Research Papers
Issue RP0139
September 2012

Impact on Soil and Coasts Division (ISC)

A brief introduction to the concept of return period for univariate variables

By **Renata Vezzoli**
Impact on Soil and Coasts Division (ISC), Euro-Mediterranean Center on Climate Change (CMCC), via Maiorise s.n.c, I-81043, Capua
renata.vezzoli@cmcc.it

Paola Mercogliano
Impact on Soil and Coasts Division (ISC), Euro-Mediterranean Center on Climate Change (CMCC), via Maiorise s.n.c, I-81043, Capua
pmercogliano@cira.it

Silvano Pecora
Servizio Idro Meteo Clima - Area Idrografia e Idrologia (SIMC), Agenzia Regionale per l'Ambiente Emilia Romagna (ARPA), via Garibaldi, I-43121, Parma
specora@arpa.emr.it

SUMMARY Here the concept of return period is discussed in terms of the stationarity of the process. If the process is stationarity the concept of return period is well defined and ambiguities do not arise. If the process is non stationary, as it could be any climate driven phenomenon under climate change assumption, the meaning of the return period varies because it could be read in a more or a less conservative way. In this work, we refer to the three possible alternatives of non stationary return period concept given in [5].

Keywords: Definition, stationarity, non-stationarity, univariate, return period

This report is carried on within a MOU agreement between CMCC and ARPA Emilia Romagna SIMC.

Introduction

Climate is characterized by a natural variability in precipitation and temperature and that variability reflects also in floods and droughts occurrence and intensity. The detection of climate change signals leads to the question of the reliability of the long term persistence of climate (hydrological) conditions and to need of a non stationary approach to the problem. In literature there are several studies on the non stationarity of climate variables [4, 6, 9, 1]. Changes in temperature and precipitation extremes simulated using a second-generation coupled GCM of the Canadian Centre for Climate Modelling and Analysis are studied by [4]. They estimated the return values of annual extremes from a generalized extreme value distribution with time-dependent location and scale parameters. The results show that changes in temperature extremes are expressed by changes in the location parameter of the distribution, while changes in precipitation extremes are reflected in changes in both location and scale distribution parameters. Moreover it results that the probability of extreme precipitation events almost double from the beginning to the ending of the simulations, thus the return period of the events decreased along the simulation period. The changes in extreme European winter (December - February) precipitation in the last 500 year are studied by [6] that note an high variability of the return periods of extremely wet and dry winters before and after the onset of the anthropogenic influences. The existence of a positive trend in the precipitation observations in Sao Pao (Brasil) from 1933 to 2005 with an increase of about 40 mm in the 0.99 quantile (100 yr return period) is the results of [9] analysis. More recently [1] investigates the non stationarity of daily rainfall in northeast Spain using the peaks-over-threshold (POT) approach. Results indicates that less than 5% of the stations are characterised by a linear trend at annual scale, i.e. there is a very low evidence of a generalised trend at annual scale, while at seasonal scale, Catalonia will experience an increase in spring precipitation and a decrease in winter events. Since climate change affects also the frequency of hydrological extreme phenomena (droughts and floods) there is interest in studying the climate driven impacts of floods and droughts on the socio-economical condition over an area, among the studies available in literature we cite [3, 2, 10, 11, 7, 8]. The future behavior of extremes in river discharge under climate change from simulated daily discharge are investigate by [3]. The results forecast an increase of floods and droughts frequency over many regions, with sometimes an increase in both floods and droughts frequency over the same region. [2] uses climate simulations to project in the future the impacts of climate on extreme droughts in UK. They find, that in 2070, a maximum increase in drought severity of about 125%, and a decrease in maximum drought duration up to 50%. The variability of flood risk related to climate and land use change in the Rhine basin is studied in [10, 11] that also provide an estimate of the costs. As main results they found that an increase of about 8% - 17% in Rhine extreme flood peak should be expected in 2050 for all the return period between 10 and 1250 years, [10]; and, more generally that the flood risk is a not stationary variable and might considerably increase over a period of several decades, [11]. Both studies [10, 11] agree that the area that will be more affected by floods will be located in the Lower Rhine in Germany because of the different flood protection standards used in the Netherlands and Germany. [7, 8] address the question of the non stationarity distribution of hydrologic time series in Canada using a combination of both non stationary probability distributions and deterministic rainfall-runoff models.

The non stationary behavior of daily maximum flow is modelled from the annual maximum simulated by the deterministic hydrologic model. The results show that a non-stationary distribution better estimates the peak flood quantiles (return level) than a stationary one, [7, 8].

Notwithstanding, new infrastructures are typically designed on the basis of normative values (maximum value observed or quantile) derived from historical information under the hypothesis of climate stationarity, neglecting climate change effects, [5]. At the same time adaptation strategies to climate change need to incorporate the foreseen changes in extreme values to buffer the impacts of climate change on the failure risk of infrastructure itself, [5, 10, 11].

For this reason we propose a brief discussion on the concept of return period, often used to define the normative value for infrastructure design, for stationary and non stationary processes. If the process is stationarity the concept of return period is well defined and any ambiguities arise. If the process is non stationary the concept of "average waiting time" can be interpreted in more or less conservative way. We present the possible interpretations given in [5] and we suggest an alternative way to provide a complete description of the return level for a non stationary process, i.e. providing not only its average value but also a measure of its dispersion within the reference period.

Return period in the stationary case

If a process is stationary the *return period* of a given event is defined as the average time elapsing between two successive realizations of the event itself or alternatively the return level is the value expected to be exceeded on average, once every return period, or with probability 1/(return period) in any given year, [5]. If the variable of interest is expressed as exceedence over a threshold (also known as POT analysis in hydrology) the return period T can be expressed as a function of the probability distribution function F_X and of the average waiting time between two events μ_T:

$$T = \frac{\mu_T}{1 - F_X(x)} \tag{1}$$

For example in hydrology, in to the case of peak flood analysis the sample size is increased by considering all the events above a certain threshold, so the sample x is composed by all the observations above the threshold and $\mu_T > 1$.

If the variable of interest are annual maxima (minimum) values μ_T is equal to 1 yr/event and equation (1) simplifies as:

$$T = \frac{1}{1 - F_X(x)} \tag{2}$$

Theoretically, from equations (1) and (2) is possible to estimate long return period, but these values are highly affected by the uncertainties associated to the size of the sample and the shape of the tails of the estimated distribution, i.e. to the capacity of the distribution to well describe the data. The confidence in the estimated values of the return period rapidly decays for return period more than twice the length of the observation period, [5]. For example, with reference to equation (2), the return period associated to $F_X(x) = 0.99$ and $F_X(x) = 0.995$ are 100 yr and 200 yr respectively; in this case an uncertainty of 5% in the value of the distribution function corresponds to an error of 100% in the estimated return period.

BOX 1: SOME DEFINITIONS OF STATIONARITY

A stochastic process is defined stationary if its mean, variance, and auto-covariance are time independent. There are different types of stationarity according to the different interpretation of the dependence concept:

Weak stationarity the mean and variance of the process are constant, and for any t, $h \geq 1$ $Cov(x_t, x_{t+h})$ depends only on h and not on t;

Weakly dependent for increasing values of h, x_t and x_{t+h} are "almost independent", meaning that $\lim_{h \to +\infty} Corr(x_t, x_{t+}h) = 0$;

Higher order stationarity all the moments of the distribution are stationary, not only mean and variance.

Usually with "stationarity" is indicated the weak stationarity.

Return period in a non stationarity case

If the stationarity hypothesis does not hold the definition of return period/return level as "the average time elapsing between two successive realizations of the event itself" is no more unique according to the interpretation given to the concept of waiting time. [5] gives three alternative definitions of return level under a changing climate, that is equivalent to say under non stationary conditions:

1. the level with a probability of exceedance currently equal to $1/T$ provides an estimate of the current risk for an extreme event with magnitude at least equal to the return value, that is equivalent to work at the best of the actual knowledge;

2. the level with a probability of exceedance in any one year never greater than $1/T$ over a fixed period, that is equivalent to consider the more extreme case, since for every year a value for the return level is estimated and the one that has a probability of exceedance never greater than $1/T$ is the return level to be used, this value is the maximum among the computed ones;

3. a level with an average probability of exceedance over a fixed period equal to $1/T$. In this case, the probability of exceedance would varying along the fixed period but be in average the equal to $1/T$. This definition may be a design criteria for infrastructures because it considers their risk of failure due to an extreme event over their whole lifetime.

A further step forward will be to provide not only the average level of return but also its range (the difference between the maximum and minimum value) to evaluate the spread of the return level or, even better, the probability distribution of the return level to fully characterised the process.

The second and the third definitions require the knowledge of how the event could evolve in time, that will be possible using a climate scenario and performing numerical simulations of the process of interest. In this case it will be possible to derive a return level within each year of the reference period. The main difficulty is to probabilistic describe the variable within each year, specially if the value are annual minimum or maximum, because the cumulative density function of each year should be derived from just one data! Among the possible alternatives to avoid this problem there are (1) to run more simulations of the same year, i.e. to perform a stationary analysis of each year, to obtain a sample of the variable big enough to evaluate the cumulative density function and the return levels, see e.g. [3, 10, 11, 1] or (2) to work with probability distribution function with parameters varying in time prescribing, a priori, the evolution of the process in the time, by doing so the whole reference period is described by the same distribution function with non constant parameters. Since the parameters are generally related to the moment of the distribution their variability in time can be derived from the expected evolution in time of the variable, while the shape of the distribution function can be inferred from the available observed data see e.g. [4, 7, 8]. The mathematical relationship between the distribution function and the return period within each year is the same given in the stationary case.

Bibliography

[1] S. Beguerìa, M. Angulo-Martìnez, S.M. Vicente-Serrano, J. Ignacio López-Moreno, and A. El-Kenawy. Assessing trends in extreme precipitation events intensity and magnitude using non-stationary peaks-over-threshold analysis: a case study in northeast spain from 1930 to 2006. *Int. J. Climatol.*, 31:2102 – 2114, 2011.

[2] H.J Fowler and C.G Kilsby. Future increases in uk water resources projected by a regional cliamte model. *Hydrology: Science & practice for the 21st century*, 1:15 – 21, 2004.

[3] Y. Hirabashi, S. Kanae, S. Emori, T. Oki, and M. Kimoto. Global projections of changing risks of floods and droughts in a changing climate. *Hydrological Sciences Journal*, 53(4):754 – 772, 2008.

[4] V.V. Kharin and F.W. Zwiers. Estimating extremes in transient climate change simulations. *Journal of Climate*, 18:1156 – 1173, 2005.

[5] A.M.G. Klein Tank, F.W. Zwiers, and X. Zhang. Guidelines on analysis of extremes in a changing climate in support of informed decisions for adaptation. Technical Report 72, World Meteorological Organization, 2009.

[6] A. Pauling and H. Paeth. On the variability of return periods of european winter precipitation extremes over the last five centuries. *Clim. Past Discuss.*, 2:157 – 189, 2006.

[7] O. Seidou, A. Ramsay, and I. Nistor. Climate change impacts on extreme floods i: combining imperfect deterministic simulations and non-stationary frequency analysis. *Nat. Hazard*, 61:647 – 657, 2012.

[8] O. Seidou, A. Ramsay, and I. Nistor. Climate change impacts on extreme floods ii: improving flood future peaks simulation using non-stationary frequency analysis. *Nat. Hazard*, 60:715 – 726, 2012.

[9] S. Sugahara, R.P. da Rocha, and R. Silveira. Non-stationary frequency analysis of extreme daily rainfall in sao paulo, brazil. *Int. J. Climatol.*, 29:1339 – 1349, 2009.

[10] A.H. te Linde, J.C.J.H. Aerts, A.M.R. Bakker, and J.C.J. Kwadijk. Simulating low probability peak discharges for the rhine basin using resampled climate modeling data. *Water Resour. Res.*, 46:W04512, 2010.

[11] A.H. te Linde, P. Bubeck, J.E.C. Dekkers, H. de Moel, and J.C.J.H. Aerts. Future flood risk estimates along the river rhine. *NHESS*, 11:459 – 473, 2011.