WATER RESOURCES RESEARCH, VOL. ???, XXXX, DOI:10.1002/,

100 years of Return Period: Strengths and limitations

E. Volpi,[1] A. Fiori,[1] S. Grimaldi,[23] F. Lombardo,[1] and D. Koutsoyiannis[4]

Corresponding author: E. Volpi, Department of Engineering, University of Roma Tre, Via Vito Volterra, 62, 00146 Roma, Italy (elena.volpi@uniroma3.it)

[1]Department of Engineering, University of Roma Tre, Via V. Volterra, 62, 00146 Rome, Italy

[2]Department for Innovation in Biological, Agro-food and Forest systems (DIBAF), University of Tuscia, Via San Camillo De Lellis snc, 01100 Viterbo, Italy

[3]Honors Center of Italian Universities (H2CU), Sapienza University of Rome, Via Eudossiana 18, 00184 Roma, Italy

[4]Department of Water Resources, Faculty of Civil Engineering, National Technical University of Athens, Heroon Polytechneiou 5, 15780 Zographou, Greece

Abstract. 100 years from its original definition by *Fuller* [1914], the probabilistic concept of return period is widely used in hydrology as well as in other disciplines of geosciences to give an indication on critical event rareness. This concept gains its popularity, especially in engineering practice for design and risk assessment, due to its ease of use and understanding; however, return period relies on some basic assumptions that should be satisfied for a correct application of this statistical tool. Indeed, conventional frequency analysis in hydrology is performed by assuming as necessary conditions that extreme events arise from a stationary distribution and are independent of one another. The main objective of this paper is to investigate the properties of return period when the independence condition is omitted; hence, we explore how the different definitions of return period available in literature affect results of frequency analysis for processes correlated in time. We demonstrate that, for stationary processes, the independence condition is not necessary in order to apply the classical equation of return period (i.e. the inverse of exceedance probability). On the other hand, we show that the time-correlation structure of hydrological processes modifies the shape of the distribution function of which the return period represents the first moment. This implies that, in the context of time-dependent processes, the return period might not represent an exhaustive measure of the probability of failure, and that its blind application could lead to misleading results. To overcome this problem, we introduce the concept of Equivalent Return Period, which

controls the probability of failure still preserving the virtue of effectively communicating the event rareness.

1. Introduction

"The storm event had a return period of 30 years" or "this dam spillway was designed for a 1000-year return period discharge" are two classical statements that one could read or hear everyday. High-school students could read them in newspapers, housewives could hear them at the market or hydrologists could write them in a technical report. This simple example recalls that the return period is the most ubiquitous statistical concept adopted in hydrology but also in many other disciplines (seismology, oceanography, geology, etc...).

It appears that the concept of return period was first introduced by *Fuller [1914]* who pioneered statistical flood frequency analysis in the USA. Return period finds wide popularity mainly because it is a simple statistical tool taken from engineering practices [*Gumbel*, 1958]. For example, engineers who work on flood control are interested in the expected time interval at which an event of given magnitude is exceeded for the first time, which gives a definition of the return period. Another common definition is the average of the time intervals between two exceedances of a given threshold of river discharge. From a logical standpoint, the first definition is as justifiable as the second one; they generally differ, even though they become practically indistinguishable if consecutive events are independent in time. Both are used in hydrology [*Fernández and Salas*, 1999a, b] and, in this paper, we will show how they may affect the frequency analysis applications under certain conditions.

The return period is inversely related to the probability of exceedance of a specific value of the variable under consideration (e.g. river discharge). For example, the annual maximum flood-flow exceeded with a 1% probability in any year is called the 100-year flood. Therefore, a T-year return period does not mean that one and only one T-year event should occur every T years, but rather that the probability of the T-year flood being exceeded is $1/T$ in every year [*Stedinger et al.*, 1993].

The traditional methods for determining the return period of extreme hydrologic events assume as key conditions that extreme events (i) arise from a stationary distribution, and (ii) are independent of one another. The hypotheses of stationarity and independence are commonly assumed as necessary conditions to proceed with conventional frequency analysis in hydrology [*Chow et al.*, 1988]. Recently, the former assumption has been questioned by several researchers [e.g. *Cooley*, 2013; *Salas and Obeysekera*, 2014; *Du et al.*, 2015; *Read and Vogel*, 2015]. However, we endorse herein the following important statement by *Gumbel* [1941] about the general validity of stationarity assumption. "In order to apply any theory we have to suppose that the data are homogeneous, i.e. that no systematical change of climate and no important change in the basin have occurred within the observation period and that no such changes will take place in the period for which extrapolations are made. It is only under these obvious conditions that forecasts can be made". The reader is also referred to *Koutsoyiannis and Montanari* [2015] and *Montanari and Koutsoyiannis* [2014], where it can be noted that many have lately questioned the stationarity assumption, but careful investigation of claims made would reveal that they mostly arise from the confusion of dependence in time with nonstationarity.

The purpose of this paper is to investigate the properties of return period when the independence condition is omitted. In hydrology, indeed, dependence has been recognized by many scientists to be the rule rather than the exception since a long time [e.g. *Hurst*, 1951; *Mandelbrot and Wallis*, 1968]. The concept of dependence in extreme events relates to the fact that the occurrence of a high or low value for the variable of interest (e.g. river discharge) has some influence on the value of any succeeding observation. *Leadbetter* [1983] found that the type of the limiting distribution for maxima is unaltered for weakly dependent occurrences of extreme events. We demonstrate that, under general dependence conditions, the classical relationship between the return period and the exceedance probability is again unaltered. On the other hand, we investigate the impact of the dependence structure on the shape of the distribution function of which the return period represents the first moment.

Based on the papers by *Fernández and Salas* [1999a], *Sen* [1999], and *Douglas et al.* [2002] we first summarize in Section 2 the available definitions of return periods (average occurrence interval - and - average recurrence interval) specifying the mass function equations and the related return period formulae. Moreover, in Section 2.2 and 2.3 the independent and time-dependent cases are analyzed in detail, while an Appendix provides the proof that the widely used return period equation (average recurrence interval) is not affected by the dependence structure of the process of interest. However, in Section 2.3 it is pointed out that the time-dependence influences the shape of the interarrival time distribution function and the probability of failure.

Two illustrative examples, i.e. using a two-state Markov process and an autoregressive process, are described in Section 3 and results are discussed in Section 4 in order to investigate further the theoretical premises depicted in Sections 2.2 and 2.3. Besides, to overcome the difficulties that arise from the application of the return period concept in a time-dependent context, we propose in Section 4.1 the adoption of an Equivalent Return Period (ERP), which resembles the classical definition of return period in the case of independence while it is able to control the probability of failure under the time-dependence condition. The ERP can be useful to avoid introducing the concept of probability of failure in engineering practice. Indeed, the latter may not be as simple to understand as the return period, which is a well-established concept in applications, routinely employed by practitioners.

Concluding remarks discuss the obtained results by stressing caution against using the concept of return period blindly given that multiple definitions exist. However, we confirm the virtue of return period showing that the classical formulation is insensitive to the time-dependence condition.

2. Return period and probability of failure

2.1. Mathematical framework

Let $Z(\tau)$ be a stochastic process that characterizes a natural process typically evolving in continuous time τ. As observations of $Z(\tau)$ are only made in discrete time, it is assumed here that the observations are made at constant time intervals $\Delta\tau$, and this interval is considered the unit of time. Hence, we consider the corresponding discrete-time process that is obtained by sampling $Z(\tau)$ at spacing $\Delta\tau$, i.e. $Z_j = Z(j\Delta\tau)$ where j $(= 1, 2, ...)$ denotes discrete time. For convenience, herein we express discrete time as $t = j - j_0$, where j_0 is the current time step; therefore the discrete-time process is indicated as Z_t and $t = 0$ denotes the present. We assume that Z_t is a stationary process [*Papoulis*, 1991]; thus, it is fully described up to the second order properties by its marginal probability function and its autocorrelation structure. Generally, in this paper we use upper case letters for random variables or events, and lower case letters for values, parameters, or constants.

We are interested in the occurrence of possible excursions of Z_t above/below a high/low level (threshold) z, which may determine the failure of a structure or system. In particular, we define a dangerous event as $A = \{Z > z\}$, which is an extreme maximum; anyway, A could be any type of extreme event, i.e. maximum or minimum. In the following we denote by p the probability of the event $B = \{Z \leq z\}$, which is the complement of A; the probability of the event A is given by $1 - p = \Pr\{Z > z\} = \Pr A$.

In hydrological applications, it is usually assumed that the event A will occur on average once every return period T, where T is a time interval and, for annual observations (i.e., $\Delta\tau = 1$ year), a number of years. In other words, the average time until the threshold z is exceeded equals T years [*Stedinger et al.*, 1993], such as

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

where X is the number of discrete time steps to the occurrence of an event A, $f_X(t) = \Pr\{X = t\}$ is its probability mass function (pmf) and E[.] denotes expectation. The definition of the return period leads to the formulation of the so-called *probability of failure* $R(l)$ (also known in literature as "risk", even if it does not account for damages) which measures the probability that the event A occurs at least once over a specified period of time: the design life l (e.g. in years) of a system or structure, where $l/\Delta\tau$ is a positive integer. Mathematically, we have

$$R(l) = \Pr\{X \leq l/\Delta\tau\} = \sum_{t=1}^{l/\Delta\tau} f_X(t) \tag{2}$$

Thus, the probability of failure is nothing else than the distribution function $F_X(t)$ computed at $t = l/\Delta\tau$.

As mentioned in the Introduction, two different definitions of the return period are available in the hydrological literature [see, e.g., *Fernández and Salas*, 1999a and *Douglas et al.*, 2002]. The return period T may be defined as:

(i) the mean time interval required to the *first occurrence* of the event A,

(ii) the mean time interval between *any two successive occurrences* of the event A.

Definition (i) assumes that an event A occurred in the past (at $t < 0$); the discrete time elapsed since the last event A to the current time step $t = 0$ is defined as *elapsing time* and it is denoted here as t_e; the sketch in Figure 1 illustrates the variables used in the present analysis. In this work, we assume that time t_e can be either deterministically known or unknown and investigate implications of both conditions on the analytical formulation of the return period. Under definition (i), the return period is based on the *waiting time* (W), i.e. the number of time steps between $t = 0$ and the next occurrence of A (see Figure 1). The sum of the waiting time and the elapsing time is denoted as *interarrival time* $N = W + t_e$.

If we assume that t_e is unknown, the probability mass function of the waiting time is given by the joint probability of the sequence of events $(B_1, B_2, ..B_{t-1}, A_t)$ (see, e.g., *Fernández and Salas*, 1999a)

$$f_W(t) = \Pr(B_1, B_2, ..B_{t-1}, A_t) \tag{3}$$

where A_t (B_t) is the event A (B) occurred at time t. Instead, if t_e is deterministically known, the pmf of the waiting time is given by the joint probability of the sequence of events $(B_1, B_2, ..B_{t-1}, A_t)$ conditioned to the realization of the events $(A_{-t_e}, B_{-t_e+1}, ..B_{-1}, B_0)$ occurred at $t \leq 0$, i.e.

$$f_{W|t_e}(t) = \Pr(B_1, B_2, ..B_{t-1}, A_t | A_{-t_e}, B_{-t_e+1}, ..B_{-1}, B_0) \quad (4)$$
$$= \frac{\Pr(A_{-t_e}, B_{-t_e+1}, ..B_{-1}, B_0, B_1, B_2, ..B_{t-1}, A_t)}{\Pr(A_{-t_e}, B_{-t_e+1}, ..B_{-1}, B_0)}$$

Definition (ii) assumes that an event A has just occurred at $t = 0$. In such a case $t_e = 0$ and the waiting time W is identical to the interarrival time N. The pmf of the interarrival time f_N is therefore a special case of equation (4), for $t_e = 0$, i.e.

$$f_N(t) = \Pr(B_1, B_2, ..B_{t-1}, A_t | A_0) \quad (5)$$
$$= \frac{\Pr(A_0, B_1, B_2, ..B_{t-1}, A_t)}{\Pr A_0}$$

Note that Figure 1 depicts a more general case than the one represented by equation (5). In the Figure, we assume that two successive occurrences of the dangerous event A are at times $-t_e$ and t. Then, N is the time elapsed between the two. As stated above, the specific case expressed by equation (5) can be obtained by setting $t_e = 0$. Moreover, we stress here that the relation $N = W + t_e$ in the Figure holds only in the case the elapsing time t_e is known, i.e. when we account for the conditional waiting time $W|t_e$.

It is interesting to note that the probability distributions of the unconditional (W, equation (3)) and conditional ($W|t_e$, equation (4)) waiting time are interrelated. In Appendix A we derive some useful relations between the return periods T_W, $T_{W|t_e}$ and T_N.

Substituting f_W (equation (3)), $f_{W|t_e}$ (equation (4)) or f_N (equation (5)) to f_X in (1) and (2), we obtain the expressions of the return periods T_W, $T_{W|t_e}$ and T_N and of the corresponding probabilities of failure R_W, $R_{W|t_e}$ and R_N, respectively. In general, the probability mass functions given by equations (3) to (5) are expected to have different shapes, leading to different values of the return period of the event A. In the following, we illustrate and discuss the differences among the above definitions when varying the correlation structure of the process Z_t; specifically, we study first the independent case, which is customary in hydrological applications, and then the more general case with some positive correlation in time (persistent case).

2.2. Independent case

If Z_t is a purely random process, then its random variables are mutually independent and their joint probability distribution equals the product of marginal ones. Therefore, we may write e.g. $\Pr\left(B_0, B_1, B_2, ..B_{t-1}, A_t\right) = \Pr B_0 \Pr B_1 ... \Pr A_t$. Substituting in equations (3), (4) and (5) the products of the marginal exceedance or non-exceedance probabilities and thanks to the stationarity assumption (that implies $\Pr A_t = 1 - p$ and $\Pr B_t = p$ for any t), we can derive the same geometric distribution in all cases. Therefore, $f_W = f_{W|t_e} = f_N = f$, with

$$f\left(t\right) = p^{t-1}\left(1 - p\right) \tag{6}$$

It follows from equation (1) that the return period T ($T = T_W = T_{W|t_e} = T_N$) is given by

$$\frac{T}{\Delta\tau} = \frac{1}{1-p} \tag{7}$$

while the variance of the pmf (6) is $v = p/\left(1-p\right)^2$. From equation (6), it also follows that the probability of failure given by equation (2) becomes

$$R(l) = 1 - \left(1 - \frac{\Delta\tau}{T}\right)^{l/\Delta\tau} = 1 - p^{l/\Delta\tau} \quad (8)$$

where again $R = R_W = R_{W|t_e} = R_N$.

Thus, for the independent case all the definitions of return period collapse to the same expression (7). This result, which is well known in the literature [e.g. *Stedinger et al.*, 1993], builds on the fact that in the independent case the occurrence of an event at any time $t \leq 0$ does not influence what happens afterwards.

2.3. Persistent case

Although independence of Z_t is usually invoked for the derivation of equation (7) [e.g. *Kottegoda and Rosso*, 1997, p. 190], it is possible to show that the mean interarrival time T_N is equal to (7) also in case of processes correlated in time; the general proof, which is given here for the first time, is illustrated in detail in Appendix B. The same property was shown by *Lloyd* [1970] for the particular case of a Markov chain process. As shown in Appendix B, equation (7) for the mean interarrival time holds true, regardless of the type of the correlation structure of Z_t.

Even though the dependence structure of the process Z_t does not affect the expected value of N (i.e., T_N), we show that this is not the case with its pmf f_N (see equation (5)). Let us consider a process characterized by a positive correlation in time. If a dangerous event A occurs at $t = 0$, then the conditional probability of occurrence of another dangerous event at $t = 1$ will be greater than $1 - p$ (independent case); this yields that the probability mass function $f_N(t)$ will have a larger mass for $t = 1$ and a lower mass elsewhere with respect to the independent case (equation (6)). Hence, while the mean value remains the same, the variance of the interarrival time N is larger than that of the independent case and it increases with the temporal correlation. This implies that the probability of failure R_N (following equation (2)) is strongly affected by the time-dependence structure of the process.

Conversely, the return periods T_W and $T_{W|t_e}$ do account for the temporal correlation of Z_t. Recalling that $(1-p) = 1/\,\mathrm{E}\,[N]$ (see equations (7) and (1)), it can be shown that (see Appendix A, equation (A8))

$$\frac{T_W}{T_N} = \frac{1}{2}\left(\frac{\mathrm{E}\,[N^2]}{\mathrm{E}\,[N]^2} + \frac{1}{\mathrm{E}\,[N]}\right) \tag{9}$$

Equation (9) shows that T_W is greater than or equal to T_N. It is easy to check that $T_W = T_N$ for independent processes, in line with the discussion reported under Section 2.2. When the process is correlated in time, the term $\mathrm{E}\,[N^2]\,/\,\mathrm{E}\,[N]^2$ is expected to increase with the autocorrelation of the process, thus resulting in the inequality $T_W > T_N$. Hence, the mean waiting time is generally larger than the mean interarrival time for temporally correlated processes.

In the following Sections we will examine the pmfs of the waiting times W and $W|t_e$ and the interarrival time N, as well as their average values (T_W, $T_{W|t_e}$ and T_N), as functions of the temporal correlation of the process. To this end, we make use of two different illustrative examples, the first is based on a Markov chain, while the second uses an AR(1) model. For convenience - and without loss of generality - $\Delta\tau$ is set equal to one.

3. Illustrative examples

3.1. Example 1: two state Markov-dependent process

We consider here a stochastic process Z_t which is based on a Markov chain Y_t. This process is considered here since it allows to easily derive the analytical expressions of the probability mass functions of the waiting and interarrival times, as done in previous literature works by *Lloyd* [1970], *Rosbjerg* [1977] and *Fernández and Salas* [1999a]. The Markov chain Y_t has two states, which here represent the events $A_t = \{Z_t > z\}$ and $B_t = \{Z_t \leq z\}$ with probability $1-p$ and p, respectively. For the Markov property, the probability of a state at a given time t depends solely on the state at the previous time step $t-1$, e.g. $\Pr(B_t|B_{t-1}...B_0) = \Pr(B_t|B_{t-1})$. Applying the chain rule to the Markov property (e.g. *Papoulis*, 1991, p. 636), it follows that the joint probability of a sequence of states, e.g. $\Pr(B_1, B_2, ..B_t) = \Pr\{Z_1 \leq z, Z_2 \leq z, ..., Z_t \leq z\}$, can be written as $\Pr(B_1)\Pr(B_2|B_1)...\Pr(B_t|B_{t-1}) = \Pr\{Z_1 \leq z\}\Pr\{Z_2 \leq z|Z_1 \leq z\}...\Pr\{Z_t \leq z|Z_{t-1} \leq z\}$.

The process Z_t described above is indicated in the following as two state Markov-dependent process and denoted by 2Mp. For each value of p (i.e. of z) Z_t is fully characterized by the marginal probabilities of the states A and B ($1-p$ and p) and by the transition probability matrix, $M = [[\Pr(A_{t+1}|A_t), \Pr(A_{t+1}|B_t)], [\Pr(B_{t+1}|A_t), \Pr(B_{t+1}|B_t)]]$ where $\Pr(A_{t+1}|A_t) + \Pr(B_{t+1}|A_t) = 1$ and $\Pr(A_{t+1}|B_t) + \Pr(B_{t+1}|B_t) = 1$. We denote by q the joint probability of non-exceedance of the threshold value z for two successive events, i.e. $q = \Pr(B_{t+1}, B_t)$ for any t; it ensues that $M = [[1-(p-q)/(1-p), 1-q/p], [(p-q)/(1-p), q/p]]$.

The probability mass function of the unconditional waiting time f_W (equation (3)) becomes

$$f_W(t) = \begin{cases} 1-p & (t=1) \\ p\left(\frac{q}{p}\right)^{t-2}\left(1-\frac{q}{p}\right) & (t \geq 2) \end{cases} \tag{10}$$

with mean given by

$$T_W = 1 + \frac{p^2}{(p-q)} \tag{11}$$

and variance $\mathrm{var}\,[W] = p^2\,(p - p^2 + q)\,/\,(p-q)^2$. After substituting equation (10) in (2), the probability of failure in a period of length l is given by

$$R_W\,(l) = 1 - p\left(\frac{q}{p}\right)^{l-1} \tag{12}$$

while the pmf of the conditional waiting time $f_{W|t_e}$ (equation (4)) for $t_e > 0$ reduces to

$$f_{W|t_e}\,(t) = \left(\frac{q}{p}\right)^{t-1}\left(1 - \frac{q}{p}\right) \tag{13}$$

with mean

$$T_{W|t_e} = \frac{p}{(p-q)} \tag{14}$$

and variance $\mathrm{var}\,[W|t_e] = pq/(p-q)^2$. The probability of failure based on the conditional waiting time is given by

$$R_{W|t_e}\,(l) = 1 - \left(\frac{q}{p}\right)^{l-1} \tag{15}$$

Equation (14) shows how for the 2Mp model the mean waiting time distribution is not affected by the value of t_e. This builds upon the fact that the conditional non-exceedance probability at t depends only on that at $t-1$, due to the property of the Markov chain.

Finally, the pmf of the interarrival time N (equation (5)) assumes the following expression

$$f_N(t) = \begin{cases} 1 - (p-q)/(1-p) & (t=1) \\ \frac{(p-q)}{(1-p)} \left(\frac{q}{p}\right)^{t-2} \left(1 - \frac{q}{p}\right) & (t \geq 2) \end{cases} \tag{16}$$

while its mean is given by equation (7) with $\Delta\tau = 1$ (following the general proof given in Appendix B), and the variance is equal to $\mathrm{var}[N] = p(p-2p^2+q)/\left[(p-1)^2(p-q)\right]$. The probability of failure in a period of length l is given by

$$R_N(l) = 1 - \frac{p-q}{1-p}\left(\frac{q}{p}\right)^{l-1} \tag{17}$$

The joint probability q may assume values in the range $[\max(2p-1,0), p]$: the lower and upper bounds correspond to perfect negative and positive correlations in time, respectively; in the independent case, $q = p^2$. We consider here only processes positively correlated (i.e. persistent), as it is commonly the case in hydrology (e.g. rainfall and discharge); thus, $q \in [p^2, p]$. Furthermore, we assume that Z_t is a standard Gaussian process and that the joint probability q is ruled by a bivariate Gaussian distribution; under the latter assumption, q can be described in terms of the lag-1 autocorrelation coefficient ρ. Specifically, q is computed as $q = \Pr\{Z_{t+1} \leq z, Z_t \leq z\} = \int_{-\infty}^{z}\int_{-\infty}^{z} f_{\mathbf{Z}}(z_t z_{t+1}; \mathbf{0}, \mathbf{\Sigma}_2)\, dz_{t+1} dz_t$ where $f_{\mathbf{Z}}$ is the probability density function of the bivariate Gaussian distribution $\mathcal{N}_2(\mathbf{Z}; \mathbf{0}, \mathbf{\Sigma}_2)$ with zero mean and $\mathbf{\Sigma}_2 = \{\{1,\rho\},\{\rho,1\}\}$, with $\rho \in [0,1]$. Note that ρ denotes the correlation in the parent process Z_t and not that between the events exceeding the threshold, i.e. $A = \{Z > z\}$. The correlation between the extremes is ruled by the shape of the parent bivariate distribution, which is assumed here to be Gaussian; the latter assumption implies that the correlation between the events A is negligible to null for high threshold values, since the Gaussian process is asymptotic independent.

3.2. Example 2: AR(1) process

We now assume that Z_t follows an AR(1) process (first-order autoregressive process), i.e. $Z_t = \rho Z_{t-1} + \alpha_t$ where ρ is the lag-1 correlation coefficient and $\alpha_t \sim \mathcal{N}\left(0, \sqrt{1-\rho^2}\right)$, such that the process is characterized by a multivariate Gaussian distributions $\mathcal{N}_t\left(\mathbf{Z}; \mathbf{0}, \mathbf{\Sigma}_t\right)$ with $\mathbf{Z} = \{Z_1, Z_2..Z_t\}$ and $\mathbf{\Sigma}_t = \left\{\rho^{|i-k|}\right\}$, $i,k = 1..t$. We assume again $\rho \in [0,1]$.

Even if conceptually simple and similar to the 2Mp (see e.g. *Saldarriaga and Yevjevich*, 1970), AR(1) is rather different in terms of the pmfs f_W, $f_{W|t_e}$ and f_N. Both the processes are based on the Markov property; however, in AR(1) the Markov property applies to the continuos random variable Z and not to the state $Y = \{Z \lesseqgtr z\}$. It means that in AR(1) the joint probability $f_{\mathbf{Z}}(z_1, z_2..z_t)$ can be expressed as $f_Z(z_1) f_Z(z_2|z_1)..f_Z(z_t|z_{t-1})$, while the same simplification cannot apply to the joint probability of a sequence of states, e.g. $\Pr\left(B_1, B_2, ..B_t\right) = \Pr\{Z_1 \leq z, Z_2 \leq z, ..., Z_t \leq z\}$, as for 2Mp. The joint probability of any sequence can be estimated by proper integration of the joint pdf of the multivariate Gaussian distribution $\mathcal{N}_t$. This entails that the pmfs f_W, $f_{W|t_e}$ and f_N, given by equations (3), (4) and (5) respectively, cannot be simplified as in the case of 2Mp, but they can be written as

$$f_W(t) = \int_{-\infty}^{z} \int_{-\infty}^{z} .. \int_{z}^{+\infty} f_{\mathbf{Z}}\left(z_1, z_2..z_t; \mathbf{0}, \mathbf{\Sigma}_t\right) dz_1 dz_2..dz_t \tag{18}$$

$$f_{W|t_e}(t) = \frac{\int_{z}^{+\infty} \int_{-\infty}^{z} .. \int_{z}^{+\infty} f_{\mathbf{Z}}\left(z_{-t_e} z_{-t_e+1}..z_t; \mathbf{0}, \mathbf{\Sigma}_{t+t_e}\right) dz_{-t_e} dz_{-t_e+1}..dz_t}{\int_{z}^{+\infty} \int_{-\infty}^{z} .. \int_{-\infty}^{z} f_{\mathbf{Z}}\left(z_{-t_e} z_{-t_e+1}..z_0; \mathbf{0}, \mathbf{\Sigma}_{t_e}\right) dz_{-t_e}..dz_0} \tag{19}$$

while f_N can be derived from the latter under the assumption $t_e = 0$. Finally, substituting the previous expressions in (1) and (2) we get the corresponding return periods and probabilities of failure.

Interestingly enough, unlike the 2Mp, $f_{W|t_e}$ (19) depends on t_e, i.e. the elapsing time. This relies on the fact that the conditional non-exceedance probability at t, i.e. $\Pr(B_t|B_{t-1}...B_0)$, generally depends on the whole sequence of previous events for AR(1), while it only depends on that at $t-1$ for the 2Mp. In such a sense, AR(1) is more correlated than 2Mp.

4. Results and discussion

We start this Section by discussing the effects of temporal correlation on the probability mass functions f_W (equation (10)) and f_N (equation (16)), and the related return periods T_W, T_N (equation (11) and (7) with $\Delta\tau = 1$, respectively) for the two state Markov-dependent process (2Mp).

Figure 2 illustrates T_W and T_N as functions of the independent return period T (i.e. of the non-exceedance marginal probability p) for several values of the correlation coefficient ρ. It is seen that T_N equals T, being independent of ρ as demonstrated in Appendix B; for $\rho = 0$ (black line) it is always $T_N = T_W = T = (1-p)^{-1}$. Conversely, the mean waiting time T_W increases with ρ (equation (9)); T_W is always greater than the mean interarrival time T_N, which thus represents a lower bound for the return period (Figure 2a). Specifically, for values of T around 5, T_W is roughly eight times larger than T_N for $\rho = 0.99$ and about twice for $\rho = 0.75$; for small and very large values of T (i.e. for small and high values of the threshold z, respectively) T_W tends to the independent limit $T = (1-p)^{-1}$ (Figure 2b).

As discussed in Section 2.3, although $T_N = T$ for any ρ, the pmf f_N (as well as f_W) may be significantly influenced by the correlation structure of the Z_t process. The distribution functions of W and N are illustrated in Figure 3, for various values of ρ and $p = 0.9$.

The mean values for each distribution (i.e. the return periods normalized with respect to $\Delta\tau = 1$) are denoted by the vertical dashed lines. The broadness of both distributions increases with ρ, as also indicated by the increase of their variance and skewness (not shown).

Figure 3a shows that the distribution function computed at T_W, which corresponds to the probability of failure in the period T_W (see equation (2)), is independent of ρ taking approximately the value 0.63 for high values of p [*Stedinger et al.*, 1993].

On the other hand, F_N changes dramatically when increasing temporal correlation ρ. This may result in very high values of the probability of failure for the same T_N, even for small time intervals t (Figure 3b). Thus, although the return period T_N remains the same for correlated and independent processes (all the vertical dashed lines corresponding to the different values of ρ collapse into a unique line, depicted in black), the probability that the threshold z is exceeded in the period T_N can be much larger for the former than for the latter (up to about 0.9 for the limit case $\rho = 0.99$).

We now illustrate and discuss the probability functions for W, $W|t_e$ and N for the AR(1) process, as well as the corresponding mean values, as functions of the lag-1 autocorrelation coefficient ρ. Results are compared to those obtained for the previously analyzed 2Mp case.

The probability mass functions f_W (equations (18)) and f_N (equation (19) for $t_e = 0$) for AR(1) are similar to those for 2Mp, even if they are characterized by a much larger dispersion, and thus they are not shown here. Their averages T_W and T_N are depicted in Figure 4, as function of the independent return period T, for $\rho = 0.75$ and $\rho = 0.99$. T_W and T_N for AR(1) (continuos lines) are also compared to those pertaining to the 2Mp (dashed lines). The mean waiting times T_W for the two models are similar, although T_W is generally larger for AR(1); since the two processes have the same ρ, this result is a direct consequence of the stronger correlation of AR(1) with respect to 2Mp, as explained in previous Section. Larger differences are expected for even more persistent processes, i.e. processes characterized by a longer range persistence with respect to the AR(1).

As mentioned in the previous Section, the stronger correlation of AR(1) also influences the mean conditional waiting time $T_{W|t_e}$, which depends on the elapsing time t_e in contrast to that of 2Mp. $T_{W|t_e}(t_e)$ is illustrated in Figure 5 for $p = 0.9$ and for a few values of the correlation coefficient ρ. For each value of ρ, $T_{W|t_e}$ is by definition equal to the mean inter-arrival time T_N for $t_e = 0$ (see equation (4)); $T_{W|t_e}$ increases with t_e tending to an asymptotic value that is greater than T_W (dashed lines). This behaviour arises from the fact that the conditional non-exceedance probability $(B_1, B_2, ..B_{t-1}, A_t|A_{-t_e}, B_{-t_e+1}, ..B_{-1}, B_0)$ (eq. 4) depends on the whole sequence of previous events. However, as t_e becomes very high the previous dangerous event A_{-t_e} has occurred too distant in time to significantly affect the realization of the next event at time t; the latter is mainly controlled by a sequence of antecedent events whose length strictly depends on the shape of the autocorrelation function of the underling process Z_t. Due to the exponential shape of the AR(1) autocorrelation function, i.e. $\rho_t(t) = \rho^t$, $T_{W|t_e}$ is expected to approach the asymptotic value when t_e becomes larger than the integral scale of the process, $\lambda(\rho) = 1/(1-\rho)$.

Conversely, $T_{W|t_e}$ for 2Mp maintains a constant value for any $t_e > 0$ since the conditional joint probability in equation (4) $\Pr(B_1, B_2, ..B_{t-1}, A_t|A_{-t_e}, B_{-t_e+1}, ..B_{-1}, B_0)$ depends only on the state at $t = 0$, due to the Markov property of the Y_t chain (as already discussed in Section 3.1); moreover, being influenced by a longer sequence of safe events (B), both $T_{W|t_e}$ and T_W of AR(1) are larger than those of 2Mp (results not shown).

We finally explore how the probabilities of failure $R_W\left(T_W\right)$, $R_{W|t_e}\left(T_{W|t_e}\right)$ and $R_N\left(T_N\right)$ behave as functions of the correlation coefficient ρ; results are summarized in Figure 6 for the processes 2Mp and AR(1) and compared to the independent case. For both processes, the probability of failure based on the interarrival time (N) may assume values much larger than the independent case; $R_N\left(T_N\right)$ significantly increases with the autocorrelation of the process ρ, (compare e.g. 2Mp for $\rho = 0.75$ and $\rho = 0.99$) and, more generally, with the correlation structure of the process (compare AR(1) and 2Mp for the same value of ρ). On the contrary, when we consider the waiting time W (conditional and unconditional), the probability of failure is less than the independent case. This reduction is significant when we account for the elapsing time t_e, thus when we add information about the last dangerous event occurred in the past. Note that Figure 6 specifically refers to the cases $t_e = 10$ for AR(1) while it is representative of any $t_e > 0$ for 2Mp. As for AR(1), $R_{W|t_e}\left(T_{W|t_e}\right)$ reduces with respect to the independent case when t_e is much larger than the integral scale of the process, i.e. $t_e > \lambda$ when $\rho = 0.75$ ($\lambda = 4$) (Fig. 6a); conversely, when the event A has happened in the recent past (when $\rho = 0.99$, we have $t_e < \lambda$ with $\lambda = 100$), the conditional waiting time for high p has a behaviour which approaches that of the interarrival time (i.e. with higher probability of failure than the independent case, as in Figure 6b).

4.1. Equivalent Return Period (ERP)

The return period is a means of expressing the exceedance probability. Despite being a standard term in engineering applications (in engineering hydrology in particular), the concept of return period is not always an adequate measure of the probability of failure and has been sometimes incorrectly understood and misused [*Serinaldi*, 2014]. The results discussed in previous Section strengthen the above message, extending it to correlated Z_t fields (with Markovian dependence); for the cases examined here, the statistics of the waiting or interarrival time show negligible differences with respect to the independent case for small values of ρ, while they are strongly affected by the autocorrelation when $\rho \gtrsim 0.5$ (see Figures 2 and 5). Consequently, using directly the probability of failure in engineering practice could be a better choice under the latter condition. However, although more effective and appropriate, the probability of failure may not be as simple to understand as the return period, which is already an established concept in applications and routinely employed by practitioners.

To overcome this problem, we introduce the concept of "equivalent" return period (ERP). Its aim is to retain the relative simplicity of the return period concept and extend it to temporally correlated hydrological variables; for correlated processes, ERP is defined to be the period that would lead to the same probability of failure pertaining to a given return period T in the framework of classical statistics (independent case). Hence, ERP resembles the classical definition of return period in the case of independence, thus preserving its simplicity and strength in indicating the event rareness; in addition it is able to control the probability of failure under the time-dependence condition.

ERP can be defined starting from the concept of interarrival time (N) or waiting time (W). Practitioners should adopt the most appropriate definition according to the circumstances, the task and the data available. If the time t_e elapsed since the last dangerous event is known, it could be adopted the definition based on the conditional waiting time, or that based on the interarrival time in the case $t_e = 0$; the latter could be the case where an existing structure failed because of an event A and the immediate construction of another structure is needed (as discussed by *Fernández and Salas* [1999a]).

In the case we are accounting for the interarrival time (N), ERP can be calculated assuming $R_N\left(ERP\right) = R\left(T\right)$ where R_N is the probability of failure based on the interarrival time (equation (2) for $f_X = f_N$), while $R\left(T\right)$ is given by equation (8) for $l = T$. For the 2Mp R_N is given by equation (17) (where $\Delta\tau = 1$) when $l = ERP$; thus, the analytical formulation of ERP can be easily derived as

$$ERP = 1 + \frac{\ln \frac{1-p}{p-q} + \frac{1}{1-p} \ln p}{\ln \frac{q}{p}} \tag{20}$$

For the AR(1), R_N can be numerically computed by substituting equation (19) in (2). In the case of more complex models for the simulation of hydrological quantities, ERP could be computed directly by numerical Monte Carlo simulations.

Figure 7 depicts the behaviour of ERP as function of T, for both the AR(1) (continuous lines) and 2Mp processes (dashed lines; equation (20) with $p = 1 - 1/T$). The figure shows that the values of ERP and T tend to coincide asymptotically; this is especially so for small correlation coefficients. For a given T, the value of ERP is always smaller (sometimes much smaller) than T; differences increase with the correlation coefficient ρ and with the correlation structure of the process (compare AR(1) to 2Mp). Recalling that $T = 1/(1-p)$, Figure 7 can be used either to determine ERP when the p-th quantile z is known (i.e., for a given event $A = \{Z > z\}$ that will be exceeded with probability $1 - p$) in risk assessment problems, or to determine the design variable (i.e. the threshold z) in terms of p once the ERP is fixed in design problems; in the latter case we choose ERP

and then calculate the design variable z, such that the probability of failure is equal to that we should have in the independent case.

We emphasize that results shown here are obtained under several assumptions, such as the type of temporal correlation, bivariate Gaussian distribution, etc.; this implies that, for example, a different distribution may result in larger differences between the independent and time-correlated conditions (due, e.g., to asymptotic dependence). Hence, further work is needed to generalize the above results.

5. Conclusions

The return period is a critical parameter largely adopted in hydrology for risk assessment and design. It is defined as the mean value of the waiting time to the next dangerous event (T_W) or the interarrival time between successive dangerous events (T_N). As shown in previous literature, both definitions lead to the same result in the case of time independence of the underlying process. However, in cases of time-persistent processes the two definitions lead to different expressions. Hence, we reexamine herein the above definitions in the context of temporally correlated processes; furthermore, by making use of two illustrative examples we discuss the effects of the temporal correlation ρ of the parent process on the return period and the probability of failure. The examples proposed here are based on a two state Markov-dependent process (2Mp), and an AR(1) process; even if the two processes share the Markov property, they are characterized by rather different time distributions.

The main conclusions drawn in this paper are listed below.

- We provide a unitary framework for the estimation of the return periods T_W, T_N and the related probabilities of failure R_W, R_N in the context of persistent processes:

we provide general relationships for the probability functions of the waiting time W (unconditional and conditional on the time t_e elapsed since the last dangerous event) and the interarrival time N. The choice between W and N in applications depends on the available information on past events and the type of structure.

• We demonstrate that the mean interarrival time T_N is not affected by the time-dependence structure of the process, e.g. the correlation coefficient ρ. Thus, the well known formula for independent processes is valid for any process, temporally correlated or not.

• Although T_N is not affected by ρ, for persistent processes the corresponding probability of failure can be much larger than that pertaining to the independent case, which is itself not negligible. Hence, the mean interarrival time T_N can easily provide a biased and wrong perception of the risk of failure, especially in the presence of temporally correlated hydrological variables.

• On the other hand, the mean waiting times effectively account for the correlation structure of the hydrological process. T_W is always larger than the mean interarrival time T_N, which acts as a lower bound. If the time t_e from the last dangerous event is deterministically known, we can use that information to condition the waiting time W to the next occurrence.

• The return periods T_W and $T_{W|t_e}$ typically increase with the correlation ρ. Specifically, they depend on the overall correlation structure of the process, as highlighted by comparing results for 2Mp and AR(1); in the case of processes characterized by a longer range persistence with respect to the AR(1), we may expect even stronger differences.

• The analyses carried out here provide some further insight into the overall meaning and significance of the return period, especially in view of hydrological applications, but also in other geophysical fields. Despite being a simple and easy to implement metric, the return period should be used with caution in the presence of time-correlated processes. Indeed, the probability of failure depends on the whole shape of the probability function, which in turn may strongly depend on ρ, and the return period is just the first order moment; the latter may not be relevant when in presence of asymmetric and skewed distributions, like e.g. some of those displayed in Figure 3.

• To partially overcome the above limitations, we propose to adopt in the time-dependent context the Equivalent Return Period (ERP), which preserves the virtue of the classical return period of effectively communicating the event rareness. ERP resembles the classical definition of return period in the case of independence, while it is able to control the probability of failure under the time-dependence condition.

We conclude with a note on the practical implications of the present analysis. Results shown here highlight that the independence condition is not necessary for the application of the classical return period equation; notwithstanding this, practitioners should take care of the time-persistence structure of the process when estimating risk from data, to correctly evaluate the probability of failure (e.g. through ERP). However, it is interesting to stress that the differences between the correlated and uncorrelated case are small to negligible when $\rho \lesssim 0.5$. Thus, the temporal correlation of the process may be safely disregarded in such cases, as far as the return period is concerned.

Acknowledgments. We thank the Editor and the three anonymous Reviewers for their thoughtful comments. The research has been partially funded by the Italian Ministry of University and Research through the projects PRIN 2010JHF437 and PRIN 20102AXKAJ. No data was used in producing this manuscript.

Appendix A: General relationships between f_W, $f_{W|t_e}$ and f_N

Since we can write that $\Pr\left(A_{-t_e}, B_{-t_e+1}, ..B_{-1}, B_0\right) = \Pr\left(B_0, B_1, ..B_{t_e-1}, A_{t_e}\right) = f_W\left(t_e+1\right)$, the probability mass function of the conditional waiting time, $f_{W|t_e}\left(t\right)$, can be expressed as function of f_W and f_N as in the following

$$\begin{aligned} f_{W|t_e}\left(t\right) &= \frac{\Pr\left(A_{-t_e}, B_{-t_e+1}, ..B_{-1}, B_0, B_1, B_2, ..B_{t-1}, A_t\right)}{\Pr\left(A_{-t_e}, B_{-t_e+1}, ..B_{-1}, B_0\right)} \\ &= \frac{\Pr\left(B_{-t_e+1}, ..B_{-1}, B_0, B_1, B_2, ..B_{t-1}, A_t | A_{-t_e}\right) \Pr A_{-t_e}}{\Pr\left(A_{-t_e}, B_{-t_e+1}, ..B_{-1}, B_0\right)} \\ &= \frac{(1-p)}{f_W\left(t_e+1\right)} f_N\left(t+t_e\right) \end{aligned} \tag{A1}$$

By making use of the simple identity $\Pr(C) = \Pr(AC) + \Pr(BC)$, which is valid for any events A and C (with B always denoting the complement of A), f_W can be expressed as function of f_N

$$\begin{aligned} f_W\left(t\right) &= \Pr\left(B_1, ..B_{t-1}, A_t\right) \\ &= \Pr\left(A_0, B_1, ..B_{t-1}, A_t\right) + \Pr\left(B_0, B_1, ..B_{t-1}, A_t\right) \\ &= \Pr\left(B_1, ..B_{t-1}, A_t | A_0\right) \Pr A_0 + \Pr\left(B_0, B_1, ..B_{t-1}, A_t\right) \\ &= f_N\left(t\right)\left(1-p\right) + f_W\left(t+1\right) \end{aligned} \tag{A2}$$

by solving equation (A2) for f_N and substituting the resulting expression in (A1) we obtain

$$f_{W|t_e}\left(t\right) = \frac{1}{f_W\left(t_e+1\right)}\left[f_W\left(t+t_e\right) - f_W\left(t+t_e+1\right)\right] \tag{A3}$$

Since f_N is a special case of $f_{W|t_e}$, when $t_e = 0$ equation (A3)

$$f_N(t) = \frac{1}{1-p}\left[f_W(t) - f_W(t+1)\right] \tag{A4}$$

Moreover, if we exploit the recursive property of equation (A2), we can write

$$\begin{aligned}
f_W(2) &= f_W(1) - (1-p) f_N(1) \\
f_W(3) &= f_W(2) - (1-p) f_N(2) \\
&= f_W(1) - (1-p) f_N(1) - (1-p) f_N(2) \\
f_W(4) &= f_W(3) - (1-p) f_N(3) \\
&= f_W(1) - (1-p) f_N(1) - (1-p) f_N(2) - (1-p) f_N(3) \\
&\dots
\end{aligned} \tag{A5}$$

thus obtaining

$$\begin{aligned}
f_W(t+1) &= f_W(1) - (1-p)\sum_{k=1}^{t} f_N(k) \\
&= (1-p)\left[1 - \sum_{k=1}^{t} f_N(k)\right] \\
&= (1-p)\left[1 - F_N(t)\right] \\
&= (1-p)\overline{F}_N(t)
\end{aligned} \tag{A6}$$

where we used $f_W(1) = \Pr A_1 = 1-p$ and the survival function of N, i.e. $\overline{F}_N(t) = 1 - F_N(t) = 1 - \sum_{k=1}^{t} f_N(k) = \sum_{k=t+1}^{\infty} f_N(k)$. The relationship between $f_{W|t_e}$ and f_N is obtained by substituting equations (A4) and (A6) into (A3)

$$f_{W|t_e}(t) = \frac{f_N(t+t_e)}{\overline{F}_N(t_e)} \tag{A7}$$

We adopt equation (A6) to derive the analytical expression of the return period T_W as function of f_N

$$\begin{aligned}\frac{T_W}{\Delta\tau} &= \sum_{t=1}^{\infty} t\left(1-p\right)\overline{F}_N\left(t-1\right) = \left(1-p\right)\sum_{t=1}^{\infty} t\overline{F}_N\left(t-1\right) \\ &= \left(1-p\right)\sum_{t=1}^{\infty} t\sum_{k=t}^{\infty} f_N\left(k\right) = \left(1-p\right)\sum_{k=1}^{\infty} f_N\left(k\right)\sum_{t=1}^{k} t \\ &= \left(1-p\right)\sum_{k=1}^{\infty}\frac{k\left(k+1\right)}{2}f_N\left(k\right) \\ &= \left(1-p\right)\left[\sum_{k=1}^{\infty}\frac{k^2}{2}f_N\left(k\right)+\sum_{k=1}^{\infty}\frac{k}{2}f_N\left(k\right)\right] \\ &= \frac{\left(1-p\right)}{2}\left(\mathrm{E}\left[N^2\right]+\mathrm{E}\left[N\right]\right)\end{aligned} \tag{A8}$$

Finally, substituting equation (A7) into (1) we obtain $T_{W|t_e}$ as function of f_N

$$\begin{aligned}\frac{T_{W|t_e}}{\Delta\tau} &= \sum_{t=1}^{\infty} t\frac{f_N\left(t+t_e\right)}{\overline{F}_N\left(t_e\right)} \\ &= \frac{1}{\overline{F}_N\left(t_e\right)}\sum_{t=1}^{\infty}\left[\left(t+t_e\right)f_N\left(t+t_e\right)-t_e\ f_N\left(t+t_e\right)\right] \\ &= \frac{1}{\overline{F}_N\left(t_e\right)}\left[\sum_{k=t_e+1}^{\infty} kf_N\left(k\right)-t_e\ \sum_{k=t_e+1}^{\infty} f_N\left(k\right)\right] \\ &= \frac{1}{\overline{F}_N\left(t_e\right)}\left[\sum_{k=t_e+1}^{\infty} kf_N\left(k\right)-t_e\ \overline{F}_N\left(t_e\right)\right] \\ &= \sum_{k=t_e+1}^{\infty}\frac{kf_N\left(k\right)}{\overline{F}_N\left(t_e\right)}-t_e\end{aligned} \tag{A9}$$

Appendix B: Mean interarrival time, T_N

Substituting equation (5), which is of general validity, in (1) we have

$$\begin{aligned}\frac{T_N}{\Delta\tau} &= \sum_{t=1}^{\infty} t\ f_N(t) = 1\Pr\{N=1\} + 2\Pr\{N=2\} + ... \\ &= \Pr(A_1|A_0) + 2\Pr(B_1, A_2|A_0) + 3\Pr(B_1, B_2, A_3|A_0) + ... \\ &= \frac{1}{\Pr A_0}\left[\Pr(A_0, A_1) + 2\Pr(A_0, B_1, A_2) + 3\Pr(A_0, B_1, B_2, A_3) + ...\right] \\ &= \frac{1}{1-p}\left[\Pr(A_0, A_1) + 2\Pr(A_0, B_1, A_2) + 3\Pr(A_0, B_1, B_2, A_3) + ...\right]\end{aligned} \tag{B1}$$

By making use again of the identity $\Pr(CA) = \Pr(C) - \Pr(CB)$, where B always denotes the opposite event of A, we obtain

$$\begin{aligned}\frac{T_N}{\Delta\tau} &= \frac{1}{1-p}\left[(\Pr A_0 - \Pr(A_0, B_1)) + 2(\Pr(A_0, B_1) - \Pr(A_0, B_1, B_2))\right. \\ &\left. +3(\Pr(A_0, B_1, B_2) - \Pr(A_0, B_1, B_2, B_3)) + ..\right] \\ &= \frac{1}{1-p}\left[\Pr A_0 + \Pr(A_0, B_1) + \Pr(A_0, B_1, B_2) + \Pr(A_0, B_1, B_2, B_3) + ..\right]\end{aligned} \tag{B2}$$

Using once more the same identity, we find

$$\begin{aligned}\frac{T_N}{\Delta\tau} &= \frac{1}{1-p}\left[(1 - \Pr B_0) + (\Pr B_1 - \Pr(B_0, B_1))\right. \\ &\left. + (\Pr(B_1, B_2) - \Pr(B_0, B_1, B_2)) + ..\right] \\ &= \frac{1}{1-p}\end{aligned} \tag{B3}$$

which proves to be valid because of stationarity, i.e. $\Pr B_0 = \Pr B_1$, $\Pr(B_0, B_1) = \Pr(B_1, B_2)$, etc.

References

Chow, V. T., D. R. Maidment, and L. W. Mays (1988), *Applied hydrology*, McGraw-HiII, New York.

Cooley, D. (2013), Return periods and return levels under climate change, in *Extremes in a Changing Climate*, pp. 97–114, Springer Netherlands.

Douglas, E. M., R. M. Vogel, and C. N. Kroll (2002), Impact of Streamflow Persistence on Hydrologic Design, *Journal of Hydrologic Engineering*, *7*(3), 220–227, doi: 10.1061/(ASCE)1084-0699(2002)7:3(220)

Du, T., L. Xiong, C. Xu, C. Gippel, S. Guo, and P. Liu (2015), Return Period and Risk Analysis of Nonstationary Low-flow Series under Climate Change, *Journal of Hydrology*, *527*, 220–227, doi:10.1016/j.jhydrol.2015.04.041.

Fernández, B., and J. D. Salas (1999a), Return period and risk of hydrologic events. II: Applications, *Journal of Hydrologic Engineering*, *4*(4), 308–316, doi:10.1061/(ASCE)1084-0699(1999)4:4(308).

Fernández, B., and J. D. Salas (1999b), 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).

Fuller, W. (1914), Flood flows, *Transactions of the American Society of Civil Engineers*, *77*, 564–617.

Gumbel, E. J. (1941), The return period of flood flows, *The annals of mathematical statistics*, *12*(2), 163–190.

Gumbel, E. J. (1958), *Statistics of Extremes*, Columbia University Press, New York.

Hurst, H. E. (1951), Long term storage capacities of reservoirs, *Transactions of the American Society of Civil Engineers*, *116*(776-808).

Kottegoda, N. T., and R. Rosso (1997), *Probability, statistics, and reliability for civil and environmental engineers*, McGraw-Hill, Milan.

Koutsoyiannis, D., and A. Montanari (2015), Negligent killing of scientific concepts: the stationarity case, *Hydrological Sciences Journal*, *60*(7-8), 2–22, doi: 10.1080/02626667.2014.959959.

Leadbetter, M. R. (1983), Extremes and local dependence in stationary sequences, *Probability Theory and Related Fields*, *65*(2), 291–306.

Lloyd, E. H. (1970), Return periods in the presence of persistence, *Journal of Hydrology*, *10*(3), 291–298.

Mandelbrot, B. B., and J. R. Wallis (1968), Noah, Joseph and operational hydrology, *Water Resources Research*, *4*(5), 909–918.

Montanari, A., and D. Koutsoyiannis (2014), Modeling and mitigating natural hazards: Stationarity is immortal!, *Water Resources Research*, *50*, 9748–9756, doi: 10.1002/2014WR016092.

Papoulis, A. (1991), *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, New York.

Read, L. K., and R. M. Vogel (2015), Reliability, Return Periods, and Risk under Nonstationarity, *Water Resources Research*, doi:10.1002/2015WR017089.Accepted.

Rosbjerg, D. (1977), Crossing and Extremes in Dependent Annual Series, *Nordic Hydrology*, *8*, 257–266.

Salas, J. D., and J. Obeysekera (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.

Saldarriaga, J., and V. Yevjevich (1970), *Application of run-lengths to hydrologic series, Hydrology Paper N.40*, Colorado State University, Fort Collins.

Sen, Z. (1999), Simple risk calculations in dependent hydrological series, *Hydrological Sciences Journal*, *44*(6), 871–878, doi:10.1080/02626669909492286.

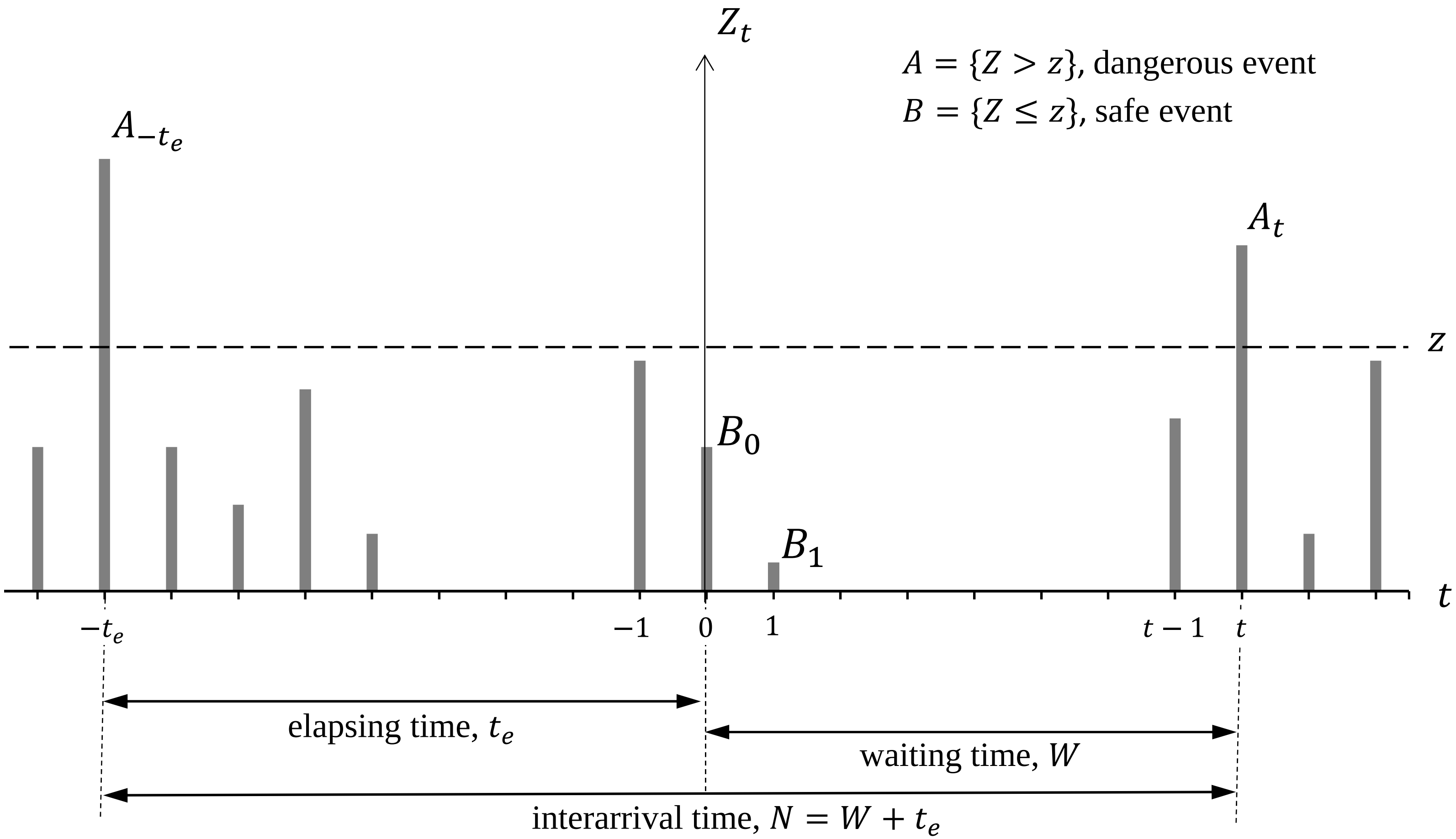
Serinaldi, F. (2014), Dismissing return periods!, *Stochastic Environmental Research and Risk Assessment*, pp. 1–11, doi:10.1007/s00477-014-0916-1.

Stedinger, J. R., R. M. Vogel, and E. Foufoula-Georgiou (1993), Frequency analysis of extreme events, in *Handbook of Hydrology*, edited by D. Maidment, chap. 18, McGraw-Hill, New York.

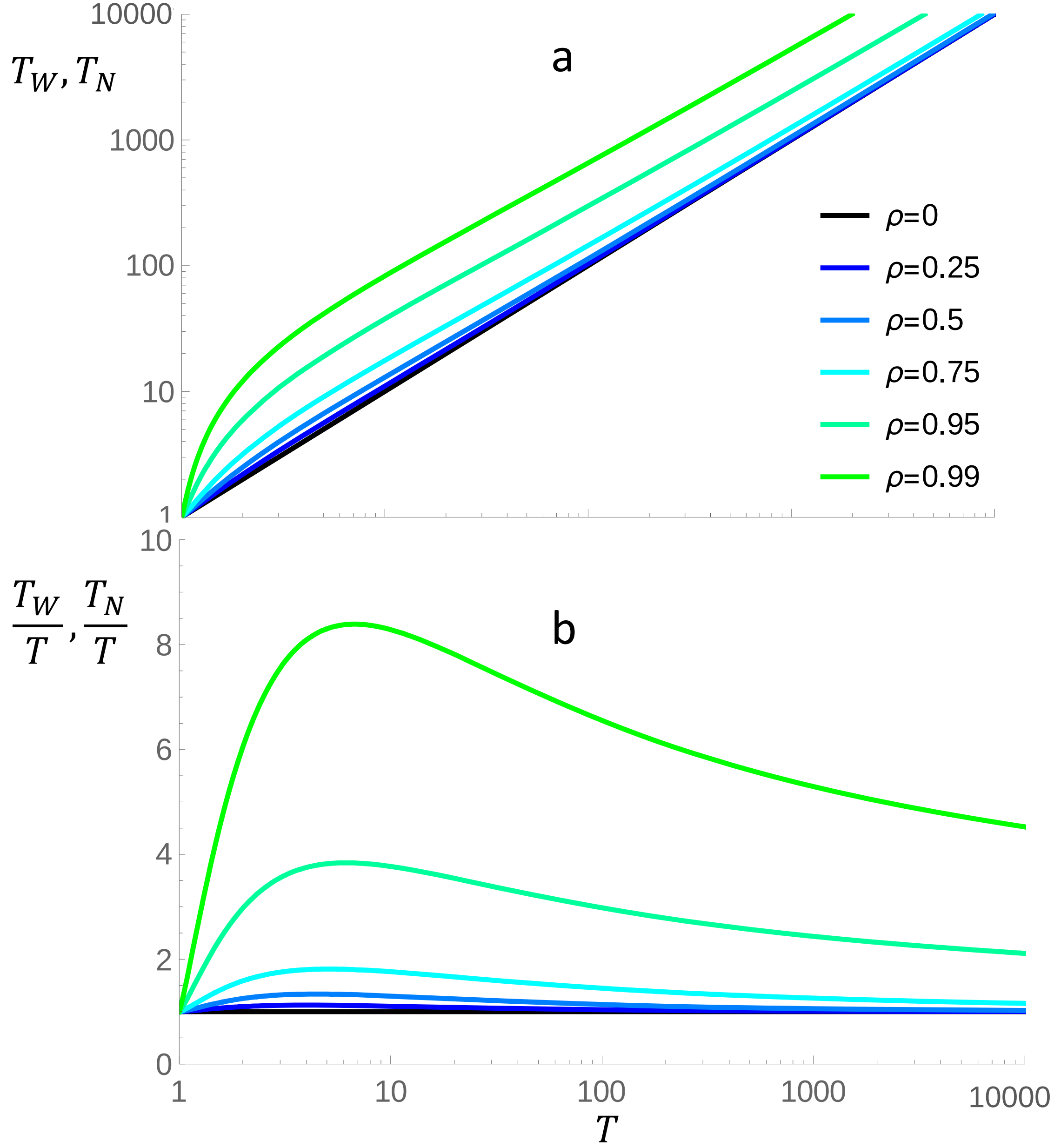
List of Figures

refer to the cases $\rho = 0.75$ (a, c) and $\rho = 0.99$ (b, d). Note that $R_{W|t_e}\left(T_{W|t_e}\right)$ of 2Mp is valid for any $t_e > 0$. Results are compared to the independent case (black line).

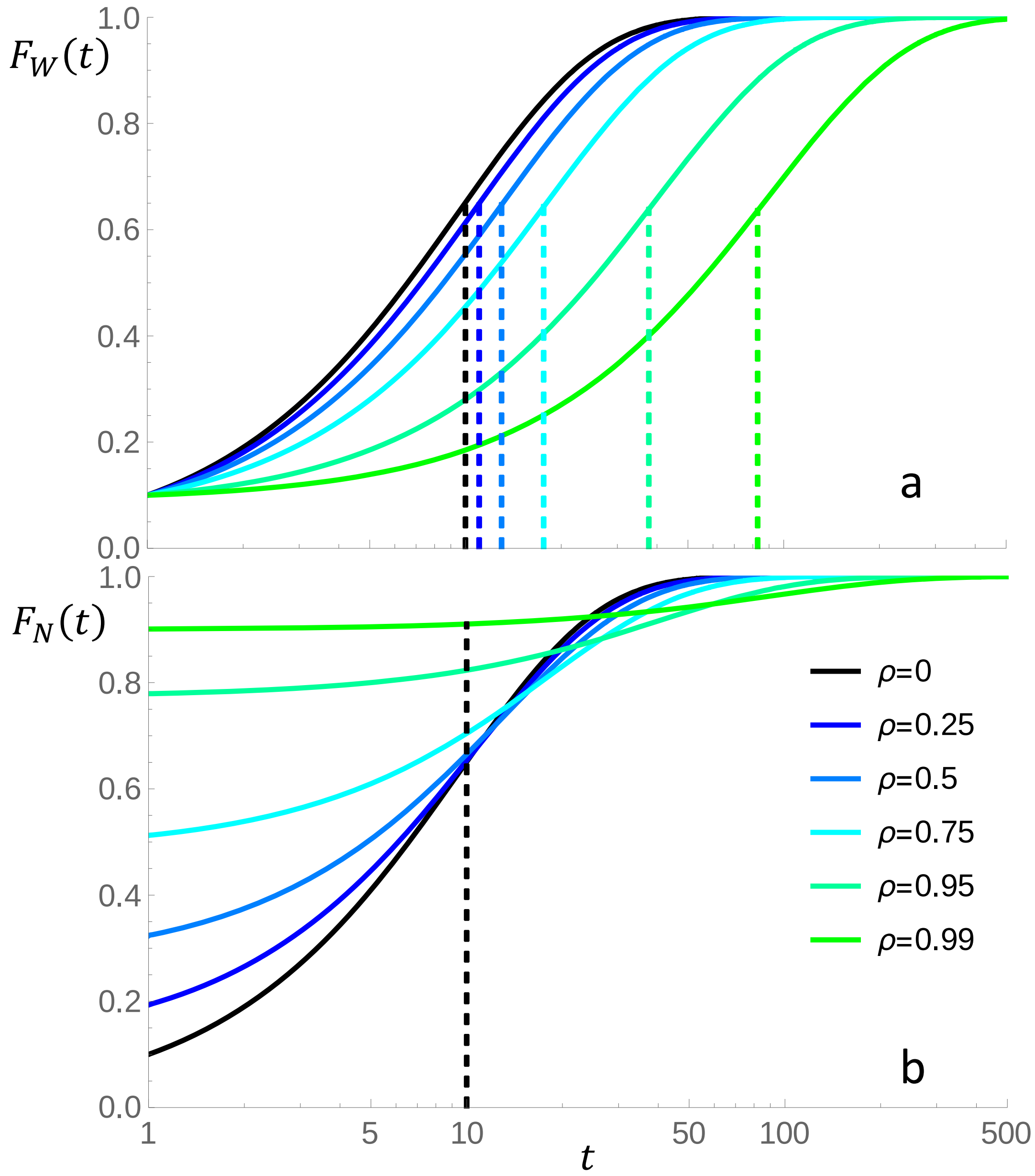
7. Equivalent Return Period (ERP), based on the interarrival time N, as function of the independent return period T for several values of the lag-1 correlation coefficient ρ; curves for $\rho < 0.75$ are not shown because the differences between ERP and T are small to negligible.



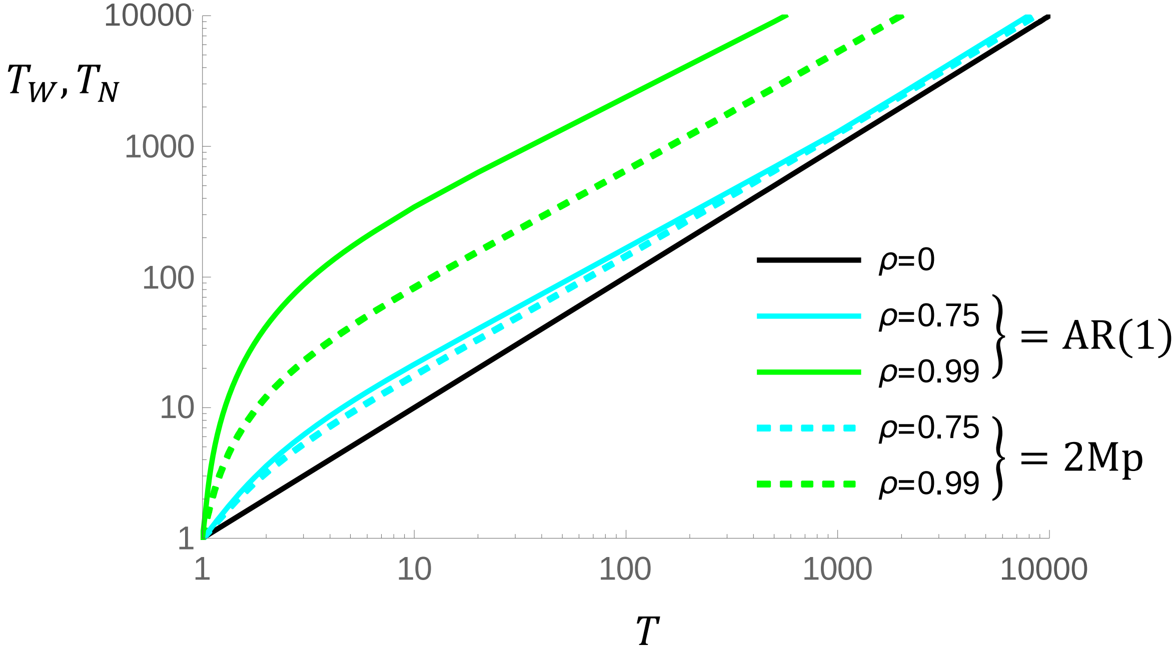
Z_t
$A = \{Z > z\}$, dangerous event
$B = \{Z \leq z\}$, safe event
A_{-t_e}
A_t
z
B_0
B_1
t
$-t_e$
-1
0
1
$t-1$
t
elapsing time, t_e
waiting time, W
interarrival time, $N = W + t_e$



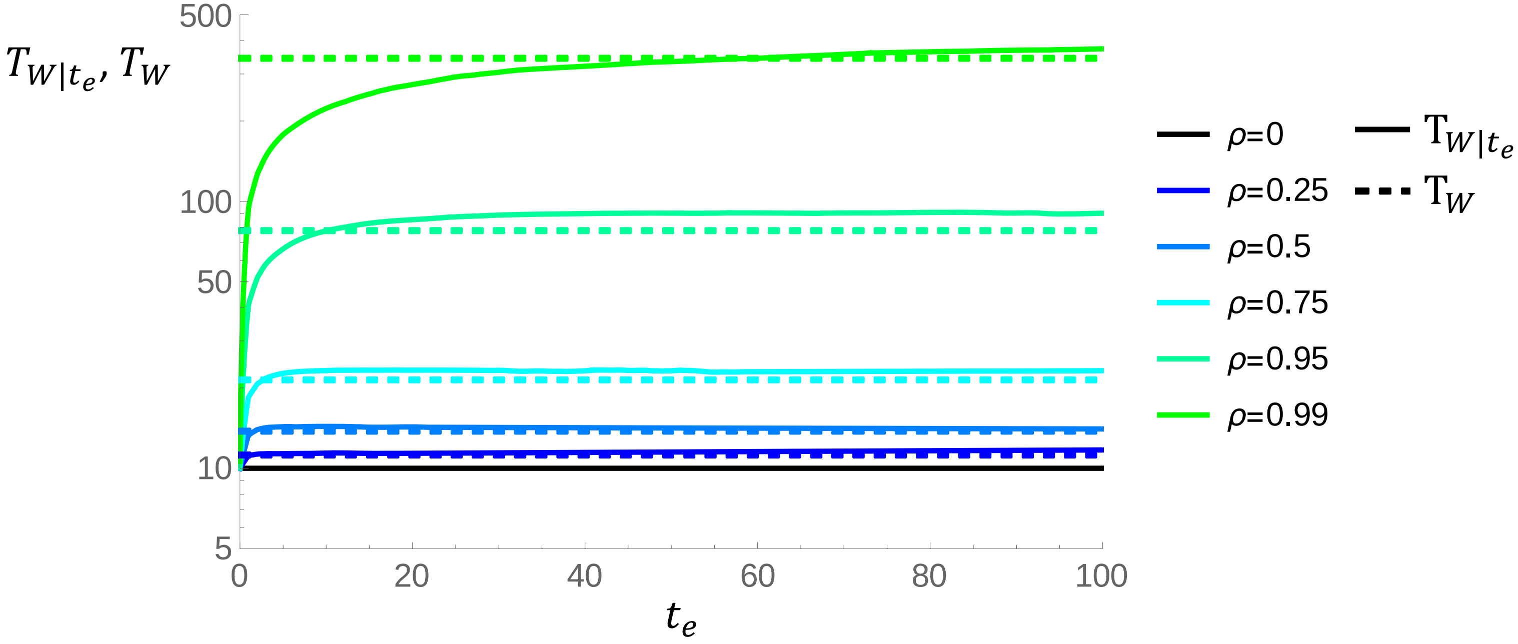
a
b
T_W, T_N
$\frac{T_W}{T}, \frac{T_N}{T}$
T
ρ=0
ρ=0.25
ρ=0.5
ρ=0.75
ρ=0.95
ρ=0.99
1
10
100
1000
10000
0
2
4
6
8



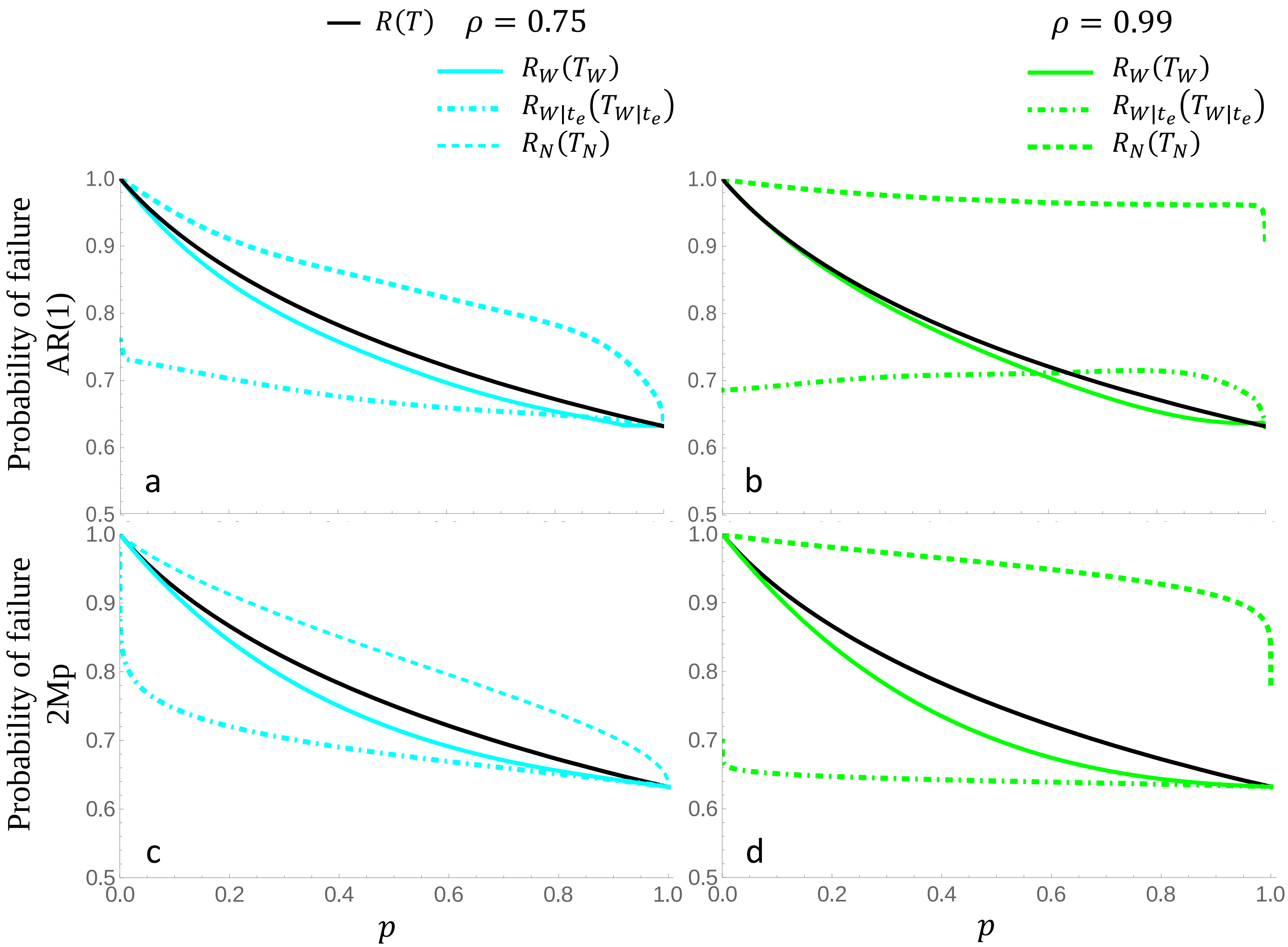
$F_W(t)$
1.0
0.8
0.6
0.4
0.2
0.0
a
$F_N(t)$
1.0
0.8
0.6
0.4
0.2
0.0
b
ρ=0
ρ=0.25
ρ=0.5
ρ=0.75
ρ=0.95
ρ=0.99
1
5
10
50
100
500
t



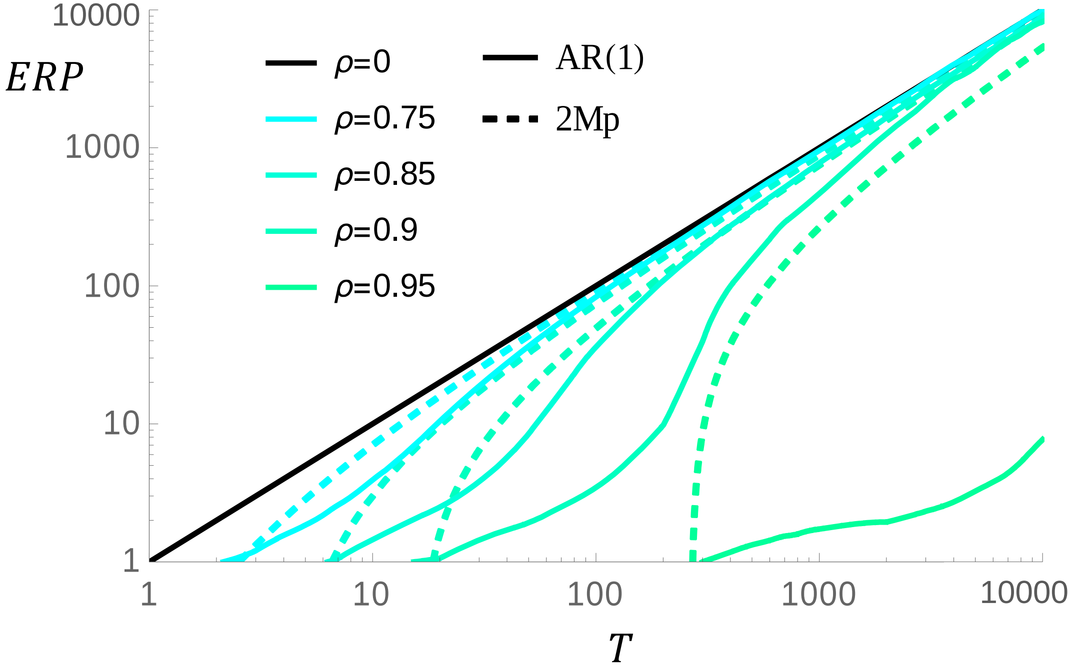
10000
1000
100
10
1
T_W, T_N
1
10
100
1000
10000
T
ρ=0
ρ=0.75
ρ=0.99
= AR(1)
ρ=0.75
ρ=0.99
= 2Mp



$T_{W|t_e}, T_W$
500
100
50
10
5
0
20
40
60
80
100
t_e
ρ=0
ρ=0.25
ρ=0.5
ρ=0.75
ρ=0.95
ρ=0.99
$T_{W|t_e}$
T_W



$R(T)$ $\rho = 0.75$
$R_W(T_W)$
$R_{W|t_e}(T_{W|t_e})$
$R_N(T_N)$
$\rho = 0.99$
$R_W(T_W)$
$R_{W|t_e}(T_{W|t_e})$
$R_N(T_N)$
Probability of failure AR(1)
Probability of failure 2Mp
a
b
c
d
1.0
0.9
0.8
0.7
0.6
0.5
0.0
0.2
0.4
0.6
0.8
1.0
p



10000
1000
100
10
1
ERP
ρ=0
ρ=0.75
ρ=0.85
ρ=0.9
ρ=0.95
AR(1)
2Mp
1
10
100
1000
10000
T