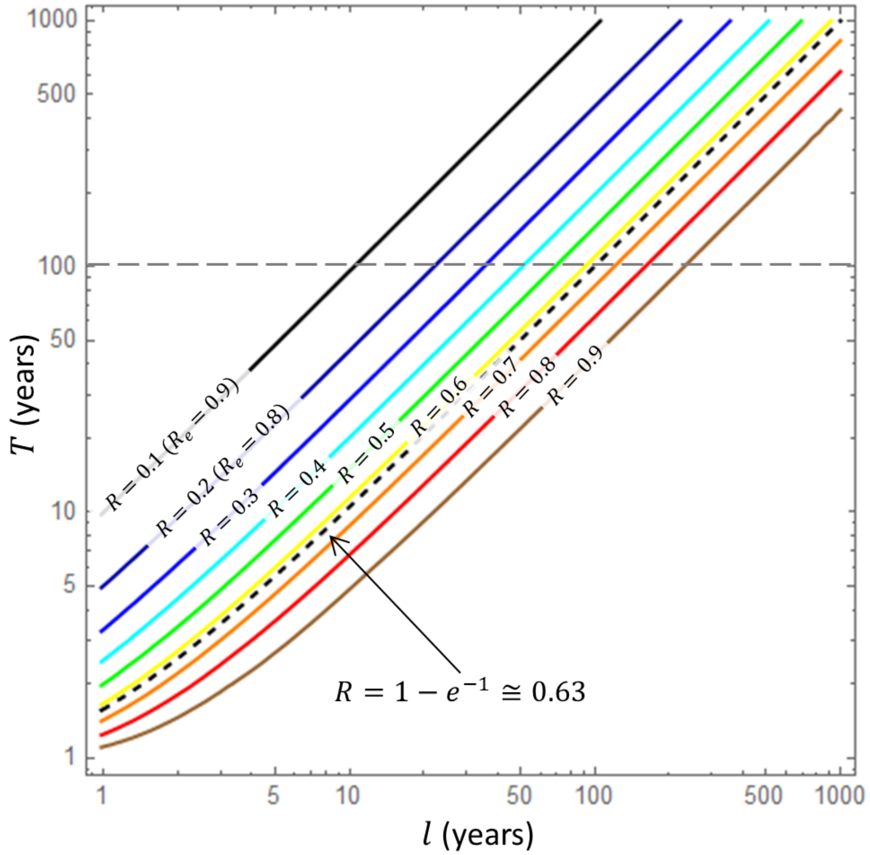
It follows that the **probability of failure** is given by

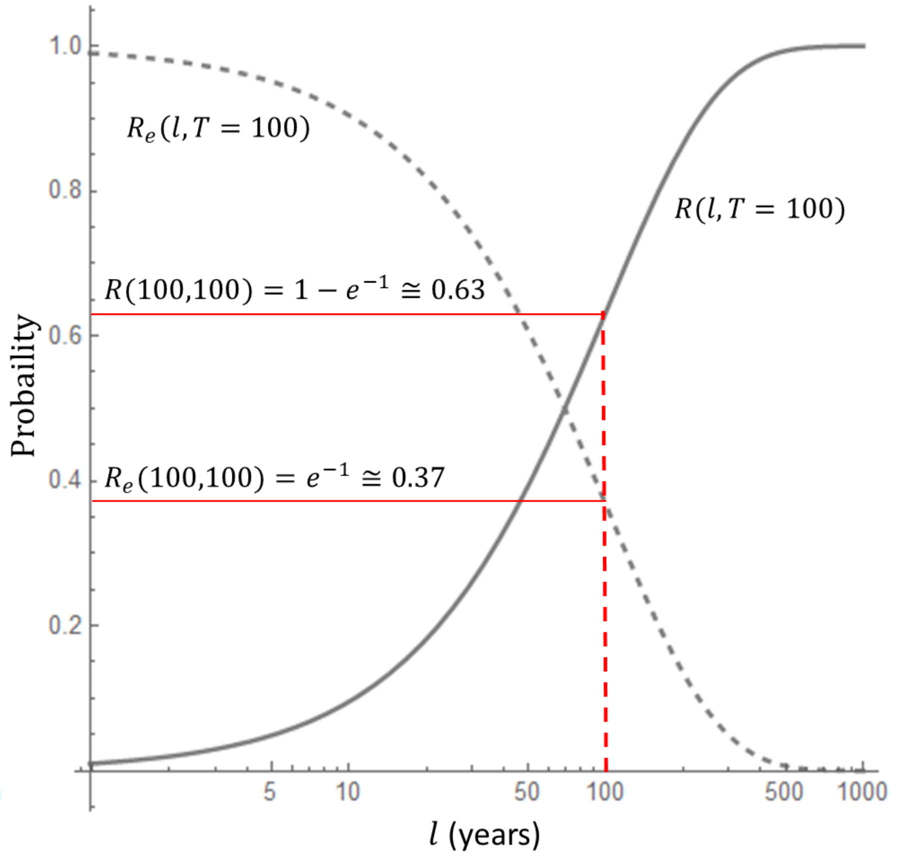
$$R(z,l) = 1 - \left(1 - \frac{\Delta\tau}{T(z)}\right)^{l/\Delta\tau} \quad (2)$$

While **reliability** is

$R_e = 1 - R$

(Read and Vogel,2015)





➢ Read, L. K., & Vogel, R. M. (2015). Reliability, return periods, and risk under nonstationarity. Water Resources Research. doi:10.1002/2015WR017089

4.1.1 Time-dependent (stationary) processes: return period T

In many other geophysical fields of application, time-dependence has been recognized to be the rule rather than the exception since a long time (e.g., Eichner et al., 2011)

When the independence condition is omitted W ($W|t_e$) and N follow different probability distributions (Fernández and Salas, 1999):

- it is not possible to derive expressions of general validity for the probability mass functions $f_X(x)$ of W ($W|t_e$) and N because the joint probabilities cannot be generally simplified, apart from a limited number of very simple processes;

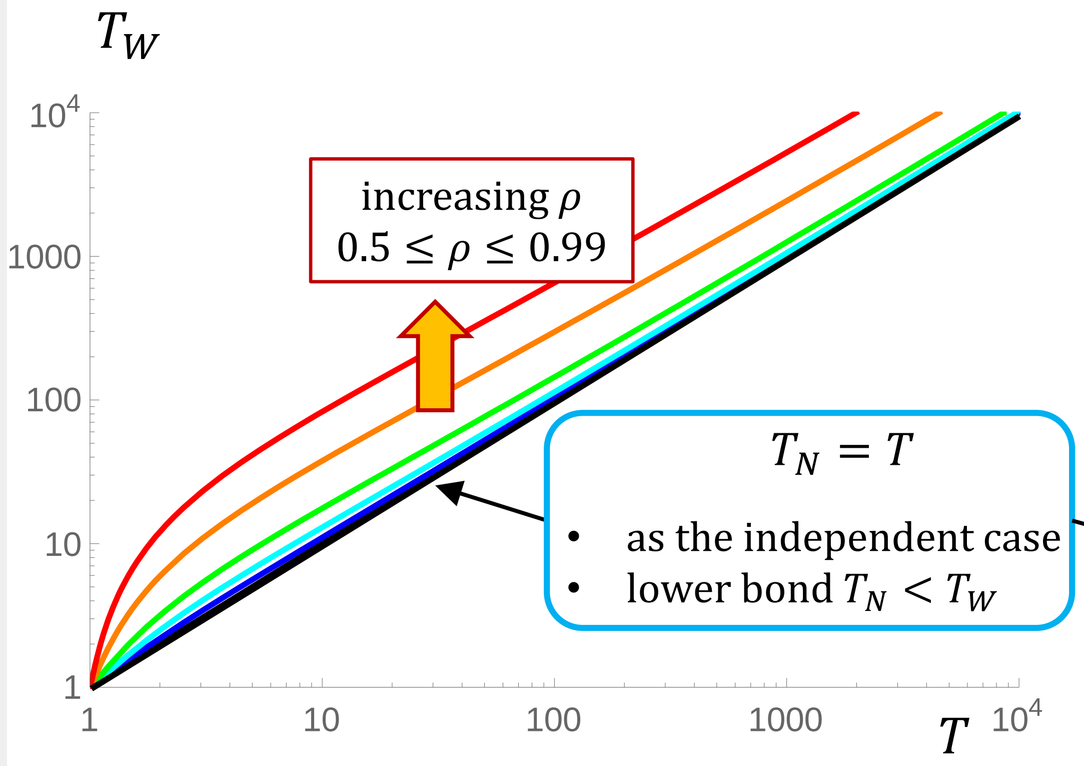
- The **average waiting time** ($T_W = \mathrm{E}[W]$) is affected by process persistence, while the **average interarrival time** ($T_N = \mathrm{E}[N]$) is always expressed by the classical return period formula (Bunde et al., 2003; Volpi et al., 2015)

$$\boxed{T_N(z) = \frac{\Delta\tau}{1 - P_Z(z)}}$$

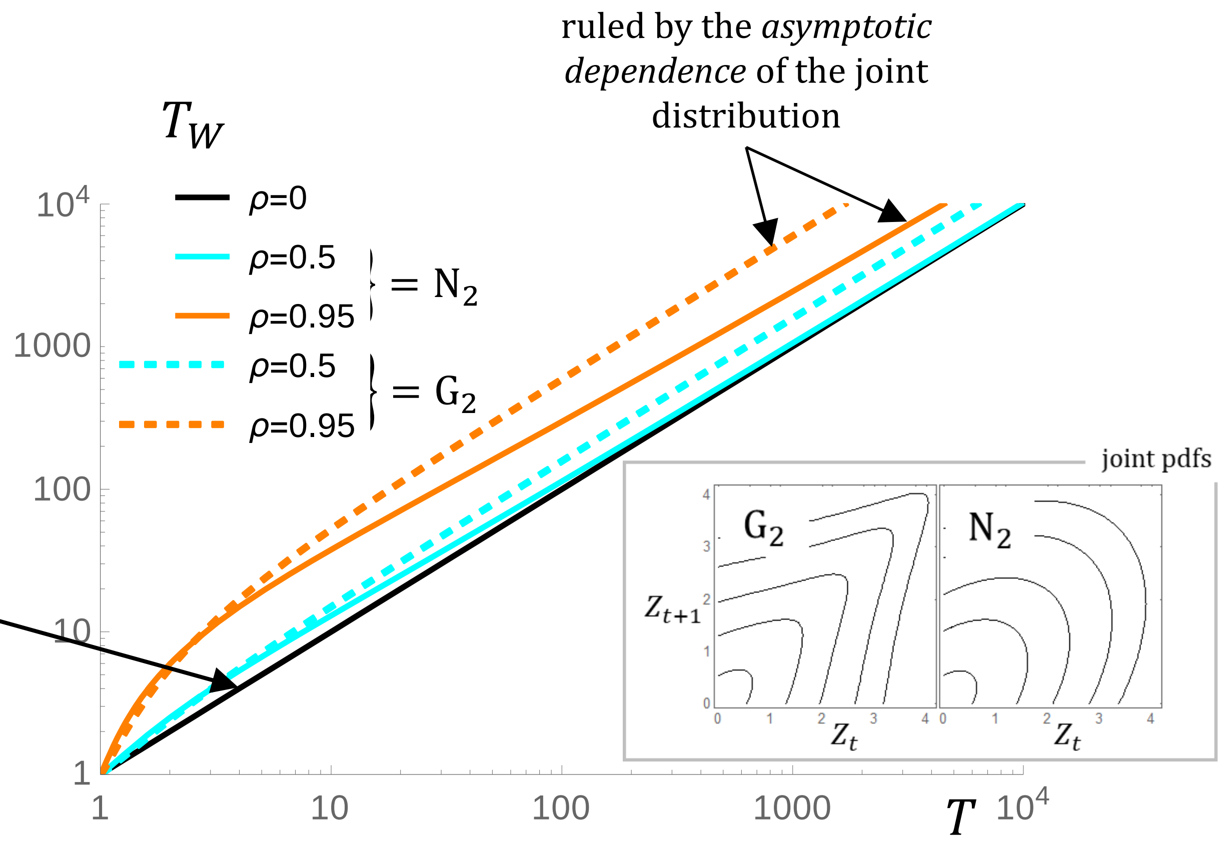
- Eichner, J. F., Kantelhardt, J. W., Bunde, A., & Havlin, S. (2011). The Statistics of return intervals, maxima, and centennial events under the influence of long term correlation. In extremis, 3-43, Berlin Heidelberg: Springer-Verlag
- Bunde, A., Eichner, J. F., Havlin, S., & Kantelhardt, J. W. (2003). The effect of long-term correlations on the return periods of rare events. Physica A: Statistical Mechanics and its Applications, 330, 1(2), 1-7, doi:10.1016/j.physa.2003.08.004
- Volpi, E., Fiori, A., Grimaldi, S., Lombardo, F., & Koutsoyiannis, D. (2015). One hundred years of return period: Strengths and limitations. Water Resources Research(51), 1-16, doi: 10.1002/2015WR017820

4.1.2 Time-dependent (stationary) processes: return period T

Illustrative examples



Z_t, two state Markov-dependent model, 2Mp
$\Pr(Z_t, Z_{t+1}) = \mathrm{N}_2(\mathbf{0}, \mathbf{1}; \rho)$
ρ, lag-1 correlation coefficient of the parent process



Z_t, two state Markov-dependent model, 2Mp
$\Pr(Z_t, Z_{t+1}) = \mathrm{G}_2(\mathbf{0}, \mathbf{1};\ \theta_\rho)$
ρ, lag-1 correlation coefficient of the parent process

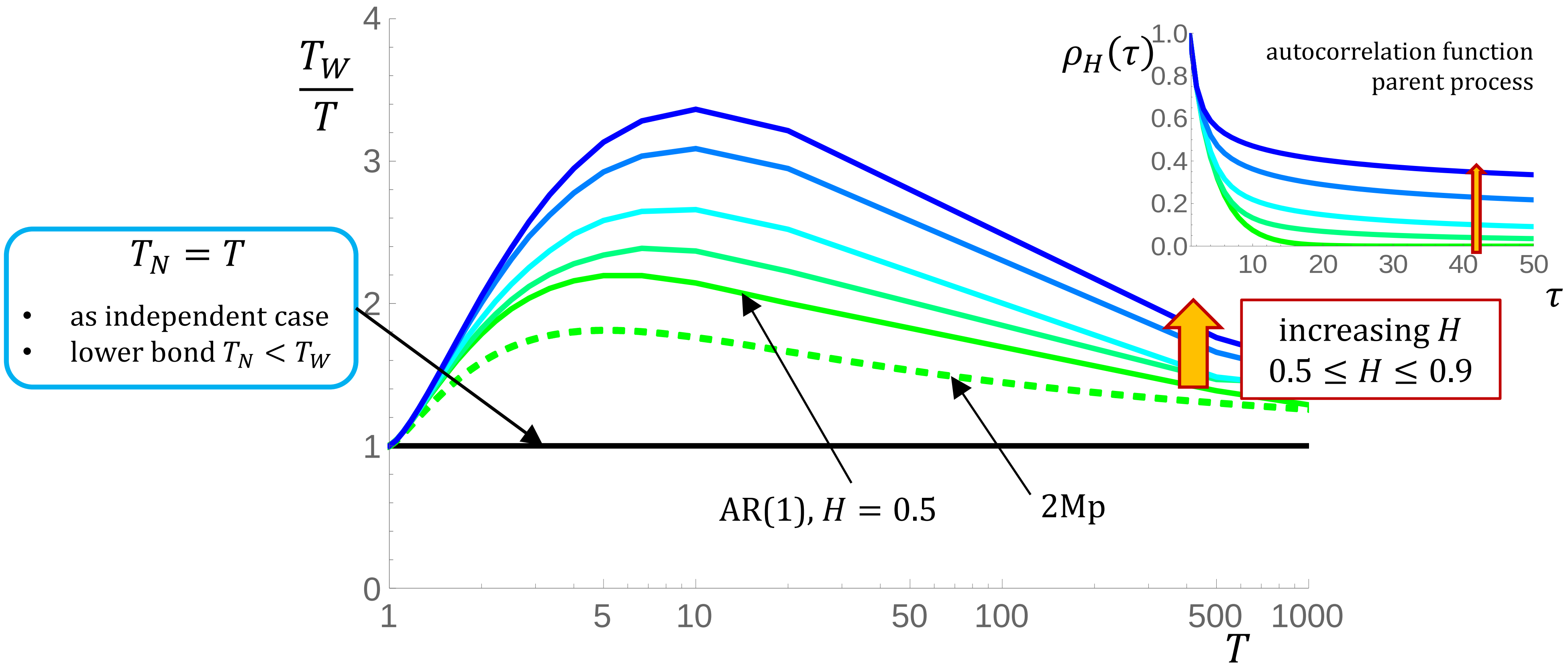
4.1.2 Time-dependent (stationary) processes: return period T

Illustrative examples

Z_t, fractionally integrated autoregressive process, FAR(1,H)

$\Pr(Z_t, .. Z_{t+\tau}) = \mathrm{N}_\tau(\mathbf{0}, \mathbf{1}; \rho_H(\tau))$

$\rho = 0.75$, lag-1 correlation coefficient of the parent process

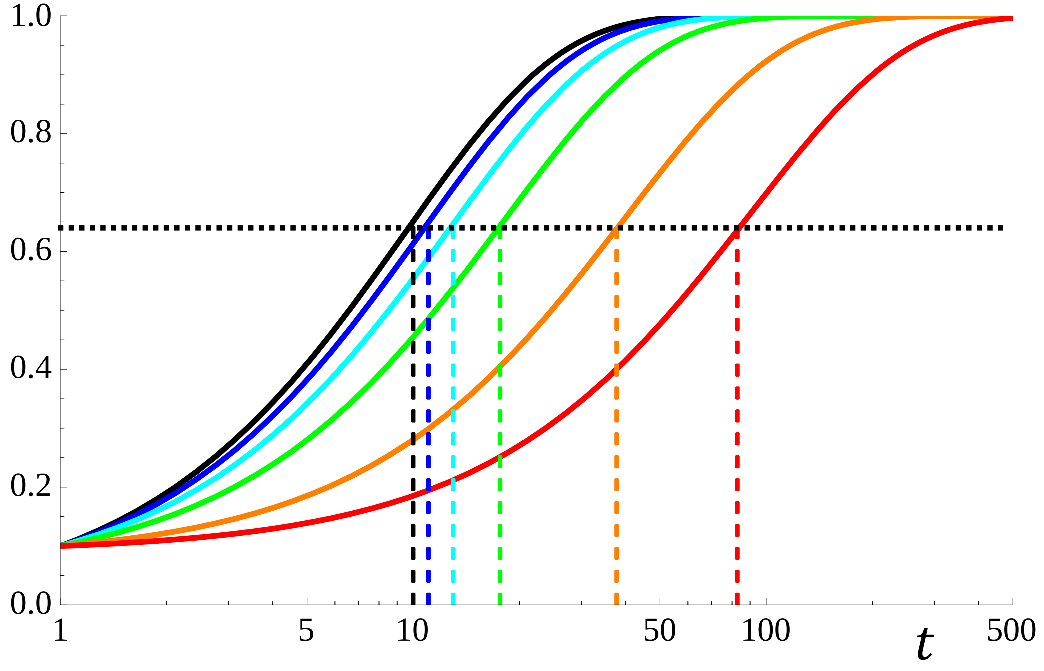


4.1.3 Time-dependent (stationary) processes: probability of failure R

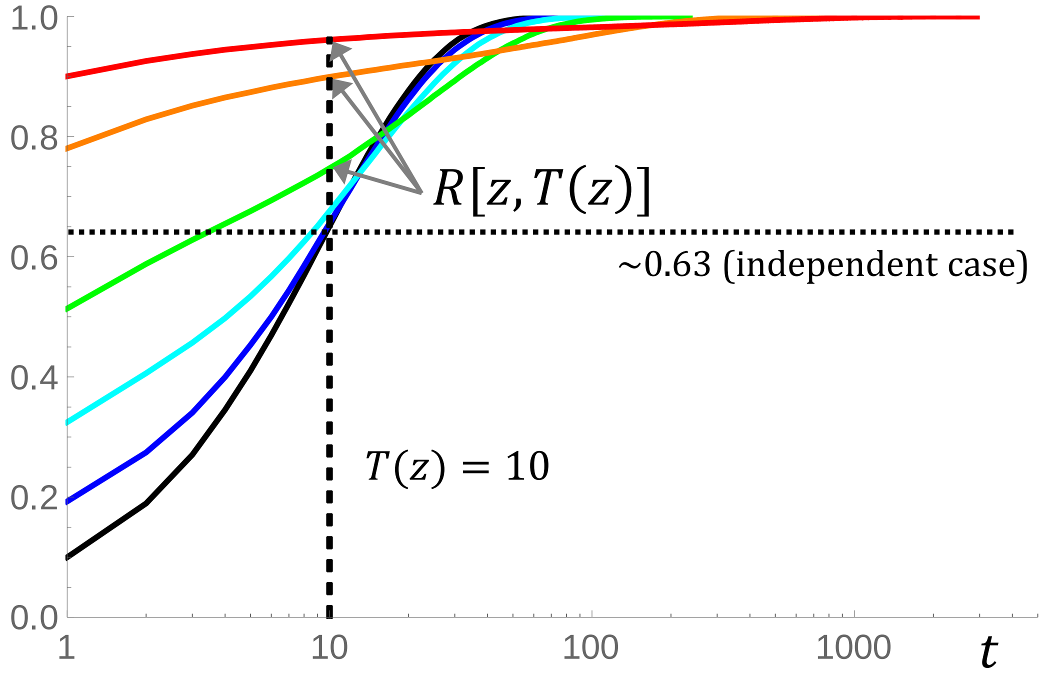
Illustrative examples

- Both the probability functions F_W and F_N <u>are affected by the autocorrelation structure of the process</u>

$F_W(t) \equiv R(z,t)$ - waiting time



$F_N(t) \equiv R(z,t)$ - interarrival time



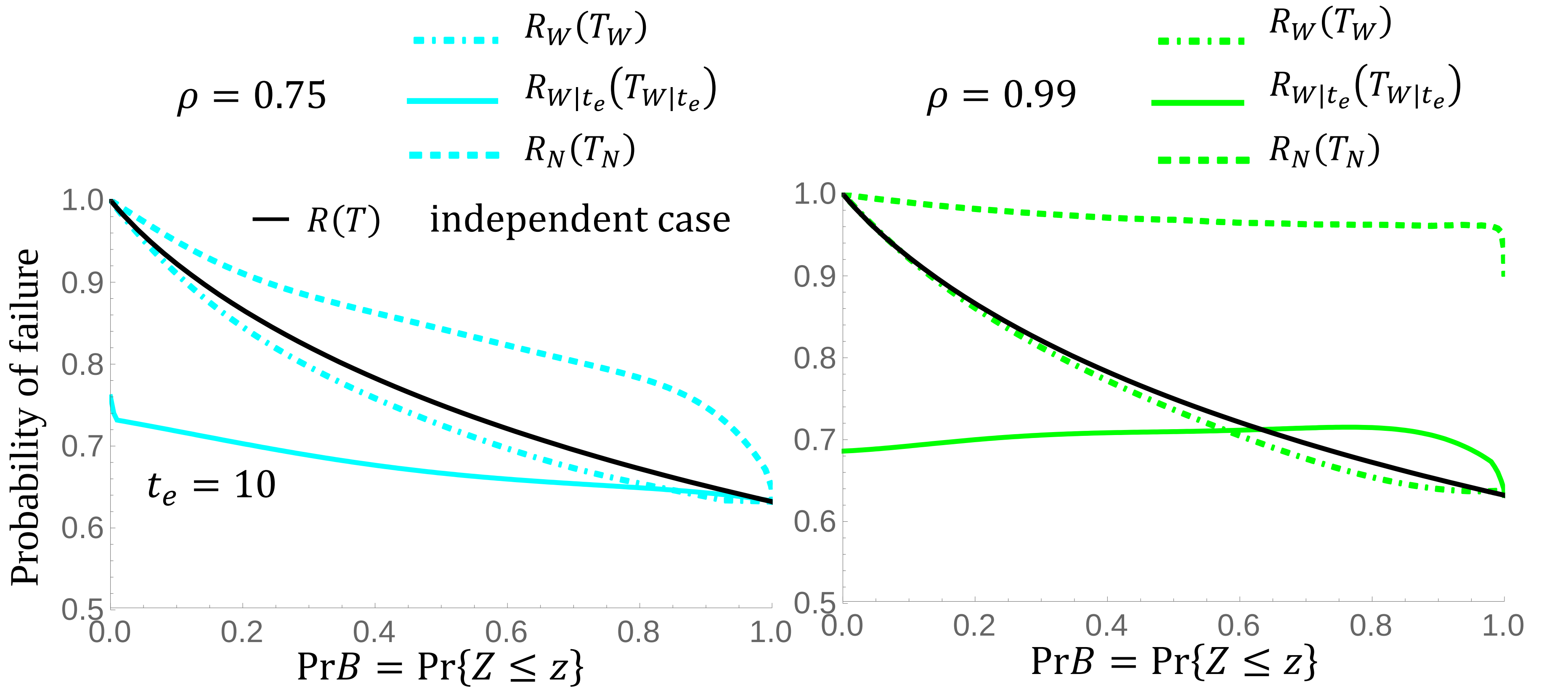
Z_t, autoregressive process, AR(1)

$\Pr(Z_t, .. Z_{t+\tau}) = \mathrm{N}_\tau(\mathbf{0}, \mathbf{1}; \rho(\tau))$

ρ, lag-1 correlation coefficient of the <u>parent process</u>

Illustrative examples

- Both the probability functions F_W and F_N are affected by the autocorrelation structure of the process



Z_t, autoregressive process, AR(1)

$\Pr(Z_t, .. Z_{t+\tau}) = \mathrm{N}_\tau(\mathbf{0}, \mathbf{1}; \rho(\tau))$

ρ, lag-1 correlation coefficient of the parent process

4.2.1 Non-stationary processes: Introduction

Many hydro-climatological records exhibit some forms of up- or downward tendency over time (trends or local shifts) that are ascribed to human interventions at the local and global scales.

Even if change in the observed data does not necessarily imply a non-stationary underlying process (Koutsoyiannis and Montanari, 2015), it is often described by assuming **non-stationarity**.

Remarks:

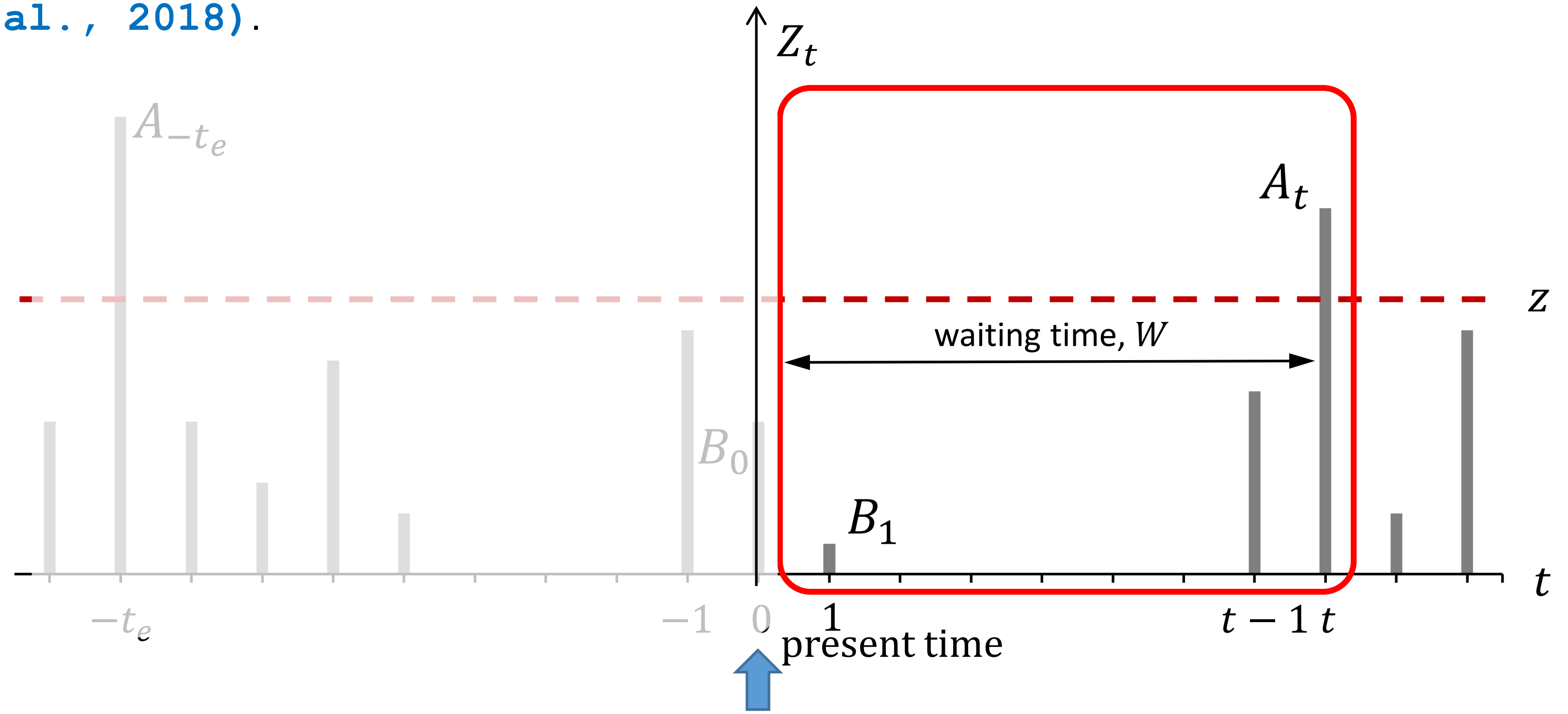
- Stationarity is a prerequisite to make inference from data.
- A non-stationary framework cannot be generally inferred from the observed data alone (Koutsoyiannis, 2016) without introducing an additional source of uncertainty, thus preventing for a practical enhancement of the credibility and accuracy of the predicted T-year event (Serinaldi & Kilsby, 2015).
- Estimates based on historical records are typically used to extrapolate projections for planning and design purposes up to very large values of T, by supposing that future will statistically behave as past.
- The extrapolation to the future should be done with caution and only if the future can be predicted in deterministic terms by using additional prior physical knowledge on the process (thus reducing uncertainty).

- Koutsoyiannis, D., & Montanari, A. (2015). Negligent killing of scientific concepts: the stationarity case. Hydrological Sciences Journal, 60(7-8), 2-22. doi: 10.1080/02626667.2014.959959
- Koutsoyiannis, D. (2016). Generic and parsimonious stochastic modelling for hydrology and beyond. Hydrological Sciences Journal, 61(2), 225-244, doi:10.1080/02626667.2015.1016950
- Serinaldi, F., & Kilsby, C. (2015). Stationarity is undead: Uncertainty dominates the distribution of extremes. Advances in Water Resources, 77 , 17-36, doi: 10.1016/j.advwatres.2014.12.013

4.2.2 Non-stationary (independent) processes: waiting time $f_X(x)$

Let's consider a non-stationary and independent processes, with positive or negative trend (Salas and Obeysekera, 2014; Salas et al., 2018).

Probability mass function, $f_X(x)$



Unconditional waiting time, W

$$f_W(t) = \Pr(B_1, B_2, ..A_t)$$

$$= \left[\prod_{x=1}^{t-1} \Pr B_x\right] \Pr A_t$$

with $x = 1,2,\dots,x_{\max}$ and where the exceedance probability $\Pr A_x = \Pr\{Z_x > z\} = 1 - P_{Z_x}(z)$ reaches the unity (zero) value at $x = x_{\max}$ in the case of positive (negative) trend.

- Salas, J. D., & Obeysekera, J. (2014). Revisiting the Concepts of Return Period and Risk for Nonstationary Hydrologic Extreme Events. Journal of Hydrologic Engineering, 19 (3),554-568. doi:10.1061/(ASCE)HE.1943-5584.0000820
- Salas, J. D., Obeysekera, J., & Vogel, R. (2018). Techniques for assessing water infrastructure for nonstationary extreme events: a review. Hydrological Sciences Journal, 63 (3), 325-352. doi: 10.1080/02626667.2018.1426858

4.2.3 Non-stationary (independent) processes: return period $T(z)$

The return period $T_W = \mathrm{E}[W]$ computed at present time (i.e. $t = 0$) is a constant quantity that accounts for the variability from present time to infinity of the probability distribution on average, by simply summarizing the average annual probability of exceedance **(Serinaldi, 2015)**.

$$T_W(z) = \frac{\Delta\tau}{1 - \overline{P_{Z_t}(z)}}$$

where

$$\overline{P_{Z_t}(z)} = \sum_{x=1}^{x_{\max}} P_{Z_x}(z)$$

Conversely, the T-year event changes in time according to the probability distribution function $P_{Z_t}(z)$.

$$z_T(t) = P_{Z_t}^{-1}\left(1 - \frac{\Delta\tau}{T}\right)$$

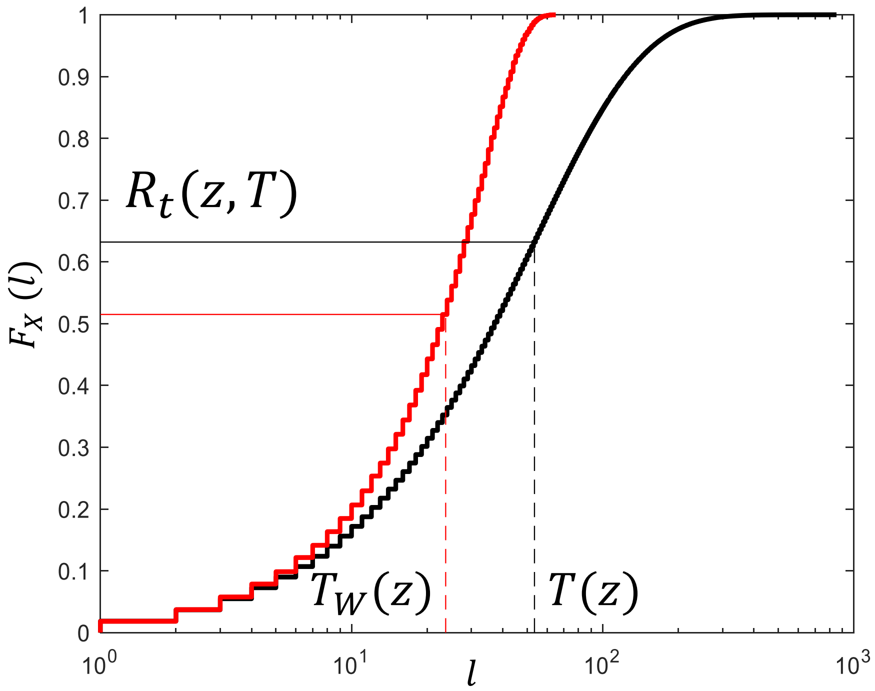
➢ Serinaldi, F. (2015). Dismissing return periods! Stochastic Environmental Research and Risk Assessment, 1{11. doi: 10.1007/s00477-014-0916-1

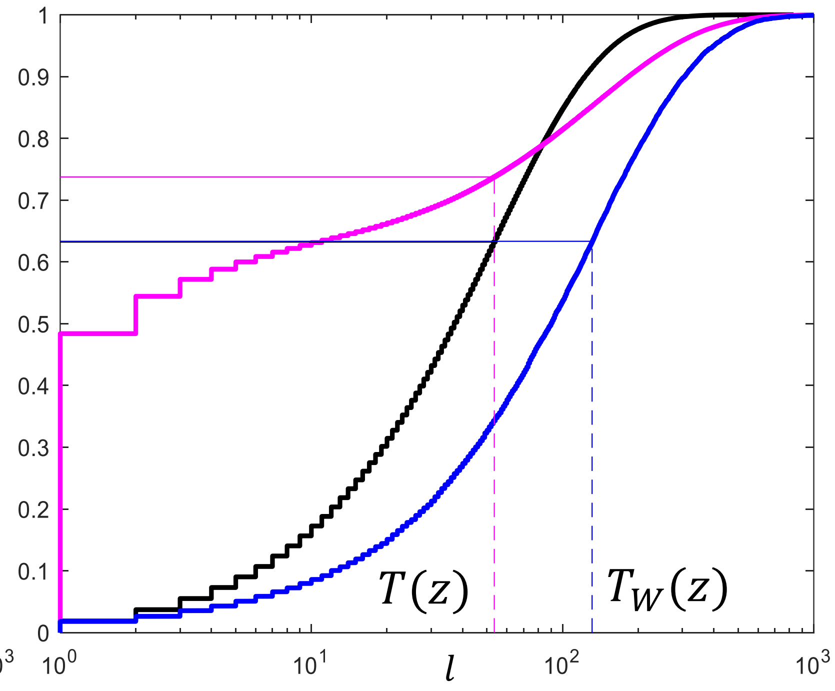
4.2.4 Non-stationary (independent) processes: probability of failure R

The relationship among T, R and l depends of the specific process **(Read & Vogel, 2015)**.

Illustrative example

- Stationary and independent ($\alpha, \rho = 0$)
- Non-stationary and independent ($\alpha = 0.05\%, \rho = 0$): waiting time , W
- Stationary and persistent ($\alpha = 0, \rho = 0.8$): waiting time, W
- Stationary and persistent ($\alpha = 0, \rho = 0.8$): interarrival time , N





Z_t, autoregressive process, AR(1)

$\Pr(Z_t, .. Z_{t+\tau}) = \mathrm{LN}_\tau\big(\boldsymbol{\alpha t} + \boldsymbol{\beta}, \mathbf{1}; \rho(\tau)\big)$

ρ, lag-1 correlation coefficient of the <u>parent process</u>

5.1 Remarks on return period and probability of failure estimation

- The problem of estimating the rareness of potentially dangerous events (i.e. $Z > z$) is generally solved by first inferring a model for $P_Z(z)$ to a stationary and independent series of observations of the random process of interest.

- $T(z)$ and the T-year event are *dynamic* quantities, subject to redefinition when new observed events add to the historical record.

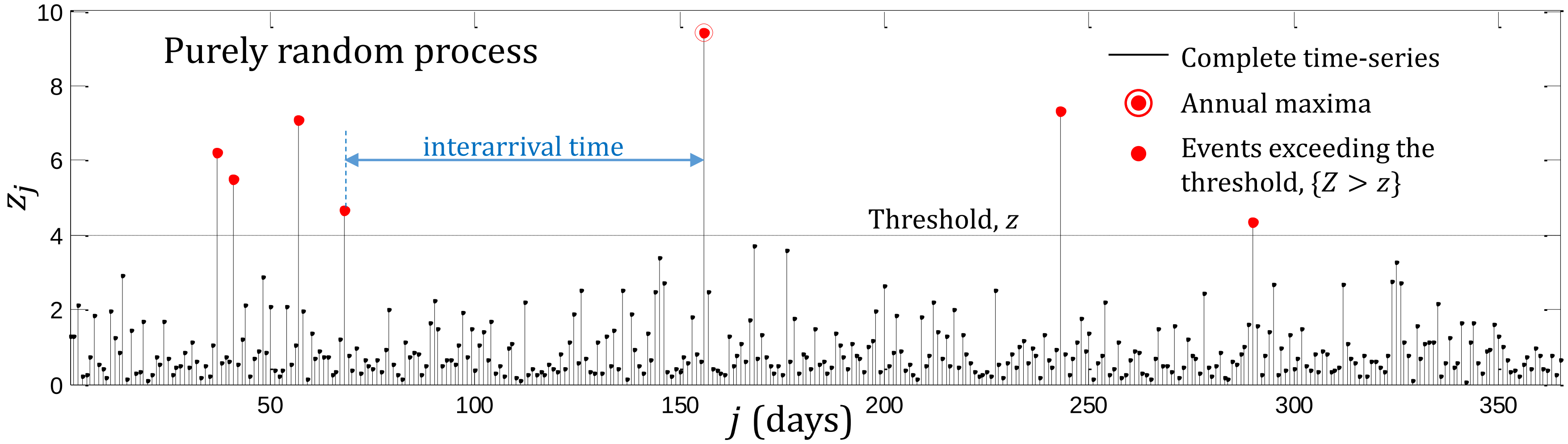
- The observed time-series is generally short (covering a period of time rarely exceeding 100 years!) with severe consequences on the reliability (accuracy and uncertainty) of the estimate of P_Z and $T(z)$: this especially occurs in the **extrapolation range** (i.e. $T \geq 100$ years, of interest in many risk assessment problems).

- The limited length of the time-series might be due to **data selection**: part of the available information in the original data-set is discarded to fulfill the stationarity and independence assumptions.

- Extending the concept of return period to non-stationary and time-dependent conditions allows exploiting all the available information; however, solving the inference problem of fitting a non-stationary model to a series of potentially time-dependent data requires additional efforts with respect to the stationary and independent, traditional yet not trivial case.

5.2 Persistent processes: on data selection and return period estimation

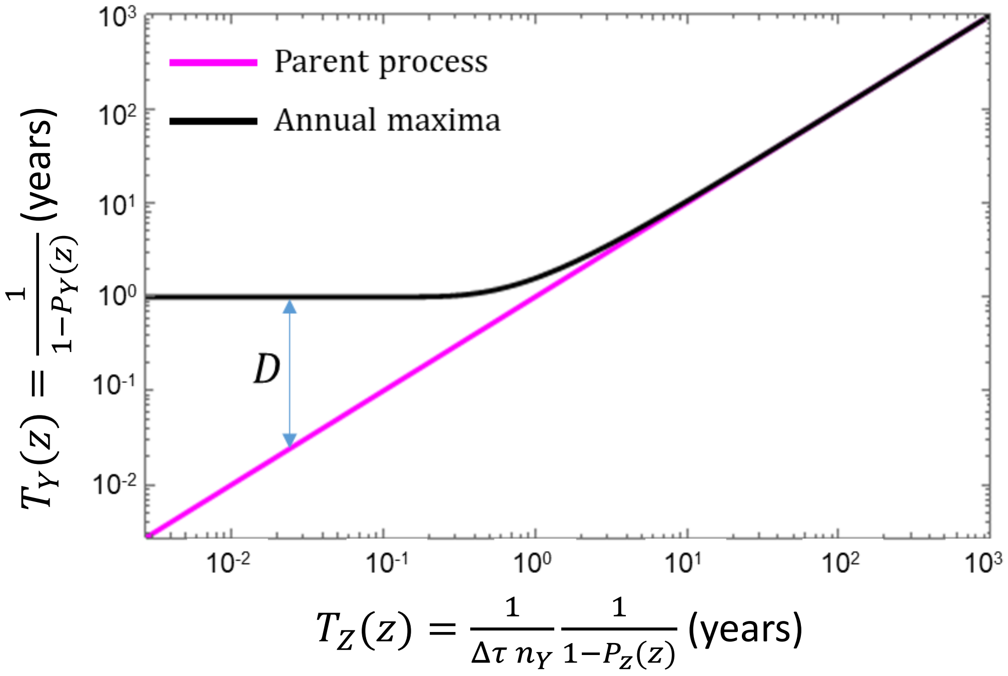
Illustrative example



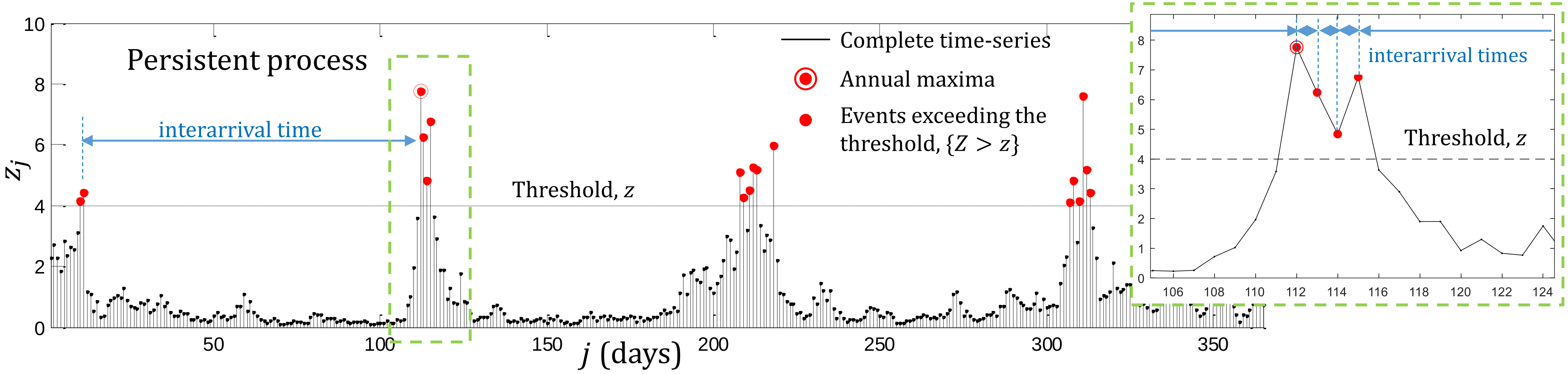
Probability distribution of annual maxima, $Y = \max_N(Z)$

$$\boxed{P_Y(z) = P_Z^N(z)}$$

- Different from that of the *parent process* (complete time series)
- **Overestimation** of return period is due to wastage of information



5.2 Persistent processes: on data selection and return period estimation



Annual maxima

$$T_Y(z) = \frac{1}{1 - P_Y(z)}$$

Complete time-series

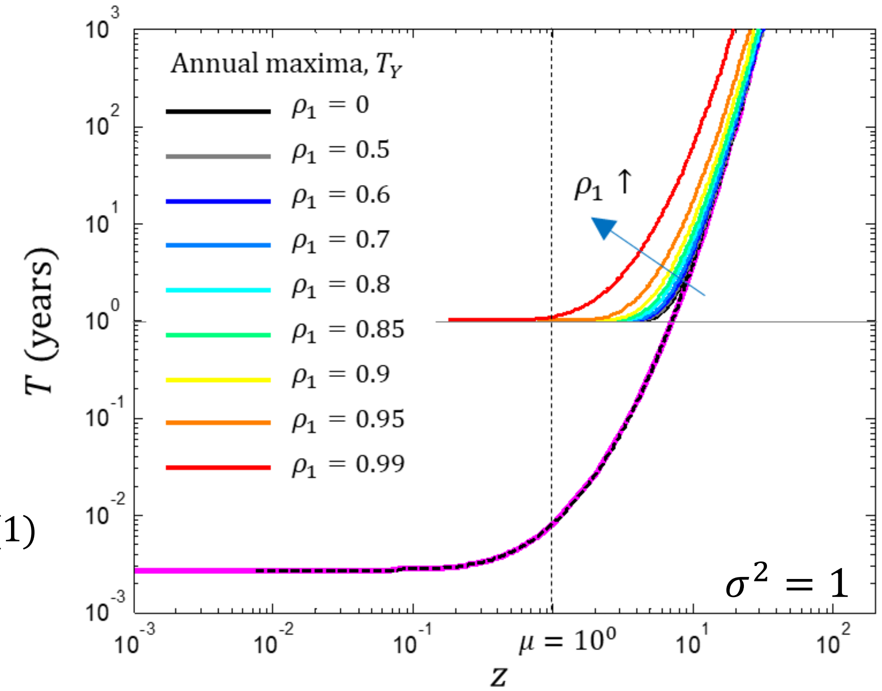
$$T_Z(z) = \frac{1}{\Delta\tau\, n_Y} \frac{1}{1 - P_Z(z)}$$

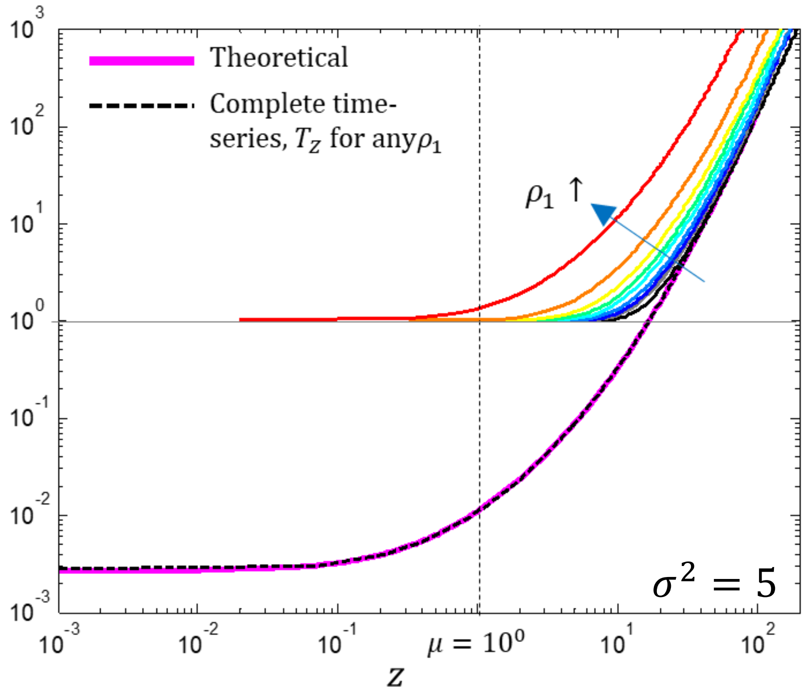
$$P_Y(z) \neq P_Z^N(z)$$

Z_t, "daily" autoregressive process, AR(1)

$\Pr(Z_t, .. Z_{t+\tau}) = \mathrm{LN}_\tau(\mathbf{1}, \sigma^2; \rho_1)$

$\rho = [0 \div 0.99]$, lag-1 correlation coefficient of the parent process





6 Summary and conclusions

Recent literature works aim at extending the return period and probability of failure concepts to more general conditions, in order to consider multivariability, non-stationarity and persistence.

Different definitions of return period are available in the literature, leading to different results when the independence or stationarity assumptions are omitted:

- the mean interarrival time T_N is not affected by the time-dependence structure of the process while mean waiting time T_W is;
- the mean interarrival time is not an exhaustive measure of the probability of failure for time-dependent processes, yet it represents a lower bound with respect to the mean waiting time for persistent processes;
- in a non-stationary framework the average waiting time T_W (determined at present time) is a constant value, while the corresponding T-year event depends explicitly on time, thus posing additional problems for decisional making.

The general framework allows exploiting all the available information provided by the observed records of data (e.g. by exploiting the property of the mean interarrival time of being insensible to the time-dependence structure of the process) with the aim of improving our estimates in terms of accuracy and reduced uncertainty.

The record of exceptionally extreme events is still very poor and a large uncertainty characterizes the extrapolation range. And data rich hydrology?

Innovative approaches and new perspectives should be proposed in the broad field of risk assessment and management, including incorporating all the available information, specifically more hydrological knowledge, into $T(z)$ and the T-year event estimation.