Weather and Climate Extremes 4 (2014) 1–10

Contents lists available at ScienceDirect

Weather and Climate Extremes

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

Projections of precipitation extremes based on a regional, non-stationary peaks-over-threshold approach: A case study for the Netherlands and north-western Germany

M. Roth [a,b,*], T.A. Buishand [a], G. Jongbloed [b,c], A.M.G. Klein Tank [a], J.H. van Zanten [b,d]

[a] Royal Netherlands Meteorological Institute (KNMI), Utrechtseweg 297, NL-3731 GA, De Bilt, The Netherlands
[b] EURANDOM, Eindhoven University of Technology, Den Dolech 2, 5612 AZ, Eindhoven, The Netherlands
[c] Delft Institute of Applied Mathematics, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands
[d] Korteweg-de Vries Institute for Mathematics, University of Amsterdam, 1098 XH Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 16 August 2013
Received in revised form
12 December 2013
Accepted 7 January 2014
Available online 23 January 2014

Keywords:
Regional frequency analysis
Peaks-over-threshold
Non-stationary
Threshold choice
Climate model data

ABSTRACT

Projections of extreme precipitation are of great importance, considering the potential severe impacts on society. In this study a recently developed regional, non-stationary peaks-over-threshold approach is applied to two transient simulations of the RACMO2 regional climate model for the period 1950–2100, driven by two different general circulation models. The regional approach reduces the estimation uncertainty compared to at-site approaches. The selection of a threshold for the peaks-over-threshold model is tackled from a new perspective, taking advantage of the regional setting. Further, a regional quantile regression model using a common relative trend in the threshold is introduced. A considerable bias in the extreme return levels is found with respect to gridded observations. This bias is corrected for by adjusting the parameters in the peaks-over-threshold model.

In summer a significant increase in extreme precipitation over the study area is found for both RACMO2 simulations, mainly as a result of an increase of the variability of the excesses over the threshold. However, the magnitude of this trend in extreme summer precipitation depends on the driving general circulation model. In winter an increase in extreme precipitation corresponding with an increase in mean precipitation is found for both simulations. This trend is due to an increase of the threshold.

1. Introduction

Information on extreme precipitation is crucial for various societal sectors, e.g. for the design of sewage and drainage systems, roads and tunnels, farming, and the insurance industry. Consensus is growing that the characteristics of extreme precipitation may alter owing to climate change. In order to project the change in extreme precipitation, climate model data have been analyzed and compared to observations. Often this evaluation is done in terms of index values, such as the empirical annual 90% quantile of the precipitation amounts for each year or the 1-day or 5-day maximum precipitation amount in a year, see e.g. Klein Tank and Können (2003), and Turco and Llasat (2011). This approach shows the evolution of precipitation extremes over time. However, the indices have mostly a return period of not more than 1 year, which is of minor importance for the planning of hydraulic infrastructure, that usually has to withstand events with much longer return periods. To estimate the changes in these rare events, extreme-value distributions have been fitted to the extremes for two subsets of the data representing current (e.g. 1980–2010) and future (e.g. 2070–2100) climate, assuming stationarity within the time slices, see e.g. Fowler et al. (2005), Ekström et al. (2005), and Kyselý and Beranová (2009). Considering only two time slices does not give a picture of the evolution of the extremes, which is e.g. necessary if one is interested in the risk of failure of a hydraulic structure during its expected lifetime. Moreover, the selection of the time slices introduces additional uncertainty. A small shift of the time slices may have large influence on the estimated change. As an alternative, extreme value distributions with time-dependent parameters, which allow the consideration of the full time period, have been used, see e.g. Coles (2001), El Adlouni et al. (2007), Sugahara et al. (2009), Kyselý et al. (2010), Beguería et al. (2011), and Tramblay et al. (2013).

The estimation of changes in rare extremes is subject to large uncertainty. A general way to reduce the estimation uncertainty is regional frequency analysis (RFA), where the similarities between

* Corresponding author at: Royal Netherlands Meteorological Institute (KNMI), Utrechtseweg 297, NL-3731 GA, De Bilt, The Netherlands. Tel.: +31 302206382.
E-mail addresses: m.roth@tue.nl (M. Roth), buishand@knmi.nl (T.A. Buishand), G.Jongbloed@tudelft.nl (G. Jongbloed), kleintan@knmi.nl (A.M.G. Klein Tank), hvzanten@uva.nl (J.H. van Zanten).

http://dx.doi.org/10.1016/j.wace.2014.01.001
2212-0947

different sites in a region are exploited (Hosking and Wallis, 1997). RFA is mostly applied to (annual) block maxima (BM). An alternative to BM is to consider all peaks over a (high) threshold (POT), which is often preferable, owing to the more efficient use of the data.

A regional peaks-over-threshold model, combining the RFA approach and POT data, which can be used to analyze precipitation extremes in a changing climate, was developed by Roth et al. (2012). In this model a temporally varying threshold, which is determined by quantile regression, is used to account for changes in the frequency of precipitation extremes. The marginal distributions of the excesses are described by generalized Pareto distributions (GPD), with parameters, that may vary over time and their spatial variation is modeled by the index flood (IF) approach. For a detailed introduction to the index flood method, see Hosking and Wallis (1997).

The selection of the threshold is a crucial step in the application of the POT approach. However, there is still no standard procedure for this, and usually one relies on visual tools. Among these the plotting of the change of the estimated GPD shape parameter or the mean excess of the exceedances against the height of the threshold is popular. Unfortunately these plots rarely give clear indications of which quantile should be used for the threshold. In the present study, the individual plots are averaged over the region in order to make the desired constant or linear structure more apparent.

Daily precipitation from two simulations of a regional climate model (RCM) driven by different general circulation models (GCM) and from gridded observational data is analyzed for the Netherlands and north-western Germany. Instead of linking the POT model parameters to time, a temperature-based covariate is used to include the evolution of climate, see also Hanel et al. (2009), and Van Oldenborgh et al. (2009). Bias correction of the return levels from the regional climate model simulations is discussed. In addition to the changes in return levels, we present a risk-based design level that was recently introduced by Rootzén and Katz (2013).

Section 2 outlines the methods and Section 3 introduces the data used. Results and discussion are given in Section 4, followed by the conclusion in Section 5.

2. Methods

2.1. Introduction to the peaks-over-threshold model

To study the extremes of independent and identically distributed random variables X_i, one can consider the excesses $Y_i = X_i - u$ over a (high) threshold u. The *Balkema, De Haan, and Pickands* theorem states that the distribution of the excesses Y, conditioned on $Y \geq 0$, can be approximated by a generalized Pareto distribution (GPD), if the threshold u is sufficiently high and certain regularity conditions hold, see e.g. Reiss and Thomas (2007):

$$P(Y \leq y | Y \geq 0) = G_{\xi,\sigma}(y) = \begin{cases} 1 - \left(1 + \dfrac{\xi y}{\sigma}\right)^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp\left(-\dfrac{y}{\sigma}\right), & \xi = 0, \end{cases}$$

for $y \geq 0$ if $\xi \geq 0$ and $0 \leq y \leq -\sigma/\xi$ if $\xi < 0$, where σ and ξ are the scale and the shape parameter respectively. For $\xi = 0$ the GPD reduces to the exponential distribution. The independence requirement can be weakened (e.g. Leadbetter et al., 1983). In the case of short-range dependence the GPD approximation applies if one considers declustered excesses, i.e. the excesses of the local maxima (peaks) in a cluster of exceedances only. Several studies have considered the GPD also for non-stationary data, using temporally varying parameters, see for recent examples Sugahara et al. (2009), Kyselý et al. (2010), and Beguería et al. (2011).

2.2. Temporal dependence and declustering

Daily precipitation exhibits temporal dependence, also at high levels. This dependence is generally stronger in winter than in summer (due to the convective nature of most extremes in summer). As mentioned in Section 2.1 one can account for this temporal dependence by considering declustered excesses. This is usually achieved by specifying a minimum separation time t_{sep} between exceedances over the threshold, where t_{sep} is determined by the temporal dependence in the data at high levels. Here, we follow this approach but decluster all data and not the excesses only.

Let $x_s(t)$ be the rainfall at site $s \in \{1, \dots, S\}$ and day $t \in \{1, \dots, T\}$. To determine t_{sep} we compute first the 95% quantile for each site s and calculate the number of clusters of length $n \geq 2$. A cluster of length n is defined as n consecutive exceedances of the 95% quantile. The number of clusters decreases usually very fast with the length n. The separation time t_{sep} is set to n, if the number of clusters of length $n+2$ is sufficiently low. After this initial step we obtain the declustered data by replacing $x_s(t)$ with zero, if it is not a maximum in the subset $x_s(t - t_{\text{sep}}), \dots, x_s(t), \dots, x_s(t + t_{\text{sep}})$. From this it is clear that also the excesses obtained from the declustered data are separated by at least t_{sep} days.

2.3. Index flood approach

Roth et al. (2012) introduced a regional approach for multi-site, non-stationary POT rainfall data:

$$y_s(t) = x_s(t) - u_s(t),$$

where $u_s(t)$ is a suitable threshold value for site s and day t. The approach is based on the index flood (IF) assumption, i.e. that the non-stationary POT data have, after scaling by a time and site dependent index variable (or index rainfall) $\eta_s(t)$, a common excess distribution. If the site-specific excess distributions are GPD with shape parameter $\xi_s(t)$ and scale parameter $\sigma_s(t)$, then we have for the scaled excesses:

$$P\left(\frac{Y_s(t)}{\eta_s(t)} \leq y \,\middle|\, Y_s(t) > 0\right) = G_{\xi(t),\gamma(t)}(y), \tag{1}$$

with $Y_s(t)$ the excess at site s and day t and $\gamma(t) = \sigma_s(t)/\eta_s(t)$ a dimensionless dispersion coefficient. The IF assumption thus implies that this coefficient and the shape parameter are constant over the region of interest. Roth et al. (2012) used the threshold $u_s(t)$ as index variable:

$$\eta_s(t) = u_s(t).$$

The mean number λ of the excesses over $u_s(t)$ in this approach is constant over time and space, which was achieved by using quantile regression to determine $u_s(t)$. With Eq. (1) we can compute for each site s and day t the value $r_{s,t}(\alpha)$ that is exceeded on average α times in a season:

$$r_{s,t}(\alpha) = \begin{cases} u_s(t)\left(1 - \dfrac{\gamma(t)}{\xi(t)}\left[1 - \left(\dfrac{\lambda}{\alpha}\right)^{\xi(t)}\right]\right), & \xi(t) \neq 0, \\ u_s(t)(1 + \gamma(t)\ln(\lambda/\alpha)), & \xi(t) = 0. \end{cases} \tag{2}$$

In analogy with a stationary setting, the quantity $r_{s,t}(\alpha)$ is termed the $1/\alpha$-year return level, although $1/\alpha$ no longer gives the expected waiting time between exceedances of $r_{s,t}(\alpha)$.

2.4. Determination of the threshold

The non-stationary threshold is estimated via quantile regression. However, we have to select an appropriate quantile, i.e. the value of the threshold has to be high enough to justify the GPD assumption.