Research Communication

USE OF THE SPATIAL EXTREMOGRAM TO FORM A HOMOGENEOUS REGION CENTERED ON A TARGET SITE FOR THE REGIONAL FREQUENCY ANALYSIS OF EXTREME STORM SURGES

Y. HAMDI, C-M. DULUC, L. BARDET & V. REBOUR
Institut de Radioprotection et de Sûreté Nucléaire, France.

ABSTRACT

Nuclear power plants in France are designed for low probabilities of failure. Nevertheless, some exceptional surges, considered as outliers, are not properly addressed by statistical models. Regional information may be used to mitigate the paucity of data and the influence of outliers. A regional frequency analysis (RFA) model assumes a homogenous behavior of the variable of interest at a regional scale. The first stage in an RFA is the delineation of homogeneous regions. We propose herein a new approach to form a homogenous region centered on a target site and using the spatial extremal dependence between observations (the spatial extremogram) to measure the neighborhood between sites. Skew surge data set collated at a total of 19 sites located on the French coast (Atlantic and English Channel) were used as the case study for this contribution (with La Rochelle as a target site). Once a physically plausible region of interest is defined, a related issue regards the statistical homogeneity of the region of interest. The L-moment-based homogeneity tests of Hosking and Wallis, widely used in hydrology, are used in this paper. The principle of the extremogram allows us to form a physically and statistically homogenous group of sites centered on a target site. It also allows to overcome the problem of the so-called "border effect" and these are two key original points of the developed concept.

Keywords: extreme surges, homogenous region, regional frequency analysis, spatial extremogram, target site.

1 INTRODUCTION

Relating extreme events to their frequency of occurrence using probability distributions has been a common issue since the 1950s [1, 2]. The at-site frequency analysis (FA) corresponding to long return periods is based on the extreme value theory [3]. The objective of an FA is to analyze records so as to estimate the probability of future occurrence. Since the presence of an outlier (induce by an exceptional event) often leads to poor frequency estimations, it must be properly addressed in the FA. The probabilistic and statistical treatment of surge data containing outliers is limited in the literature. Additional data may be used to mitigate the lack of data and the presence of the outlier. The additional data may have a spatial dimension (making use of regional data). A regional frequency analysis (RFA) may be used to make data sets longer and to increase the representativeness of the outlier in the sample. The objective of an RFA is to use information, within a homogeneous region, from gauged sites to an ungauged or poorly gauged target site. RFAs are commonly used and have been extensively studied [1, 4–6]. The stages of a standard RFA are: (i) delineation of homogeneous regions; (ii) regional frequency estimation of the quantiles of interest. A comparative study of various RFAs was presented by GREHYS [7, 8]. One of the simplest and most popular approaches, privileged by engineers and widely used in hydrology, is the index flood method [1]. Its major hypothesis is that the probability distributions at different sites in the region are identical,

ISSN: 2041-9031 (paper format), ISSN: 2041-904X (online), http://www.witpress.com/journals
DOI: 10.2495/SAFE-V6-N4-777-781

except for a scale parameter. Recently, the RFAs have been applied to coastal hazards [9, 10]. However, these applications were limited by the problem of the inter-site dependence issue and its impact on the variance of local and regional quantiles and on the statistical homogeneity. Recent developments go a step further and involve a procedure to consider the spatial dependence structure using copulas [11, 12] and to form physically homogenous regions. Weiss [12] proposed a criterion (related to the spatial propagation of storms) to form physically homogeneous regions.

The existing delineation procedures do not give specific weight to the target site which may lead to a loss of some relevant local effects. Furthermore, the delineation of a homogenous region usually leads to the problem of the so-called "border effect" (information at the site located on the other side of the region is excluded even though both sites have similar asymptotic properties). In our view, these elements motivate the development of a method that overcomes these drawbacks. In this communication, we put forward a new approach to form a physically homogeneous region for further discussion to enrich the debate on RFA models. The approach is based on the empirical spatial extremogram to define a homogenous region around a target site. The neighborhood between sites is measured by a degree of physical and statistical inter-site tail dependence. The region constituting the neighboring sites moves from one target site to the other. The principle of moving regions around target sites allow us to overcome the "border effect" problem and is one of the key original points of the developed model. The formation of a homogenous region centered on the La Rochelle site in France is performed with the approach proposed herein. Skew surge data sets at a total of 19 sites located on the French coast (Atlantic and English Channel) were used as the case study for this contribution.

2 CONCEPT OF THE SPATIAL EXTREMOGRAM

As a starting point for the definition of the neighborhood between sites, we consider the neighborhood threshold N_t and the extremal dependence. These two parameters are defined for any site proposed to be neighboring to the target site and then be part of the region of interest. The extremal dependence can be viewed as the probability that a surge observed at a site is extreme (greater than a fixed threshold u) given that the surge observed at the same time at the target site belongs to an extreme set. This is the principle of the extremogram proposed by Davis and Mikosch [13] and Chavez and Davison [14]. Indeed, this extremogram is defined as a correlogram of extreme events and is developed to be a versatile and flexible tool for assessing extremal dependence in stationary time series. Let S_i and S_j be the time series of storm surges at two sites i and j, the upper tail dependence coefficient is given by

$$\varphi\left(S_i,S_j\right)=\lim_{u\to\infty}P(S_i>u|S_j>u) \tag{1}$$

The larger is the φ, the larger is the extremal dependence between the sequences S_i and S_j (of course $\varphi\in[0,1]$). The two sequences are considered asymptotically independent when $\varphi=0$ and perfectly dependent when $\varphi=1$. The tail dependence coefficient φ is often applied to pairs of one-dimensional stationary time series to study the serial extremal dependence in a temporal sequence (analog to the autocorrelation function ACF). We refer, for instance, to the discussions in Ledford and Tawn [15] on the tail dependence coefficient. In the present contribution, we define the notion of the empirical spatial extremogram (ESE) centered on a target site (defined in the spatial dimension when the data are observed at many sites). Unlike the ESE proposed by Cho *et al.* [16], the ESE centered on a target site does not estimate the extremal dependence based on distances and lags but the pairwise extremal dependences (between any site and a target site) are used separately. We refer to the discussion in Cho *et al.* [16] on the ESE. On the other hand, once a physically plausible region of interest is defined, it is reasonable to assess whether the proposed region is meaningful and may be accepted as being homogenous (since homogeneity is a prerequisite for the FA). It is also desirable to assess whether one or more sites are consistent with the principle of similarity of at-sites frequency distributions in a homogenous region and whether they should be removed from the region by increasing the neighborhood threshold N_t. The L-moment-based homogeneity tests of Hosking and Wallis [4, 5], widely used in hydrology, are used herein.

3 APPLICATION

3.1 Data

Surge data sets were used for 19 sites located on the French coast (Atlantic and English Channel) as the case study (Fig. 1). The developed RFA is performed at La Rochelle (as a target site). One of the most important features of the La Rochelle site is the fact that it has experienced significant storms during the last two decades (Martin in 1999 and the Xynthia in 2010). Fig. 1 displays the geographic location of the whole region and the site of La Rochelle. As depicted in the left-hand panel of Fig. 1, the available surge data were generally recorded in the period from 1940 to 2015 except for Brest where recording began in 1846. The figure also shows that the vast majority of sites are poorly gauged. The La Rochelle site (1941–2010) has an effective duration of only 26 years.

3.2 The homogeneous region centered on the La Rochelle Site

The scheme for obtaining the extremal dependence between the La Rochelle site and all the other sites is applied herein. From the ESE based on the pairwise extremal dependence, depicted in the right-hand panel of Fig. 1, a geographically coherent region of interest is obtained. Fig. 1 presents the distributions of the ESE with a neighborhood threshold $N_t = 0.6$. This means that in 60% of the time in the common observation periods, the region of interest

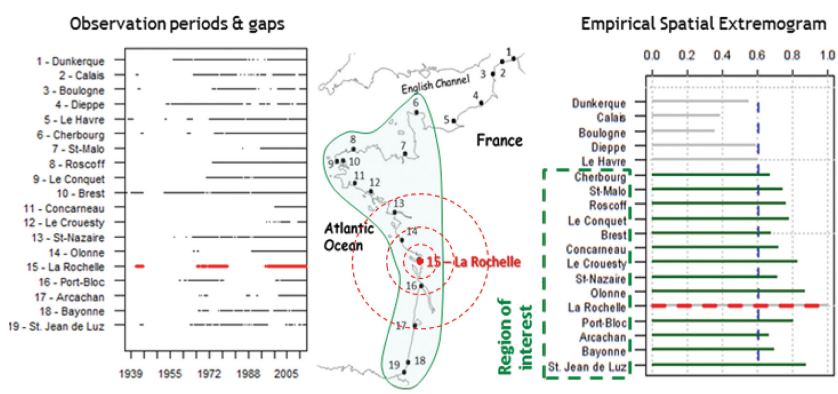

Figure 1: Location of sites and region of interest (middle); data and gap periods at each site The ESE centered on the target site $N_t = 60\%$; $u = 0,75$.

has experienced an extreme surge at the same time. From a physical point of view, these extreme surges are likely to be induced simultaneously by the same storms. Regarding the extreme surge definition and extraction, a standardized value $u = 0.75$ is used as a sufficiently high threshold above which a surge is considered extreme.

The ESE is performed on all the available sites to define the region of interest around the target site. It is noteworthy that the propagation of a storm between sites is taken into account in the definition of simultaneous occurrences of extremes. Indeed, identifying extreme simultaneous surges is performed in a 1-day time window. As stated earlier the statistical homogeneity can be verified by applying the regional heterogeneity and the at-site discordancy measures proposed by Hosking and Wallis [4, 5]. The calculation and the analysis of these statistical measures confirm that the hypothesis of homogeneity of the region is accepted.

4 CONCLUSION AND FUTURE WORK

A new approach based on the spatial extremal dependence and the empirical spatial extremogram (ESE), aiming at forming a homogenous region centered on a target site for the RFA, has been proposed and applied to extreme surges at a poorly gauged target site (La Rochelle), which is located on the French coast (Atlantic and English Channel). The asymptotic dependence and ESE based approach leads to a homogenous region centered on the target site and avoids the problem of the "border effect". Indeed, the developed approach assigns a significant weight to the data at the target site (often poorly gauged) which allow us to consider some relevant local effects in the RFA. The approach presented in this communication is thus promising to estimate extreme events at poorly gauged sites. One question that nevertheless remains open is related to the robustness of the proposed approach. An in-depth study could help to thoroughly analyze these results by conducting an RFA with the obtained homogenous region. This could generate several important perspectives and questions. An extremal dependence threshold of 0.6 and an arbitrary high threshold of 0.75 to extract extreme normalized surges were assumed. The extreme quantile corresponds to a probability of exceedance of 1% or 2% may be used to extract normalized extremes. Sensitivity analyses to these settings (since they are directly related to the size and the extent of the region of interest and to the extremogram) and further research on other target sites could provide us with solutions and answers to these problems and questions.

ACKNOWLEDGEMENTS

The authors are grateful to the editor for the importance he gave to the present work. Special thanks are addressed to Yves Deville (consulting engineer) for its theoretical support and guidance. The authors are also grateful to the SHOM (Service Hydrographique et Océanographique de la Marine) for providing data.

REFERENCES

[1] Dalrymple, T., *Flood-Frequency Analyses. Manual of Hydrology: Flood-Flow Techniques*, Geological Survey Water-Supply: Washington, 1960.

[2] Rao, A.R. & Hamed, K.H. (eds), *Flood Frequency Analysis*, CRC Press, Boca Raton, Florida, USA, 2000.

[3] Coles, S. (ed), *An Introduction to Statistical Modeling of Extreme Values*, Springer, Berlin, 2001.

[4] Hosking, J.R.M. & Wallis, J.R., Some statistics useful in regional frequency analysis. *Water Resources Research*, **29**, pp. 271–282, 1993.
https://doi.org/10.1029/92WR01980
[5] Hosking, J.R.M. & Wallis, J.R., *Regional Frequency Analysis: An Approach Based on L-Moments,* Cambridge University Press: Cambridge, U.K, 1997.
https://doi.org/10.1017/CBO9780511529443
[6] Ouarda, T.B.M.J., Girard, C., Cavadias, G.S. & Bobée, B., Regional flood frequency estimation with canonical correlation analysis. *Journal of Hydrology*, **254**, pp. 157–173, 2001.
https://doi.org/10.1016/S0022-1694(01)00488-7
[7] Groupe de Recherche en Hydrologie Statistique, Presentation and review of some methods for regional flood frequency analysis. *Journal of Hydrology*, **186**, pp. 63–84, 1996a.
https://doi.org/10.1016/S0022-1694(96)03042-9
[8] Groupe de Recherche en Hydrologie Statistique, Inter-comparison of regional flood frequency procedures for Canadian rivers. *Journal Hydrology*, **186**, pp. 85–103, 1996b.
https://doi.org/10.1016/S0022-1694(96)03043-0
[9] Bardet, L. & Duluc, C.-M., Apport et limites d'une analyse statistique régionale pour l'estimation de surcotes extrêmes en France. *Proceeding of the SHF Conference*, Paris, France, 2012.
[10] Bernardara, P., Andreewsky, M. & Benoit, M., Application of regional frequency analysis to the estimation of extreme storm surges. *Journal of Geophysical Research*, **116**, C02008, 2011.
https://doi.org/10.1029/2010JC006229
[11] Renard, B., A Bayesian hierarchical approach to regional frequency analysis. *Water Resources Research*, **47**(11), W11513, 2011.
https://doi.org/10.1029/2010WR010089
[12] Weiss, J., Analyse régionale des aléas maritimes extrêmes. Thèse de doctorat. Sciences de l'ingénieur, Université Paris-Est, France, p. 256, 2014.
[13] Davis, R.A. & Mikosch, T., The extremogram: a correlogram for extreme events. *Bernoulli*, **15**(4), pp. 977–1009, 2009.
https://doi.org/10.3150/09-BEJ213
[14] Chavez, D.V. & Davison, A.C., Modelling time series extremes. *Statistical Journal,* **10**(1), pp. 109–133, 2012.
[15] Ledford, A.W. & Tawn, J.A., Diagnostics for dependence within time series extremes. *Journal of the Royal Statistical Society Series B,* **65**, pp. 521–543, 2003.
https://doi.org/10.1111/1467-9868.00400
[16] Cho, Y.B., Davis, R.A., & Ghosh, S., Asymptotic Properties of the Empirical Spatial Extremogram. *Scand J Statist*, **43**, pp. 757–773, 2016.
https://doi.org/10.1111/sjos.12202