Frequency Analysis of Extreme Sub-Daily Precipitation under Stationary and Non-Stationary Conditions across Two Contrasting Hydroclimatic Environments

Eleonora M.C. Demaria[1], David Goodrich[1], Timothy Keefer[1]

[1]Southwest Watershed Research Center, USDA-ARS, 2000 E Allen Rd, Tucson, 85719 AZ, USA

Correspondence to: Eleonora M.C. Demaria (eleonora.demaria@ars.usda.gov)

Abstract. Observed sub-daily precipitation intensities from contrasting hydroclimatic environments in the USA are used to evaluate temporal trends and to develop Intensity-Duration-Frequency (IDF) curves under stationary and non-stationary climatic conditions. Analyses are based on observations from two United States Department of Agriculture (USDA)-Agricultural Research Service (ARS) experimental watersheds located in a semi-arid and a temperate environment. We use an Annual Maximum Series (AMS) and a Partial Duration Series (PDS) approach to identify temporal trends in maximum intensities for durations ranging from 5- to 1440-minutes. A Bayesian approach with Monte Carlo techniques is used to incorporate the effect of non-stationary climatic assumptions in the IDF curves. The results show increasing trends in observed AMS sub-daily intensities in both watersheds whereas trends in the PDS observations are mostly positive in the semi-arid site and a mix of positive and negative in the temperate site. Stationary climate assumptions lead to much lower estimated sub-daily intensities than those under non-stationary assumptions with larger absolute differences found for shorter durations and smaller return periods. The risk of failure (R) of a hydraulic structure is increased for non-stationary effects over those of stationary effects, with absolute differences of 25% for a 100-year return period (T) and a project life (*n*) of 100 years. The study highlights the importance of considering non-stationarity, due to natural variability or to climate change, in storm design.

1 Introduction

Extreme rainfall events are linked to flooding, property damage, and potential loss of life. As the atmosphere gets warmer (Hartmann et al., 2013), changes in its water vapor holding capacity and shifts in circulation dynamics are expected to lead to changes in the characteristics of precipitation. Positive trends in daily precipitation extremes have been reported globally and within the U.S. (Groisman et al., 2005; Kunkel et al., 2012; Mishra et al., 2012) with indications that sub-daily precipitation extremes are increasing beyond what is expected from the Clausius-Clapeyron relation (Lenderink and van Meijgaard, 2008; Westra et al., 2014). Climate models project substantial increases in water vapor concentration during the 21st century and it has been hypothesized that this will change the magnitude and frequency of intense precipitation events (Emori, 2005; Janssen et al., 2014; Kunkel et al., 2013; Min et al., 2011). A contrast between positive trends in daily

precipitation extremes in the eastern U.S. and negative trends in the western U.S. has been documented (Kunkel et al., 1999; Pryor et al., 2009; Yu et al., 2016). There are indications that the southwest U.S. has become drier during the last several decades (Prein et al., 2016; Seager, 2007) and that in the arid and semi-arid regions located in subtropical zones an expansion of the sub-tropical highs (Seidel et al., 2008) and the poleward shift of the mid-latitude jet (Archer and Caldeira, 2008) are likely to increase heat and aridity in the future. In contrast, in the Northeast and Midwest regions observed increases in precipitation are projected to continue during the 21st century (Horton et al., 2014; Pryor et al., 2014).

Infrastructure design is commonly based on Intensity-Duration-Frequency (IDF) curves that are developed using rain gauge observations for different durations ranging from 5 minutes to 60 days. However, observations are not often available at sub-daily time steps and design storms are estimated indirectly from daily totals using linear scaling factors as implemented in the National Oceanic and Atmospheric Administration (NOAA) Atlas 14 (Bonnin, G.M. et al., 2011). Although the intensification of precipitation events has been widely evaluated by the scientific community, the engineering community relies on the concept of return period (T) (also referred as average recurrence intervals or annual return period) to design hydraulic and storm water structures. An important assumption in IDF curves development is the stationarity of the model, which implies that the probability of occurrence of events is not expected to significantly change in time. However, low-frequency components of climate variability such as El Nino Southern Oscillation (ENSO) and decadal and multi-decadal oscillations such as the Pacific Decadal Oscillation (PDO) and Atlantic Multi-decadal Oscillation (AMO) can introduce increasing or decreasing trends in the hydrologic records (Salas and Obeysekera, 2014). In addition, positive increases in observed precipitation linked to a warmer atmosphere also bring into question the validity of the IDF curves developed under stationary assumptions and has prompted the development of new methods for determining return periods and risk that take into account the effect of non-stationary in climate extremes (Bhatkoti et al., 2016; Condon et al., 2015; Du et al., 2015; Khaliq et al., 2006; Obeysekera and Salas, 2014; Salas and Obeysekera, 2014; Serinaldi and Kilsby, 2015). A few climate studies have shown that design storm estimates from stationary models are significantly lower than estimates from non-stationary models (Cheng and AghaKouchak, 2014; Sarhadi and Soulis, 2017; Wi et al., 2016). However, the lack of observed sub-daily precipitation limits the studies to using scaling factors to temporally disaggregate daily observations or use merged radar-rain gauge data (Cheng and AghaKouchak, 2014; Yu et al., 2016). The United States Department of Agriculture (USDA)-Agricultural Research Service (ARS) network of experimental watersheds has been systematically measuring sub-daily precipitation since the mid-twentieth century and its database represents a unique opportunity to evaluate temporal changes and develop IDF curves in different hydroclimatic environments. The goals of this study are: (1) to use observed sub-daily summer (June-October) precipitation intensities from two ARS sites, located in a semi-arid and a temperate climate, to test for evidence of temporal trends; (2) to build IDF curves using Annual Maximum Series (AMS) and Partial Duration Series (PDS) approaches and a Bayesian method that takes into account the non-stationarity of the time series; and (3) to estimate the risk of structural failure associated with design storms under stationary and non-stationary assumptions.

2 Data and methods

2.1 Study area and data used

Sub-daily precipitation intensities over the 1961-2014 period are selected for two ARS watersheds located in distinct hydroclimatic environments in the U.S.: the Walnut Gulch Experimental Watershed (WGEW) located in a semi-arid environment in Arizona and the North Appalachian Experimental Watershed (NAEW) located in a temperate environment in

5   Ohio (Fig. 1 (a)). Table 1 shows the main characteristics of the watersheds. The semi-arid site is equipped with 56 rain gauges in order to capture the highly localized and scattered nature of summer convective storms in the southwestern United States which can have, on average, a spatial extent of 70 km^2 (Syed et al., 2003); the temperate site has available observations from 3 rain gauges. Both sites are equipped with Belfort 8-inch (0.2032 m) unshielded weighing-bucket rain

10 gauges. Originally, a mechanical system recorded accumulated precipitation on paper charts and rainfall totals were later computed by manually digitizing the charts. In the mid-1990s, the analog-recording system was replaced by an electronic-weighing digital system that recorded rainfall with a 1/100-inch (0.254 mm) detection accuracy (Keefer et al., 2008; Owens et al., 2010). In order to build confidence in future analysis of the WGEW precipitation record, a direct comparison of event precipitation for nine pairs of co-located analog and digital rain gauges was undertaken for a 5 year period, 2000-2004

15 (Keefer et al., 2008). For an average of 260 events per pair, event total precipitation amount and peak intensities from 1 minute to 30 minutes were found to be equivalent. This analysis indicates that no artificial discontinuities were introduced by the switch in the recording system in a water-limited environment where highly localized, short duration, and high intensity summer storms are difficult to accurately observe. No comparable analysis has been carried out in the temperate watershed.

20 TABLE 1 HERE

The total seasonal precipitation (June-October) for the watersheds is presented in Fig. 1 (b) and (c). The semi-arid site shows large decadal variability with relatively drier seventies, wetter conditions in the eighties, and drier conditions in the beginning of the 21st century. There is a tendency towards drier conditions during the 54 years of record, however the trend

25 is not statistically significant at a 1% significance level. Conversely, seasonal precipitation has steadily increased in the temperate site despite the strong natural variability that dominated the seventies (wet period) and the nineties (dry period). A statistically positive trend (α 0.1) of 2.47 mm/yr was found for the period (Fig. 1 (c)).

FIGURE 1 HERE

30 ### 2.2 Time series and trend analysis

Two approaches were used to build IDF curves: 1) Annual Maximum Series (AMS) approach, also known as Block Maxima Approach, and 2) Partial Duration Series (PDS), also referred to as the Peak-over-Threshold approach. The AMS consists of

selecting the largest precipitation intensity for each year, thus the time series will have as many values as years in the record. This approach is widely used for extreme value analysis and for hydrologic design. The probability distribution of precipitation events selected with this method is well described by the Generalized Extreme Value (GEV) distribution function (Coles, 2001) given by:

$$F(x) = exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{\frac{-1}{\xi}}\right\} \tag{1}$$

where u is the location parameter, σ is the scale parameter and ξ is the shape parameter under stationary conditions. Estimates of extreme quantiles (z_T) used to estimate a design value or a T-year return level of the GEV distribution are given by:

$$z_T = \mu - \frac{\sigma}{\xi}\left[1 - \left\{-log\left(1 - \frac{1}{T}\right)\right\}^{\xi}\right] \tag{2}$$

10

The PDS is based on sampling all the observations that exceed an *a priori* specified threshold μ. Unlike the AMS approach, the PDS approach allows one to select more than one value per year and it provides additional flexibility to describe the natural variability of precipitation. Similarly to the AMS being described by GEV, the distribution of excess over the threshold y = x - μ has an approximate Generalized Pareto distribution (GPD) with the following cumulative distribution function (Katz et al., 2005):

$$F(y) = 1 - \left[1 + \xi\left(\frac{x-\mu}{\sigma^*}\right)\right] \tag{3}$$

where ξ is the shape parameter, and σ* is the scale parameter which as for the GEV controls the spread of the distribution. The m-observational return level is defined as:

$$x_m = \mu + \frac{\sigma}{\xi}\left[(m\zeta_\mu)^\xi - 1\right] \tag{4}$$

where $\zeta\mu = Pr(x>\mu)$ and m is the number of observations above the threshold.

For AMS the return period is defined as the inverse of the exceedance probability (P) and it is easy to interpret since data are taken at fixed intervals. Thus the 1000-year flood has 0.001 chance, on average, to be exceeded in any given year. Conversely, for PDS, since multiple observations can be found in a year, the average number of observations (r) must be considered in the definition of T (Soong et al., 2004):

$$T = \frac{1}{rP} \ (in \ years) \tag{5}$$

The PDS approach requires the selection of a high threshold to include extreme events in the analysis. Unfortunately, there is no reliable automated method to select the threshold in such a way that the independence criterion is satisfied and, at the same time, the sample size is equal to or larger than the AMS sample. Finding the right threshold is always a tradeoff between finding a value high enough to fit the GPD function to the data but not so high that too few observations are

selected which makes the statistical estimation of the distribution function parameters challenging (Coles, 2001). Conversely, a low threshold might lead to the selection of observations that are not statistically independent. The declustering methods available in the literature to meet the independence requirement make use of a fixed number of observations exceeding the threshold (Yilmaz et al., 2014), graphical approaches (Coles, 2001), or temporal windows

5   (Begueria, 2005). The reader is referred to Lang et al. (1999) for a review of available methods. In this study, we select the threshold based on three criteria: first the number of events exceeding the threshold to be larger than the number of years in the record; second the events are statistically independent; and third the goodness of the fitting of the GPD to the observed data. The GEV and GPD parameters were determined using the Maximum Likelihood (ML) method since likelihood-based techniques have shown better adaptability to the extreme value variation observed in natural systems (Coles, 2001; Hosking

10   and Wallis, 2005).

The presence of monotonic linear trends in precipitation extremes was evaluated with the Mann-Kendall (M-K) non-parametric test for each duration and for each rain gauge (Kendall, 1948; Mann, 1945; Sen, 1968). The magnitude of the trend was estimated with the Sen's method which has shown to be more robust since it is insensitive to outliers (Sen, 1968).

15   Trends were deemed statistically significant at a 0.1 significance level. Whenever the significance level is higher than the probability value (p) of the test statistic, the null hypothesis of no monotonic trend (trend absence) is rejected. The M-K test assumes independence (i.e. observations are not serially correlated over time). The presence of spatial (cross) correlation in precipitation data might affect the ability of statistical methods to assess the significance of a trend. When two rain gauges are highly correlated, the second one merely duplicates the information provided by the first one and might also lead to the

20   rejection of the null hypothesis (no trend) when in reality there was no trend. The Regional Average Mann-Kendall (RAMK) test incorporates the spatial (cross) correlation of the rain gauge network in the variance estimation (Yue and Wang, 2002). This study used the RAMK test to evaluate trends in AMS time series in both sites.

2.3 Non-stationary GEV and GPD estimation

25   When time series of AMS or PDS are known to be changing in time, time variant parameters should be fitted to the observations. The commonly used approach to incorporate non-stationarity in time series into the frequency analysis is to vary one parameter of the theoretical probability distribution while the other two are kept constant (Liuzzo and Freni, 2015; Wi et al., 2016). In this study we used a Bayesian approach to estimate distribution parameters under stationary and non-stationary assumptions (Cheng and AghaKouchak, 2014). A Monte Carlo sampling approach is implemented to infer the

30   posterior parameter distribution which allows uncertainty bounds to be obtained for estimated precipitation intensity values for different return periods. Parameters were estimated with the ML method and the goodness of the model fit was assessed with the $\hat{R}$ criterion (Gelman and Shirley, 2011). An initial set of 50,000 random samples were used to estimate the GEV and GPD posterior parameter distributions for each duration. From these, 10,000 samples were not included in the analysis

(burn-in) which reduced the final number of parameters to 40,000. The observed non-stationarity in precipitation extremes was modeled with a time-variant linear model of the location parameter $\mu(t) = \mu_1(t) + \mu_o$ in the GEV. For the GPD model, the threshold parameter was assumed to be stationary and only a linear time-variant model of the scale parameter (ξ) was used. The observed trends were used as a proxy for future trend magnitudes and utilized to compute the return levels as the

5 median of the 40,000 parameter samples multiplied by the linear trend over a 100-year period into the future. This approach implicitly assumes that the magnitude of the observed trend will remain constant in the future.

To measure the goodness of the fitting of observations to the theoretical distribution functions we used quantile-quantile (QQ) plots (Katz et al., 2002). To this end, the Weibull plotting positions were used to compute the empirical quantiles and

10 the theoretical quantiles were obtained from the distributions fitted to observed precipitation intensities. This method was only implemented for the stationary models since the parameters of the non-stationary models change with time. To prove that the non-stationary extreme value distributions appropriately described the observed precipitation intensities, both stationary and non-stationary models were tested with the deviance statistic (D). Considering the stationary model M_0 as a restrictive case of non-stationary model M_1 where $\mu_1 = 0$ for the GEV model and $\xi_1 = 0$ for the GPD model, it is possible to

15 test the null hypothesis of M_0 fitting to the data better than M_1 with D which is defined as:

$$D = 2\{l_1(M_1) - l_0(M_0)\} \hspace{3cm} (7)$$

where $l_l M_1$ and $l_0 M_0$ are the maximized log-likelihoods of models M_1 and M_0, respectively. Large D values indicate that model M_1 explains more of the variation in the data hence the increase in the model size improves its skill to explain

20 observed precipitation intensities. The validity of model M_0 relative to M_1 is tested at the α-level if D > c_α, where c_α is the (1-α) quantile of a χ_k^2 distribution and k is the difference in the dimensionality between the stationary and non-stationary models (Coles, 2001; Delgado et al., 2014).

The traditional risk analysis used for hydrologic design uses the concept of return period (i.e. the inverse of the exceedance

25 probability for a given magnitude) which is assumed to be constant in time during the period of interest. However, when observations show a temporal trend the exceedance probability varies with time and it can be used to extend the concept of risk analysis used in stationary risk design to non-stationary conditions. Non-stationary return periods based on time varying exceedance probability are computed using waiting time theory which incorporates the upward trend in observations. The hydrologic risk associated with stationary and time-varying precipitation intensities can be computed as:

30

6

Hydrology and
Earth System
Sciences

Open Access

Discussions

$$R = 1 - \left(1 - \frac{1}{T}\right)^n \quad or \quad R = 1 - \prod_{t=1}^{n}\left(1 - \frac{1}{T_t}\right) \tag{8}$$

where n is the length of time of the project life, T is the stationary return period, and T_t is the corresponding non-stationary return period.

3 Results

3.1 Trend analysis of precipitation extremes

For the analysis of temporal trends, 56 rain gauges were used in the semi-arid watershed and 3 rain gauges in the temperate watershed. Maximum rainfall intensities for each day were obtained for short: 5-, 15-, 30-minute, intermediate: 60-minute, and long: 1440-(1 day) minute durations for the warm months (June-October). We generated a time series of AMS at each gauge by selecting the maximum intensity for each year and each duration. The time series of PDS intensities were obtained by selecting the 95th percentile for all maximum precipitation intensities, this percentile value (assumed constant in a non-stationary climate) allowed an average sample size of 1.6 events per year. Previous studies have used percentiles ranging from the 90th to the 98th to define what constitutes an extreme event (Begueria et al., 2011; Groisman et al., 2005; Villarini et al., 2013). A 24-hour window was imposed between two events to ensure the statistical independence of PDS intensities. Outliers in the observations were removed using the Adjusted Boxplot method (Hubert and Vandervieren, 2008). The number of outliers was, on average 2(± 3.1) and 10(± 2.0) for all the rain gauges for the semi-arid and temperate sites, respectively. The effect of removing outliers on the trend analysis was negligible (not shown). Table 2 shows the number of rain gauges with positive(negative) trends, both statistically and not statistically significant, in precipitation intensities for the semi-arid(temperate) site. Both sites show a mix of positive and negative trends for AMS and PDS intensities for all durations. Overall, the semi-arid watershed shows positive trends in 72 %(68 %) of the gauges for AMS(PDS). However, the number of rain gauges with statistically significant trends is lower at 11 %(15 %), respectively. The regional trend analysis with RAMK shows that the 5- and 1440-min durations are statistically significant. Note that RAMK only allows computation of the trend for the AMS since it requires time series of the same length. In the temperate watershed, positive trends in AMS maximum intensities are statistically significant for 15-, 30-, and 1440-minute durations whereas PDS maximum intensities are mostly negative for short and intermediate durations. The RAMK test indicates that trends are statistically positive for 15-, 30-, and 1440-minute durations. The lack of a clear signal in the direction of the trends has been previously reported for other regions (Sarhadi and Soulis, 2017) and might be linked to the weak power of statistical methods when the data are affected by high natural variability as is the case of precipitation (Renard et al., 2008).

From the available rain gauges in each watershed, we selected one gauge to perform the IDF analysis for all the durations. Dimensionless trends were also computed as the magnitude of the trend in mm/hr, estimated with the Sen's method, divided

by the maximum intensity. These trends enable comparison of their magnitudes between the different durations. Figure 2 shows linear trends in AMS and PDS precipitation intensities. Positive trends ranged from 5.01 mm/hr per decade for the 5-min durations to 0.11 mm/hr for the 1440-min durations in the semi-arid site. Even though the magnitude of the trends declined with increasing durations, the smaller dimensionless trends were found for the 30- and 60-min durations and the largest values for the 1440-min duration. In the temperate site, trends ranged from 3.21 to 0.19 mm/hr per decade for between the 5-min and 1440-min durations, however the smallest(largest) dimensionless trend corresponded to the 15-min(1440-min) duration. For the PDS observations, the trend analysis in the semi-arid site revealed increasing statistically significant trends for durations 15-, 30- and 1440-min. A noticeable difference between the AMS and PDS trend magnitudes at the temperate site is the identification of negative trends in PDS extremes. The decreasing trends in precipitation extremes are likely related to the natural variability of precipitation with wetter-than-normal conditions in the beginning of the record.

FIGURE 2 HERE

3.2 Extreme analysis of maximum precipitation intensities

Figure 3 shows the fitting of GEV and GPD models to precipitation intensities in the semi-arid ((a) and (b)) and temperate ((c) and (d)) watersheds, respectively. The fitting of the theoretical distribution functions under stationary conditions was found to be statistically significant in all rain gauges (5 % significance level) with the Anderson-Darling (A-D) and Kolmogorov-Smirnov (K-S) non-parametric tests. The GEV distribution, fitted to observations with the ML method, is able to appropriately represent the distribution of annual maximum intensities for the different durations. The ML method was selected for the analysis since it easily adapts to different models including non-stationary models (Coles, 2001). The PDS extremes in both watersheds were also tested for the GPD with the ML estimator (Fig. 3 (b) and (d)). The results of the A-D and K-S tests suggested that this distribution can properly model the observed distribution of extremes over the 95th percentile in the different hydroclimatic environments.

FIGURE 3 HERE

A visual evaluation of the fit of the GEV or GPD distributions to the time series of precipitation intensities is provided by the QQ plots (Fig. 4). For the GEV distribution based on Eq. (1) the QQ plot indicates that the fit is reasonably adequate although there is strong evidence of a heavy tail for large intensities in both sites that the theoretical distribution might not be able to represent (Fig. 4 (a) through (e)). When the PDS approach is used, the fit of the GPD model to observations based on Eq. (3) appears to be a better model to represent time series of maximum intensities in the semi-arid site (Fig. 4 (f) through (j)). Note that the temperate site is not included in the analysis since no statistically significant trends were identified for that site. For all 40,000 Monte Carlo realizations, we obtained the maximized log-likelihoods for the stationary and non-

stationary models to compute the deviance statistic D (Eq. (10)). The significance level (c_α), obtained from the χ^2_k distribution for α= 0.05, is equal to 3.84. Large D values indicate that the non-stationary model explains more of the variation in precipitation intensities than the stationary model (Table 3). In the semi-arid site, small values of D for all durations indicate that the non-stationary GEV model, with a linear trend in the location parameter, might not be a

5   significant improvement over the stationary version of the model despite the statistically significant positive trends found in Section 3.1. The D values for the temperate site exceed c_α for the 5-, 30-, and 1440-min supporting the incorporation of a linear trend in μ in the GEV models. There is strong evidence, D values exceeding the significance level, to support a non-stationary GPD model with a time-variant scale parameter for the semi-arid site.

10   FIGURE 4 and Table 3 HERE

3.3 Design storms and risk under stationary and non-stationary conditions

In this section, we address the issue of how design storms differ under stationary and non-stationary assumptions? An
15   answer to this question is relevant since it might inform practitioners about the uncertainty in the estimates when designing storm water infrastructure. In Figure 2, by examining storm durations, we found that a subset of sub-daily intensities show statistically significant increasing linear trends. With this in mind, a new set of IDF curves were developed with a framework that incorporates temporal changes in precipitation intensities. For both extreme value models, the magnitudes of the trends observed in the study period are assumed to continue at the same rate in the next 100 years into the future. Estimated
20   precipitation intensities (return levels) for different return periods, and the corresponding credibility interval (this term is called confidence interval in Frequentist context (Renard et al., 2008)) are shown in Fig. 5. In the semi-arid and temperate sites, the GEV estimated intensities under stationary climatic conditions are much lower than those under the non-stationary assumptions (Fig. 5 (a) to (j)), i.e., extreme precipitation intensities are likely to occur more frequently if time-variant conditions are incorporated in the analysis. Note that durations with no statistically significant trends were excluded from
25   the analysis, therefore both curves fall on top of each other in the plot. Conversely, the GPD model does not show appreciable difference between both climatic conditions in the semi-arid watershed. This could be the result of less steep trends (compared to the AMS data) despite the larger D values found (Table 3). These results indicate that a specific precipitation intensity will be more likely to occur, i.e., shorter return period, in the future if the observed trends are incorporated in the IDF design.
30

FIGURE 5 HERE

The IDF curves for each site are shown in Fig. 6. The largest absolute differences in GEV-intensities are found for short durations (changes are computed as the absolute difference between non-stationary and stationary) at both sites. For example, for the semi-arid site the estimated 2-year 5-min storm (5-minute duration with a 2-year return period) is 99.3 mm/hr under stationary assumptions and 122.3 mm/hr under non-stationary conditions, which represents an absolute

5   difference of 22.9 mm/hr, while for the estimated 100-year 5-min storm estimates are 189(204) mm/hr for stationary(non-stationary) which represents an absolute change of 15.5 mm/hr. For longer event durations, the absolute differences between stationary and nonstationary become smaller in absolute terms, ranging from 0.75 mm/hr to 0.59 mm/hr for the 2-year and 100-year return periods, respectively. In the semi-arid site, differences in GPD-estimated intensities are almost negligible. These results indicate that larger changes in design storms are likely to happen for shorter durations and smaller return

10   periods under warmer-than-normal climate conditions in agreement with previously found evidence that hourly and sub-hourly precipitation intensity is more sensitive to atmospheric temperature changes see Westra et al. (2014) and the references therein. This demonstrates that the stationary framework currently used for structural design systematically underestimate short-duration precipitation extremes which might lead to more frequent infrastructure damage.

15   FIGURE 6 HERE

We have shown in the previous sections that there is evidence of increasing precipitation intensities for certain durations during the period 1961-2014 (54 years). This implies that the exceedance probability of a certain event occurring in a given year will also vary with time and hence the risk of failure (R) for a specific T and project life (n) will change accordingly.

20   The difference in R for the semi-arid and temperate sites is shown in Fig. 7 for the 15- and 1440-min duration AMS precipitation intensities, for T= 10, 50, and 100-yr under stationary conditions and the corresponding T_t under non-stationary assumptions. The insets in each panel show that Tt is systematically smaller than T at both sites and for the two selected durations. For instance, for the 15-min duration intensities the Tt value for a 50-year return period is 36(31) years for the semi-arid(temperate) whereas a Tt of 46 years corresponds to T= 100-year in both sites. Under stationary conditions, R

25   increases in accordance with the project life durations for the three return periods. However, the risk of failure is larger under the non-stationary conditions as the project life increases. For example, in the semi-arid site for a project life n=50 years and a return period T=10 years (T_t= 8.8 years) the absolute differences between stationary and non-stationary conditions are negligible. Conversely, for n=100 years and a return period T=100 years (T_t = 46 years) the risks are 63%(89%) for stationary(non-stationary) conditions in the semi-arid watershed. This represents a difference in risk levels of 25%. Note that

30   we are using absolute differences no relative differences in this analysis. The risk of failure is very similar for the temperate site for smaller n, for a larger project life the absolute difference decreases to 18% for T=100 years (Fig. 7 (b)). The shape of the curves for the different durations differ between sites and no clear pattern between durations and the magnitude of R is found. This behavior can be a function of the parameters of the GEV distribution fitted to the observations and the

magnitude of the observed trend. However, despite these differences, the shape of the curves and the magnitude of the R are comparable for both sites.

FIGURE 7 HERE

5

The magnitude range of R for all the considered project life values (n =1 to 100 years) and different return periods is shown Fig. 8. As expected, the probability of precipitation intensities exceeding the given design storm during the n-year of the project decreases for larger return periods since more intense precipitation intensities are less frequently observed. Changes in the median R (i.e., differences between non-stationary and stationary conditions) increase from 0.25 % for T=10 years to

10   27 %(18 %) for T = 100 years in the semi-arid(temperate) site. The uncertainty, reflected in the inter-quantile distance (25th to 75th percentiles), grows as the return period increases. The temperate watershed shows slightly larger differences between stationary and non-stationary durations, however the magnitudes are comparable for both sites. There is a close agreement between the magnitude of non-stationary R and the magnitude of the dimensionless trend shown in Fig. 2. These results indicate that despite the differences in precipitation intensities between stationary and non-stationary conditions are larger

15   for shorter storm durations (Fig. 5), more intense-short duration events under non-stationary climate conditions are not necessarily linked to a larger risk of failure perhaps because these differences are not large enough to be captured by Eq. (8).

FIGURE 8 HERE

20   When R is plotted over a range of precipitation intensities for the semi-arid site (Fig. 9), as expected R increases as a power function of the project life (Eq. (8)) and decreases with precipitation intensity (large intensities correspond to larger T or less frequent events). Note that precipitation intensity values are computed with the fitted GEV for set return period values. For the smaller n (10-year), the differences in the median of R between stationary and non-stationary conditions do not exceed 10 % for the different event durations. As n becomes larger, the differences increase up to 27 % for all durations. Overall, the

25   increased risk from the stationary condition to the non-stationary case is between 10 to 27 % for precipitation intensities ranging from 5 to 200 mm/hr as the project life grows.

FIGURE 9 HERE

30   **4 Summary and conclusions**

Engineering planning and water infrastructure design rely on IDF curves that have been developed under the assumption that climate conditions are stationary. This study shows that in two contrasting hydroclimatic environments the changing nature

of precipitation intensities over time leads to the underestimation of extreme storms that are used for hydrologic design. Our analysis is unique since it incorporates observed precipitation intensities at sub-daily durations and non-stationarity into the design IDF curves. Sub-daily precipitation intensities (period 1961-2014) at two USDA-ARS watersheds located in the United States: the semi-arid climate of Arizona and the temperate climate of Ohio are used to generate Annual Maximum

5   Series (AMS) and Partial Duration Series (PDS) for durations ranging from 5- to 1440-minutes. For each duration, first the presence of temporal linear trends is assessed with a non-parametric test and second, in the cases of statistically significant trends, IDF curves and the hydrologic risk of failure under stationary and non-stationary assumptions are computed using a Bayesian approach that incorporates time-variant parameters.

10   We find increasing trends in observed AMS sub-daily intensities in both sites. The PDS observations show positive trends in the semi-arid site, and a mix of positive and negative not statistically significant trends in the temperate site, perhaps due to long-term variability. Trends in AMS(PDS) observations are incorporated in a GEV(GPD) model with time-variant parameters expressed as a linear function of time to estimate precipitation intensities. In general, non-stationary models have improved skills in explaining the variability of precipitation intensities in both watersheds.

15

Using the derived GEV and GPD stationary and non-stationary models, we generate IDF curves for the different event durations and return periods. Despite the hydroclimatic differences between the semi-arid and the temperate site, the absolute differences in GEV-estimated precipitation intensities between stationary and non-stationary conditions are comparable for both sites. Estimated sub-daily intensities under stationary climatic conditions are much lower than those under the non-

20   stationary assumptions. For instance, for the semi-arid site the absolute difference in the 2-year 5-min storm is 22.9 mm/hr while for the 100-year 5-min storm the difference was 15.5 mm/hr. For longer event durations, the absolute difference become smaller. These changes in design storms under non-stationary conditions might lead to under-design of critical infrastructure not able to withstand future storms occuring more frequently than planned.

25   The risk of failure for non-stationarity climate conditions, due to natural variability or to climate change, exceeds that for stationary assumptions at both sites. The differences in risk under both climate conditions increase as a function of n (project life), independently of the event duration, and for large return periods. The absolute difference in risk levels for n=50 years and T= 10 years is almost negligible, conversely the difference is 25 % for n=100 years and T= 100 years. These results indicate that the increase in precipitation risk under non-stationary climate conditions might be considered for projects

30   requiring a greater level of protection.

The results are valid for the two sites included in the study; however, the methodology can be implemented in different hydroclimatic environments where sub-daily precipitation observations are available to update IDF curves used for infrastructure design. However, our approach has one caveat. We base the analysis on the assumption that historical trends in

precipitation intensities are expected to continue at the same rate in the future. Even though we agree that this is not necessarily expected and that the projected trend magnitude, and its uncertainty, can be estimated from climate model projections, we argue that the historic work provides a baseline for the magnitude of trends in precipitation intensities. Our results can be used as a basis to incorporate non-stationarity in the design storm process and could be further updated with

5 more observations or as climate model efforts acquire more skill, and would move the design community forward where design life could be coupled with intensity trend projections to obtain more realistic design storms.

Data and code availability

The precipitation data can be found at http://www.tucson.ars.ag.gov/dap/ and the Matlab codes used for the analysis are available upon request from the corresponding author.

10 **Competing interests**

The authors declare that they have no conflict of interest.

15

References

Archer, C. L. and Caldeira, K.: Historical trends in the jet streams, Geophys. Res. Lett., 35(8), L08803, doi:10.1029/2008GL033614, 2008.

Begueria, S.: Uncertainties in partial duration series modelling of extremes related to the choice of the threshold value, J.
20 Hydrol., 303(1–4), 215–230, doi:10.1016/j.jhydrol.2004.07.015, 2005.

Begueria, S., Angulo-Martinez, M., Vicente-Serrano, S. M., Ignacio Lopez-Moreno, J. and El-Kenawy, A.: 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(14), 2102–2114, doi:10.1002/joc.2218, 2011.

Bhatkoti, R., Moglen, G. E., Murray-Tuite, P. M. and Triantis, K. P.: Changes to Bridge Flood Risk under Climate Change,
25 J. Hydrol. Eng., 4016045, doi:10.1061/(ASCE)HE.1943-5584.0001448, 2016.

Bonnin, G.M., Martin, D., Lin, B., Tye Parzybok, Tye, P., Yekta, M. and Riley, D.: NOAA atlas 14 precipitation- frequency atlas of the United States volume 1 version 5. 0: Semiarid southwest (southeast California, Arizona, New Mexico, Utah, Nevada), NOAA., 2011.

Cheng, L. and AghaKouchak, A.: Nonstationary Precipitation Intensity-Duration-Frequency Curves for Infrastructure
30 Design in a Changing Climate, Sci. Rep., 4, 7093, doi:10.1038/srep07093, 2014.

Coles, S.: An Introduction to Statistical Modeling of Extreme Values, Springer London, London. [online] Available from: http://link.springer.com/10.1007/978-1-4471-3675-0 (Accessed 29 March 2016), 2001.

Condon, L. E., Gangopadhyay, S. and Pruitt, T.: Climate change and non-stationary flood risk for the upper Truckee River basin, Hydrol. Earth Syst. Sci., 19(1), 159–175, doi:10.5194/hess-19-159-2015, 2015.

5   Delgado, J. M., Merz, B. and Apel, H.: Projecting flood hazard under climate change: an alternative approach to model chains, Nat. Hazards Earth Syst. Sci., 14(6), 1579–1589, doi:10.5194/nhess-14-1579-2014, 2014.

Du, T., Xiong, L., Xu, C.-Y., Gippel, C. J., Guo, S. and Liu, P.: Return period and risk analysis of nonstationary low-flow series under climate change, J. Hydrol., 527, 234–250, doi:10.1016/j.jhydrol.2015.04.041, 2015.

Emori, S.: Dynamic and thermodynamic changes in mean and extreme precipitation under changed climate, Geophys. Res.
10   Lett., 32(17), doi:10.1029/2005GL023272, 2005.

Gelman, A. and Shirley, K.: Inference from Simulations and Monitoring Convergence in Handbook of Markov Chain Monte Carlo, Chapman and Hall/CRC., 2011.

Groisman, P. Y., Knight, R. W., Easterling, D. R., Karl, T. R., Hegerl, G. C. and Razuvaev, V. N.: Trends in Intense Precipitation in the Climate Record, J. Clim., 18(9), 1326–1350, doi:10.1175/JCLI3339.1, 2005.

15   Hartmann, D. L., Klein Tank, A. M. G., Rusticucci, M., Alexander, L. V., Brönnimann, S., Charabi, Y., Dentener, F. J., Dlugokencky, E. J., Easterling, D. R., Kaplan, A., Soden, B. J., Thorne, P., Wild, M. and Zhai, P. M.: Observations: Atmosphere and Surface. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate
20   Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]., Cambridge, United Kingdom and New York, NY, USA., 2013.

Horton, R., Yohe, G., Easterling, W., Kates, R., Khaliq, M., Sussman, E., Whelchel, A., Wolfe, D. and Lipschultz, F.: Climate Change Impacts in the United States: The Third National Climate Assessment. Ch. 16: Northeast, U.S. Global Change Research Program., 2014.

25   Hosking, J. R. M. and Wallis: Regional Frequency Analysis, Cambridge University Press. [online] Available from: http://www.cambridge.org/us/academic/subjects/earth-and-environmental-science/environmental-science/regional-frequency-analysis-approach-based-l-moments?format=PB (Accessed 29 March 2016), 2005.

Hubert, M. and Vandervieren, E.: An adjusted boxplot for skewed distributions, Comput. Stat. Data Anal., 52(12), 5186–5201, doi:10.1016/j.csda.2007.11.008, 2008.

30   Janssen, E., Wuebbles, D. J., Kunkel, K. E., Olsen, S. C. and Goodman, A.: Observational- and model-based trends and projections of extreme precipitation over the contiguous United States, Earths Future, 2(2), 99–113, doi:10.1002/2013EF000185, 2014.

Katz, R. W., Parlange, M. B. and Naveau, P.: Statistics of extremes in hydrology, Adv. Water Resour., 25(8–12), 1287–1304, doi:10.1016/S0309-1708(02)00056-8, 2002.

35   Katz, R. W., Brush, G. S. and Parlange, M. B.: Statistics of extremes: Modeling ecological disturbances, Ecology, 86(5), 1124–1134, doi:10.1890/04-0606, 2005.

Hydrology and
Earth System
Sciences

Open Access

EGU

Discussions

Keefer, T. O., Unkrich, C. L., Smith, J. R., Goodrich, D. C., Moran, M. S. and Simanton, J. R.: An event-based comparison of two types of automated-recording, weighing bucket rain gauges, Water Resour. Res., 44(5), W05S12, doi:10.1029/2006WR005841, 2008.

Kendall, M. G.: Rank correlation methods, 5th ed., Oxford University Press., 1948.

5   Khaliq, M. N., Ouarda, T. B. M. J., Ondo, J.-C., Gachon, P. and Bobée, B.: Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: A review, J. Hydrol., 329(3–4), 534–552, doi:10.1016/j.jhydrol.2006.03.004, 2006.

Kunkel, K. E., Andsager, K. and Easterling, D. R.: Long-term trends in extreme precipitation events over the conterminous United States and Canada, J. Clim., 12(8), 2515–2527, doi:10.1175/1520-0442(1999), 1999.

10   Kunkel, K. E., Easterling, D. R., Kristovich, D. A. R., Gleason, B., Stoecker, L. and Smith, R.: Meteorological Causes of the Secular Variations in Observed Extreme Precipitation Events for the Conterminous United States, J. Hydrometeorol., 13(3), 1131–1141, doi:10.1175/JHM-D-11-0108.1, 2012.

Kunkel, K. E., Karl, T. R., Easterling, D. R., Redmond, K., Young, J., Yin, X. and Hennon, P.: Probable maximum precipitation and climate change, Geophys. Res. Lett., 40(7), 1402–1408, doi:10.1002/grl.50334, 2013.

15   Lang, M., Ouarda, T. and Bobee, B.: Towards operational guidelines for over-threshold modeling, J. Hydrol., 225(3–4), 103–117, doi:10.1016/S0022-1694(99)00167-5, 1999.

Lenderink, G. and van Meijgaard, E.: Increase in hourly precipitation extremes beyond expectations from temperature changes, Nat. Geosci., 1(8), 511–514, doi:10.1038/ngeo262, 2008.

Liuzzo, L. and Freni, G.: Analysis of Extreme Rainfall Trends in Sicily for the Evaluation of Depth-Duration-Frequency
20   Curves in Climate Change Scenarios, J. Hydrol. Eng., 20(12), doi:10.1061/(ASCE)HE.1943-5584.0001230, 2015.

Mann, H. B.: Nonparametric Tests Against Trend, Econometrica, 13(3), 245–259, doi:10.2307/1907187, 1945.

Min, S.-K., Zhang, X., Zwiers, F. W. and Hegerl, G. C.: Human contribution to more-intense precipitation extremes, Nature, 470(7334), 378–381, doi:10.1038/nature09763, 2011.

Mishra, V., Wallace, J. M. and Lettenmaier, D. P.: Relationship between hourly extreme precipitation and local air
25   temperature in the United States: Extreme Precipitation and Temperature, Geophys. Res. Lett., 39(16), n/a-n/a, doi:10.1029/2012GL052790, 2012.

Obeysekera, J. and Salas, J. D.: Quantifying the Uncertainty of Design Floods under Nonstationary Conditions, J. Hydrol. Eng., 19(7), 1438–1446, doi:10.1061/(ASCE)HE.1943-5584.0000931, 2014.

Owens, L. B., Bonta, J. V. and Shipitalo, M. J.: USDA-ARS North Appalachian Experimental Watershed: 70-Year
30   Hydrologic, Soil Erosion, and Water Quality Database, Soil Sci. Soc. Am. J., 74(2), 619–623, doi:10.2136/sssaj2009.0196N, 2010.

Prein, A. F., Holland, G. J., Rasmussen, R. M., Clark, M. P. and Tye, M. R.: Running dry: The U.S. Southwest's drift into a drier climate state, Geophys. Res. Lett., 43(3), 2015GL066727, doi:10.1002/2015GL066727, 2016.

Pryor, S. C., Howe, J. A. and Kunkel, K. E.: How spatially coherent and statistically robust are temporal changes in extreme
35   precipitation in the contiguous USA?, Int. J. Climatol., 29(1), 31–45, doi:10.1002/joc.1696, 2009.

Pryor, S. C., Scavia, D., Downer, C., Iverson, L., Nordstrom, R., Patz, J. and Robertson, G. P.: Climate Change Impacts in the United States: The Third National Climate Assessment. Ch. 18:Midwest, U.S. Global Change Research Program., 2014.

Renard, B., Lang, M., Bois, P., Dupeyrat, A., Mestre, O., Niel, H., Sauquet, E., Prudhomme, C., Parey, S., Paquet, E., Neppel, L. and Gailhard, J.: Regional methods for trend detection: Assessing field significance and regional consistency,
5 Water Resour. Res., 44(8), W08419, doi:10.1029/2007WR006268, 2008.

Salas, J. D. and Obeysekera, J.: Revisiting the Concepts of Return Period and Risk for Nonstationary Hydrologic Extreme Events, J. Hydrol. Eng., 19(3), 554–568, doi:10.1061/(ASCE)HE.1943-5584.0000820, 2014.

Sarhadi, A. and Soulis, E. D.: Time-varying extreme rainfall intensity-duration-frequency curves in a changing climate, Geophys. Res. Lett., doi:10.1002/2016GL072201, 2017.

10 Seager, R.: The Turn of the Century North American Drought: Global Context, Dynamics, and Past Analogs, J. Clim., 20(22), 5527–5552, doi:10.1175/2007JCLI1529.1, 2007.

Seidel, D. J., Fu, Q., Randel, W. J. and Reichler, T. J.: Widening of the tropical belt in a changing climate, Nat. Geosci., 1(1), 21–24, doi:10.1038/ngeo.2007.38, 2008.

Sen, P.: Estimates of Regression Coefficient Based on Kendalls Tau, J. Am. Stat. Assoc., 63(324), 1379-, 1968.

15 Serinaldi, F. and Kilsby, C. G.: Stationarity is undead: Uncertainty dominates the distribution of extremes, Adv. Water Resour., 77, 17–36, doi:10.1016/j.advwatres.2014.12.013, 2015.

Soong, D., Ishii, A., Sharpe, J. and Avery, C.: Estimating Flood-Peak Discharge Magnitudes and Frequencies for Rural Streams in Illinois, 2004.

Syed, K. H., Goodrich, D. C., Myers, D. E. and Sorooshian, S.: Spatial characteristics of thunderstorm rainfall fields and
20 their relation to runoff, J. Hydrol., 271(1–4), 1–21, doi:10.1016/S0022-1694(02)00311-6, 2003.

Villarini, G., Smith, J. A. and Vecchi, G. A.: Changing Frequency of Heavy Rainfall over the Central United States, J. Clim., 26(1), 351–357, doi:10.1175/JCLI-D-12-00043.1, 2013.

Westra, S., Fowler, H. J., Evans, J. P., Alexander, L. V., Berg, P., Johnson, F., Kendon, E. J., Lenderink, G. and Roberts, N. M.: Future changes to the intensity and frequency of short-duration extreme rainfall: future intensity of sub-daily rainfall,
25 Rev. Geophys., 52(3), 522–555, doi:10.1002/2014RG000464, 2014.

Wi, S., Valdes, J. B., Steinschneider, S. and Kim, T.-W.: Non-stationary frequency analysis of extreme precipitation in South Korea using peaks-over-threshold and annual maxima, Stoch. Environ. Res. Risk Assess., 30(2), 583–606, doi:10.1007/s00477-015-1180-8, 2016.

Yilmaz, A. G., Hossain, I. and Perera, B. J. C.: Effect of climate change and variability on extreme rainfall intensity-
30 frequency-duration relationships: a case study of Melbourne, Hydrol. Earth Syst. Sci., 18(10), 4065–4076, doi:10.5194/hess-18-4065-2014, 2014.

Yu, L., Zhong, S., Pei, L., Bian, X. and Heilman, W. E.: Contribution of large-scale circulation anomalies to changes in extreme precipitation frequency in the United States, Environ. Res. Lett., 11(4), 44003–44003, 2016.

Yue, S. and Wang, C. Y.: Regional streamflow trend detection with consideration of both temporal and spatial correlation,
35 Int. J. Climatol., 22(8), 933–946, doi:10.1002/joc.781, 2002.

Hydrology and
Earth System
Sciences
Discussions

TABLES

Table 1. Selected watershed for the extreme event analysis

Experimental Watershed	State	Lat.	Lon.	Elev. [m]	Annual Prec. [mm]	Temp [C]	Rain gauges	Area [km²]	Climate Characteristics
Walnut Gulch	AZ	31.72	-110.01	1,200	312	17.7	56	149	Semi-arid climate with high-intensity summer convective precipitation
North Appalachian	OH	40.37	-81.8	359	944	9.7	3	4.2	Temperate Continental climate with convective thunderstorms in late spring and summer

5

Table 2. Number of rain gauges showing positive/negative and statistically significant positive/negative trends in
10 precipitation intensities for different durations. Totals are computed as percentage of the number of rain gauges in each
watershed. NA (not applicable) indicates that the method was not used for these time series.

			Positive		Significant Positive		Negative		Significant Negative		RAMK	
	Duration (min)		AMS	PDS	AMS	PDS	AMS	PDS	AMS	PDS	AMS	PDS
semi-arid		5	41	32	9	10	15	19	0	6	SS	NA
	short	15	38	37	5	6	20	21	0	1	-	NA
		30	40	41	2	6	17	17	1	1	-	NA
	intermediate	60	41	38	5	7	17	20	1	1	-	NA
	long	1440	50	48	12	14	6	7	0	1	SS	NA
	Total [%]		72	68	11	15	26	29	1	3		
temperate		5	2	1	0	0	0	2	0	1	-	NA
	short	15	3	0	3	0	0	1	0	0	SS	NA
		30	3	1	2	0	0	2	0	0	SS	NA
	intermediate	60	3	0	0	0	0	3	0	3	-	NA
	long	1440	3	3	3	1	0	0	0	0	SS	NA
	Total [%]		93	33	53	7	0	53	0	27		

15

Hydrology and
Earth System
Sciences

Open Access

Discussions

Table 3. Deviance statistic (D) between stationary and non-stationary models. The significance level of D is computed with a χ_k^2 distribution ($\alpha = 0.05$). Values of D larger than the 3.84 significance level are denoted in bold type.

		5-min	15-min	30-min	60-min	1440-min
GEV	semi-arid	0.5	2.2	0.4	1.4	3.4
	temperate	**9.5**	3.3	**4.6**	1.2	**6.7**
GPD	semi-arid	**8.5**	**4.0**	**6.6**	**5.2**	1.5
	temperature	-	-	-	-	-

5

10

FIGURE CAPTIONS

Figure 1. Geographic location of the study sites (a). Total seasonal precipitation for the warm months (June-October) for the period 1961-2014 for: (b) the semi-arid site (Walnut Gulch Experimental Watershed) and (c) the temperate site (North Appalachian Experimental Watershed). Thin gray lines denote a rain gauge, the blue line denotes the watershed-averaged values, the black line denotes the 11-year watershed-averaged running mean, and the dotted blue line shows the linear trend.

Figure 2. Linear trends in AMS and PDS time series for two selected rain gauges. The numbers in each plot indicate: the magnitude of the slope per decade, and dimensionless trends defined as the magnitude of the slope divided by the largest precipitation intensity. Bold numbers denote statistically significant (alpha 0.1) trends.

Figure 3. Theoretical models for precipitation intensities in: semi-arid watershed top panels and temperate watershed bottom panels for 5-, 15-, 30-, 60-, and 1440-min durations. GEV (a) and (c) and; GPD (b) and (d), respectively.

Figure 4. Model diagnosis using quantile-quantile (QQ) plots for stationary models: GEV top panels, GPD bottom panels.

Figure 5. Return periods vs. precipitation intensity return levels under stationary and non-stationary assumptions. The top two panels for GEV model: plots a. to e. semi-arid site, and plots f. to j temperate site. The bottom panels for the GPD model: plots k. to o. semi-arid site. The shaded band represents the 5th-95th percentiles of the posterior distribution (credibility interval). The solid orange/green lines denote the median of the return level posterior distribution.

Figure 6. Stationary and non-stationary IDF curves for different return periods and durations at the semi-arid (Arizona) and the temperate (Ohio) watersheds. Plots in the right most panels (p. to r.) show changes between estimated non-stationary and stationary intensities as percentage of the stationary estimates (delta= $(X_{non-sta}-X_{sta})/X_{sta} * 100$).

Figure 7. Risk of failure as a function of project life (years) assuming stationary return periods (solid lines) T= 10, 50, and 100-year and the corresponding non-stationary Tt return periods (symbols). a. semi-arid site and b. temperate site. In both cases the 15- and 1440-min duration intensities are represented with the GEV model. The subplot inset shows the variation of Tt as a function of T.

Figure 8. Probability of precipitation (risk) for different project life (n: 1 to 100-year) versus return periods in years. Precipitation intensities for durations ranging from 5- to 1440-min are represented with the GEV model. Orange(green) lines show stationary(non-stationary) conditions. a) semi-arid site and b) temperate site. The notch extremes correspond to the 25th and 75th percentiles, respectively, and the extent of the box correspond to the maximum and minimum values.

Figure 9. Risk of failure as a function of precipitation intensity in the semi-arid site for four project life: a) 10-year, b) 50-year, and c) 75-year and d) 100-year under stationary (orange) and non-stationary (green) conditions. Risk is computed for 1- to 100-yr return periods. Precipitation intensities are represented with the GEV model.
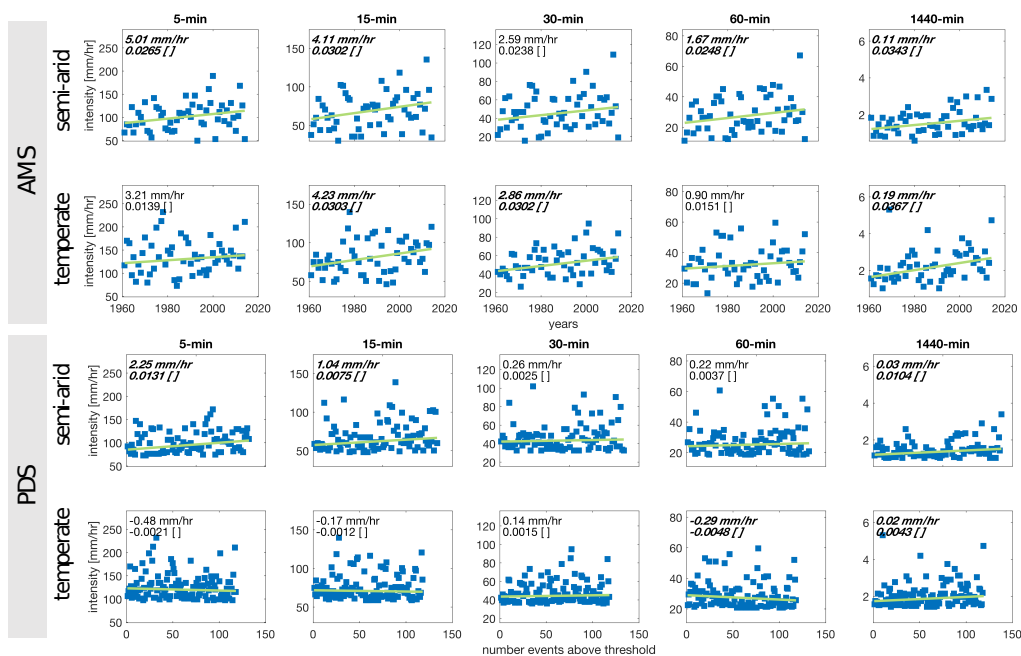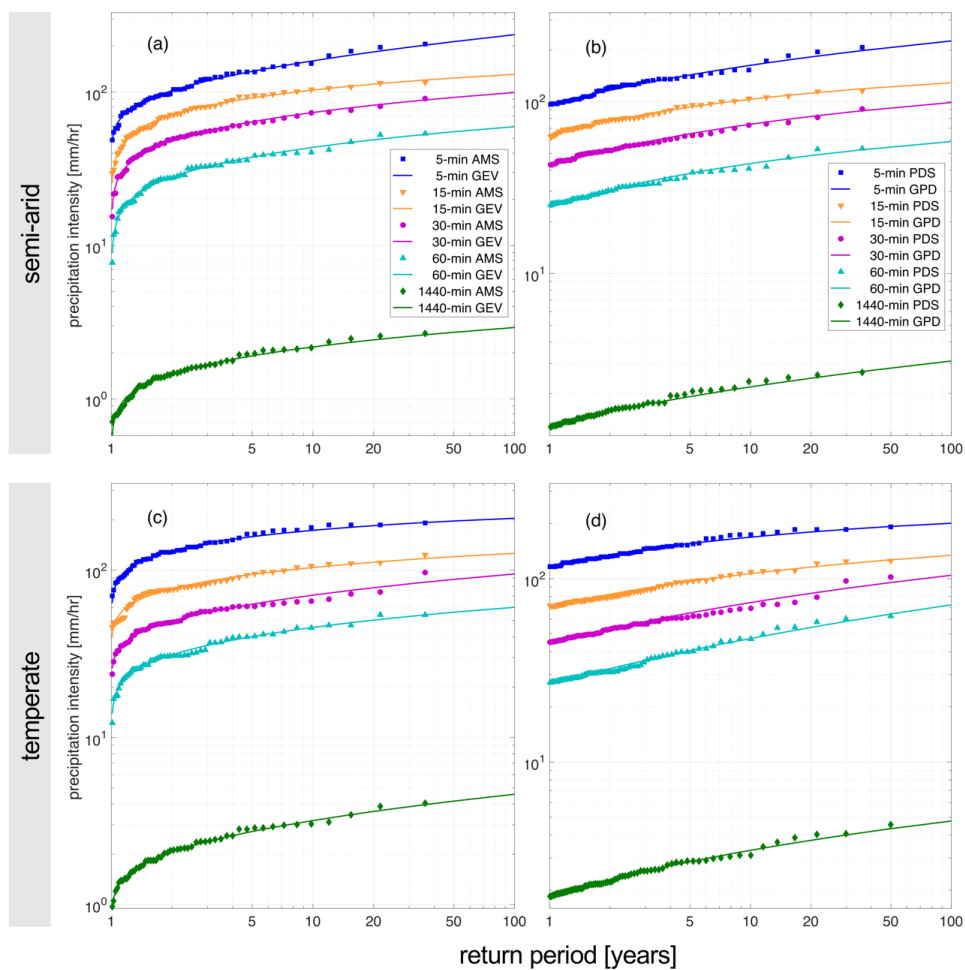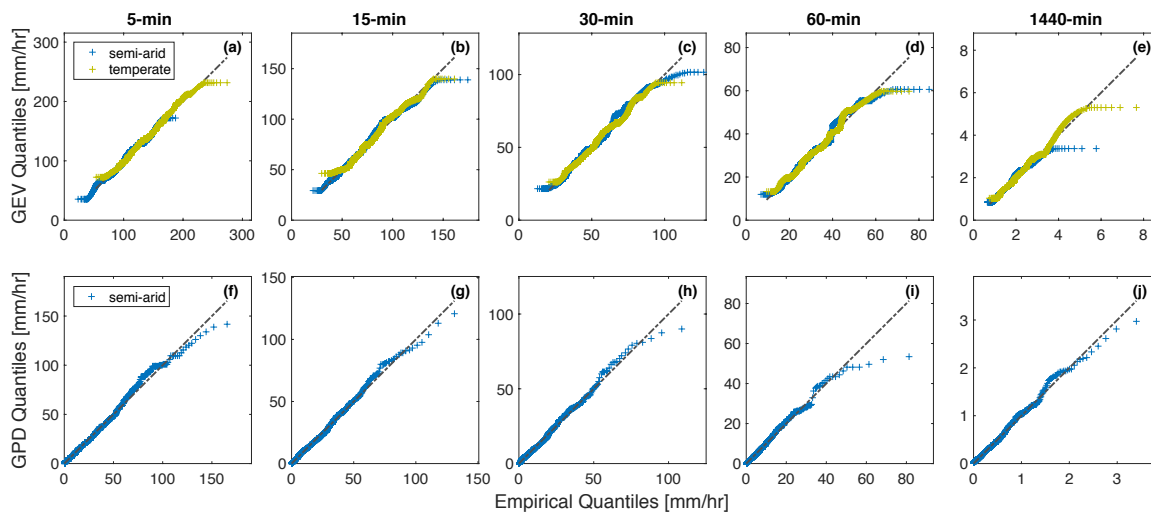
FIGURES

Figure 1. Geographic location of the study sites (a). Total seasonal precipitation for the warm months (June-October) for the period 1961-2014 for: (b) the semi-arid site (Walnut Gulch Experimental Watershed) and (c) the temperate site (North Appalachian Experimental Watershed). Thin gray lines denote a rain gauge, the blue line denotes the watershed-averaged values, the black line denotes the 11-year watershed-averaged running mean, and the dotted blue line shows the linear trend.

Figure 2. Linear trends in AMS and PDS time series for two selected rain gauges. The numbers in each plot indicate: the magnitude of the slope per decade, and dimensionless trends defined as the magnitude of the slope divided by the largest precipitation intensity. Bold numbers denote statistically significant (alpha 0.1) trends.

Figure 3. Theoretical models for precipitation intensities in: semi-arid watershed top panels and, temperate watershed bottom panels for 5-, 15-, 30-, 60-, and 1440-min durations. GEV (a) and (c) and; GPD (b) and (d), respectively.

5

Figure 4. Model diagnosis using quantile-quantile (QQ) plots for stationary models: GEV top panels, GPD bottom panels.
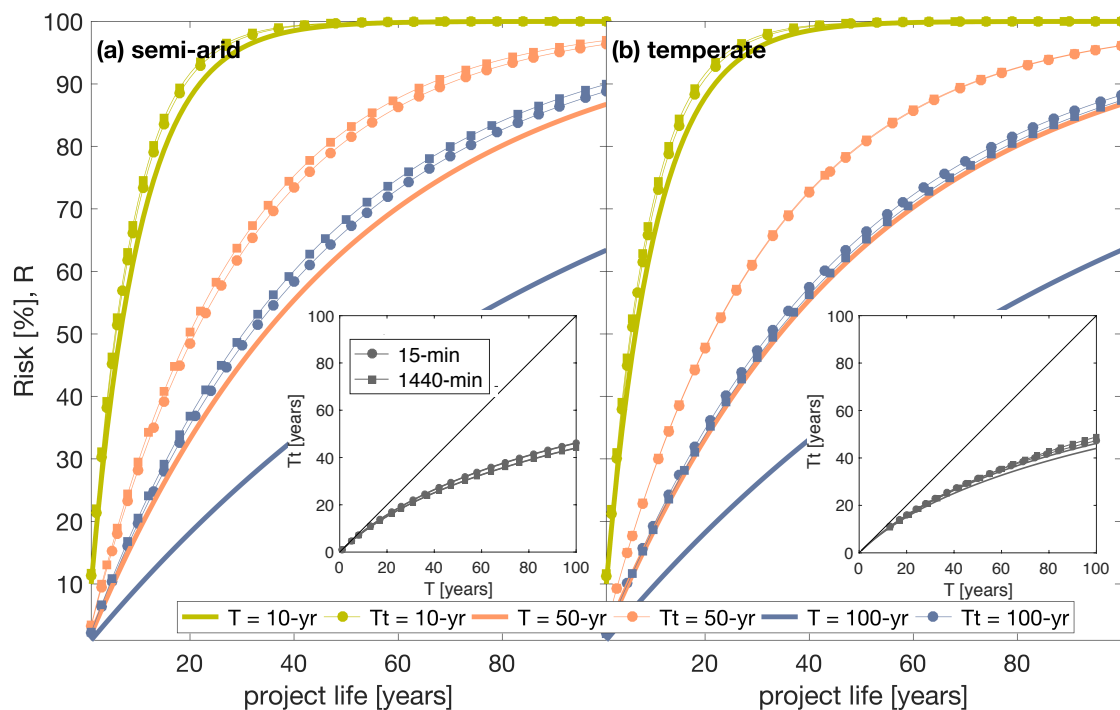
5

10

15
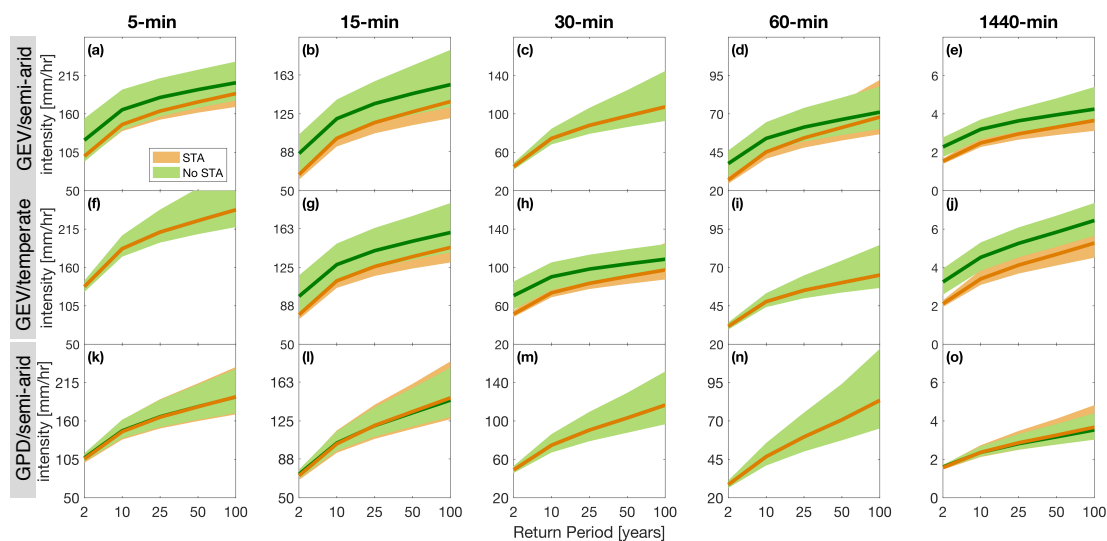
Figure 5. Return periods vs. precipitation intensity return levels under stationary and non-stationary assumptions. The top two panels for GEV model: plots a. to e. semi-arid site, and plots f. to j temperate site. The bottom panels for the GPD model: plots k. to o. semi-arid site. The shaded band represent the 5th-95th percentiles of the posterior distribution (credibility interval). The solid orange/green lines denote the median of the return level posterior distribution.
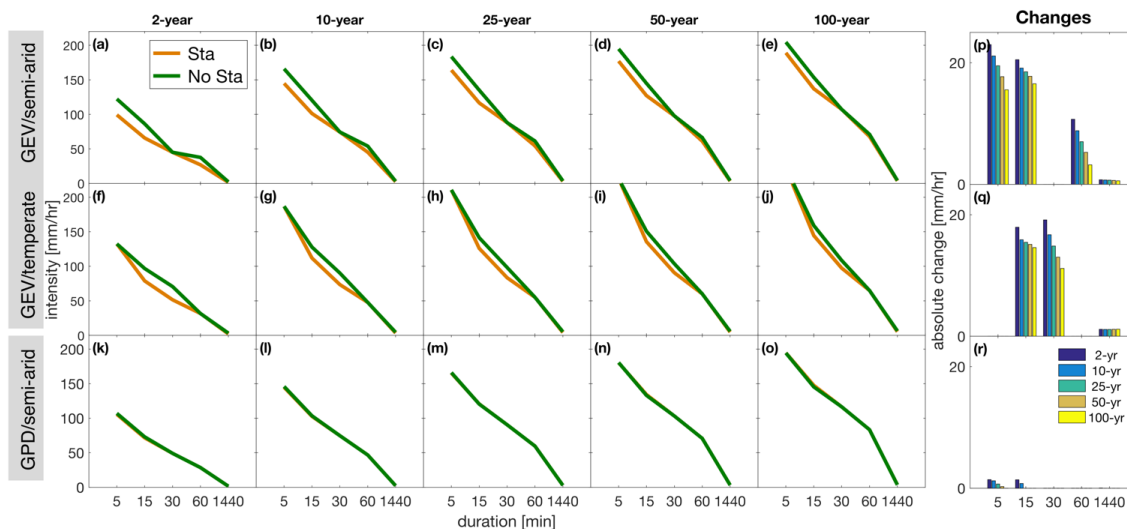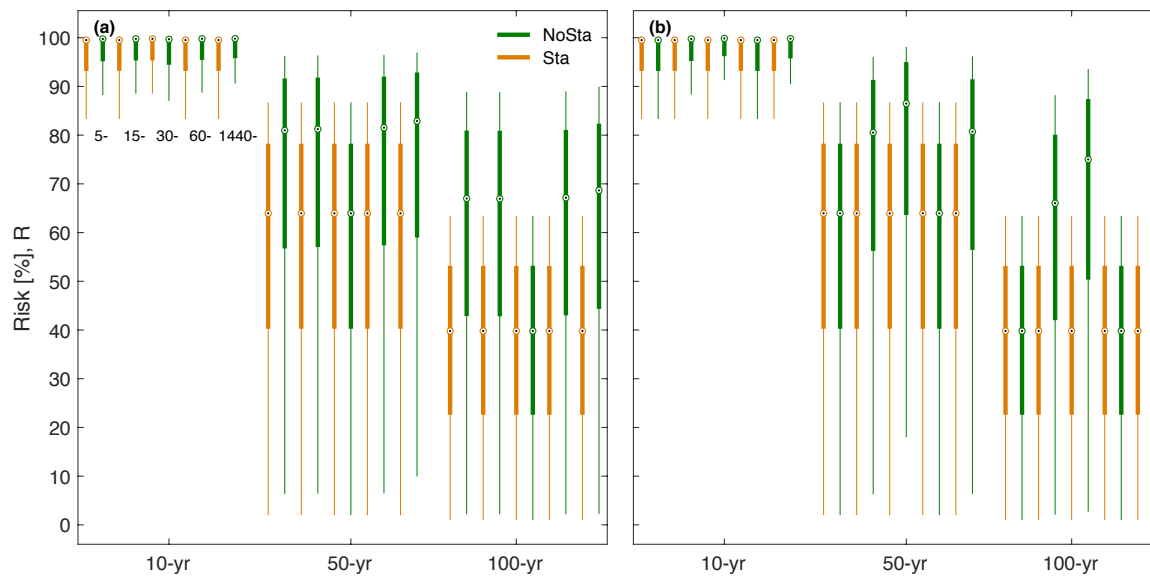
Figure 6. Stationary and non-stationary IDF curves for different return periods and durations at the semi-arid (Arizona) and the temperate (Ohio) watersheds. Plots in the right most panels (p. to r.) show changes between estimated non-stationary and stationary intensities as percentage of the stationary estimates (delta= $(X_{non-sta}-X_{sta})/X_{sta} * 100$).
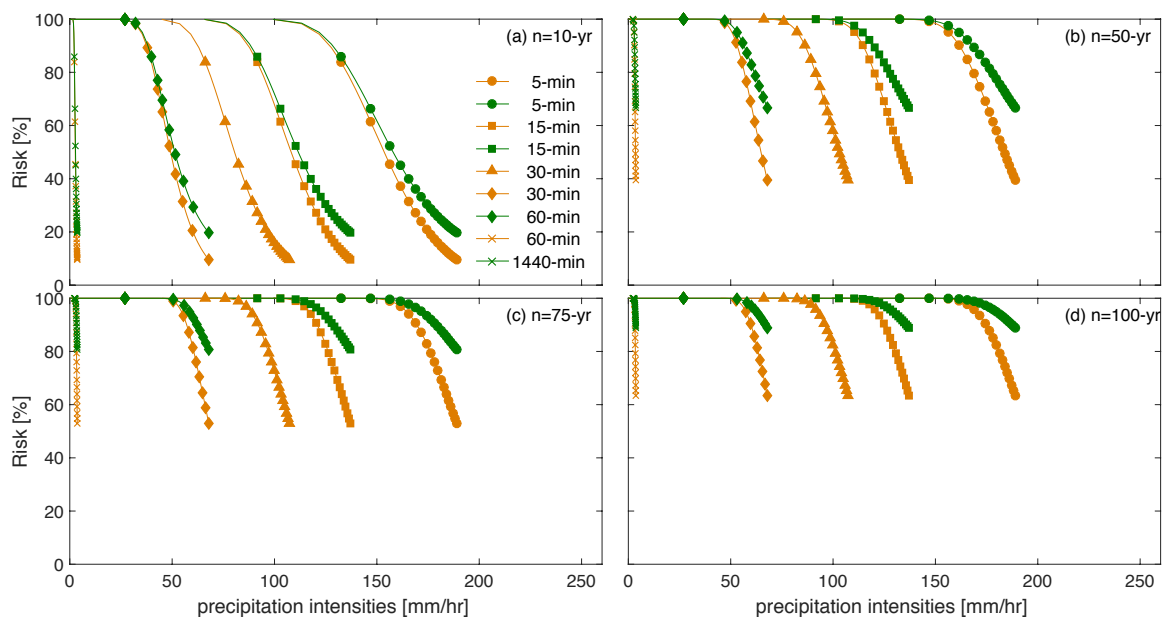
5

Figure 7. Risk of failure as a function of project life (years) assuming stationary return periods (solid lines) T= 10, 50, and 100-year and the corresponding non-stationary Tt return periods (symbols). a. semi-arid site and b. temperate site. In both cases the 15- and 1440-min duration intensities are represented with the GEV model. The subplot inset shows the variation of Tt as a function of T.

Figure 8. Probability of precipitation (risk) for different project life (n: 1 to 100-year) versus return periods in years. Precipitation intensities for durations ranging from 5- to 1440-min are represented with the GEV model. Orange(green) lines show stationary(non-stationary) conditions. a) semi-arid site and b) temperate site. The notch extremes correspond to the 25th and 75th percentiles, respectively, and the extent of the box correspond to the maximum and minimum values.

Figure 9. Risk of failure as a function of precipitation intensity in the semi-arid site for four project life: a) 10-year, b) 50-year, and c) 75-year and d) 100-year under stationary (orange) and non-stationary (green) conditions. Risk is computed for 1-
5 to 100-yr return periods. Precipitation intensities are represented with the GEV model.