

Responses to Reviewer #1 on “Does Nonstationarity in Rainfall Requires Nonstationary Intensity-Duration-Frequency Curves? By Poulomi Ganguli and Paulin Coulibaly

We thank Referee #1 for reviewing our manuscript and providing constructive feedback, which improves the quality of the manuscript. Our responses are embedded within the comments (in BLACK) in BLUE. The new additions to the revised manuscript are embedded below in GREEN.

The manuscript presents an interesting topic, and discuss the crucial question of whether there is enough evidence of changes in hydrometric series to warrant a change in the IDF curves used for the design and maintenance of hydraulic structures. Although the topic discussed is interesting and worthy, the paper is quite inconclusive and does not manage, in my opinion, to provide a clear point of view on the matter. The authors have definitely done a lot of work and have looked very carefully at the data, but they fail to summarize their finding in any useful way and simply provide a lot (too much maybe) of information. The presentation of the methods and results is quite unclear and it has several opaque points. The statistical methods are often presented with some imperfections and in general the paper could greatly benefit from some proof-reading and re-organisation. In particular the authors should make more of an attempt to summarise their findings from all the non-parametric tests in a way that is more informative.

Response: Thanks for the feedback. The reviewer comments are well appreciated. In our case, a series of statistical tests are necessary to assess nonstationarity in design rainfall, as echoed in earlier literature (Sadri et al., 2016; Yilmaz et al., 2014, 2017). A single statistical test may not be reliable enough to detect signatures of nonstationarity in hydrometeorological time series. Further, we note that multiple tests allow a more rigorous assessment of overall trend in the time series since certain tests are complimentary to each other. Therefore, we explored various statistical tests, starting from testing auto-correlation, the presence of monotonic (using trend tests) or abrupt change (using single point change detection algorithm) at different statistical significance levels in practice. Further, we have presented a flowchart of complete methodology in Figure 2 to comprehend the overall analysis. Now coming to statistical methods, we have significantly revised the manuscript to correct any miss perfections as pointed by the reviewer. While we are highly appreciative of the suggestions and comments by the reviewers, we do have one minor point to make which may come across as a slight disagreement with one set of comments. We sense a sentiment shared in one of the comments that our presentation of methods and results are quite unclear. We do not agree with this sentiment even though we agree that the various nuances were not clearly explained in the previous version of the manuscript. Since the focus of the work is insight driven, we have discussed methodologies thoroughly in Supplements to avoid the distraction of audience by over-emphasizing the methodologies.

However, we have attempted to improve the presentation of methods and re-organized our manuscript in light of the reviewer's comments. As suggested we have made the following changes in the revised manuscript:

- We have expanded Section 3.2 in Methods to include rationale for the inclusion of multiple tests for detecting nonstationarity. We argue that some of the tests are complimentary to each other. Further, multiple tests allows a robust assessment of overall trend, shifts and nonstationarity in the time series as suggested in the literature (Sadri et al., 2016; Yilmaz et al., 2014, 2017).
- We have reorganized Section 3.3 to include mathematical formulations of GEV distribution and associated time varying covariates to model nonstationary GEV parameters.
- We have re-written the Methodology section and re-organized the Supplements into different sections to present it in a more coherent and clearer way to the readers.
- We have summarized the results of trend detection tests in detail in Page 13, lines 11 – 23.
- We have included Bayes-factor criterion in addition to AIC statistics for small sample to evaluate fit of the nonstationary model.
- We have restricted our analysis to Bayesian fit for stationary and nonstationary model.
- We have recalculated 95% credible intervals for all sites from 0.025 and 0.975 quantiles of the simulated posterior samples.

The title of the manuscript indicate that IDF curves are the main topic, although the authors limit themselves to the (hard) task of fitting different frequency curves to the each series with different duration separately. This could result in non-consistent estimates eventually. The type of studies the authors perform is laudable and would be the first step to take to assess whether new IDF curves would need to be derived.

Response: Here we slightly disagree with the reviewer. First, we fitted both stationary and nonstationary frequency curves corresponding standard durations, commonly used in practice for infrastructure in design. We also test the hypothesis whether we need nonstationary frequency curves for the moderately and densely populated urbanized locations across Southern Ontario. We discussed motivation of our study in detail and extended literature review in the revision. Next, we compared the design storm estimates using simple z-statistics considering a range of uncertainty as assessed by 95% credible interval to find out whether statistically significant differences exist between nonstationary versus stationary method. Finally, we presented updated IDF curves for all nine locations across Southern Ontario, which is of interest to stakeholders' of

the region. We further compared updated versus EC-generated IDFs considering both nonstationary and stationary (Figures 6 and 7) conditions.

The authors do a lot (a lot!) of tests to the data series of each duration - definitely the issue of multiple comparisons arise and it is to be expected that some tests will turn out to be significant just by randomness.

Response: We appreciate the reviewer's point. However, we would stress that multiple tests are needed to detect the presence of monotonic trends or abrupt shifts, and nonstationarity in the time series since a selected or cherry-picked number of tests may not be sufficient to detect plausible changes and nonstationarity in the time series. Multiple tests were also performed in earlier studies (Sadri et al., 2016; Yilmaz et al., 2014, 2017) to detect temporal changes in the time series. For example, we employ both Mann-Whitney and Pettitt method to find an abrupt shift in mean in the time series, whereas Mann-Kendall test was employed to detect the monotonic trend in the time series.

Previous studies (Xie et al., 2014; Yue and Wang, 2002) have found that the rank-based nonparametric Mann-Whitney test is not really distribution free and the power of the test is often affected by the properties of sampled data. In practice, when real change point is unknown, often Mann-Whitney test, in general, does not work well, and the Pettitt method can yield plausible change point location along with its statistical significance. However, the significance of the Pettitt test can be obtained using an approximated limiting distribution. As shown earlier, the p-value associated with the test statistics is evaluated following an approximate estimate (Xie et al., 2014). Further, it is also important to note that presence of nonstationarity may not be evaluated merely on the basis of trends or abrupt shifts in the time series, even if the increasing or decreasing trends are statistically significant (Yilmaz et al., 2014). Therefore, we also employed three statistical tests, namely Augmented Dickey-Fuller (ADF), Kwiatkowski-Phillips-Schmidt-Shin (KPSS) and Priestley Subbarao (PSR) test to further investigate nonstationarity in the time series. Both ADF and KPSS tests are based on autoregressive nature of time series. However, Yilmaz et al. (2014) did not observe the presence of any significant nonstationarity in short-duration extreme rainfall time series in the city of Melbourne even after employing these tests. Therefore as an alternative, we employed frequency-based PSR test, which is able to capture nonlinear dynamical nature of hydrological system than the former two tests (Ali and Mishra, 2017; Hamed and Rao, 1999). We have incorporated these points in the revised version of the manuscript in appropriate places (Page 8, lines 24-26; Page 9, lines 5 – 10; 15 - 24).

I have to say it is difficult to follow the authors in all their testing, there is very little effort made to summarise the finding in any useful way and the results are simply presented/dumped as they are in the SI.

Response: We agreed. In the revised manuscript we provided a more detailed description of the results:

- In page 12, lines 16-20, we provided results of skewness and kurtosis in Annual Maxima (AM) time series. We move results of skewness and kurtosis analysis in the form of Tables (Tables 2 and 3) in main manuscript. We have added following sentences:

“The skewness is a measure of the asymmetry in the AMP distribution. Positive values of skewness indicate that data are skewed to the right. The skewness of sub-hourly precipitation extremes varies between 0.22 and 4.45, with highest being 30-min AMP record at Hamilton and least being at Oshawa respectively (Table 2). Likewise, for hourly extremes, the skewness ranges between 0.54 and 2.54, with least being 1-hour AMP at Oshawa and highest is 1-hour AMP at Hamilton respectively (Table 3).”

- In page 13, lines 11 – 23, we summarized results of nonstationary trend detection tests. We have added following sentences in the revised manuscript:

“We find statistically significant monotonic increase and abrupt step changes, both in mean and variance in Oshawa and Trenton respectively (Table S6 and S10), whereas London show (significant) decrease (Table S9) from duration of 6-hour and more. Windsor, Kingston and Stratford show (significant) step changes as confirmed by Mann-Whitney and Mood Tests (Tables S7, S8 and S11). On the other hand, Toronto, Hamilton and Fergus Shand Dam (Tables S4, 4.1; S5, 5.1; S12) do not exhibit any statistically significant gradual or abrupt changes in the AMP time series. The ADF tests show presence of nonstationarity in all durations across the sites. To further validate results of ADF test, KPSS and PSR tests are employed. The KPSS test detects presence of nonstationarity at 3 out of 9 sites for 24-hour rainfall extreme at 5% significance level, whereas the results of PSR test indicate nonstationarity across 5 sites in 24-hour rainfall extremes. While KPSS test alone could not detect presence of nonstationarity in any of the extreme series in Oshawa and Stratford respectively, the results of PSR test did not indicate nonstationarity in any of the short-duration rainfall extreme in Windsor. Both of these tests taken together detect presence of nonstationarity in rainfall extremes across 6 out of 9 sites”.

- We have incorporated results of the nonstationary versus stationary model fit of selected airport sites, such as, Toronto, Hamilton, Windsor and London in Tables 4 – 7 in the main manuscript and explained the results in page 14, lines 5 – 9.
- We have revised the result section to include a more thorough explanation of each of the findings.

I am not sure whether the results are reliable given the authors have p-values larger than 1.

Response: Here we briefly explain computation procedure of Pettitt Test (Xie et al., 2014), which we have appended in Supplements (SI 2).

When a sequence of random variables is divided into two segments represented by x_1, \dots, x_{t_0} and $x_{t_0+1}, x_{t_0+2}, \dots, x_T$, if each segment has distribution functions, $F_1(x)$ and $F_2(x)$, where $F_1(x) \neq F_2(x)$, then change point is identified at t_0 . Thus the null hypothesis of the test is “no change”, $H_0: \tau = T$ against the alternative of “change” $H_1: 1 \leq \tau < T$. The test is based on following statistic (Serinaldi and Kilsby, 2016; Xie et al., 2014)

$$K_T = \max_{1 \leq t \leq T} |U_{t,T}|, \text{ where } U_{t,T} = \sum_{i=1}^t \sum_{j=i+1}^T \text{sgn}(X_i - X_j)$$

Where $\text{sgn}(x) = 1$, if $x > 0$, 0 if $x = 0$ and -1 if $x < 0$. The p-value associated with K_T is

approximately evaluated as (Xie et al., 2014), $p = 2 \exp\left(\frac{-6K_T^2}{T^2 + T^3}\right)$. Given a certain significance

level α , if $p < \alpha$, we reject the null hypothesis and conclude that x_t is a significant change point at level α . Since the associated *p-value* is computed following an approximate estimate of *p-value*, in few cases it exceeds the value 1, which we sense is due to analytical intractability of the estimate. In that case, we have kept the table value blank simply putting a hyphen, and added a footnote indicating the calculation of *p-value* is analytically intractable in those cases.

Response to Further Remarks by Reviewer 1

Comment 1 The beginning of Section 3.3 is very messy and should be rewritten. Distributions do not contain parameters, they are characterised by parameters. Line 25, " a value of the shape parameter equal to zero". Line 28: "In the case of a negative shape parameter, the distribution is a Weibull". Note that the Frechet is also a bounded distribution, except it has a lower bound. Overall I would write down the whole thing in a formula, specifying the limits of the distribution for the different values of the shape.

Response: Agreed. We have added following sentences in the revision:

“The GEV distribution is characterized by three parameters, the location, the scale and the shape of the distribution, which describes the center of the distribution, the deviation around the mean and the shape or the tail of the distribution (Katz et al., 2002; Katz and Brown, 1992). The cumulative distribution function of stationary (time invariant) GEV model is given by (Coles et al., 2001; Gilleland and Katz 2016):

$$G(z) = \begin{cases} \exp\left\{-\left[1 + \zeta \left(\frac{z - \mu}{\sigma}\right)_+\right]^{-1/\zeta}\right\} & \text{if } \zeta \neq 0 \\ \exp\left\{-\exp\left(-\frac{z - \mu}{\sigma}\right)_+\right\} & \text{if } \zeta \rightarrow 0 \end{cases} \quad (3.1)$$

Where, $y_+ = \max\{y, 0\}$, and

$z \in [(\mu - \sigma)/\zeta, +\infty)$ when $\zeta > 0$; $z \in (-\infty, (\mu - \sigma)/\zeta]$ when $\zeta < 0$; and $z \in (-\infty, +\infty)$ when $\zeta = 0$

μ is a location parameter, σ is a scale parameter and ζ is a shape parameter determining the heaviness of the tail. The shape parameter ζ , determines the higher moments of the density function and also the skew in the probability mass. The ‘+’ sign indicates positive part of the argument. The Eq. (3.1) encompasses three types of DFs based on the sign of the shape parameter, ζ : (i) the Fréchet, with a finite lower bound of $(\mu - \sigma)/\zeta$ and an unbounded, heavy upper tail, ($\zeta > 0$), (ii) the Weibull, unbounded below and with a finite upper bound of $(\mu - \sigma)/\zeta$, ($\zeta < 0$) and (iii) the Gumbel, unbounded below and above with a light upper tail $\zeta = 0$, formally obtained by taking limit as $\zeta \rightarrow 0$. The Gumbel distribution is described by an unbounded light tailed distribution and the tail decreases rapidly following an exponential decay. The Fréchet distribution is a heavy-tailed distribution, and the tail drops relatively slowly following a polynomial decay (Towler et al., 2010). On the other hand, the Weibull distribution is a bounded distribution”.

Comment 2 Page 2 line 13. It is often the case though that IDF curves are derived not only from at-site data but using a pooled set of stations see for Svensson and Jones (2010, doi:10.1111/j.1753-318X.2010.01079.x) for a review of methods used in several countries.

Response: Agreed. The approach can be implemented locally (at Site; or SFA) or regionally (RFA or pooled). The regional frequency analysis is used when available record length is short or at locations where no observed data are available (Castellarin et al., 2012; Komi et al., 2016). However, various RFA estimation methods have certain drawbacks, such as Index flood method is sensitive to the homogeneity assumption and formation of regions; in Bayesian method of regionalization, the prior distributions of parameters are often not precise enough and do not add precision to the estimates; in Hierarchical approach, the method may produce abrupt changes in the parameters from one site to another. Komi et al. (2016) summarize the limitations and advantages of some of the widely used RFA techniques. In our case, the available records across all sites range between 47 and 66 years, which are more than the climatology (often over time periods of 30-years) of a region. Hence, we employ SFA method in our study. The rationale of incorporating at-site frequency method to derive IDF curves in the present study is discussed briefly in page 3, lines 20 – 28. This also allows a consistent comparison with the EC-IDFs that have been used in practice in the study area.

Comment 3 Page 3 - line 8-9: the authors seem to imply that the Gumbel distribution is symmetric - which is not the case, as it is easy to see by plotting the pdf of a Gumbel distribution.

Response: We agree. This was a mistake. We revise the sentence as follows:

EV1 distribution has certain limitations, such that it is a non-heavy tailed distribution and characterized by constant skewness and kurtosis coefficients.

Comment 4 Section 3.1: I think the information of the percentage of missing values of each station/duration should be given somewhere - ideally in the main text and not in SI. I can not judge whether the MCR technique is the most appropriate one, as this is too far away from my area of expertise.

Response: We agree. We have moved Table S1 from Supplement to the main manuscript as Table 1. We have also added an extra column in Table 1 indicating information of missing years and durations at each station.

Comment 5 Page 8 - line 8: if the 5% and 95% quantile of the posterior samples are taken then a 90% credibility interval is constructed. A 95% interval is taken to be one that contains 95% of the distribution.

Response: Agreed! As suggested we have re-analyzed our data to incorporate 2.5% and 97.5% quantiles of the posterior sample to construct a 95% credible interval.

Comment 6 Section 3.3: it is not clear to me why the authors go through the trouble of fitting both an ML and Bayesian fit for the stationary model if they only use a Bayesian model for the non-stationary models. Just use the Bayesian methods and embrace Bayesian Inference.

Response: We appreciate the reviewer's comment. As suggested, we have presented the results only using Bayesian model and exclude ML method.

Also, seeing in Table SI16-S24 that the more complex non-stationary model GEVII is often selected I wonder whether the authors have tried to only fit models with the scale taken as the only varying function?

Response: We have revised our results in light of the above comments. However, from the revised set of results, we noted that in a few cases GEV II model (nonstationary in location and scale parameter), performed better than GEV I model (nonstationary in location only). The above results are not uncommon given the highly nonstationary nature of precipitation extremes as observed from the Figure 3. Similar findings were also noted by (Gu et al., 2017) in a flood frequency analysis of Pearl River basin in China, where the authors have analyzed 28 stream gauge locations. The results of their analysis suggested in 5 out of 28 sites GEVII performed better as compared to the stationary and GEV I models.

Lastly, why not to formally test stationary/non-stationary model is better by using a Bayesian factor or some pre-set rule on the 95% credibility interval not-containing zero?

Response: Agreed! We have incorporated Bayes factor, AIC statistics for small sample and probability-probability ($P-P$) plot to evaluate model fit.

Comment 7 Section 3.3: what do you do with the results of the Pettitt test? One could use it to build a model with a step-change rather than a continuous function of time. In general, why doing all the non-parametric test AND the parametric models? What is the use of the non-parametric tests exactly?

Response: This is indeed a good point raised by the reviewer. Here, we used three different tests, Pettitt, Mann-Whitney and Mood tests to identify abrupt step changes in the time series, which is different from monotonic or gradual trends in the time series. We have implemented a series of statistical tests since a single statistical test may not be able to capture full ranges of nonstationarity in highly nonlinear dynamical system, such as short-duration extreme precipitation. As we discussed earlier, the rank-based nonparametric Mann-Whitney test is not really a distribution free and the power of the test is often affected by the properties of sampled data. In practice, when real change point is unknown, often Mann-Whitney test, in general, does not work well and the Pettitt method can yield plausible change point location along with its

statistical significance. However, the significance of the Pettitt test can be obtained using an approximated limiting distribution. Therefore, above tests were needed in the current setting.

Further, we applied nonparametric tests due to their robustness to non-normality, which usually appears in the hydroclimatic time series. Further, in order to reduce the number of underlying assumption required for testing a hypothesis, such as the presence of specific kind of trend or change point in the data set, nonparametric tests were employed. We discussed each of these issues in the revised manuscript.

Comment 8. Page 8 - line 14: it is very good that the authors verify the goodness of fit by using PP, but it is unclear to me how they "select the model with fewer parameters as the best model when two models have comparable performances.". This is exactly what the AIC should do, so even if the AIC does not indicate that a simpler model should be used the authors might cull a non-stationary model out if the stationary model give a better fit in the PP plot?

Response: We have reanalyzed the data and new results are different from the previous ones.

Comment 9. Page 8 - line 25-26: a positive skewness is just an indication of an asymmetric/skewed distribution, it doesn't necessary indicate a change in the distribution. I mean "extreme values are more frequent in the time series" compared to what?

Response: We have revised this sentence in page 12 (line 17-18) as follows:

Positive values of skewness indicate that data are skewed to the right.

Comment 10. Page 9, line 29: Bayesian measures of uncertainty are normally called credibility and not confidence intervals. Also as I mentioned above - unclear if the 95% or the 90% intervals are derived.

Response: We appreciate reviewer's feedback. As suggested we have replaced the word with credibility interval wherever it is appropriate. We have constructed 95% credibility intervals from the 2.5th and 97.5th percentiles of the simulated posterior samples.

Comment 11. Page 10/Figure 4: how are the DSI calculated for the non-stationary models? Is the last value of the parameters used to compute the quantiles? Why do you show boxplots of the posterior sample and not a 95% credibility interval? As I said I would drop the estimation using ML completely, but if you do use it, you could show confidence intervals based on the delta-method (see Coles, 2001).

Response: We estimated parameters using Bayesian inference (BI) coupled with Differential Evaluation Markov Chain (DE-MC) simulation as in (Cheng and AghaKouchak, 2014; Cheng et

al., 2014). DE-MC is an adaptive Monte Carlo Markov Chain (MCMC) algorithm (Ter Braak and Vrugt, 2008; Ter Braak, 2006), in which multiple chains (here, we fix chain length ‘ n ’ as 5) are run in parallel. The resulting MC simulations are then run to an equilibrium (often referred to as the *burn-in* period). It is a standard practice to discard the initial iterations of simulated samples since they are strongly influenced by starting values and do not provide usable information of the target distribution. Here we run DE-MC simulations for 3000 iterations and kept the 2001-3000th iterations of each chain. The convergence of MC simulation is checked by the “potential scale reduction factor (\widehat{R})” as in (Gelman et al., 2011), which suggests the value of \widehat{R} should remain below the threshold value of 1.1. The post burn-in random draws from posterior distribution are then used to construct predictive distributions. For annual maxima time series of each duration, the mean and associated 95% credibility intervals of parameters ($\mu(t), \sigma(t)$) are derived by computing 50th (the median), 2.5th and 97.5th (bounds) percentiles of post *burn-in* random draw (for example, 50th percentile of $\mu(t_1), \dots, \mu(t_{100})$). The derived model parameters are then used to compute corresponding design rainfall quantiles at T -year return period and corresponding credibility interval. We calculated the median value of design storm by computing 50th percentiles of the post-burn in simulated posterior quantiles for the nonstationary model. We have constructed 95% credibility intervals from the 2.5th and 97.5th percentiles of the posterior samples.

In the manuscript, the boxplots are shown for 95% credibility interval and not with posterior samples. To avoid further ambiguity we have revised corresponding figure caption (Figure 4) as, “DSI estimates of the median (horizontal line within the box plot) and 95% credible intervals for 100-year return periods of stationary versus nonstationary models across nine sites (a - i). The boxplots indicate the uncertainty in estimated DSI using Bayesian inference”. As suggested we have dropped ML method completely in the revised manuscript.

Comment 12. Page 11/Figure 6: has any assessment been done on whether the stationary version of the fitted curves has a good overlap to the EC-curves? Surely if these two curves are very different, any mis-match between the non-stationary results and the EC-curves could be due to the fact that the EC curve doesn’t fully fit the data of a site. This links to a comment on the statements in page 13 between line 20-25: you are saying that from the comparison of stationary to non-stationary models there seems to be no indication of a need to update DSI, but when comparing the outputs of a non-stationary model to the EC-curves (obtained assuming stationarity) then the evidence is that we should update the DSI. This points in the direction of the EC-curves being different from the at-site stationary curves.

Response: As suggested we have compared the stationary version of the fitted curve with EC curves. Associated results are presented in Figures 8 and S15. We discuss following results in the revision:

“In order to distinguish between the stationary and nonstationary method of analysis, we also present updated IDF assuming stationary condition relative to EC IDF in the same plot (in top panel). The comparisons of remaining sites are presented in Figure S15. Thus we made the first attempt to compare the results of updated versus EC-generated IDFs considering both nonstationary and stationary conditions, which are part of contemporary Design Standards and widely used by the stakeholders and practitioners. Overall, the updated IDFs closely follow the pattern of trends analogous to EC-generated IDFs, except for the 100-year return period. The difference is more pronounced considering nonstationary condition, especially at Toronto International Airport (Figure 8), Oshawa WPCP and Stratford WWTP (Figure S15). At longer durations and higher return periods, stations in metropolitan areas (such as Toronto International Airport, Hamilton Airport, Oshawa WPCP and Windsor Airport) show large differences in DSIs, whereas moderately populated locations such as Kingston P. station and Fergus Shand dam show relatively smaller changes. Considering, nonstationary condition, the maximum increase in Furgas Shand dam is noted as 18.7% for the 2-hour storm duration and 100-year return period, whereas an increase of around 44.5% is shown for 12-hour storm duration at Toronto Airport”.

Comment 13. Page 14: I don't understand what the last sentence of the paper means.

Response: We have revised the sentence as follows:

“Given that these findings are for the current period (e.g. historical extreme rainfall time series), we recommend a careful extrapolation of the findings with regards to future climate projections, in which frequency and magnitude of extreme rainfall are expected to intensify (Mailhot et al., 2012; Deng et al., 2016; Fischer and Knutti, 2016; Prein et al., 2016; Pfahl et al., 2017). Further work should consider nonstationary methods for deriving future IDFs in Southern Ontario.”

Comment 14. SI3: I would give the lower and upper bound of the GEV in a formula to give a simpler indication of the effect of the value of the shape parameter.

Response: Agreed, we add following expressions to indicate effect of shape parameter in GEV distribution:

$$G(z) = \begin{cases} \exp \left\{ - \left[1 + \zeta \left(\frac{z - \mu}{\sigma} \right)_+ \right]^{-1/\zeta} \right\} & \text{if } \zeta \neq 0 \\ \exp \left\{ - \exp \left(- \frac{z - \mu}{\sigma} \right)_+ \right\} & \text{if } \zeta \rightarrow 0 \end{cases} \quad (3.1)$$

Where, $y_+ = \max \{y, 0\}$, and

$z \in [(\mu - \sigma)/\zeta, +\infty)$ when $\zeta > 0$; $z \in (-\infty, (\mu - \sigma)/\zeta]$ when $\zeta < 0$; and $z \in (-\infty, +\infty)$ when $\zeta = 0$

μ is a location parameter, σ is a scale parameter and ζ is a shape parameter determining the heaviness of the tail. The shape parameter ζ , determines the higher moments of the density function and also the skew in the probability mass. The ‘+’ sign indicates positive part of the argument. The Eq. (3.1) encompasses three types of DFs based on the sign of the shape parameter, ζ : (i) the Fréchet, with a finite lower bound of $(\mu - \sigma)/\zeta$ and an unbounded, heavy upper tail, ($\zeta > 0$), (ii) the Weibull, unbounded below and with a finite upper bound of $(\mu - \sigma)/\zeta$, ($\zeta < 0$) and (iii) the Gumbel, unbounded below and above with a light upper tail $\zeta = 0$, formally obtained by taking limit as $\zeta \rightarrow 0$.

Comment 15. SI3.1: why using ML in one case and Bayesian methods for another?

Response: Agreed. As suggested we have excluded the results of ML estimate.

Comment 16. SI3.1, paragraph after equation 3.8: $p(y|\lambda, x)$ does not give information on the parameters. The formulation of the sentence seem to imply that the likelihood $p(y|\lambda, x)$ gives information on the parameters under non-stationarity, which is not the case.

Response: Agreed. To avoid any ambiguity, we have revised the sentence as:

“The posterior distributions, $p(\omega|y)$ and $p(y|\lambda, x)$ indicate likelihood functions, which infer parameters $\omega = \{\mu, \sigma, \zeta\}$ considering stationarity, and $\lambda = \{\mu_1, \mu_0, \sigma_1, \sigma_0, \zeta\}$ assuming nonstationarity conditions, respectively.”

Comment 17. SI4.1 - the definition in eq 4.1 for the Akaike information criterion is not correct (or better it is correct for a normal model, but not for a GEV). AIC is generally defined as $AIC = -2\log(L(\omega, x)) + 2m$. That’s how the two references cited by the authors define the AIC as well. From what I understand from the explanation of the observed/expected values the authors are doing a model selection using AIC based on the quantiles, which is not made explicit in section 3.3. If that’s the case, which quantiles are used?

Response: Here we cannot concur with the reviewer. We also point to the reviewer that we have used a least square version of Akaike Information Criterion (AIC), which is calculated as the largest deviation between the observed (empirical in this case, obtained from rank-based plotting position formula) and modelled cumulative distribution. This form of AIC is widely used in hydrology in general and multivariate statistics in particular (Dawson et al., 2007; Deepthi Rajsekhar et al., 2015; Ganguli and Reddy, 2012; Hu, 2007; Janga Reddy and Ganguli, 2012; Karmakar and Simonovic, 2007, 2009). Further, we point that this form does not correspond to a normal model. For calculation of AIC statistics, we consider median of the DE-MC sampled parameters, which can be considered as an average or expected value of risk in the historical observation. We have added this in detail in section 3.3 as suggested by the reviewer.

Comment 18. Equation 5.1 and 5.2, what happens if $\zeta = 0$?

Response: When $\zeta \rightarrow 0$, the GEV distribution reduces to Gumbel distribution (or Extreme Value Type I). In that case, the return period is obtained by calculating frequency factor. We add following sentences SI 4, page 40 in the revised version of the manuscript:

“When $\zeta \rightarrow 0$, the GEV distribution reduces to Gumbel distribution (or Extreme Value Type I). It should be noted that Gumbel Extreme value distribution has been commonly used to estimate design storm by Environment Canada (CSA, 2010). The Gumbel probability distribution has following form (Wang et al., 2015)

$$q_p = \mu + K_p \sigma$$

Where K_p denotes frequency factor depending on the return period T , which is obtained using following relationship (Wang et al., 2015)

$$K_p = \frac{-\sqrt{6}}{\pi} \left[0.5772 + \ln \left(\ln \left(\frac{T}{T-1} \right) \right) \right]$$

Environment Canada uses this method to estimate rainfall frequency at a given duration and obtain nationwide IDF curves”.

Comment 19 Table S7, 24-hours, the p-value for the pettitt test is larger than 1 - this cannot be right. (see also S9 30min, S11 15min to 2hr, S14 15min to 2hr, S15 12hr)

Response: Agreed. As explained before, the significance of the Pettitt test can be obtained using an approximated limiting distribution, the p-value of certain durations could not be computed accurately due to analytical intractability. We have kept those places as blank (-) in the revised manuscript. We have added a footnote at the end of Table S4 explaining this point.

Comment 20 Table SI16 - not sure if the red and blue are right in all stations.

Response: We have revised our analysis and revised results are different from earlier.

Comment 21 Pg 14 Supplement : the definition of return level has the word expected in the wrong place. ... often referred as return level in the literature is the expected value to be exceeded on an average once in every... should be ... often referred as return level in the literature, is the value which is expected to be exceeded on an average once in every... - see Coles, 2001 - end of section 3.1.3 (pg 49 in my edition).

Response: Agreed. We have revised the definition in the current version as suggested.

Comment 22 I also find some of the Figures - and in particular their captions - could be improved.

Response: Agreed. We have revised captions of the figures wherever appropriate to enhance clarity. By doing so, we have also incorporated changes as suggested by the reviewer.

Comment 22.1 Figure 3 caption

- Durations higher than an hour are also shown I would say "Spatial distribution of trends, change points and non-stationarities in rainfall extremes of several durations in nine urbanized locations, Southern Ontario"
- Drop the information on the population - it's in Figure 2 and in the text (several times)
- Drop the information on the tests performed or at least reduce it since it's given in the text (for example drop the references)
- Include information on the color coding in the legend.
- If tests are performed at 5% and 10% - what is considered statistically significant? p-values < 0.05 or p-value < 0.1 ?

Response: Agreed and incorporated in the revision. Further, p-values < 0.1 is considered to be statistically significant. The same has been incorporated in the revision.

Comment 22.2 Figure 4 caption: drop the list of the name of the station - it is given in the plot.

Response: Agreed and incorporated in the revision. Also, we have revised the figure caption in light of comment no. 11.

Comment 22.3 Figure 5 caption: add the information on the cyan shading representing the site with significant autocorrelation in the legend and drop from the legend. The second last sentence grammar is not correct.

Response: Agreed and incorporated in the revision. We have revised the grammar of the second last sentence.

Comment 22.4 Figure 7: I would include the information on solid/dotted lines in the legend.

Response: Agreed and incorporated in the revision.

Comment 23 The paper has several grammar mistakes, with articles missing or appearing in the wrong place and several sentences which have non-concordant subject and verb. I list here a minuscule sample of the typos/mistakes I found

Response: We have thoroughly checked the manuscript, corrected all typos. We have revised the manuscript in places as they were suggested.

Comment 23.1 Page 3, line 16 slowly or varying are not antonyms. Line 18-19 does should have a singular subject (not signal). Same in line 25-26.

Response: Agreed. We have revised this to gradual or monotonic changes. We have revised the sentence in line 18-19. We have revised the grammar in line 25-26.

Comment 23.2 Page 3 line 23-24: The structure of the sentence is confusing. It is not the signatures that necessitate IDF. Maybe use "...make necessary the use..."

Response: Agreed and incorporated in the revision.

Comment 23.3 Page 5: line 4-5 more repeated twice.

Response: Agreed and we have revised the sentence as suggested.

Comment 23.4 Page 8: Line 27-28: the sentence is not complete.

Response: We apologized for this. We have corrected all incomplete sentences including this one in the revision.

Comment 23.5 Page 10 - line 16: less uncertainty (not lesser).

Response: Agreed and incorporated in the revision.

Comment 23.6 Page 11 - line 17: More genrally - and the sentence has a singular subject so line 19 should be is not are.

Response: Agreed and incorporated in the revision.

Comment 23.7 Page 12 - line 2: smaller, not lesser.

Response: Agreed and incorporated in the revision.

Comment 23.8 Page 12 - line 17: It? I think you need a "We"?

Response: Agreed and incorporated in the revision.

Comment 23.9 Page 13 - line 6: does/is?

Response: Agreed and incorporated in the revision.

Comment 23.10 Page 13 - line 12: several studies HAVE.

Response: Agreed and incorporated in the revision.

Comment 24. Further inconsistencies I identified:

- **Comment 24.1** Page 4 Line 10: the ref to Jien and Gough is missing in the reference list and I think is not needed since it states a basic fact about the geography of Canada.

Response: Agreed and the citation is excluded from the revised version.

- **Comment 24.2** Page 9, Line 28 - ξ , instead of ζ used in the SI, for the shape parameter of the GEV.
Response: Agreed and incorporated in the revised version of the manuscript.
- **Comment 24.3** Reference list: Cheng, L. and AghaKouchak, A. 2014 - just give the doi, not the ncbi link.
Response: Agreed and incorporated in the revision.
- **Comment 24.3** Supplement references: Coles and Tawn (1996) cited in text missing in the ref. Anyway, for that formula Coles, 2001 is probably enough as a citation.
Response: The citation Coles and Tawn (1996) is included in the revised version.
- **Comment 24.4** The citation to Coles 2001, An introduction to statistical modelling of extreme values, Springer in the supplementary material is wrong, as it has additional authors other than Coles.
Response: Agreed and incorporated in the revision.

References

- Ali, H. and Mishra, V.: Contrasting response of rainfall extremes to increase in surface air and dewpoint temperatures at urban locations in India, *Sci. Rep.*, 7(1), 1228, doi:10.1038/s41598-017-01306-1, 2017.
- Castellarin, A., Kohnová, S., Gaál, L., Fleig, A., Salinas, J. L., Toumazis, A., Kjeldsen, T. R. and Macdonald, N.: Review of applied-statistical methods for flood-frequency analysis in Europe, Available from: <http://nora.nerc.ac.uk/19286/>, 2012.
- Cheng, L. and AghaKouchak, A.: Nonstationary precipitation intensity-duration-frequency curves for infrastructure design in a changing climate, *Sci. Rep.*, 4, doi: 10.1038/srep07093, 2014.
- Cheng, L., AghaKouchak, A., Gilleland, E. and Katz, R. W.: Non-stationary extreme value analysis in a changing climate, *Clim. Change*, 127(2), 353–369, 2014.
- Coles, S. G. and Tawn, J. A.: A Bayesian Analysis of Extreme Rainfall Data, *J. R. Stat. Soc. Ser. C Appl. Stat.*, 45(4), 463–478, doi:10.2307/2986068, 1996.
- Coles, S.: An introduction to statistical modeling of extreme values, Springer, 2001.
- CSA (Canadian Standards Association): Technical Guide – Development, Interpretation and Use of Rainfall Intensity-duration-frequency (IDF) Information: Guideline for Canadian Water Resources Practitioners, 2010.
- Dawson, C. W., Abrahart, R. J. and See, L. M.: HydroTest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts, *Environ. Model. Softw.*, 22(7), 1034–1052, doi:10.1016/j.envsoft.2006.06.008, 2007.

- Deepthi Rajsekhar, Vijay P. Singh and Ashok K. Mishra: Hydrologic Drought Atlas for Texas, *J. Hydrol. Eng.*, 20(7), doi:10.1061/(ASCE)HE.1943-5584.0001074, 2015.
- Deng, Z., Qiu, X., Liu, J., Madras, N., Wang, X. and Zhu, H.: Trend in frequency of extreme precipitation events over Ontario from ensembles of multiple GCMs, *Clim. Dyn.*, 46(9–10), 2909–2921, 2016.
- Fischer, E. M. and Knutti, R.: Observed heavy precipitation increase confirms theory and early models, *Nat. Clim. Change*, 6(11), 986–991, 2016.
- Ganguli, P. and Reddy, M. J.: Probabilistic assessment of flood risks using trivariate copulas, *Theor. Appl. Climatol.*, 111(1–2), 341–360, doi:10.1007/s00704-012-0664-4, 2012.
- Gelman, A., Shirley, K. and others: Inference from simulations and monitoring convergence, *Handb. Markov Chain Monte Carlo*, 163–174, 2011.
- Gilleland, E. and Katz, R. W.: Extremes 2.0: an extreme value analysis package in r, *J. Stat. Softw.*, 72(8), 2016.
- Gu, X., Zhang, Q., Singh, V. P., Xiao, M. and Cheng, J.: Nonstationarity-based evaluation of flood risk in the Pearl River basin: changing patterns, causes and implications, *Hydrol. Sci. J.*, 62(2), 246–258, 2017.
- Hamed, K. H. and Rao, A. R.: A modified Mann-Kendall trend test for autocorrelated data, *J. Hydrol.*, 204(1), 182–196, 1998.
- Hu, S.: Akaike information criterion, *Cent. Res. Sci. Comput.*, North Carolina State University. Available from: http://www4.ncsu.edu/~shu3/Presentation/AIC_2012.pdf, 2007.
- Janga Reddy, M. and Ganguli, P.: Application of copulas for derivation of drought severity–duration–frequency curves, *Hydrol. Process.*, 26(11), 1672–1685, doi:10.1002/hyp.8287, 2012.
- Karmakar, S. and Simonovic, S. p.: Bivariate flood frequency analysis. Part 2: a copula-based approach with mixed marginal distributions, *J. Flood Risk Manag.*, 2(1), 32–44, doi:10.1111/j.1753-318X.2009.01020.x, 2009.
- Karmakar, S. and Simonovic, S.: Flood Frequency Analysis Using Copula with Mixed Marginal Distributions, *Water Resour. Res. Rep.* Available from: <http://ir.lib.uwo.ca/wrrr/19>, 2007.
- Katz, R. W. and Brown, B. G.: Extreme events in a changing climate: variability is more important than averages, *Clim. Change*, 21(3), 289–302, 1992.
- Katz, R. W., Parlange, M. B. and Naveau, P.: Statistics of extremes in hydrology, *Adv. Water Resour.*, 25(8), 1287–1304, 2002.
- Komi, K., Amisigo, B. A., Diekkrüger, B. and Hountondji, F. C.: Regional Flood Frequency Analysis in the Volta River Basin, West Africa, *Hydrology*, 3(1), 5, 2016.
- Mailhot, A., Duchesne, S., Caya, D. and Talbot, G.: Assessment of future change in intensity–duration–frequency (IDF) curves for Southern Quebec using the Canadian Regional Climate Model (CRCM), *J. Hydrol.*, 347(1), 197–210, 2007.
- Prein, A. F., Rasmussen, R. M., Ikeda, K., Liu, C., Clark, M. P. and Holland, G. J.: The future intensification of hourly precipitation extremes, *Nat. Clim. Change*, advance online publication, doi:10.1038/nclimate3168, 2016.
- Sadri, S., Kam, J. and Sheffield, J.: Nonstationarity of low flows and their timing in the eastern United States, *Hydrol Earth Syst Sci*, 20(2), 633–649, 2016.
- Serinaldi, F. and Kilsby, C. G.: Stationarity is undead: Uncertainty dominates the distribution of extremes, *Adv. Water Resour.*, 77, 17–36, 2015.

- Ter Braak, C. J. and Vrugt, J. A.: Differential evolution Markov chain with snooker updater and fewer chains, *Stat. Comput.*, 18(4), 435–446, 2008.
- Ter Braak, C. J.: A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces, *Stat. Comput.*, 16(3), 239–249, 2006.
- Wang, X., Huang, G., Liu, J., Li, Z. and Zhao, S.: Ensemble projections of regional climatic changes over Ontario, Canada, *J. Clim.*, 28(18), 7327–7346, 2015.
- Wang, X., Huang, G., Liu, J., Li, Z. and Zhao, S.: Ensemble projections of regional climatic changes over Ontario, Canada, *J. Clim.*, 28(18), 7327–7346, 2015.
- Xie, H., Li, D. and Xiong, L.: Exploring the ability of the Pettitt method for detecting change point by Monte Carlo simulation, *Stoch. Environ. Res. Risk Assess.*, 28(7), 1643–1655, 2014.
- Yilmaz, A. G., Hossain, I. and Perera, B. J. C.: Effect of climate change and variability on extreme rainfall intensity–frequency–duration relationships: a case study of Melbourne, *Hydrol. Earth Syst. Sci.*, 18(10), 4065–4076, 2014.
- Yilmaz, A. G., Imteaz, M. A. and Perera, B. J. C.: Investigation of non-stationarity of extreme rainfalls and spatial variability of rainfall intensity–frequency–duration relationships: a case study of Victoria, Australia, *Int. J. Climatol.*, 37(1), 430–442, doi:10.1002/joc.4716, 2017.
- Yue, S. and Wang, C. Y.: Power of the Mann–Whitney test for detecting a shift in median or mean of hydro-meteorological data, *Stoch. Environ. Res. Risk Assess.*, 16(4), 307–323, 2002.

Responses to Reviewer #2 on “Does Nonstationarity in Rainfall Requires Nonstationary Intensity-Duration-Frequency Curves? By Poulomi Ganguli and Paulin Coulibaly

We thank Referee #2 for reviewing our manuscript and providing constructive feedback. Our responses are embedded within the comments (in BLACK) in BLUE. The new additions to the revised manuscript are embedded below in GREEN.

Reviewer #2

Comment 1. The manuscript could do with a good proof read and rewrite. There are lots of little mistakes which makes the paper very difficult to read. I was constantly stopped in my train of reading by small errors or references to figures/tables which weren't explained. The supplementary material is 66 pages and has 37 Tables. This is huge and difficult to come to terms with – I couldn't follow it all. As I don't believe a specific structure is required can I recommend the following? Group the supplementary text and figures and tables into sections. That way you will have separate sections to refer to in the main text. You can then go sequentially through the text. S1 is the infilling, S2 is the autocorrelation method and results, S3 non-stationarity test method and results, S4 GEV fitting. I may have got the headings incorrect but I hope what I mean is clear. Then with the results you can just reference a section for detailed results and focus on discussing the figures in the main text. Trying to interpret 37 tables (some split into two) – almost all which are referenced in the main text - it is like trying to read a thesis.

Response: This is indeed a good point and we have revised the supplementary section and reorganized the material into various sections as suggested. We discussed corresponding results in the form of tables and figures under each subsection making it more coherent and easier to read. Also, we have moved some of the Tables (for example, Table S1) from supplements to main manuscript reducing the length of the Supplements to 57 pages with 30 tables all together.

Comment 2. Moving Table S1 to the main text, and maybe removing Figure S1 altogether will make the manuscript more standalone and easier to read. This manuscript is a bit short on doing justice to some of the previous work done in this area.

Response: Here we partially agree with the reviewer's comment. We have moved Table S1 to the main text. However, we have retained the Figure S1 in the Supplement since the figure provides a conceptual representation of changes in probability density functions of extremes in a nonstationary environment. We feel the figure will help readers in understanding how the nonstationarity may lead to changes in the distribution of extremes, which can potentially lead to the changes in the frequency of extremes.

Comment 3. Page 2 – Line 21: This is the only line discussing previous work to do with non-stationary IDFs. I think this work deserves more attention given that the focus of this manuscript is non-stationary IDFs. My recommendation is as follows:

In Page 1 – Line 23: “In a warming climate . . .” I would be a bit more careful here and expand this. I would cite Lenderink and van Meijgaard (2008) and Wasko and Sharma (2015) as papers that link temperature increases to intensifying rainfall. Most of the papers cited at the end of this sentence deal with temporal precipitation trends (and not necessarily links to temperature). It is important to make that distinction.

The reason I make the above point is the covariate used for non-stationarity is important. The authors don't raise this till the second last of their manuscript citing Mondal and Mujumdar (2015). This needs to come up in the introduction to put this manuscript novelty in context. There are more papers in this space. For example Agilan and Umamahesh (2017) and Ali and Mishra (2017) who argue for temperature to be used as a covariate (and not necessarily time). Indeed Wasko and Sharma (2017) show that temperature is a good covariate when predicting future rainfall. Other work by Agilan and Umamahesh may also be relevant and should be discussed. Finally, I am pretty sure at least one of the Yilmaz papers suggests not much evidence (if any) for using non-stationarity so in the introduction this is not cited correctly (though I note in the discussion it is). To summarise – the literature review needs to be expanded on the above point.

Response: Agreed. We expanded the literature review section in the revision. We add following sentences in the revision:

“For sub-hourly and up to six-hourly extreme precipitation, increases at or above the C-C rate have been found in the Netherlands (Lenderink and van Meijgaard, 2008; Lenderink et al., 2017), Switzerland (Ban et al., 2014), Germany (Berg et al., 2013), the UK (Blenkinsop et al., 2015), the Mediterranean (Drobinski et al., 2016), most of Australia (Wasko and Sharma, 2015, 2017), North America (Shaw et al., 2011) and China (Miao et al., 2016), while in India (Ali and Mishra, 2017) and northern Australia (Hardwick Jones et al., 2010) negative rates have been observed. The extent of urbanization also contributes to extreme regional precipitation through urban heat island effect and aerosol concentration (Dixon and Mote, 2003; Mölders and Olson, 2004; Nihongi et al. 2007; Mohsen and Gough, 2012; Wang et al., 2015). Agilan and Umamahesh (2017) incorporated six physical processes, namely, time, urbanization, local temperature changes, annual global temperature anomaly (as an indicator of global warming), El Niño-Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD) as covariates in the nonstationary GEV models for analyzing extreme precipitation in the city of Hyderabad, India. Their analysis indicated that the local processes, urbanization and local temperature changes are the best covariates for short-duration rainfall, whereas global processes, such as global warming, ENSO cycle and IOD are the best covariates for the long duration rainfall. In their study, however, time was never qualified as the best covariate for modeling local scale extreme rainfall intensity. Singh et al. (2016) performed nonstationary frequency analysis of Indian Summer Monsoon Rainfall extreme (ISMR; defined as

cumulative rainfall over continental India during 1 June to 30 September) and found evidence of significant nonstationarity in ISMR extremes in urbanizing/developing-urban areas (transitioning from rural to urban), as compared to completely urbanized or rural areas. However, their analysis was performed at a spatial resolution of 1° using gridded daily precipitation data obtained from Indian Meteorological Department (IMD). Ali and Mishra (2017) showed that a strong (higher than C-C rate) positive relationship exists between 3-hourly and daily rainfall extremes and dew point, and tropospheric temperature (T850; or the temperature in the upper troposphere at 850 hPa) over 23 urban locations in India. The latter two were subsequently used as covariates for nonstationary design storm estimates. The results indicated an increase in rainfall maxima at a majority of locations assuming nonstationary conditions over stationary atmospheric conditions. In contrast, in other studies, over Melbourne and Victoria, in Australia, Yilmaz et al. (2014; 2017) found superiority of stationary models over nonstationary models. Yilmaz et al. (2014; 2017), considered both nonstationarity in time and large scale climate oscillations affecting Australian rainfall in their analyses. However, most of these previous studies have analyzed changes in expected point estimates of nonstationary versus stationary Design Storm Intensity (hereafter referred as DSI), but have not reported the statistical significance of the difference between two methods of estimates. To our best knowledge, no thorough comparison of stationary vs. nonstationary methods for deriving IDF statistics has been conducted in Southern Ontario.”

Comment 4. Another problem I have is with the paragraph on Page 3 that starts with “secondly” – I don’t think any of the research questions actually address the “secondly” point. Reading page 7 it seems you adopt the GEV and don’t necessarily test this is a better fit than other distribution. This is fine – but the way this paragraph sets up the reader for something else. Either omit the “secondly” paragraph altogether or add another point to the bottom of Page 3 saying you use a GEV and the reason for doing so.

Response: This is indeed a good point. Agreed! We have re-organized this section and moved limitations of GEV in subsection 3.3 (Page 10; lines 3 – 7) in section 3. The choice of the GEV was based on a previous study where various distribution functions were compared in the study area (Switzman et al. 2017).

Comment 5. You introduce the EC data without context – so I had no idea why it was there until I got to page 11.

Response: We appreciate the reviewer’s point. We have introduced few sentences in the introduction section (page 5, lines 1-7) to highlight the rationale behind the inclusion of EC data. We argue that:

“... so far very few studies have reported the difference between the updated versus EC generated IDF, taking into account nonstationarity in design consideration. Simonovic and Peck (2009) compared updated versus EC IDF for the city of London, Ontario and reported EC IDF curves shows a difference of the order of around 20%. However, their analysis was

based on the stationarity assumption of precipitation extremes. Similarly, Coulibaly et al. 2015 have compared EC-IDF with stationary GEV based IDF across southern Ontario, no nonstationary methods were investigated.”

Comment 6. Top of Page 11 reads like a discussion and seems squished between the presentation of results in Figure 5 and 6. You could consider a separate discussion section and reordering of the text.

Response: Agreed. We have moved this part of the text to Discussion and Conclusion section.

Other comments:

Page 2 – Line 16: If you are to introduce an abbreviation (TBRG) it helps to capitalise the first letter in each word before the abbreviation. This happens at several points in the text – I won’t comment on the other occurrences.

Response: Agreed. We have capitalized the first letter in each word before the abbreviation for TBRG and other words in the revised version of the manuscript.

Page 2 – Line 22: “The nonstationary behaviour. . .” I think I would expand this sentence to just state what places/regions the citations have studied. Reason being – in the abstract and following sentences you are referring only to Canada – so when I get to this point I am not sure if you are being Canada specific or not. Maybe this should be the start of a new paragraph and expanded a bit.

Response: Agreed. As suggested we have expanded this sentence to include a list of regions where the citations have studied in Page 4, lines 10 – 18. We also started this in a new paragraph as suggested. We have added following sentences in the revision:

“The nonstationary behavior of rainfall extremes is already being reflected in the increase in frequency or magnitude of such events, resulting in a shift of its distribution [Figure SPM 0.3 in Intergovernmental Panel on Climate Change Special Report on Extremes, IPCC SREX Report: Field, 2012; Fig S1: IPCC AR5 working Group Report, (Stocker et al., 2013)]. For instance, seasonal and annual extreme precipitation events in north-central and eastern US in 2013 (Knutson et al., 2014); extreme rainfall events in the Golden Bay region in New Zealand (Dean et al., 2013); increase in precipitation rate in northern Europe (Yiou and Cattiaux, 2013), successive winter storm events in southern England in 2013/2014 leading to severe winter floods (Schaller et al., 2016), are primarily attributable to intrinsic natural variability and partly to anthropogenic influences.”

Also, in Page 2, lines 1-6, we list the places where increase/decrease in extreme precipitation is linked to C-C scaling. We have added following sentences:

“For sub-hourly and up to six-hourly extreme precipitation, increases at or above the C-C rate have been found in the Netherlands (Lenderink and van Meijgaard, 2008; Lenderink et al., 2017), Switzerland (Ban et al., 2014), Germany (Berg et al., 2013), the UK (Blenkinsop et al., 2015), the Mediterranean (Drobinski et al., 2016), most of Australia (Wasko and Sharma, 2015, 2017), North America (Shaw et al., 2011) and China (Miao et al., 2016), while in India (Ali and Mishra, 2017) and northern Australia (Hardwick Jones et al., 2010) negative rates have been observed.”

Page 2 – Line 26: What result? This sentence doesn’t make sense – maybe some expansion of the sentences here would help.

Response: Agreed. We have revised this sentence as:

“The asymmetric changes in the distribution of extremes owing to climate change have been subsequently validated for winter temperature extremes over the northern hemisphere (Kodra and Ganguly, 2014), and regional short duration precipitation extremes in India and Australia (Mondal and Mujumdar, 2015; Westra and Sisson, 2011)”.

Page 3 – Line 7: Replace “secondly” with “The second drawback of IDF curves is”. You have written too much to have just the word “secondly” here. Stylistically, I don’t think “first”, “second” etc need to be in italics. Particularly at the bottom at Page 3 – if you are that keen on this maybe a bullet point list would be better?

Response: We have revised this section in the current version of the manuscript.

Page 4 - Line 1: Remove “secondly”.

Response: Agreed and incorporated as suggested.

Page 5 – Line 6: The reference to Table S1 doesn’t belong here. I also believe Table S1 belongs in the main text.

Response: Agreed. Table S1 is moved to the main manuscript.

Figure 1 – Are the record lengths for daily or sub-daily? I don’t think the caption says which.

Response: Agreed and we have revised the caption accordingly. This includes hourly, sub-hourly and daily record, which we together termed as short-duration Annual Maxima Precipitation (AMP) record.

Page 5 – Line 26 – “Imputation” isn’t the correct word I don’t think. Infilling maybe?

Response: Agreed and incorporated as suggested.

Page 6 – Line 21 – Stylistically, why don't you just say "Tables S2-S4"? I do feel if you composed the supplementary material in sections you could say section S1 and be done with it.

Response: Agreed and incorporated.

Page 6 – Line 24 – "Figure 2 shows . . ." You are repeating a previous a sentence Section 3.2 – Is the KPSS test in Figure 2?

Response: Agreed. We have included KPSS test in the flowchart.

Page 8 – Line 4 – who else makes this assumption that only the location and scale parameter vary? I know other authors make this assumption so this assumption needs to be put in context of the other work done in this area.

Response: Agreed. We have included list of references that assumes location and scale parameter(s) vary. We have added the following sentences in page 11, line 8 in the revised manuscript:

"For nonstationary model, the shape parameter is assumed as constant throughout. Here it should be noted that for modeling temporal changes in ζ requires long-term observations, which are often not available in practice (Cheng et al., 2014). Hence, following previous studies (Cannon, 2010; Cheng et al., 2014; El Adlouni et al., 2007; Gu et al., 2017) we incorporated time-varying covariates into GEV location (GEV_t-I), and both in location and scale parameters (GEV_t-II) respectively, to describe trends as a function of time".

Page 8 – Line 18 – So I went to the supplementary material as the text recommends and I saw four models fitted for each duration but I wasn't sure which model was which. Could this section in the main text be rewritten (maybe use some sort of list?) to say what models were fitted and clearly state their abbreviation

Response: This section has been revised. Further, as suggested the abbreviations of models are included in page 11, line 12 and in the footnote of Table 4.

Page 8 – Line 26: I disagree. Skewness of a distribution does not indicate a temporal trend. This a good example of a vague sentence with a Figure in brackets (in this case Figure 3) but no mention of what I am meant to get out of looking Figure 3 in reference to this sentence. This happens throughout the text.

Response: Agreed. We have revised the sentence as follows:

"The skewness is a measure of the asymmetry in the AMP distribution. Positive values of skewness indicate that data are skewed to the right."

Further, we have removed such inconsistencies in the revised version of the manuscript.

Figure 3 – your caption says hourly and sub-hourly. The headings in the captions go up to daily. You say you did statistical tests at 5 and 10% but don't say which final significance is presented in the plot. A legend wouldn't go astray . . .

Response: We have revised the Figure 3 caption as suggested.

Figure 4 – is there a particular time used for the non-stationary plots?

Response: This comment was not clear to us. Nevertheless, we have revised the Figure 4 caption to avoid any ambiguity. We have revised our figure caption as:

“DSI estimates of the median (horizontal line within the box plot) and 95% credible intervals for 100-year return periods of stationary versus nonstationary models (a - i). The boxplots indicate the uncertainty in estimated DSI using Bayesian inference.”

Page 10 – Maybe I missed this somewhere but what is the “z-statistic”? Is this the statistical test for the difference between two means?

Response: The reviewer is correct. The z-statistic is the test score for the difference between two means. We clarify this procedure in the Supplementary section of the revised manuscript.

Figure 6 – Should this have a negative scale too? Are there some sites which decrease?

Response: The reviewer is correct. We have added color map for negative scale too in the revised manuscript.

References

- Agilan, V. and Umamahesh, N. V.: What are the best covariates for developing non-stationary rainfall Intensity-Duration-Frequency relationship?, *Adv. Water Resour.*, 101, 11–22, 2017.
- Ali, H. and Mishra, V.: Contrasting response of rainfall extremes to increase in surface air and dewpoint temperatures at urban locations in India, *Sci. Rep.*, 7(1), 1228, doi:10.1038/s41598-017-01306-1, 2017.
- Ban, N., Schmidli, J. and Schär, C.: Evaluation of the convection-resolving regional climate modeling approach in decade-long simulations, *J. Geophys. Res. Atmospheres*, 119(13), 7889–7907, doi:10.1002/2014JD021478, 2014.
- Berg, P., Moseley, C. and Haerter, J. O.: Strong increase in convective precipitation in response to higher temperatures, *Nat. Geosci.*, 6(3), 181–185, doi:10.1038/ngeo1731, 2013.
- Blenkinsop, S., Chan, S. C., Kendon, E. J., Roberts, N. M. and Fowler, H. J.: Temperature influences on intense UK hourly precipitation and dependency on large-scale circulation, *Environ. Res. Lett.*, 10(5), 054021, doi:10.1088/1748-9326/10/5/054021, 2015.
- Cannon, A. J.: A flexible nonlinear modelling framework for nonstationary generalized extreme value analysis in hydroclimatology, *Hydrol. Process.*, 24(6), 673–685, doi:10.1002/hyp.7506, 2010.
- Cheng, L., AghaKouchak, A., Gilleland, E. and Katz, R. W.: Non-stationary extreme value analysis in a changing climate, *Clim. Change*, 127(2), 353–369, 2014.

- Coulibaly, P., Burn, D., Switzman, H., Henderson, J. and Fausto, E.: A comparison of future IDF curves for Southern Ontario, Technical Report, McMaster University, Hamilton., 2015.
- Dean, S. M., Rosier, S., Carey-Smith, T. and Stott, P. A.: The role of climate change in the two-day extreme rainfall in Golden Bay, New Zealand, December 2011, *Bull. Am. Meteorol. Soc.*, 94(9), S61, 2013.
- Dixon, P. G. and Mote, T. L.: Patterns and Causes of Atlanta's Urban Heat Island–Initiated Precipitation, *J. Appl. Meteorol.*, 42(9), 1273–1284, 2003.
- Drobinski, P., Da Silva, N., Panthou, G., Bastin, S., Muller, C., Ahrens, B., Borga, M., Conte, D., Fosser, G., Giorgi, F. and others: Scaling precipitation extremes with temperature in the Mediterranean: past climate assessment and projection in anthropogenic scenarios, *Clim. Dyn.*, 1–21, 2016.
- El Adlouni, S., Ouarda, T. B. M. J., Zhang, X., Roy, R. and Bobée, B.: Generalized maximum likelihood estimators for the nonstationary generalized extreme value model, *Water Resour. Res.*, 43(3), W03410, doi:10.1029/2005WR004545, 2007.
- Gu, X., Zhang, Q., Singh, V. P., Xiao, M. and Cheng, J.: Nonstationarity-based evaluation of flood risk in the Pearl River basin: changing patterns, causes and implications, *Hydrol. Sci. J.*, 62(2), 246–258, 2017.
- Hardwick Jones, R., Westra, S. and Sharma, A.: Observed relationships between extreme sub-daily precipitation, surface temperature, and relative humidity, *Geophys. Res. Lett.*, 37(22), L22805, doi:10.1029/2010GL045081, 2010.
- Knutson, T. R., Zeng, F. and Wittenberg, A. T.: Seasonal and annual mean precipitation extremes occurring during 2013: A US focused analysis, *Bull. Am. Meteorol. Soc.*, 95(9), S19, 2014.
- Kodra, E. and Ganguly, A. R.: Asymmetry of projected increases in extreme temperature distributions, *Sci. Rep.*, 4, 5884, doi:10.1038/srep05884, 2014.
- 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.
- Lenderink, G., Barbero, R., Loriaux, J. M. and Fowler, H. J.: Super-Clausius–Clapeyron Scaling of Extreme Hourly Convective Precipitation and Its Relation to Large-Scale Atmospheric Conditions, *J. Clim.*, 30(15), 6037–6052, doi:10.1175/JCLI-D-16-0808.1, 2017.
- Miao, C., Sun, Q., Borthwick, A. G. L. and Duan, Q.: Linkage Between Hourly Precipitation Events and Atmospheric Temperature Changes over China during the Warm Season, *Sci. Rep.*, 6, srep22543, doi:10.1038/srep22543, 2016.
- Mohsin, T. and W.A. Gough, 2012. Characterization and estimation of Urban Heat Island at Toronto: impact of the choice of rural sites. *Theoretical and Applied Climatology*, 108(1-2): 105-117.
- Mölders, N. and Olson, M. A.: Impact of Urban Effects on Precipitation in High Latitudes, *J. Hydrometeorol.*, 5(3), 409–429, 2004.
- Mondal, A. and Mujumdar, P. P.: Modeling non-stationarity in intensity, duration and frequency of extreme rainfall over India, *J. Hydrol.*, 521, 217–231, doi:10.1016/j.jhydrol.2014.11.071, 2015.
- Schaller, N., Kay, A. L., Lamb, R., Massey, N. R., van Oldenborgh, G. J., Otto, F. E. L., Sparrow, S. N., Vautard, R., Yiou, P., Ashpole, I., Bowery, A., Crooks, S. M., Haustein, K., Huntingford, C., Ingram, W. J., Jones, R. G., Legg, T., Miller, J., Skeggs, J., Wallom, D., Weisheimer, A.,

- Wilson, S., Stott, P. A. and Allen, M. R.: Human influence on climate in the 2014 southern England winter floods and their impacts, *Nat. Clim. Change*, 6(6), 627–634, 2016.
- Shaw, S. B., Royem, A. A. and Riha, S. J.: The relationship between extreme hourly precipitation and surface temperature in different hydroclimatic regions of the United States, *J. Hydrometeorol.*, 12(2), 319–325, 2011.
- Singh, J., Vittal, H., Karmakar, S., Ghosh, S. and Niyogi, D.: Urbanization causes nonstationarity in Indian Summer Monsoon Rainfall extremes, *Geophys. Res. Lett.*, 43(21), 2016.
- Stocker, T. F., Qin, D., Plattner, G. K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V. and Midgley, P. M.: Climate change 2013: the physical science basis. Intergovernmental panel on climate change, working group I Contribution to the IPCC fifth assessment report (AR5), N. Y., 2013.
- Switzman, H., Razavi, T., Traore, S., Coulibaly, P., Burn, D.H., Henderson, J., Fausto, F., Ness, R. Variability of Future Extreme Rainfall Statistics: Comparison of Multiple IDF Projections. *ASCE Journal of Hydrologic Engineering*, 22(10), 1-20, 2017
- Wang, X., Huang, G., Liu, J., Li, Z. and Zhao, S.: Ensemble projections of regional climatic changes over Ontario, Canada, *J. Clim.*, 28(18), 7327–7346, 2015.
- Wasko, C. and Sharma, A.: Continuous rainfall generation for a warmer climate using observed temperature sensitivities, *J. Hydrol.*, 544, 575–590, doi:10.1016/j.jhydrol.2016.12.002, 2017.
- Wasko, C. and Sharma, A.: Steeper temporal distribution of rain intensity at higher temperatures within Australian storms, *Nat. Geosci.*, 8(7), 527–529, doi:10.1038/ngeo2456, 2015.
- Westra, S. and Sisson, S. A.: Detection of non-stationarity in precipitation extremes using a max-stable process model, *J. Hydrol.*, 406(1–2), 119–128, 2011.
- Yilmaz, A. G., Hossain, I. and Perera, B. J. C.: Effect of climate change and variability on extreme rainfall intensity–frequency–duration relationships: a case study of Melbourne, *Hydrol. Earth Syst. Sci.*, 18(10), 4065–4076, 2014.
- Yilmaz, A. G., Imteaz, M. A. and Perera, B. J. C.: Investigation of non-stationarity of extreme rainfalls and spatial variability of rainfall intensity–frequency–duration relationships: a case study of Victoria, Australia, *Int. J. Climatol.*, 37(1), 430–442, doi:10.1002/joc.4716, 2017.
- Yiou, P. and Cattiaux, J.: Contribution of atmospheric circulation to wet north European summer precipitation of 2012, *Bull. Am. Meteorol. Soc.*, 94(9), S39, 2013.

Does Nonstationarity in Rainfall Requires Nonstationary Intensity-Duration-Frequency Curves?

Poulomi Ganguli^{1,2}, Paulin Coulibaly¹

¹Department of Civil Engineering, McMaster Water Resources and Hydrologic Modelling Group, McMaster University, 1280 Main Street West, Hamilton, ON L8S 4L7, Canada

²Present Address: GFZ German Research Centre for Geosciences, Section 5.4 Hydrology, 14473 Potsdam, Germany

Correspondence to: Poulomi Ganguli (poulomi.ganguli@alumnimail.iitkgp.ac.in; gangulip@mcmaster.ca)

Abstract. In Canada, increased risk of flooding due to heavy rainfall has risen in recent decades; most notable recent examples include July 2013 storm in Greater Toronto region and May 2017 flood of Toronto Island. We investigate nonstationarity and trends in the short-duration precipitation extremes in selected urbanized locations in Southern Ontario, Canada, and evaluate the potential of nonstationary Intensity-Duration-Frequency (IDF) curves, which form an input to civil infrastructural design. Despite apparent signals of nonstationarity in precipitation extremes in all locations, the stationary versus nonstationary models do not exhibit any significant differences in the design storm intensity. The signatures of nonstationarity in rainfall extremes do not necessarily imply the use of nonstationary IDFs for design considerations. When comparing the proposed IDFs with current design standards, for return periods (10-year or less) typical for urban drainage design, current design standards require an update up to ~~744~~%, whereas for longer recurrence intervals (50 - 100-year), ideal for critical civil infrastructural design, updates ranging between ~ 2 to ~~3044~~% are suggested. We further emphasize that above findings need re-evaluation in light of climate change projections since intensity and frequency of extreme precipitation are expected to intensify due to global warming.

1 Introduction

Short-duration extreme rainfall events can have devastating consequences, damage to crops and infrastructures, leading to severe societal and economic losses in Canada (~~Canadian Climate Forum~~, 2013; ~~Toronto Region Conservation Authority~~, 2013). In a warming climate, extreme precipitation events are expected to intensify due to moistening of the atmosphere (~~Donat et al., 2016; Fischer and Knutti, 2016; Pendergrass et al., 2015; Prein et al., 2016; Pfahl et al., 2017~~). Using observational record, review of the literature suggests a dependency between mean and extreme precipitation on temperature (O’Gorman, 2015). The increased water-holding capacity of warmer air,

Formatted: Top: 0.6", Bottom: 0.93", Footer distance from edge: 0.51"

Formatted: Superscript

Formatted: French (France)

Field Code Changed

Formatted: English (United States)

Formatted: English (United States)

as governed by the Clausius-Clapeyron (C-C) relation (Lenderink and van Meijgaard, 2008; O’Gorman and Schneider, 2009; Wasko and Sharma, 2015, 2017), intensifies heavy rainfall at a rate of approximately 7-8%°C⁻¹ of warming. On a local scale, for sub-hourly and up to six-hourly extreme precipitation, increases at or above the C-C rate have been found in the Netherlands (Lenderink and van Meijgaard, 2008; Lenderink et al., 2017), Switzerland (Ban et al., 2014), Germany (Berg et al., 2013), the UK (Blenkinsop et al., 2015), the Mediterranean (Drobinski et al., 2016), most of Australia (Wasko and Sharma, 2015, 2017), North America (Shaw et al., 2011) and China (Miao et al., 2016), while in India (Ali and Mishra, 2017) and northern Australia (Hardwick Jones et al., 2010) negative rates have been observed. The extent of urbanization also contributes to extreme regional precipitation through urban heat island effect and aerosol concentration (Dixon and Mote, 2003; Mölders, Mölders and Olson, 2004; Niyogi et al., 2007; Mohsin, Mohsen and Gough, 2012; Wang et al., 2015). Agilan and Umamahesh processes, namely, time, urbanization, local temperature changes, annual global temperature anomaly (as an indicator of global warming), El Niño-Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD) as covariates for the nonstationary extreme precipitation analysis in the city of Hyderabad, India. Their analysis indicated that the local processes, urbanization and local temperature changes are the best covariates for short-duration rainfall, whereas global processes, such as global warming, ENSO cycle and IOD are the best covariates for the long duration rainfall. In their study, time was never qualified as the best covariate for modeling local scale extreme rainfall intensity. Singh et al. (2016) performed nonstationary frequency analysis of Indian Summer Monsoon Rainfall extreme (ISMR; defined as cumulative rainfall over continental India during 1 June to 30 September) and found evidence of significant nonstationarity in ISMR extremes in urbanizing/developing-urban areas (transitioning from rural to urban), as compared to completely urbanized or rural areas. However, their analysis was performed at a spatial resolution of 1° using gridded daily precipitation data obtained from Indian Meteorological Department (IMD). Ali and Mishra (2017) showed that a strong (higher than C-C rate) positive relationship exists between 3-hourly and daily rainfall extremes and dew point, and tropospheric temperature (T850; or the temperature in the upper troposphere at 850 hPa) over 23 urban locations in India. The latter two were subsequently used as covariates for nonstationary design storm estimates. The results indicated an increase in rainfall maxima at a majority of locations assuming nonstationary conditions over stationary atmospheric conditions. In contrast, in another studies, over Melbourne and Victoria, in Australia, Yilmaz et al. (2014; 2017) found superiority of stationary models over nonstationary models. For developing nonstationary models, authors (Yilmaz et al. 2014; 2017), considered both the time dependency and dependency to large scale climate oscillations affecting Australian rainfall. However, most of these previous studies have analyzed changes in expected point

estimates of nonstationary versus stationary Design Storm Intensity (hereafter referred as DSI), but have not reported the statistical significance of the difference between the two methods of estimates. To our best knowledge, no thorough comparison of stationary vs. nonstationary methods for deriving IDF statistics has been conducted in Southern Ontario. For densely populated Southern Ontario, Canada, observations and multiple climate models suggest an increasing trends in regional surface temperature and extreme precipitation in recent decades (Stone et al., 2000; Paixao et al., 2011; Mailhot et al., 2012; De Carolis, 2012; Burn and Taleghani, 2013; Shephard et al., 2014; Deng et al., 2016). A recent study shows an increase in local surface temperature of 3.06 ± 0.18 °C/century in Greater Toronto Area (GTA) since the 1960s (Berkeley Earth, 2017). In July 2013, a single storm event has resulted in 126 mm of rainfall in GTA causing total insured losses of around \$940 million and claimed to be the third-most expensive weather-related event in Canada (CDD, 2015; TRCA, 2013).

Extreme rainfall statistics are often mathematically expressed using the concept of exceedance probability or T -year return period [*i.e.*, $T = 1 / (1 - Fp(P))$, where $Fp(P)$ is the cumulative probability of the underlying distribution], and graphically as a decision relevant metrics in the form of Intensity-Duration-Frequency (IDF) curves (or relations) (ASCE, 2006; CSA, 2010; EC, 2012). These curves are based on a comprehensive statistical analysis of historical rainfall records and widely used for the design and operation of storm-water and sewerage systems, and other engineered hydraulic structures (Coulibaly and Shi, 2005; Durrans and Brown, 2001; Lima et al., 2016; Madsen et al., 2009; Rana et al., 2013; Sandink et al., 2016; Yilmaz et al., 2014a). At given return period and the storm duration, the average design storm intensity (hereafter referred as DSI) is determined from the at-site-IDF relationships. The IDF curves are based on fitting a theoretical probability distribution to short-duration (sub-hourly, hourly- and sub-hourly daily) Annual Maximum Precipitation (AMP). The approach can be locally (at site) or regionally [Svensson and Jones, 2010; Regional Frequency Analysis (RFA) or pooled]. The RFA is used when available record lengths are short or at locations where no observed data are available (Castellarin et al., 2012; Komi et al., 2016). However, various RFA estimation methods have certain drawbacks; for instance, the index flood method is sensitive to the homogeneity assumption and formation of regions; in a Bayesian method of regionalization, the prior distributions of parameters are often not precise enough and do not add precision to the estimates. Komi et al. (2016) summarize the limitations and advantages of some of the widely used RFA techniques. In the present study, the available records across all sites range between 47 and 66 years, which are more than the climatology (often over time periods of 30-years) of a region. Therefore, we employ at-site frequency analysis herein. This also allows a consistent comparison with the Environment Canada (EC) IDFs that have been used in

Field Code Changed

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Field Code Changed

practice in the study area. For Canada, information for preparation of IDF curves and nation-wide IDF curves (EC at ~~Environment Canada~~ (EC) Engineering database (Environment Canada, 2012; http://climate.weather.gc.ca/prods_servs/engineering_e.html), which are produced based on short-duration available rainfall records from ~~the Tipping-Bucket Rain Gauges~~ (TBRG). ~~Nevertheless, the~~ methodology to existing IDF curves has certain drawbacks, ~~first, such as~~, the current IDF curves in Canada are based on the of stationarity, which implies statistical properties of hydroclimatic time series will remain same over the period of time. ~~However,~~ ~~However,~~ impact of urbanization and human-induced climate changes (Field, 2012; Milly et al., Villarini et al., 2009) raises the question whether the stationarity assumption to derive IDF curves is still reliable for urban infrastructural planning (Sarhadi and Soulis, 2017; Cheng and AghaKouchak, 2014; Jakob, 2013; Yilmaz et al., 2014a; Yilmaz and Perera, 2013).

The nonstationary behavior of rainfall extremes is already being reflected in the increase in frequency or magnitude of such events, ~~resulting in a shift of its distribution~~ [Figure SPM 0.3 in Intergovernmental Panel on Climate Change Special Report on Extremes, IPCC SREX Report; Field, 2012; Fig S1: IPCC AR5 working Group Report, (Stocker et al., 2013)]. ~~recently~~ For instance, seasonal and annual extreme precipitation in north-central and eastern US in 2013 (Knutson et al., 2014); extreme rainfall in the Golden Bay region in New Zealand (Dean et al., 2013); increase in summer precipitation rate in northern Europe (Yiou and Cattiaux, 2013); successive winter storm events in southern England in 2013/2014 leading to severe winter floods (Schaller et al., 2016), are primarily attributable to intrinsic natural variability and partly to anthropogenic influences, (Dixon and Mote, 2003; Guo et al., 2006; Mölders and Olson, 2004), resulting in a shift of its distribution [Figure SPM 0.3 in Intergovernmental Panel on for winter temperature extremes over the northern hemisphere (Kodra and Ganguly, 2014), and regional ~~short~~ duration precipitation extremes in India and Australia (Mondal and Mujumdar, 2015; Westra and Sisson, 2011). Two of the recent studies (Deng et al., 2016; Mailhot et al., 2012) analyzed large ensemble of CMIP3 Global Climate Model (GCM) runs and a sub-set of regional climate models that are part of North American Regional Climate Change Assessment Program (NARCCAP) in terms of impact-relevant metrics over Canada. Both studies confirmed a relative increase in intensity and magnitude of rainfall extremes, especially over Southern Ontario. This issue has come to attention in the Guidelines for Canadian Water Resources Practitioner (CSA, 2010), that urges the need for updated IDF calculations: "...climate change will likely result in an increase in the intensity and frequency of extreme precipitation events in most regions in the future. As a result, IDF values will optimally need to be updated more frequently than in the past".

~~Secondly~~Furthermore, so far very few studies have reported the difference between the updated versus EC taking into account nonstationarity in design consideration. Simonovic and Peck (2009) compared updated versus EC IDF_s for the city of London, Ontario and reported EC IDF curves shows a difference of the order of around 20%. However, their analysis was based on the stationarity assumption of precipitation extremes. Similarly, Coulibaly et al. (2015) have compared EC IDF_s with stationary GEV based IDF curves across Southern Ontario, no nonstationary methods were investigated. for the development of IDF curves, a particular family of distribution questions pertained to short-duration precipitation extremes in Southern Ontario, ~~Canada~~, to improve pro-active management of storm-induced urban flooding. First, is there any signature of statistically significant nonstationary trends (~~gradual or slowly monotonic or varying changes~~), change points or regime shifts (occurrence of any abrupt mean/variance of the distribution) in short-duration AMP in densely and moderately populated urbanized locations across Southern Ontario? Second, does ~~signals of nonstationarity in the time series necessitates the use of~~ barring economic consideration and mathematical complexity involved in the design? Third, how can we use this knowledge to assess the credibility of existing EC-generated IDFs in the backdrop of a changing climate? We do not attempt to provide a methodological comparison ~~of~~ EC-generated versus current approach but will focus on differences in estimated DSI values between the updated and EC-IDF. Further, to this end, we test the hypothesis that signatures of nonstationarity in rainfall extremes do not necessitate the use of nonstationary IDFs for design considerations. In general, urban drainage areas have substantial proportions of impervious or semi-impervious land cover, which significantly reduces response time to extreme precipitation and increases the peak flow, resulting into storm-induced floods (Miller et al., 2014). Hence, it is the short-duration precipitation extremes, which controls the design of urban infrastructure (Mishra et al., 2012). Therefore, we focus our analysis on AMP intensity. We select Southern Ontario as a test bed because of the majority of stations with more than 30-years of available rainfall record (Adamowski and Bougadis, 2003; Deng et al., 2016; Shephard et al., 2014). ~~Secondly,~~ have indicated ~~that~~ the region is more vulnerable to climate change than any other part of Canada (Deng et al., 2016; Mailhot et al., 2012). Furthermore, southern Ontario is one of the prominent economic hubs with largest population concentration in Canada (Bourne and Simmons, 2003; Kerr, 1965; Partridge et al., 2007). In this context, we explore a robust statistical framework to evaluate possible nonstationary trends, analyze the frequency of urban precipitation extremes and assess the risk of severe rain-induced urban flooding in Southern Ontario (Table S1).

2.1 Study Area

Southern Ontario is situated on a Southwest-northeast transect, in the southernmost Canadian region, and separated from the United States by lakes Erie, Huron, and Ontario (Jien and Gough, 2013; Figure 1). The study includes nine densely and moderately populated urbanized and anthropogenically altered locations of the Windsor - Kingston corridor in Southern Ontario. The specific sites include (in the order from southwest to northeast): Windsor Airport, London International Airport, Stratford ~~w~~astewater ~~t~~reatment ~~p~~lant (WWTP), Shand Dam in Fergus on the Grand River, Hamilton Airport, Toronto International Airport, Oshawa Water Pollution Control Plant (WPCP), Trenton Airport, and Kingston Pumping Station (Figure 1; Table S1). ~~The last column in Table 1 shows a list of missing years and AMP values for each duration at each station.~~ The Digital Elevation Model (DEM) of the study area was derived from Shuttle Radar Topography Mission (SRTM) 90-m Digital Elevation Database v4.1 (Jarvis et al., 2008), which indicates a shallow slope with ~~a~~ maximum altitude of 670 m above ~~m~~ean ~~s~~ea ~~l~~evel (MSL). The proximity to Great Lakes and topographic effect, especially in areas to the lee of Lakes Erie, Lake Ontario, and the Georgian Bay significantly modifies the climate in the region (Baldwin et al., 2011). Convective showers and thunderstorms primarily modulate the summer rainfall, but fall rainfall is dominated by reduced convective activity and increased lake effect precipitation (Lapen and Hayhoe, 2003). Further, the topographic features and associated westerly winds in the Niagara Escarpment and the Oak Ridge Moraine, play a significant role in modulating rainfall in Toronto region. On the other hand, Windsor metropolitan area, the southernmost urbanized location in the region, has ~~a the~~ humid continental climate, which results in warm summer temperature (30°C or higher) with the greatest precipitation in the spring and summer seasons, and lowest in the fall and winter (Sanderson and Gorski, 1978). Moreover, because of the part of Windsor-Detroit international transborder agglomeration, the extreme summer precipitation in the city of Windsor is primarily influenced by convection and urban heat island effect (Sanderson and Gorski, 1978; De Carolis, 2012).

2.2 Hydrometeorological Data

We identified the station locations (Figure 1b) based on the quality of long-range rainfall records (*e.g.*, ~~more than~~ 30 years or more) and 2011 Census information archived at Statistics Canada website (<https://www12.statcan.gc.ca>). The geographic areas of these locations are extracted from 2011 census digital boundary shape files (<https://www12.statcan.gc.ca/census-recensement/2011/geo/bound-limit/bound-limit-2011-eng.cfm>). The Toronto metropolitan area is the most populous (over 5 million population) and known to be one of the fastest growing population base in Canada (<http://torontosvitalsigns.ca/main-sections/demographics/>), while

Formatted: Font: Not Italic

Formatted: Font: Not Italic

Formatted: Font color: Custom Color(0,0,204)

Fergus is the least populated (population of around 19,000) (Table S1) city. The other cities have population ranges between ~ 500,000 (Hamilton) and 30,000 (Stratford) (Table S1). We obtained AMP observations at particular durations (15-, 30- minutes, 1-, 2-, 6-, 12- and 24-hours) with a few data gaps from Canada's National Climate Data Archive, maintained by the EC (http://climate.weather.gc.ca/prods_servs/documentation_index_e.html). The rainfall records collected from TBRG are thoroughly quality controlled (Shephard et al., 2014). These records have been previously analyzed for the assessment of national extreme rainfall trends (Burn and Taleghani, 2013; Shephard et al., 2014). We consider seven storm durations ranging from 15-, 30- minutes (the typical time of concentration for small urban catchments), and 1-, 2-, 6-, 12-, and 24- hours (the standard time of concentration for larger watersheds) following a previous study (Bougadis and Adamowski, 2006). Except for a few stations (for example, Toronto International Airport and Trenton Airport), for most of the sites, the AMP observation is available either until the year 2007 or before (Table S1). Also, we found missing values in the AMP time series in all sites. We obtained daily and hourly rainfall records from the EC website and Toronto Region Conservation Authority (TRCA).

3 Methods

Figure 2 shows schematics of the overall analysis. In subsequent subsection, we will discuss each of these steps in detail.

We infilled missing values and updated the AMP records by successively disaggregating daily rainfall values to hourly and sub-hourly time steps using a Multiplicative Random Cascade (MRC)-based disaggregation tool. The Cascade-based disaggregation model for continuous rainfall time series was suggested by (Olsson, 1995, 1998). The technique was later successfully implemented by (Güntner et al., 2001; Jebari et al., 2012; Rana et al., 2013) for temporal disaggregation of point rainfall and the development of IDF-curves from short-duration rainfall extremes. Due to freezing weather conditions during winter, most of the TBRGs' are inoperative from early November to late April of the following year. Therefore, when short-duration rainfall records were not available, the AMP values over moving windows of n - durations (n varies from 15-, 30- minutes and 1-,2-,6-,12- and 24-hours) are extracted from May to October (warm season) disaggregated rainfall volumes for remaining years. There are several reasons for selecting warm periods: first, extreme rainfall events mostly occur in the study area during the warm season (Cheng et al., 2010); second, the focus of our analysis is an investigation of extreme rainfall

Formatted: Font: Not Italic

Formatted: Font: Not Italic

related flood risks and development of IDF curves using extreme rainfall statistics. We adjusted the occasional overestimation of extreme values at a higher order cascade step by a statistical post-processing method. We employed ~~Quantile~~ ~~Matching~~ (QM) approach (Li et al., 2010), which claims to outperform other simple bias correction methods and corrects not only the mean but also the variance of the distribution of interest (Gudmundsson et al., 2012; Teutschbein and Seibert, 2012). QM is based on equidistant cumulative probability distribution matching of observed and disaggregated AMP time series using three-parameter Generalized Extreme Value (GEV) distribution. Although like other statistical post-processing technique, QM relies on the stationarity assumption of the time series, in our case, we applied QM to entire time series of both observed and disaggregated AMP, which comes from the same station location (or similar spatial resolution) and a similar period. Therefore, we avoid potential consequences of inflation by quantile mapping (Maraun, 2013) in our analysis. We discuss the implementations of MRC, adjustment of extremes and associated model fits in more details in the Supplementary Information (SI_1; ~~Table S2; Figures S2-S8 and Tables S3-S4~~).

3.2 Detection of Nonstationarity

A series of statistical tests ~~are~~ employed to detect the presence of nonstationary trends and abrupt shifts in the short-duration AMP before frequency analysis. ~~The multiple tests allow a more rigorous and comprehensive assessment of overall trend in the time series since certain tests are complementary to each other (Sadri et al., 2016; Yilmaz et al., 2014, 2017).~~ Figure 2 shows schematics of the overall analysis. Most of the trend and change-point detection algorithms assume observations are mutually independent. The presence of autocorrelation over/underestimates the statistical significance of trend and change-point detection algorithms (Serinaldi and Kilsby, 2016; von Storch and Navarra, 1999). We employed a Ljung-Box test with 20 lags to the short-duration AMP time series of each site to check if they show statistically significant autocorrelation (at 5% and 10% significance levels). For the time series with no serial autocorrelation, we test for trending behavior and nonstationarity. ~~It is also important to note that presence of nonstationarity may not be evaluated merely on the basis of trends or abrupt shifts in the time series, even if the increasing or decreasing trends are statistically significant (Yilmaz et al., 2014).~~ ~~First~~First, we check for a presence of nonstationarity in the time series by employing unit root-based Augmented Dickey-Fuller (ADF; Dickey and Fuller, 1981) test. However, the test may have a low power against stationary near unit root processes (Dritsakis, 2004; Chowdhury and Mavrotas, 2006). Therefore, as a complementary to unit root test, KPSS test (Kwiatkowski et al., 1992) is employed to validate the results of the ADF test. Since both ADF and KPSS tests assume linear regression or normality of the distribution;

Formatted: Space After: 0 pt

Formatted: Font: 11 pt

Formatted: Font: Not Italic

Formatted: Font: Not Italic

alternatively, a log-transformation can convert a possible exponential trend present in the data into a linear trend. Therefore, following previous studies (Gimeno et al., 1999; Van Gelder et al., 2006), AMP time series is log-transformed before applying stationarity tests. However, Yilmaz et al. (2014) did not observe the presence of any significant nonstationarity in extreme rainfall time series in the city of Melbourne even after employing ADF and KPSS tests. Therefore as an alternative, we also employed frequency-based Priestley and Subbarao test [‘PSR’-test; (Priestley and Rao, 1969), which is able to better capture nonlinear dynamical nature of hydrological system than the former two tests (Ali and Mishra, 2017; Hamed and Rao, 1999). The other test we employed is frequency-based Priestley and Subbarao test [‘PSR’-test; (Priestley and Rao, 1969) for nonstationarity. Next, we detected the presence of smooth and abrupt changes in the time series. -The continuous or monotonic trends in short-duration rainfall extremes are identified using non-parametric Mann-Kendall trend statistics with correction for ties (Hamed and Rao, 1998; Reddy and Ganguli, 2013) at 5 and 10% significance levels. In general, the abrupt change (or change point) in the time series occur at a single point in the record and bifurcate the time series into two halves, either with different means, variances, or both dissimilar means and variance together at each part. The change-point in location (or mean) is identified using non-parametric Pettitt’s (Pettitt, 1979) and Mann-Whitney tests (Ross et al., 2011). As indicated by previous studies (Xie et al., 2014; Yue and Wang, 2002), the rank-based nonparametric Mann-Whitney test is not really distribution free and the power of the test is often affected by the properties of sampled data. In practice, when real change point is unknown, often Mann-Whitney test, in general, does not work well, and the Pettitt method can yield plausible change point location along with its statistical significance. However, the significance of the Pettitt test can be obtained using an approximated limiting distribution (Xie et al., 2014; SI2).The shift in scale (or variance) is detected using non-parametric Mood’s Test (Ross et al., 2011; See Figure 2 for details). We applied nonparametric tests due to their robustness to non-normality, which usually appears in the hydroclimatic time series. Further, in order to reduce the number of underlying assumptions required for testing a hypothesis, such as a presence of specific kind of trend or change point in the data, nonparametric tests are employed. For the time series with significant autocorrelation, we employed a Trend-Free Pre-Whitening procedure (TPFW; SI 2) as described in (Yue et al., 2002, 2003) and later modified by (Petrow and Merz, 2009). Then, we applied trend and change point detection algorithms to the pre-whitened AMP extremes.

3.3 Extreme Value Analysis of Sub-daily and Daily Precipitation Extremes

Formatted: Font color: Custom Color(0,0,204)

Field Code Changed

Formatted: Font color: Custom Color(0,0,204)

Formatted: Font color: Custom Color(0,0,204)

Nation-wide EC IDF curves were developed using a particular family of distribution function from the extreme value theory (i.e., Gumbel distribution or Extreme Value type I, hereafter referred as EVI). However, EVI distribution has certain limitations, such that it is a non-heavy tailed distribution and characterized by constant skewness and kurtosis coefficients (Markose and Alentorn, 2005; Pinheiro and Ferrari, 2016). However, the short-duration AMP intensities often exhibit fat-tail behavior and have left asymmetries (skewed to the left relative to standard normal distribution). In fact, a few studies in the past have shown that EVI fits poorly to the historical rainfall extremes (Burn and Taleghani, 2013; Coulibaly et al., 2015). Therefore, in the present study, we perform frequency analysis of extreme precipitation using GEV distribution. The choice of GEV distribution was based on a previous studies where various distribution functions were compared in the study area (Coulibaly et al. 2015; Switzman et al., 2017). Next, we perform frequency analysis of AMP intensity using GEV distribution. GEV fitted to block or AM time series (Cheng and AghaKouchak, 2014; Katz et al., 2002; Katz and Brown, 1992). The GEV distribution is characterized by contains three parameters, the location, the scale and the shape of the which describes the center of the distribution, the deviation around the mean and the shape or the tail of the distribution (Katz et al., 2002; Katz and Brown, 1992). The cumulative distribution function of stationary (time-invariant) GEV model is given by (Coles et al., 2001):

$$G(z) = \begin{cases} \exp\left\{-\left[1 + \zeta \left(\frac{z - \mu}{\sigma}\right)_+\right]^{-1/\zeta}\right\} & \text{if } \zeta \neq 0 \\ \exp\left\{-\exp\left(-\frac{z - \mu}{\sigma}\right)\right\} & \text{if } \zeta \rightarrow 0 \end{cases} \quad (3.1)$$

Where, $y_+ = \max\{y, 0\}$, and

$z \in [(\mu - \sigma)/\zeta, +\infty)$ when $\zeta > 0$; $z \in (-\infty, (\mu - \sigma)/\zeta]$ when $\zeta < 0$; and $z \in (-\infty, +\infty)$ when $\zeta = 0$

μ is a location parameter, σ is a scale parameter and ζ is a shape parameter determining the heaviness of the tail. The shape parameter ζ , determines the higher moments of the density function and also the skew in the probability mass. The '+' sign indicates positive part of the argument. The Eq. (3.1) encompasses three types of DFs based on the sign of the shape parameter, ζ : (i) the Fréchet, with a finite lower bound of $(\mu - \sigma)/\zeta$ and an

Formatted: Font color: Custom Color(0,0,204)

Formatted: Right

Formatted: Font color: Custom Color(0,0,204)

Field Code Changed

Formatted: Font color: Custom Color(0,0,204)

Field Code Changed

Formatted: Font: 11 pt, Font color: Custom Color(0,0,204)

Formatted: Font: 11 pt, Font color: Custom Color(0,0,204)

Field Code Changed

Formatted: Font: 11 pt, Font color: Custom Color(0,0,204)

Field Code Changed

Field Code Changed

Formatted: Font: 11 pt, Font color: Custom Color(0,0,204)

Field Code Changed

Formatted: Font: 11 pt, Font color: Custom Color(0,0,204)

Formatted: Font: 11 pt, Font color: Custom Color(0,0,204)

Field Code Changed

unbounded, heavy upper tail, ($\zeta > 0$), (ii) the Weibull, unbounded below and with a finite upper bound of $(\mu - \sigma)/\zeta$ ($\zeta < 0$) and (iii) the Gumbel, unbounded below and above with a light upper tail $\zeta = 0$, formally obtained by taking limit as $\zeta \rightarrow 0$. The shape parameter zero indicates Gumbel distribution, which is described by an unbounded light tailed distribution and the tail decreases rapidly following an exponential decay. On the other hand, the positive shape parameter denotes Fréchet distribution is a heavy-tailed distribution, and the tail drops relatively slowly following a polynomial decay (Towler et al., 2010). On the other hand, the negative shape parameter represents a Weibull distribution, which is a bounded distribution. Here we compare the performance of both stationary and nonstationary form of GEV distribution. For stationary model, we estimate parameters by using maximizing the log likelihood function of the distribution (i.e., Maximum Likelihood or ML-based) and Bayesian Inference (BI) coupled with Differential Evaluation Markov Chain (DE-MC) Monte Carlo (MC) simulation as suggested by (Cheng et al., 2014; Cheng and AghaKouchak, 2014). For nonstationary model, the shape parameter is assumed as constant throughout. Here it should be noted that for modeling temporal changes in ζ requires long-term observations, which are often not available in practice (Cheng et al., 2014). However, following previous studies (Cannon, 2010; Cheng et al., 2014; El Adlouni et al., 2007; Gu et al., 2017) we incorporate time-varying covariates into GEV location (GEV-I), and both in location and scale parameters (GEV-II) respectively, to describe trends as a function of time (in years):

$$\mu(t) = \mu_1 t + \mu_0 \quad (3.2)$$

$$\sigma(t) = \sigma_1 t + \sigma_0 \quad (3.3)$$

Since the scale parameter must be positive throughout, it is often modeled using a log link function (Gilleland and Katz, 2014)

$$\ln \sigma(t) = \sigma_1 t + \sigma_0 \Rightarrow \sigma(t) = \exp(\sigma_1 t + \sigma_0) \quad (3.4)$$

Where t is the time (in years), $\lambda = \{\mu_1, \mu_0, \sigma_1, \sigma_0, \zeta\}$ are the parameters.

(See SI 3 for details). Then we estimate parameters of the nonstationary GEV distribution by integrating BI combined with DE-MC simulation. For AMP intensity, we derive the time variant parameter(s) from the 50th (the median or the mid-point of the distribution) percentiles of the DE-MC sampled parameter(s). We obtain the associated 95% confidence credible intervals (the bounds) from the 2.5th and 97.5th percentiles.

Formatted: Font: 11 pt, Font color: Custom Color(0,0,204))

Field Code Changed

Field Code Changed

Formatted: Font: 11 pt, Font color: Custom Color(0,0,204))

Field Code Changed

Formatted

Formatted

Field Code Changed

Field Code Changed

Formatted

Formatted: Font color: Custom Color(0,0,204))

Field Code Changed

Formatted: Font: 11 pt

Field Code Changed

Formatted: Font color: Custom Color(0,0,204))

Formatted: Space Before: 0 pt, After: 0 pt

Formatted: Font color: Custom Color(0,0,204))

Formatted: Font color: Custom Color(0,0,204))

Formatted: Font color: Custom Color(0,0,204))

Formatted: Superscript

Formatted: Font color: Custom Color(0,0,204))

Formatted: Font color: Custom Color(0,0,204))

Formatted: Superscript

Formatted: Font color: Custom Color(0,0,204))

Formatted: Font color: Custom Color(0,0,204))

the ~~simulated~~ ~~posterior samples~~ (See SI 3 for details). We perform the calculations following (Cheng and AghaKouchak, 2014) using an MATLAB-based software package, Nonstationary Extreme Value Analysis (NEVA, Version 2.0). The ~~Bayes factor followed by~~ Akaike information criterion (AIC) with a small sample correction (AIC_c) ~~is~~are used to identify the best model. ~~The AIC_c which~~ claims to avoid overfitting the data as compared to traditional AIC (Burnham and Anderson, 2004; Hurvich and Tsai, 1995). ~~Here we assess model fitness based on a least square sense of AIC statistics considering maximum deviation between empirical (obtained from rank-based plotting position formula) and modelled cumulative distribution (CDF) (Dawson et al., 2007; Hu, 2007; Karmakar and Simonovic, 2007, 2007). For calculation of AIC_c statistics, we consider median of the DE-MC sampled parameters, which can be considered as an average or expected value of risk in the historical observation.~~

5 Besides this, we ~~also~~ assess the performance of the models using ~~p~~Probability-~~p~~Probability (PP) plots. ~~Following a previous study (El Adlouni et al., 2007), we select the model with fewer parameters as the best model when two models have comparable performances. For example, we chose GEV I as the best fit for 15 minute and 12 hour durations rainfall extremes at Hamilton and Trenton Airport respectively (Table S17 and Table S22).~~The derived model parameters are then utilized to obtain DSI using the concept of a T -year return period. We discuss the methods to estimate DSI and T -year return periods using stationary and nonstationary methods in detail in section SI 54. To test (statistically) significant difference in the estimated DSI from the best-selected stationary versus nonstationary models, we calculate standardized z-statistics for selected return periods (Madsen et al., 2009; Mikkelsen et al., 2005). We applied the 2-sided option with ~~510% and 10%~~ significance levels to assess the statistical significance of the test statistics (See SI 6-5 for details). Finally, we compared the DSI obtained from

10 nonstationary and stationary models with existing EC-generated DSI estimates.

15

20

The extreme rainfall statistics show high standard deviation with positive skew ~~behavior~~behavior (Tables S5-2 and skewness is a measure of the asymmetry in the AMP distribution. Positive values of skewness indicate ~~that data are skewed to the right~~a shift towards an increase in the intensity of extreme events. The skewness of sub-hourly highest being 30-min AMP record at Hamilton and least being at Oshawa respectively (Table 2). Likewise, for hourly extremes, the skewness ranges between 0.54 and 2.54, with least being 1-hour AMP at Oshawa and highest is 1-hour AMP at Hamilton respectively (Table 3). ~~For example, extremes at London International Airport, Trenton S5-2 and S63~~, which indicates ~~the data have a distinct peak near the mean, which decline rapidly, and have heavy~~

25

30 We find the presence of statistically significant autocorrelation in the AMP time series of Toronto International

Formatted: Font color: Custom Color(0,0,204)

Formatted: Font: Italic

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: Italic

Formatted: Font color: Custom Color(0,0,204)

Airport, Hamilton Airport, and Fergus Shand Dam (Table S7.1, Table S8.1, and Table S15SI 2). We apply with a significant autocorrelation (Table S7S4.1, Table S8S5.1, and Table S4S5S12). However, two successive to correct the effect of autocorrelation in 12- and 24-hour duration extremes in Shand Dam. Hence we exclude those two time series from frequency analysis (Table S4S5S12). The ADF-test for nonstationarity is statistically significant in all durations, as indicated by the higher p -values. As a complementary to ADF test, we also employed KPSS and PSR tests (Figure 2; Tables S7—S15SI 2) to check significant nonstationarity. Figure 3 shows the spatial trends, change points and nonstationarity in short-duration rainfall extremes. We find co-occurrences of change points and nonstationarities in extremes at multiple locations (Figure 3). In general, the three sites in the Northeast, the Oshawa WPCP, Trenton Airport and Kingston P. Station show evidence of statistically significant upshifts and nonstationarities in the time series, whereas the rest of the sites in the Southwest exhibit downshifts and statistically significant nonstationarities (Figure 3). For 2-hour and beyond durations, London International Airport shows a presence of statistically significant downward trends with change points. An increase (decrease) in mean precipitation imply an increase in heavy precipitation and vice-versa. Further, it could also alter the shape of the right-hand tail, changing overall asymmetry in the distribution (Fig. S1), and hence affecting the nature of extremes (Stocker et al., 2013). Furthermore, the presence of opposite signs of trends within a proximity of sites are prominent in all durations, for example, except for 1-hour duration, extremes in all durations at Toronto International Airport and Oshawa WPCP show downwards and upward shifts respectively. Our findings confirm the other study (Burn and Taleghani, 2013), where authors report a lack of spatial structures and presence of different trends within a close vicinity of stations. Further, we find statistically significant monotonic increase and abrupt step changes, both in mean and variance in Oshawa and Trenton respectively (Table S6 and S10), whereas London shows (significant) decrease (Table S9) from the duration of 6-hour and more. Windsor, Kingston and Stratford show (significant) step changes as confirmed by Mann-Whitney and Mood Tests (Tables S7, S8 and S11). On the other hand, Toronto, Hamilton and Fergus Shand Dam (Tables S4, 4.1; S5, 5.1; S12) do not exhibit any statistically significant gradual or abrupt changes in the AMP time series. The ADF tests show the presence of nonstationarity in all durations across the sites. To further validate results of ADF test, KPSS and PSR tests are employed. The KPSS test detects the presence of nonstationarity at 3 out of 9 sites for 24-hour rainfall extreme at 5% significance level, whereas the results of PSR test indicate nonstationarity across 5 sites in 24-hour rainfall extremes. While KPSS test alone could not detect the presence of nonstationarity in any of the extreme series in Oshawa and Stratford respectively, the results of PSR test did not indicate nonstationarity in any of the short-duration rainfall extreme in Windsor. Both of these tests taken together detect the presence of nonstationarity in

Field Code Changed

[rainfall extremes across 6 out of 9 sites](#). We find even if trends in individual sites may not deem significant, the magnitude of trends (as measured by slope per decade, Tables ~~S7-S4 – S15~~[S12 in S12](#)) is never zero in any of the

A weak trend can also have a significant impact on the results of probability analysis (Porporato and Ridolfi, 1998).

Hence even if precipitation extremes often exhibit statistically insignificant trends in few durations, we assess the

5 performance of both nonstationary and stationary models in all sites (~~Tables S16 – S24~~). [Tables 4 – 7 lists performance of nonstationary versus stationary models for selected airport locations, whereas Tables S13 - S17 presents results of the distribution fit for the remaining stations.](#) Barring a few exceptions, the shape parameters in

most of the models range between -0.30 and +0.3, which is an acceptable range of GEV shape parameter as shown in an earlier study (Martins et al., 2000). Our results corroborate well with recent research (Papalexiou et al., 2013;

10 Wilson and Toumi, 2005), which showed that distribution with fat tails (with GEV shape parameter, $\xi - \zeta < 0$) fits

better for the precipitation extremes. [The nonstationary models are selected employing Bayes-factor and minimum AICc criterion. For example, for the 6-hour duration at Hamilton Airport, the nonstationary GEV_{-I} \(nonstationary](#)

[model with time-varying GEV location\) model performed the best as shown by both test metrics. However, in certain cases, nonstationary models does not pass Bayes-factor test. In such cases, we select the best nonstationary](#)

15 [model \(i.e., between GEV_{-I} and GEV_{-II}\) following AICc test statistics. Here, it should be noted that the objective is to compare the design storm obtained from stationary versus best nonstationary model and not to analyze the best distribution between them.](#) As a measure of uncertainty, we also report the 95% ~~confidence-credible~~

interval of design rainfall quantiles at 100-year return period as a ratio between the upper and the lower bounds, which ranges between the factor of 1.2-to-1 and ~~43.09~~

20 (Figure S9) closely follows the spatial pattern as indicated in the trend map (Figure 3). For example, Trenton Airport, which showed significant upward trends with change points and nonstationarity, is better modeled by the nonstationary GEV distributions for most of the durations. Likewise, ~~except a few cases,~~ we find that [in few cases,](#)

GEV II fits best if the time series exhibit (significant) evidence of a ~~change point in variance~~[nonstationarity as detected by PSR-test statistics, for example, 155-min and 12-hr extremes in Trenton](#)

[London and Toronto International Airport \(Tables ~~S224~~ and 7\) and Kingston P. station \(Table S20\) respectively.](#) However, in many

25 cases, the performance of nonstationary models are often comparable and even superseded by their stationary counterparts (~~Figure S9, Tables S16 – S24~~[SI 3](#)). In fact, the scatter of data points in the *PP*-plots (Figures S10 – ~~S13~~[S12](#)) suggests a close resembles between stationary and nonstationary models across all durations. Figure 4

shows the relation between DSI and durations (ranges between 15-min and 24-hr) for 100-year return periods

Formatted: Font color: Custom Color(0,0,204)

Field Code Changed

Formatted: Font: Italic

Formatted: Font: Italic, Subscript

Formatted: Font color: Custom Color(0,0,204)

Formatted: Font: Not Bold

Formatted: Font color: Custom Color(0,0,204)

Formatted: Font color: Custom Color(0,0,204)

Formatted: Font color: Blue

Formatted: Font color: Blue

5 estimated by stationary versus nonstationary GEV distributions. ~~While the~~ interquartile range of the boxplot uncertainty in estimated rainfall quantiles ~~obtained from the~~ Bayesian ~~Inference~~ analysis, the black circles simulated by the nonstationary model is found to be relatively narrower as compared to the one simulated by the stationary model for most of the sites (Figure 4), indicating less uncertainty in the estimated quantiles. For return periods of less than 100 years, such as for 10- and 5-year, the DSI from stationary versus nonstationary models, show subtle differences (Figures ~~S14~~ ~~S13~~ – ~~S15~~ ~~S14~~).

10 Figure 5 displays the differences in DSI obtained using the best performing nonstationary model relative to the ~~best selected~~ stationary models using ~~percentage changes and~~ z-statistics for different durations and return periods. ~~While percentage change indicates a magnitude of change, the z-statistics show statistical significance of the relative difference in estimated DSI using the two different methods. The percentage differences at 2- and 10-year return period are small relative to larger return periods. At 100-year return period, a maximum positive difference of around 44% is observed at 12-hour storm duration in Toronto International Airport (Table S18.1). The standardized z-statistics show positive (negative) values indicating an increase (decrease) in DSI values assuming nonstationarity in return period estimates against its stationarity counterparts. However, a comparison between T-year event estimates from both models indicates statistically indistinguishable differences in rainfall intensity. We find for all return periods and durations, z-statistics ranges between -1 and +1 for all nine sites (SI 5). Nonetheless, extreme precipitation intensity shows either positive or negative (statistically insignificant) changes in signs. The difference between DSI shows a decrease, at 1- and 2-hour storm duration in Toronto, 6-hour storm duration in London, and 15-min and 12-hour storm duration at Trenton Airport for 50- and 100-year return periods (Figure 5, SI 5). In contrast, Toronto, Windsor and London International Airport shows an increase in DSI value at 15-min duration (Figure 5; SI 5), although the increase is statistically insignificant. Further, we note, except 6- and 12-hour storm duration, the performance statistics show a comparable and in few cases even better performance of the stationary versus nonstationary GEV models across most of the sites (SI 3). At 2- and 10-year return periods, which is typical for most urban drainage planning, the differences are close to zero (Figure 5, Tables S27 and S28 in SI 5) for most of the durations.~~

20

25

~~We find for all return periods and durations, z-statistics ranges between -1 and +1 for all nine sites. Nonetheless, extreme precipitation intensity shows either positive or negative (statistically insignificant) changes in signs. The difference between DSI shows a decrease, especially between 15-min and 2-hour for Hamilton, Windsor, London International Airport, and Shand Dam for 50- and 100-year return periods~~

Formatted: Font color: Auto

(Figure 5, Tables S25–S26). These findings are in agreement with Figure 3, in which sites show a decrease in periods taking into account both stationary and nonstationary condition. The median and associated lower and upper bounds of the ratio of regional nonstationary updated versus EC-generated T -year event estimates can be analogous to most likely, minimum and maximum plausible scenarios. While the positive value of the ratio indicates a required increase in DSI, the negative value indicates a decrease in DSI estimate. Considering nonstationarity, for $T = 10$ -year return period (Figure 6, SI 4), the ratio of updated versus old estimates of DSI the order of $\sim 1.01 - 1.308$, for Oshawa WPCP, Kingston P. station, Stratford WWTP, Windsor and Trenton period, except Oshawa and Windsor, a majority of the sites show decrease in DSI for most of the storm durations (Figure 6, middle row). In contrast, the increase in the estimated ratio being is more pronounced at 50- and 100-year return periods, and which are in the order of $\sim 1.034 - 1.580$ (Figure 6, Table S29.1 – Table S37.2 SI 4). While for Toronto International Airport and Hamilton Airport, we find no increase in the short-duration rainfall extremes of less than 1-hour and 50-year return period considering nonstationary condition, the increase is more pronounced for longer durations and larger return periods (12 and 24-hour, and 50- and 100-year return period, Figure 6, bottom panel SI 4). For longer recurrence intervals, while the heat maps of minimum bounds and the most likely scenario show a lesser smaller number of stations and durations to reach a ratio of 1.5 and beyond, the maximum bounds suggest a sharp increase in the ratio across all durations and locations. Further, for return periods more than of 50-year and more, the increase in the ratio is more prominent in the upper bound of the stationary models (Figure S167, left two columns) as compared to the nonstationary models. The resulting increase in T -year event estimates is because of the relatively wider confidence interval of estimated DSIs in stationary models than that of the nonstationary models (Figures 64; S14 – S15 SI 3). In general, for larger return periods, our analysis reveals, the increase in the ratio of updated versus EC-generated rainfall intensity is more prominent in sites with statistically significant signatures of nonstationarity, upward trends, and change points. For example, the updated DSIs of Oshawa WPCP (Table S31.1), Windsor and Trenton Airport (Table S35.1) shows an increase in the ratio for most of the durations and return periods as compared to the EC-generated DSI values (SI 4). On the other hand, except for the 100-year return period events, the hourly precipitation extremes in London International Airport, in general, show a decrease in the ratio (Table S34.23.1 – 23.2) across all return period, which is predominantly due to the presence of significant downward trends with change points in the time series.

Based on the study results and in anticipation of stakeholders' participation in adaptive management, we present updated nonstationary IDF's for four selected the nine urbanized and semi-urbanized locations across Southern

Formatted: Font color: Custom Color(RGB(0,0,204))

Ontario (Figure 78). In order to distinguish between stationary and nonstationary method of analysis, we also updated IDF assuming stationary condition relative to EC IDF in the same plot (in top panel). The comparison of remaining sites is presented in Figure S15. Thus ~~It we also makes~~ the first attempt to compare the results of EC-generated IDFs considering both nonstationary and stationary conditions, which are the part of contemporary Design Standards and widely used by the stakeholders and practitioners. Overall, the updated IDFs closely follow the pattern of trends analogous to EC-generated IDFs, except for the 100-year return period. The difference is more pronounced considering nonstationary condition, especially at Toronto International Airport (Figure 8), Oshawa WPCP and Stratford WWTP (Figure S15). At longer durations and higher return periods, stations in metropolitan areas (such as Toronto International Airport, Hamilton Airport, Oshawa WPCP and Windsor Airport) show large differences in DSIs, whereas moderately populated locations such as Kingston P. station and Fergus Shand dam show relatively smaller changes. Considering, nonstationary condition, the maximum increase in Furgas Shand dam is noted as 18.7% for the 2-hour storm duration and 100-year return period, whereas an increase of around 44.5% is shown for 12-hour storm duration at Toronto Airport. Stratford WWTP, shows substantial differences in range of ~ 2 – 40% in the T-year event estimates (Table S29.1 – S37.2, S14). Meanwhile, for $T = 10$ -year return increase is in the order of ~~~1.4 to 7.2%~~ ~~1.2 – 11%~~ across several stations (*i.e.*, Oshawa WPCP, Windsor Airport, more, the required increase ranges between ~~~ 24.78 – 3044.5%~~. We find the largest increase is for the 12-hour extreme in Toronto International Airport (~~~ 320 – 3444.5%~~; Table S29.1), followed by 12-hour extreme at WWTP (~~~ 237 – 346%~~; Table S36.25.1). However, considering stationarity condition, for $T = 50$ -year and more, required increase ranges between ~ 1.4 – 26%. It should be noted that above results are based on an average risk approach for extreme value analysis by considering median of the sampled parameters in the historical observation and not considering the overall risk envelope (*i.e.*, minimum and maximum bounds). -In summary, our findings confirm that updates in the order of ~~~ 2 – 3044%~~ are required based on locations and return periods to mitigate the risk of precipitation induced urban flooding irrespective of the choice of methods used in the IDF estimation (Table 4). The results are consistent with (Simonovic and Peck, 2009), in which authors recommends an average update about 21%, with a difference, ranges between ~ 11 – 35% for the updated versus EC-IDF in London Metropolitan area. However, they assumed stationarity condition in the method of to develop at-site IDF estimation. The above need to update existing EC IDFs, which are generated using Gumbel probability distributions and do not fit the data well.

The increase could also indicate a tendency towards an increase in mean precipitation and (or) a shift in the distribution, affecting its tail behavior. However, a few caveat remains, for example, a critical question could be:

[does](#) an increase in DSI ~~is~~ potentially linked towards more frequent and more intense precipitation extremes or is an artifact of the new dataset in the update process? It is worthwhile to note that results shown here are manifestations of present-day climate using ground-based hydrometeorological observations and the specific insights are nonetheless subject to the quality of available rainfall records. -It remains an open-ended question to what extent we credibly develop IDF in a changing climate (Coulibaly et al., 2015) since there is no uniformly accepted method of generating IDF information and related projection uncertainty in light of climate change. [More](#) [In general, highlighting advantages and limitations of nonstationary versus the stationary methods of analyses \(Koutsoyiannis and Montanari, 2015; Montanari and Koutsoyiannis, 2014; Serinaldi and Kilsby, 2015\) is beyond the scope of the current study.](#) Further, several studies (Deng et al., 2016; Kunkel, 2003) [has](#)ve indicated an increase in frequency and magnitude of short-duration rainfall extremes in Southern Ontario due to global warming. Research towards this direction is currently underway for regional preparedness and to develop comprehensive adaptation strategies.

Formatted: Font color: Custom Color(0,0,204)

[4 Discussions and Conclusions](#)

This paper has sought to assess signatures of nonstationarity in densely and moderately populated urbanized locations in Southern Ontario, which is one of the major economic hubs in Canada. We update short-duration rainfall extremes with latest available ground-based observations and present a comprehensive analysis to evaluate nonstationary versus the stationary method of IDF estimation. This analysis yields two principal findings. First, despite signatures of (statistically) significant nonstationarity and trends in extremes in most of the sites, the changes in design storm intensity remain statistically indistinguishable using stationary versus nonstationary methods. [These findings pose an important question: does the presence of nonstationarities in rainfall extremes require the design of nonstationary IDF curves? We ~~argue speculate~~ that although it is crucial to recognize nonstationarity in precipitation extremes, the stationary form of IDFs can still represent the extreme rainfall statistics for the present-day climate over Southern Ontario region. Our results are consistent with \(Yilmaz et al., 2014; Yilmaz and Perera, 2013\), in which authors found despite the presence of \(statistically\) significant trends in rainfall extremes; nonstationary GEV models did not show any additional advantages over the stationary models. As supported by the previous study \(Singh et al., 2016\), we attribute that the little or no changes in extreme rainfall statistics in the urbanized setting is due to the stabilization of urban development leading to no substantial variations in the land use pattern. Hence, no significant changes in synoptic scale circulations, which in turn affect space-time pattern in rainfall extremes \(Moglen and Schwartz, 2006\).](#) Second, comparison of at-site T -year event estimates of

Formatted: Font color: Custom Color(0,0,204)

Formatted: Space Before: 6 pt, After: 0 pt

updated versus EC-generated IDF curves shows ~~for~~at $T = 10$ -year, the return period commonly used for urban drainage design, current design standards require updates up to ~~11~~7% to mitigate the risk of urban flooding. Meanwhile, for longer recurrence interval ($T = 50$ -year or more), typical for critical civil infrastructural design, comparison of updated versus EC-generated IDF curves shows a difference ~~that~~ ranges between 2% and ~~3044~~% based on 5 locations.

Preliminary investigations based on regional and global climate model simulations in the study area confirm a considerable uncertainty in the projection of short-duration and high-intensity extreme rain events (Coulibaly et al., 2015). While short-duration precipitation extremes are typically controlled by synoptic scale moisture convergence (Ruiz-Villanueva et al., 2012; Westra and Sisson, 2011), the daily extremes are often modulated by large-scale circulation patterns and local orographic factors (Carvalho et al., 2002; Gershunov and Barnett, 1998; Trenberth, 1999). Further, the role of natural variability and multidecadal modes of sea-surface temperature (SST) in modulating Canadian extreme rainfall intensity have already been shown in the past (Gan et al., 2007; Shabbar et al., 1997). The future research should include two aspects. First, investigation of physical drivers (such as temperature, decadal and multidecadal modes of SST) in influencing short-duration rainfall extremes. Second, the inclusion of these covariates in nonstationary IDF development (Mondal and Mujumdar, 2015; Yilmaz et al., 2014b). Given that these findings are for the ~~current period climate~~current period (e.g., historical extreme rainfall time series), we recommend a careful extrapolation of findings with regards to future climate projections, it is also recommended to explore hypothesis driven studies for the future time periods using high resolution climate model simulations in which frequency and magnitude of extreme rainfall are expected to intensify (Mailhot et al., 2012; Deng et al., 2016; Fischer and Knutti, 2016; Prein et al., 2016; Pfahl et al., 2017; Switzman et al., 2017). Further work should consider nonstationary methods for deriving future IDF curves in Southern Ontario.

P. Ganguli and P. Coulibaly designed the experiment. P. Ganguli carried out the experiment. P. Ganguli prepared the manuscript with contributions from P. Coulibaly.

Acknowledgments

The annual maximum rainfall data used in this study is downloaded from Environment Canada website: ftp://ftp.tor.ec.gc.ca/Pub/Engineering_Climate_Dataset/IDF/. Hourly and daily rainfall data are obtained from

Formatted: Font color: Custom Color(0,0,204)

Formatted: Font color: Custom Color(0,0,204)

Formatted: Space Before: 0 pt, After: 0 pt

Field Code Changed

Formatted: Font color: Custom Color(0,0,204)

Formatted: Font color: Custom Color(0,0,204)

Formatted: Font: Not Bold, Font color: Custom Color(0,0,204)

Formatted: Font color: Custom Color(0,0,204)

Formatted: Font color: Custom Color(0,0,204)

Toronto and Region Conservation Authority (TRCA; <https://trca.ca/>) and Environment Canada Historical Climate Data Archive (<http://climate.weather.gc.ca/>). ~~The authors would like to acknowledge financial support from the Natural Science and Engineering Research Council (NSERC) of Canada and the NSERC Canadian FloodNet. This work was supported by the Natural Science and Engineering Research Council (NSERC) Canadian FloodNet~~
5 ([Grant number: NETGP 451456](#)). The first author of the manuscript would like to thank Dr. Jonas Olsson of Swedish Meteorological and Hydrological Institute (SMHI), Norrköping, Sweden for sharing MATLAB-based random cascade disaggregation tools and implementation details through email. While change point and nonstationarity analyses were conducted in the statistical software R version 3.30 with add-on packages “trend”, “fractal” and “cpn”, the remaining analyses were performed in a MATLAB platform (MATLAB R2016a). The
10 nonstationary GEV analyses were performed using MATLAB-based NEVA toolbox, available at the University of California, Irvine website: <http://amir.eng.uci.edu/neva.php> (as accessed on May 2016). ~~The work was completed and communicated when the first author was a postdoctoral research fellow at McMaster Water Resources and Hydrologic Modeling Lab, McMaster University, Canada.~~

References

- 15 Adamowski, K., and J. Bougadis: Detection of trends in annual extreme rainfall, *Hydrol. Process.*, 17(18), 3547–3560, American Society of Civil Engineers (ASCE): *Standard Guidelines for the Design of Urban Stormwater Systems, Standard Guidelines for Installation of Urban Stormwater Systems, and Standard Guidelines for the Operation and Maintenance of Urban Stormwater Systems*, ASCE/EWRI 45-05, 46-05, and 47-05., American Society of Civil Engineers, Reston, VA, 2006.
- 20 Baldwin, D. J., J. R. Desloges, and L. E. Band: 2 Physical Geography of Ontario, *Ecol. Manag. Terr. Landsc. Patterns Process. For. Landsc. Ont.*, 12, 2011.
- ~~Agilan, V. and Umamahesh, N. V.: What are the best covariates for developing non-stationary rainfall Intensity-Duration-Frequency relationship?, *Adv. Water Resour.*, 101, 11–22, 2017.~~
- Ali, H. and Mishra, V.: Contrasting response of rainfall extremes to increase in surface air and dewpoint temperatures at urban
25 locations in India, *Sci. Rep.*, 7(1), 1228, doi:10.1038/s41598-017-01306-1, 2017.
- ~~[ASCE: Standard Guidelines for the Design of Urban Stormwater Systems, Standard Guidelines for Installation of Urban Stormwater Systems, and Standard Guidelines for the Operation and Maintenance of Urban Stormwater Systems, ASCE/EWRI 45-05, 46-05, and 47-05., American Society of Civil Engineers, Reston, VA. \(Accessed 9 December 2016\). 2006.](#)~~
- 30 ~~[ASCE: Standard Guidelines for the Design of Urban Stormwater Systems, Standard Guidelines for Installation of Urban Stormwater Systems, and Standard Guidelines for the Operation and Maintenance of Urban Stormwater Systems, ASCE/EWRI 45-05, 46-05, and 47-05., American Society of Civil Engineers, Reston, VA. \(Accessed 9 December 2016\). 2006.](#)~~
- ~~For. Landsc. Ont., 12, 2011.~~

Formatted: Font: (Default) Times New Roman, 10 pt

Formatted: Normal, Indent: Left: 0", Hanging: 0.5", Spacing: Before: 0 pt, After: 0 pt

Field Code Changed

- Ban, N., Schmidli, J. and Schär, C.: Evaluation of the convection-resolving regional climate modeling approach in decade-long simulations, *J. Geophys. Res. Atmospheres*, 119(13), 7889–7907, [doi:10.1002/2014JD021478](https://doi.org/10.1002/2014JD021478), 2014.
- Berg, P., Moseley, C. and Haerter, J. O.: Strong increase in convective precipitation in response to higher temperatures, *Nat. Geosci.*, 6(3), 181–185, [doi:10.1038/ngeo1731](https://doi.org/10.1038/ngeo1731), 2013.
- 5 Blenkinsop, S., Chan, S. C., Kendon, E. J., Roberts, N. M. and Fowler, H. J.: Temperature influences on intense UK hourly precipitation and dependency on large-scale circulation, *Environ. Res. Lett.*, 10(5), 054021, [doi:10.1088/1748-9326/10/5/054021](https://doi.org/10.1088/1748-9326/10/5/054021), 2015.
- Bougadis, J. and Adamowski, K.: Scaling model of a rainfall intensity-duration-frequency relationship, *Hydrol. Process.*, 20(17), 3747–3757, 2006.
- 10 Bourne, L. S. and Simmons, J.: New Fault Lines? Recent Trends in the Canadian Urban System and Their Implications for Planning and Public Policy, *Can. J. Urban Res.*, 12(1), 22–47, 2003.
- Burn, D. H. and Taleghani, A.: Estimates of changes in design rainfall values for Canada, *Hydrol. Process.*, 27(11), 1590–1599, 2013.
- Burnham, K. P. and Anderson, D. R.: Multimodel inference understanding AIC and BIC in model selection, *Sociol. Methods Res.*, 33(2), 261–304, 2004.
- 15 [CCF \(Canadian Climate Forum\). Extreme Weather, 1\(1\), 1- 4, Ottawa, 2013. Available at: http://www.climateforum.ca/ accessed on, December 2016.](http://www.climateforum.ca/)
- [Canadian Climate Forum: Extreme Weather, Ottawa., 2013.](http://www.climateforum.ca/)
- hydroclimatology, *Hydrol. Process.*, 24(6), 673–685, [doi:10.1002/hyp.7506](https://doi.org/10.1002/hyp.7506), 2010.
- 20 Carvalho, L. M., Jones, C. and Liebmann, B.: Extreme precipitation events in southeastern South America and large-scale convective patterns in the South Atlantic convergence zone, *J. Clim.*, 15(17), 2377–2394, 2002.
- Castellarin, A., Kohnová, S., Gaál, L., Fleig, A., Salinas, J. L., Toumazis, A., Kjeldsen, T. R. and Macdonald, N.: Review of applied-statistical methods for flood-frequency analysis in Europe, [\[online\]](http://nora.nerc.ac.uk/19286/)—Available from: <http://nora.nerc.ac.uk/19286/>, 2012.
- 25 [CDD \(Canadian Disaster Database\): Public Safety Canada. Available at: https://www.publicsafety.gc.ca/cnt/rsrscs/cndn-dsstr-dtbs/index-en.aspx, accessed on: December 2016.](https://www.publicsafety.gc.ca/cnt/rsrscs/cndn-dsstr-dtbs/index-en.aspx)
- ~~CDD: The Canadian Disaster Database. [online] Available from: https://www.publicsafety.gc.ca/cnt/rsrscs/cndn-dsstr~~
- Ontario, Canada: Potential for Climate Change Projections, *J. Appl. Meteorol. Climatol.*, 49(5), 845–866, [doi:10.1175/2010JAMC2016.1](https://doi.org/10.1175/2010JAMC2016.1), 2010.
- 30 Cheng, L. and AghaKouchak, A.: Nonstationary precipitation intensity-duration-frequency curves for infrastructure design in a changing climate, *Sci. Rep.*, 4 [\[online\]](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4235283/)—Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4235283/> (Accessed 7 September 2016), [doi: 10.1038/srep07093](https://doi.org/10.1038/srep07093), 2014.

- Cheng, L., AghaKouchak, A., Gilleland, E. and Katz, R. W.: Non-stationary extreme value analysis in a changing climate, *Clim. Change*, 127(2), 353–369, 2014.
- Chowdhury, A. and Mavrotas, G.: FDI and Growth: What Causes What?, *World Econ.*, 29(1), 9–19, [doi:10.1111/j.1467-9701.2006.00755.x](https://doi.org/10.1111/j.1467-9701.2006.00755.x), 2006.
- 5 Coles, S., Bawa, J., Trenner, L. and Dorazio, P.: An introduction to statistical modeling of extreme values, Springer—[\[online\]](#) Available from: <http://link.springer.com/content/pdf/10.1007/978-1-4471-3675-0.pdf> (Accessed 5 January 2017), 2001.
- Coulibaly, P. and Shi, X.: Identification of the effect of climate change on future design standards of drainage infrastructure in Ontario, Rep. Prep. McMaster Univ. Funding Minist. Transp. Ont., 82 [\[online\]](#)—Available from: http://www.cspi.ca/sites/default/files/download/Final_MTO_Report_June2005rv.pdf (Accessed 9 December 2016), 10 2005.
- Coulibaly, P., Burn, D., Switzman, H., Henderson, J. and Fausto, E.: A comparison of future IDF curves for Southern Ontario., Technical Report, McMaster University, Hamilton., 2015. Available from: <https://climateconnections.ca/wp-content/uploads/2014/01/IDF-Comparison-Report-and-Addendum.pdf>
- 15 CSA (Canadian Standards Association): Technical Guide – Development, Interpretation and Use of Rainfall Intensity-duration-frequency (IDF) Information: Guideline for Canadian Water Resources Practitioners., 2010.
- Dawson, C. W., Abraham, R. J. and See, L. M.: HydroTest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts, *Environ. Model. Softw.*, 22(7), 1034–1052, [doi:10.1016/j.envsoft.2006.06.008](https://doi.org/10.1016/j.envsoft.2006.06.008), 2007.
- 20 Deng, Z., Qiu, X., Liu, J., Madras, N., Wang, X. and Zhu, H.: Trend in frequency of extreme precipitation events over Ontario from ensembles of multiple GCMs, *Clim. Dyn.*, 46(9–10), 2909–2921, [doi:10.1007/s00382-015-2740-9](https://doi.org/10.1007/s00382-015-2740-9), 2016.
- [Dean, S. M., Rosier, S., Carey-Smith, T. and Stott, P. A.: The role of climate change in the two-day extreme rainfall in Golden Bay, New Zealand, December 2011, *Bull. Am. Meteorol. Soc.*, 94\(9\), S61, 2013.](#)
- 25 Soc., 1057–1072, 1981.
- Dixon, P. G. and Mote, T. L.: Patterns and Causes of Atlanta’s Urban Heat Island–Initiated Precipitation, *J. Appl. Meteorol.*, 42(9), 1273–1284, [doi:10.1175/1520-0450\(2003\)042<1273:PACOAU>2.0.CO;2](https://doi.org/10.1175/1520-0450(2003)042<1273:PACOAU>2.0.CO;2), 2003.
- Donat, M. G., Lowry, A. L., Alexander, L. V., O’Gorman, P. A. and Maher, N.: More extreme precipitation in the world’s dry and wet regions, *Nat. Clim. Change*, 6(5), 508–513, [doi:10.1038/nclimate2941](https://doi.org/10.1038/nclimate2941), 2016.
- 30 Drobinski, P., Da Silva, N., Panthou, G., Bastin, S., Muller, C., Ahrens, B., Borga, M., Conte, D., Fossier, G., Giorgi, F. and others: Scaling precipitation extremes with temperature in the Mediterranean: past climate assessment and projection in anthropogenic scenarios, *Clim. Dyn.*, 1–21, 2016.

- Durrans, S. and Brown, P.: Estimation and Internet-Based Dissemination of Extreme Rainfall Information, *Transp. Res. Rec. J. Transp. Res. Board*, 1743, 41–48, doi:10.3141/1743-06, 2001.
- EC (Environment Canada): Documentation on Environment Canada Rainfall Intensity-Duration-Frequency (IDF) Tables and Graphs. V2.20., 2012.
- 5 El Adlouni, S., Ouarda, T. B. M. J., Zhang, X., Roy, R. and Bobée, B.: Generalized maximum likelihood estimators for the nonstationary generalized extreme value model, *Water Resour. Res.*, 43(3), W03410, doi:10.1029/2005WR004545, 2007.
- Field, C. B.: Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change, Cambridge University Press. Available from: <http://ipcc-wg2.gov/SREX/report/> [online] Available from: https://books.google.ca/books?hl=en&lr=&id=nQg3SJtkOGwC&oi=fnd&pg=PR4&dq=Managing+the+Risks+of+Extreme+Events+and+Disasters+to+Advance+Climate+Change+Adaptation:+Special+Report+of+the+Intergovernmental+Panel+on+Climate+Change.+&ots=12KcsqmyRP&sig=wdWfRCjO2Hm4AB9NAo_f_d60m4 (Accessed 9 December 2016), 2012.
- 10
- 15 Fischer, E. M. and Knutti, R.: Observed heavy precipitation increase confirms theory and early models, *Nat. Clim. Change*, 6(11), 986–991, 2016.
- Gan, T. Y., Gobena, A. K. and Wang, Q.: Precipitation of southwestern Canada: Wavelet, scaling, multifractal analysis, and teleconnection to climate anomalies, *J. Geophys. Res. Atmospheres*, 112(D10), D10110, doi:10.1029/2006JD007157, 2007.
- 20 Gershunov, A. and Barnett, T. P.: ENSO influence on intraseasonal extreme rainfall and temperature frequencies in the contiguous United States: Observations and model results, *J. Clim.*, 11(7), 1575–1586, 1998.
- Gilleland, E. and Katz, R. W.: Extremes 2.0: an extreme value analysis package in r, *Submitt.-J. Stat. Softw.* [online] Available from: <https://www.rap.ucar.edu/staff/erieg/extRemes/extRemes2.pdf> (Accessed 6 January 2017), 72(8), 20146.
- Gimeno, R., Machado, B. and Mínguez, R.: Stationarity tests for financial time series, *Phys. Stat. Mech. Its Appl.*, 269(1), 72–78, 1999.
- 25 Gu, X., Zhang, Q., Singh, V. P., Xiao, M. and Cheng, J.: Nonstationarity-based evaluation of flood risk in the Pearl River basin: changing patterns, causes and implications, *Hydrol. Sci. J.*, 62(2), 246–258, 2017.
- Gudmundsson, L., Bremnes, J. B., Haugen, J. E. and Engen-Skaugen, T.: Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations—a comparison of methods, *Hydrol. Earth Syst. Sci.*, 16(9), 3383–3390, 2012.
- 30 Güntner, A., Olsson, J., Calver, A. and Gannon, B.: Cascade-based disaggregation of continuous rainfall time series: the influence of climate, *Hydrol. Earth Syst. Sci. Discuss.*, 5(2), 145–164, 2001.

- Guo, X., Fu, D. and Wang, J.: Mesoscale convective precipitation system modified by urbanization in Beijing City, *Atmospheric Res.*, 82(1–2), 112–126, doi:10.1016/j.atmosres.2005.12.007, 2006.
- Hamed, K. H. and Rao, A. R.: A modified Mann-Kendall trend test for autocorrelated data, *J. Hydrol.*, 204(1), 182–196, 1998.
- [Hamed, K. H. and Rao, A. R.: Variability in spectral characteristics of hydrologic data, IAHS AISH Publ., 129–133, 1999.](#)
- 5 Hu, S.: Akaike information criterion, *Cent. Res. Sci. Comput.*, [NC State University, Raleigh, NC, 93](#) [online] Available from: http://www4.ncsu.edu/~shu3/Presentation/AIC_2012.pdf, 2007.
- Hurvich, C. M. and Tsai, C.-L.: Model selection for extended quasi-likelihood models in small samples, *Biometrics*, 1077–1084, 1995.
- Jakob, D.: Nonstationarity in extremes and engineering design, in *Extremes in a Changing Climate*, pp. 363–417, Springer-
- 10 [\[online\] Available from: \[http://link.springer.com/chapter/10.1007/978-94-007-4479-0_13\]\(http://link.springer.com/chapter/10.1007/978-94-007-4479-0_13\) \(Accessed 9 December 2016\)](#), 2013.
- Jarvis, A., Reuter, H. I., Nelson, A., Guevara, E. and others: Hole-filled SRTM for the globe Version 4, [Available CGLAR-CSI SRTM 90m Database <https://www.cgiar-csi.org>](#) [online] Available from: <http://www.cgiar-csi.org/data/srtm-90m-digital-elevation-database-v4-1> (Accessed 13 December 2016), 2008.
- 15 Jebari, S., Berndtsson, R., Olsson, J. and Bahri, A.: Soil erosion estimation based on rainfall disaggregation, *J. Hydrol.*, 436, 102–110, 2012.
- Karmakar, S. and Simonovic, S.: Flood Frequency Analysis Using Copula with Mixed Marginal Distributions, *Water Resour. Res. Rep.* [\[online\]](#) Available from: <http://ir.lib.uwo.ca/wrrr/19>, 2007.
- Katz, R. W. and Brown, B. G.: Extreme events in a changing climate: variability is more important than averages, *Clim. Change*, 21(3), 289–302, 1992.
- 20 Katz, R. W., Parlange, M. B. and Naveau, P.: Statistics of extremes in hydrology, *Adv. Water Resour.*, 25(8), 1287–1304, 2002.
- Kerr, D.: Some Aspects of the Geography of Finance in Canada, *Can. Geogr. Géographe Can.*, 9(4), 175–192, doi:10.1111/j.1541-0064.1965.tb00825.x, 1965.
- 25 [Knutson, T. R., Zeng, F. and Wittenberg, A. T.: Seasonal and annual mean precipitation extremes occurring during 2013: A US focused analysis, *Bull. Am. Meteorol. Soc.*, 95\(9\), S19, 2014.](#)
- doi:10.1038/srep05884, 2014.
- Komi, K., Amisigo, B. A., Diekkrüger, B. and Hountondji, F. C.: Regional Flood Frequency Analysis in the Volta River Basin, West Africa, *Hydrology*, 3(1), 5, 2016.
- 30 Koutsoyiannis, D. and Montanari, A.: Negligent killing of scientific concepts: the stationarity case, *Hydrol. Sci. J.*, 60(7–8), 1174–1183, 2015.

- Kunkel, K. E.: North American Trends in Extreme Precipitation, *Nat. Hazards*, 29(2), 291–305, doi:10.1023/A:1023694115864, 2003.
- Kwiatkowski, D., Phillips, P. C., Schmidt, P. and Shin, Y.: Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?, *J. Econom.*, 54(1–3), 159–178, 1992.
- 5 Lapen, D. R. and Hayhoe, H. N.: Spatial Analysis of Seasonal and Annual Temperature and Precipitation Normals in Southern Ontario, Canada, *J. Gt. Lakes Res.*, 29(4), 529–544, doi:10.1016/S0380-1330(03)70457-2, 2003.
- 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.
- [Lenderink, G., Barbero, R., Loriaux, J. M. and Fowler, H. J.: Super-Clausius–Clapeyron Scaling of Extreme Hourly Convective Precipitation and Its Relation to Large-Scale Atmospheric Conditions, *J. Clim.*, 30\(15\), 6037–6052, doi:10.1175/JCLI-D-16-0808.1, 2017.](#)
- 10 [Panel on Climate Change AR4 models using equidistant quantile matching, *J. Geophys. Res. Atmospheres*, 115\(D10\), D10101, doi:10.1029/2009JD012882, 2010.](#)
- 15 Lima, C. H., Kwon, H.-H. and Kim, J.-Y.: A Bayesian beta distribution model for estimating rainfall IDF curves in a changing climate, *J. Hydrol.*, 540, 744–756, 2016.
- Madsen, H., Arnbjerg-Nielsen, K. and Mikkelsen, P. S.: Update of regional intensity–duration–frequency curves in Denmark: tendency towards increased storm intensities, *Atmospheric Res.*, 92(3), 343–349, 2009.
- [Mohsin, T. and W.A. Gough, 2012. Characterization and estimation of Urban Heat Island at Toronto: impact of the choice of rural sites. *Theoretical and Applied Climatology*, 108\(1-2\): 105-117.](#)
- 20 Mailhot, A., Beauregard, I., Talbot, G., Caya, D. and Biner, S.: Future changes in intense precipitation over Canada assessed from multi-model NARCCAP ensemble simulations, *Int. J. Climatol.*, 32(8), 1151–1163, 2012.
- Maraun, D.: Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue, *J. Clim.*, 26(6), 2137–2143, 2013.
- 25 [Markose, S. and Alentorn, A.: The Generalized Extreme Value \(GEV\) Distribution, Implied Tail Index and Option Pricing, *J. Deriv.*, 18\(3\), 35–60, 2005. Markose, S. and Alentorn, A.: The Generalized Extreme Value \(GEV\) Distribution, Implied Tail Index and Option Pricing 1, \[online\] Available from: <http://citeseerx.ist.psu.edu/viewdoc/citations.jsessionid=691E18DF64D87FB92BED43C1C15CB2B1?doi=10.1.1.83.4945> \(Accessed 10 December 2016\), 2005.](#)
- 30 Martins, E. S., Stedinger, J. R. and others: Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data, *Water Resour. Res.*, 36(3), 737–744, 2000.
- Miao, C., Sun, Q., Borthwick, A. G. L. and Duan, Q.: Linkage Between Hourly Precipitation Events and Atmospheric Temperature Changes over China during the Warm Season, *Sci. Rep.*, 6, srep22543, doi:10.1038/srep22543, 2016.

- Mikkelsen, P. S., Madsen, H., Arnbjerg-Nielsen, K., Rosbjerg, D. and Harremoës, P.: Selection of regional historical rainfall time series as input to urban drainage simulations at ungauged locations, *Atmospheric Res.*, 77(1–4), 4–17, doi:10.1016/j.atmosres.2004.10.016, 2005.
- 5 Miller, J. D., Kim, H., Kjeldsen, T. R., Packman, J., Grebby, S. and Dearden, R.: Assessing the impact of urbanization on storm runoff in a peri-urban catchment using historical change in impervious cover, *J. Hydrol.*, 515, 59–70, doi:10.1016/j.jhydrol.2014.04.011, 2014.
- Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P. and Stouffer, R. J.: Stationarity Is Dead: Whither Water Management?, *Science*, 319(5863), 573–574, doi:10.1126/science.1151915, 2008.
- 10 Mishra, V., Dominguez, F. and Lettenmaier, D. P.: Urban precipitation extremes: How reliable are regional climate models?, *Geophys. Res. Lett.*, 39(3), L03407, doi:10.1029/2011GL050658, 2012.
- [Moglen, G. E. and Schwartz, D. E.: Methods for adjusting US geological survey rural regression peak discharges in an urban setting, U.S. Geological Survey Scientific Investigation Report 2006 - 5270, 1-55, 2006.](#)
- 15 [Moglen, G. E. and Schwartz, D. E.: Methods for adjusting US geological survey rural regression peak discharges in an urban setting, doi:10.1175/1525-7541\(2004\)005<0409:IOUEOP>2.0.CO;2, 2004.](#)
- Mondal, A. and Mujumdar, P. P.: Modeling non-stationarity in intensity, duration and frequency of extreme rainfall over India, *J. Hydrol.*, 521, 217–231, doi:10.1016/j.jhydrol.2014.11.071, 2015.
- Montanari, A. and Koutsoyiannis, D.: Modeling and mitigating natural hazards: Stationarity is immortal!, *Water Resour. Res.*, 50(12), 9748–9756, 2014.
- 20 Olsson, J.: Limits and characteristics of the multifractal behaviour of a high-resolution rainfall time series, *Nonlinear Process. Geophys.*, 2(1), 23–29, 1995.
- Olsson, J.: Evaluation of a scaling cascade model for temporal rain-fall disaggregation, *Hydrol. Earth Syst. Sci. Discuss.*, 2(1), 19–30, 1998.
- 25 [O’Gorman, P. A.: Precipitation Extremes Under Climate Change, *Curr. Clim. Change Rep.*, 1\(2\), 49–59, doi:10.1007/s40641-015-0009-3, 2015.](#)
- distribution tails, *Hydrol Earth Syst Sci*, 17(2), 851–862, doi:10.5194/hess-17-851-2013, 2013.
- Partridge, M., Olfert, M. R. and Alasia, A.: Canadian cities as regional engines of growth: agglomeration and amenities, *Can. J. Econ. Can. Déconomique*, 40(1), 39–68, doi:10.1111/j.1365-2966.2007.00399.x, 2007.
- 30 Pendergrass, A. G., Lehner, F., Sanderson, B. M. and Xu, Y.: Does extreme precipitation intensity depend on the emissions scenario?, *Geophys. Res. Lett.*, 42(20), 8767–8774, 2015.
- Petrow, T. and Merz, B.: Trends in flood magnitude, frequency and seasonality in Germany in the period 1951–2002, *J. Hydrol.*, 371(1), 129–141, 2009.

- Pettitt, A. N.: A non-parametric approach to the change-point problem, *Appl. Stat.*, 126–135, 1979.
- [Pfahl, S., O’Gorman, P. A. and Fischer, E. M.: Understanding the regional pattern of projected future changes in extreme precipitation, *Nat. Clim. Change*, 7\(6\), 423–427, doi:10.1038/nclimate3287, 2017.](#)
- 5 Porporato, A. and Ridolfi, L.: Influence of weak trends on exceedance probability, *Stoch. Hydrol. Hydraul.*, 12(1), 1–14, 1998.
- Prein, A. F., Rasmussen, R. M., Ikeda, K., Liu, C., Clark, M. P. and Holland, G. J.: The future intensification of hourly precipitation extremes, *Nat. Clim. Change*, advance online publication, doi:10.1038/nclimate3168, 2016.
- Priestley, M. B. and Rao, T. S.: A test for non-stationarity of time-series, *J. R. Stat. Soc. Ser. B Methodol.*, 140–149, 1969.
- Rana, A., Bengtsson, L., Olsson, J. and Jothiprakash, V.: Development of IDF-curves for tropical india by random cascade modeling, *Hydrol. Earth Syst. Sci. Discuss.*, 10(4), 4709–4738, 2013.
- 10 Reddy, M. J. and Ganguli, P.: Spatio-temporal analysis and derivation of copula-based intensity–area–frequency curves for droughts in western Rajasthan (India), *Stoch. Environ. Res. Risk Assess.*, 27(8), 1975–1989, 2013.
- Ross, G. J., Tasoulis, D. K. and Adams, N. M.: Nonparametric Monitoring of Data Streams for Changes in Location and Scale, *Technometrics*, 53(4), 379–389, doi:10.1198/TECH.2011.10069, 2011.
- 15 Ruiz-Villanueva, V., Borga, M., Zoccatelli, D., Marchi, L., Gaume, E. and Ehret, U.: Extreme flood response to short-duration convective rainfall in South-West Germany, *Hydrol. Earth Syst. Sci.*, 16(5), 1543–1559, 2012.
- Sadri, S., Kam, J. and Sheffield, J.: Nonstationarity of low flows and their timing in the eastern United States, *Hydrol Earth Syst Sci*, 20(2), 633–649, doi:10.5194/hess-20-633-2016, 2016.
- Sanderson, M. and Gorski, R.: The effect of metropolitan Detroit-Windsor on precipitation, *J. Appl. Meteorol.*, 17(4), 423–427, 1978.
- 20 Sandink, D., Simonovic, S. P., Schardong, A. and Srivastav, R.: A decision support system for updating and incorporating climate change impacts into rainfall intensity-duration-frequency curves: Review of the stakeholder involvement process, *Environ. Model. Softw.*, 84, 193–209, 2016.
- Serinaldi, F. and Kilsby, C. G.: Stationarity is undead: Uncertainty dominates the distribution of extremes, *Adv. Water Resour.*, 25, 77, 17–36, 2015.
- Serinaldi, F. and Kilsby, C. G.: The importance of prewhitening in change point analysis under persistence, *Stoch. Environ. Res. Risk Assess.*, 30(2), 763–777, doi:10.1007/s00477-015-1041-5, 2016.
- Shabbar, A., Bonsal, B. and Khandekar, M.: Canadian Precipitation Patterns Associated with the Southern Oscillation, *J. Clim.*, 10(12), 3016–3027, doi:10.1175/1520-0442(1997)010<3016:CPPAWT>2.0.CO;2, 1997.
- 30 Shaw, S. B., Royem, A. A. and Riha, S. J.: The relationship between extreme hourly precipitation and surface temperature in different hydroclimatic regions of the United States, *J. Hydrometeorol.*, 12(2), 319–325, 2011.
- [Schaller, N., Kay, A. L., Lamb, R., Massey, N. R., van Oldenborgh, G. J., Otto, F. E. L., Sparrow, S. N., Vautard, R., Yiou, P., Ashpole, I., Bowery, A., Crooks, S. M., Hausteine, K., Huntingford, C., Ingram, W. J., Jones, R. G., Legg, T.,](#)

- [Miller, J., Skeggs, J., Wallom, D., Weisheimer, A., Wilson, S., Stott, P. A. and Allen, M. R.: Human influence on climate in the 2014 southern England winter floods and their impacts, *Nat. Clim. Change*, 6\(6\), 627–634, 2016.](#)
- Duration Extreme Rainfall: Including an Intensity–Duration–Frequency Perspective, *Atmosphere-Ocean*, 52(5), 398–417, 2014.
- 5 Simonovic, S. P. and Peck, A.: Updated rainfall intensity duration frequency curves for the City of London under the changing climate, Department of Civil and Environmental Engineering, The University of Western Ontario. [\[online\]](#) Available from: <http://ir.lib.uwo.ca/wrrr/29/> (Accessed 13 January 2017), 2009.
- Singh, J., Vittal, H., Karmakar, S., Ghosh, S. and Niyogi, D.: Urbanization causes nonstationarity in Indian Summer Monsoon Rainfall extremes, *Geophys. Res. Lett.*, 43(21) [\[online\]](#) Available from: <http://onlinelibrary.wiley.com/doi/10.1002/2016GL071238/full> (Accessed 21 December 2016), 2016.
- 10 Stocker, T. F., Qin, D., Plattner, G. K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V. and Midgley, P. M.: Climate change 2013: the physical science basis. Intergovernmental panel on climate change, working group I Contribution to the IPCC fifth assessment report (AR5), N. Y., 2013.
- 15 [Switzman, H., Razavi, T., Traore, S., Coulibaly, P., Burn, D. H., Henderson, J., Fausto, E. and Ness, R.: Variability of Future Extreme Rainfall Statistics: Comparison of Multiple IDF Projections, *J. Hydrol. Eng.*, 22\(10\), 04017046, 2017.](#)
- Svensson, C. and Jones, D. a.: Review of rainfall frequency estimation methods, *J. Flood Risk Manag.*, 3(4), 296–313, [doi:10.1111/j.1753-318X.2010.01079.x](https://doi.org/10.1111/j.1753-318X.2010.01079.x), 2010.
- 20 Teutschbein, C. and Seibert, J.: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods, *J. Hydrol.*, 456, 12–29, 2012.
- [TRCA \(Toronto Region Conservation Authority\): Resilient City: Preparing for Extreme Weather Events.](#), 2013.
- Towler, E., Rajagopalan, B., Gilleland, E., Summers, R. S., Yates, D. and Katz, R. W.: Modeling hydrologic and water quality extremes in a changing climate: A statistical approach based on extreme value theory, *Water Resour. Res.*, 46(11), W11504, [doi:10.1029/2009WR008876](https://doi.org/10.1029/2009WR008876), 2010.
- 25 Trenberth, K. E.: Atmospheric Moisture Recycling: Role of Advection and Local Evaporation, *J. Clim.*, 12(5), 1368–1381, [doi:10.1175/1520-0442\(1999\)012<1368:AMRROA>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<1368:AMRROA>2.0.CO;2), 1999.
- Van Gelder, P., Wang, W. and Vrijling, J. K.: Statistical estimation methods for extreme hydrological events, in *Extreme hydrological events: New concepts for security*, pp. 199–252, Springer. [\[online\]](#) Available from: http://link.springer.com/chapter/10.1007/978-1-4020-5741-0_15 (Accessed 14 December 2016), 2006.
- 30 Villarini, G., Serinaldi, F., Smith, J. A. and Krajewski, W. F.: On the stationarity of annual flood peaks in the continental United States during the 20th century, *Water Resour. Res.*, 45(8), W08417, [doi:10.1029/2008WR007645](https://doi.org/10.1029/2008WR007645), 2009.

- Wasko, C. and Sharma, A.: Steeper temporal distribution of rain intensity at higher temperatures within Australian storms, *Nat. Geosci.*, 8(7), 527–529, [doi:10.1038/ngeo2456](https://doi.org/10.1038/ngeo2456), 2015.
- Wasko, C. and Sharma, A.: Continuous rainfall generation for a warmer climate using observed temperature sensitivities, *J. Hydrol.*, 544, 575–590, [doi:10.1016/j.jhydrol.2016.12.002](https://doi.org/10.1016/j.jhydrol.2016.12.002), 2017.
- 5 [Wang, X., Huang, G., Liu, J., Li, Z. and Zhao, S.: Ensemble projections of regional climatic changes over Ontario, Canada, *J. Clim.*, 28\(18\), 7327–7346, 2015.](#)
- [Hydrol.](#), 406(1–2), 119–128, [doi:10.1016/j.jhydrol.2011.06.014](https://doi.org/10.1016/j.jhydrol.2011.06.014), 2011.
- Wilson, P. S. and Toumi, R.: A fundamental probability distribution for heavy rainfall, *Geophys. Res. Lett.*, 32(14), L14812, [doi:10.1029/2005GL022465](https://doi.org/10.1029/2005GL022465), 2005.
- 10 Xie, H., Li, D. and Xiong, L.: Exploring the ability of the Pettitt method for detecting change point by Monte Carlo simulation, *Stoch. Environ. Res. Risk Assess.*, 28(7), 1643–1655, 2014.
- Yilmaz, A. G. and Perera, B. J. C.: Extreme rainfall nonstationarity investigation and intensity–frequency–duration relationship, *J. Hydrol. Eng.*, 19(6), 1160–1172, 2013.
- 15 Yilmaz, A. G., Hossain, I. and Perera, B. J. C.: Effect of climate change and variability on extreme rainfall intensity–frequency–duration relationships: a case study of Melbourne, *Hydrol. Earth Syst. Sci.*, 18(10), 4065–4076, 2014a.
- [Yilmaz, A. G., Hossain, I. and Perera, B. J. C.: Effect of climate change and variability on extreme rainfall intensity–frequency–duration relationships: a case study of Melbourne, *Hydrol. Earth Syst. Sci.*, 18\(10\), 4065–4076, 2014b.](#)
- [Yiou, P. and Cattiaux, J.: Contribution of atmospheric circulation to wet north European summer precipitation of 2012, *Bull. Am. Meteorol. Soc.*, 94\(9\), S39, 2013.](#)
- 20 [data, *Stoch. Environ. Res. Risk Assess.*, 16\(4\), 307–323, 2002.](#)
- Yue, S., Pilon, P., Phinney, B. and Cavadas, G.: The influence of autocorrelation on the ability to detect trend in hydrological series, *Hydrol. Process.*, 16(9), 1807–1829, 2002.
- 25 Yue, S., Pilon, P. and Phinney, B. O. B.: Canadian streamflow trend detection: impacts of serial and cross-correlation, *Hydrol. Sci. J.*, 48(1), 51–63, 2003.
- [Adamowski, K. and Bougadis, J.: Detection of trends in annual extreme rainfall, *Hydrol. Process.*, 17\(18\), 3547–3560,](#)

30

Figure 1. (a) Selected urbanized sites in Southern Ontario. The Southern Ontario (41° - 44°N, 84° - 76°W) is the southernmost region of Canada and is situated on a southwest-northeast transect, bounded by lakes Huron, Erie, and Ontario. The nine locations on the map are (from southwest to northeast corner): Windsor Airport, London International Airport, Stratford Wastewater Treatment Plant (WWTP), Fergus Shand Dam, Hamilton Airport, Toronto International Airport, Oshawa Water Pollution Control Plant (WPCP), Trenton Airport, and Kingston Pumping Station. Topography map indicates maximum slope of 670 m above mean sea level. (b) The population map shows six the sites: Windsor Airport, London International Airport, Hamilton Airport, Toronto International Airport, Oshawa WPCP, and Kingston P. Station to be located either in or the vicinity of densely populated urbanized area. The remaining three sites are located in the moderately populated area. The daily and sub-daily AMP records in all locations vary between the minimum of 46 and maximum of 66 years.

~~Selected urbanized sites in Southern Ontario. The Southern Ontario (41° - 44°N, 84° - 76°W) is the southernmost region of Canada and is situated on a southwest-northeast transect, bounded by lakes Huron, Erie, and Ontario. The nine locations on the map are (from southwest to northeast corner): Windsor Airport, London International Airport, Stratford Wastewater Treatment Plant (WWTP), Fergus Shand Dam, Hamilton Airport, Toronto International Airport, Oshawa Water Pollution Control Plant (WPCP), Trenton Airport, and Kingston Pumping Station. Topography map indicates maximum slope of 670 m above mean sea level. (b) The population map shows six the sites: Windsor Airport, London International Airport, Hamilton Airport, Toronto International Airport, Oshawa WPCP, and Kingston P. Station to be located either in or the vicinity of densely populated urbanized area. The remaining three sites are located in the moderately populated area. The rainfall records in all locations vary between the minimum of 46 and maximum of 66 years.~~

Figure 2. Schematics of the process flow (Blue - input step, orange - process step, and green - decision steps). All three tests - Mann-Kendall, Pettitt's and Mann-Whitney, check for shifts in the mean. While Mann-Kendall tests for monotonic trends, the other two tests, Pettitt's and Mann-Whitney check for change point or regime shift in the time series.

~~Schematics of the process flow (Blue - input step, orange - process step, and green - decision steps). All three tests - Mann-Kendall, Pettitt's and Mann-Whitney, check for shifts in the mean. While Mann-~~

Kendall tests for monotonic trends, the other two tests, Pettitt's and Mann-Whitney check for change point or

Figure 3. Spatial distribution of trends, change points and nonstationarities in rainfall extremes of several durations in nine urbanized locations, Southern Ontario (a – g). The up and down triangles in white indicate (statistically insignificant) up and downward shifts; the up and down triangles in cyan and orange indicate shifts with change points only; the up and down triangles in the dark blue and red show presence of (statistically significant) trends including change point(s). Sites with significant nonstationarity are marked with an 'x' sign. All tests are performed at 10% significance levels, i.e., p-value < 0.10. Spatial distribution of trends, change points and nonstationarities in hourly and sub-hourly rainfall extremes in nine urbanized locations, Southern Ontario (a – g). The population estimates in and the vicinity of urbanized sites varies between 5 Million and 19,000 with highest in Toronto Metropolitan Area and the lowest in Fergus-Shand dam area. The up and down triangles in white indicate (statistically insignificant) up and downward shifts; the up and down triangles in cyan and orange indicate shifts with change points only; the up and down triangles in the dark blue and red show presence of (statistically significant) trends including change point(s). Sites with significant nonstationarity are marked with an 'x' sign. The trends in short duration AMP extremes are detected using nonparametric Mann-Kendall trend test with correction for ties. The change point in precipitation extremes are identified either as a shift in the mean or the variance in AMP time-series. We apply nonparametric Pettit (Pettitt, 1979) and Mann-Whitney (Ross et al., 2011) tests to identify change point in the mean and Mood's test (Ross et al., 2011) to detect change point in the variance respectively (See section 2 for details). All tests are performed at 5 and 10% significance levels.

Figure 4. DSI estimates of median (horizontal line within the box plot) and 95% credible intervals for 100-year return periods of stationary versus nonstationary models across nine sites (a - i). The boxplots indicate the uncertainty in estimated DSI using Bayesian inference. Uncertainty in DSI for 100-year return periods for stationary (blue) versus nonstationary (red) models with durations ranging between 15-min and 24-hr for nine sites (a – i): Toronto International Airport, Hamilton Airport, Oshawa WPCP, Windsor Airport, Kingston P. station, London International Airport, Trenton Airport, Stratford WWTP, and Fergus-Shand Dam. The boxplots indicate the uncertainty in estimated DSI from Bayesian inference, whereas the DSI obtained from maximum likelihood approach is shown as a black dot in the stationary simulation.

Figure 5. Percentage changes (in top panel) and Z-statistics (bottom panel) of at-site T-year event estimates for $T = 2$ -year to $T = 100$ -year return periods (a – d) with durations between 15-min and 24-hr in nine urbanized locations, Southern Ontario. The Z-statistic represents statistical significance of differences in DSI obtained from the best

Formatted: Font: Bold

Formatted: Font: Bold

selected nonstationary versus the stationary model. The Z-statistics is statistically significant when $|Z| > 1.64$ at 10% significance level. The shades in blue and red denote decrease and increase in Z-statistics with the strength of shading represents the magnitude of the test statistics. The cyan shading represents the site with significant autocorrelation, which we exclude from the analysis. Z-statistics of at site T-year event estimates for T = 2-year to T = 100-year return periods (a—d) with durations between 15 min and 24 hr in nine urbanized locations, Southern Ontario. The Z-statistic represents statistical significance of differences in DSI obtained from the best selected nonstationary versus the stationary model. The Z-statistics is statistically significant when $|Z| > 1.96$ and $|Z| > 1.64$ at 5 and 10% significance levels. The shades in blue and red denote decrease and increase in Z-statistics with the strength of shading represents the magnitude of the test statistics. The cyan shading represents the site with significant autocorrelation, which we exclude from the analysis.

Figure 6. Central tendency (median, **b**) and the bounds (95% credible interval, **a** and **c**) of the updated nonstationary versus EC-generated T = 2-and 10-year event estimates for DSI at selected return periods with durations between 15-min and 24-hr. The DSI and associated 95% confidence limits of EC-generated IDF is obtained from the national archive of Engineering Climate Datasets (<http://climate.weather.gc.ca/>). The shades in blue and red denote decrease and increase in DSI. The strength of shading represents the magnitude of the ratio between updated versus EC-generated DSI. Central tendency (median, **b**) and the bounds (5 and 95% range, **a** and **c**) of the updated nonstationary versus EC-generated T-year event estimates for DSI at selected return periods with durations between 15 min and 24 hr. The minimum, median and maximum T-year event estimates of nonstationary models are obtained from time variant GEV parameter(s) by computing the 5th, 50th and 95th percentiles of DE-MC sampled parameters. The DSI and associated 95% confidence limits of EC-generated IDF is obtained from the national archive of Engineering Climate Datasets maintained by the Environment Canada (<http://climate.weather.gc.ca/>). The strength of shading represents the magnitude of the ratio between updated versus EC-generated DSI with a deeper shade indicates an increase in the ratio. The cyan shading indicates the site with significant autocorrelation.

Figure 7. Central tendency (median, **b**) and the bounds (95% credible interval, **a** and **c**) of the updated nonstationary versus EC-generated T = 50-and 100-year event estimates for DSI at selected return periods with durations between 15-min and 24-hr. The DSI and associated 95% confidence limits of EC-generated IDF is obtained from the national archive of Engineering Climate Datasets (<http://climate.weather.gc.ca/>). The shades in blue and red denote decrease

Formatted: Font: Bold

Formatted: Font: Bold

Formatted: Font: Bold

and increase in DSI. The strength of shading represents the magnitude of the ratio between updated versus EC-generated DSI.

Figure 8. Estimated nonstationary versus EC-generated IDF~~s~~ for ~~return periods~~ T = 2, 5, 10, 25, 50 and 100-year ~~return periods~~ for the selected urbanized locations in Southern Ontario, Canada. ~~The nonstationary IDF~~s~~ are shown using solid lines, while EC-generated IDF~~s~~ are shown using dotted lines.~~

¹CMA and CA denote census metropolitan area and census agglomeration respectively. Statistics of Canada defines a CMA with a population of at least 100,000, where the urban core of that area has at least 50,000 people, whereas CA must have an urban core population of at least 10,000. A population Center (or urban area) is an area with at least a population of 1,000 and a density of 400 or more people per square kilometer. All population information are collected from Statistics Canada (<https://www12.statcan.gc.ca/>) website. *Missing values are infilled using observations from nearest Environment Canada station ID 6144475 (latitude 44° and longitude -81.5°) located at 111.5 km geodesic distance. Annual maxima values of missing years or durations are obtained by disaggregating daily data to hourly and sub-hourly time steps.

Formatted: Normal, Left, R
After: 6 pt

Table 2. Selected stations and their statistical properties for annual maxima time series of rainfall volume

<u>Stations</u>	<u>EC-Station</u>	<u>Analysis</u>	<u>Time Frame</u>	<u>Mean</u>	<u>Std. deviation</u>	<u>Skew</u>	<u>Excess¹</u>
	<u>ID</u>	<u>Period</u>	<u>(min)</u>	<u>(mm)</u>	<u>(mm)</u>		<u>Kurtosis</u>
	<u>6158731</u>	<u>1950 - 2013</u>	<u>15</u>	<u>16.35</u>	<u>5.88</u>	<u>0.46</u>	<u>-0.36</u>

Formatted Table

<u>Toronto Int'l. Airport</u>			<u>30</u>	<u>21.85</u>	<u>8.68</u>	<u>0.86</u>	<u>0.9</u>
<u>Hamilton Airport</u>	<u>6153194</u>	<u>1960 - 2010</u>	<u>15</u>	<u>16.12</u>	<u>5.61</u>	<u>1.26</u>	<u>1.11</u>
			<u>30</u>	<u>16.63</u>	<u>8.47</u>	<u>4.45</u>	<u>24.09</u>
<u>Oshawa WPCP</u>	<u>6155878</u>	<u>1970 - 2015</u>	<u>15</u>	<u>56.01</u>	<u>17.84</u>	<u>0.84</u>	<u>1.68</u>
			<u>30</u>	<u>36.09</u>	<u>11.40</u>	<u>0.22</u>	<u>-0.57</u>
<u>Windsor Airport</u>	<u>6139525</u>	<u>1950 - 2013</u>	<u>15</u>	<u>17.79</u>	<u>5.56</u>	<u>1.03</u>	<u>1.69</u>
			<u>30</u>	<u>23.49</u>	<u>8.20</u>	<u>0.77</u>	<u>-0.15</u>
<u>Kingston P. Station</u>	<u>6104175</u>	<u>1961 - 2007</u>	<u>15</u>	<u>12.89</u>	<u>3.79</u>	<u>0.93</u>	<u>2.15</u>
			<u>30</u>	<u>16.54</u>	<u>5.31</u>	<u>0.78</u>	<u>1.12</u>
<u>London Airport</u>	<u>6144478/75*</u>	<u>1950 - 2015</u>	<u>15</u>	<u>15.96</u>	<u>6.62</u>	<u>1.28</u>	<u>1.73</u>
			<u>30</u>	<u>21.06</u>	<u>8.55</u>	<u>1.68</u>	<u>3.94</u>
<u>Trenton Airport</u>	<u>6158875</u>	<u>1950 - 2015</u>	<u>15</u>	<u>13.30</u>	<u>6.52</u>	<u>2.90</u>	<u>10.23</u>
			<u>30</u>	<u>16.60</u>	<u>6.40</u>	<u>1.54</u>	<u>2.97</u>
<u>Stratford WWTP</u>	<u>6148100</u>	<u>1960 - 2015</u>	<u>15</u>	<u>16.33</u>	<u>5.08</u>	<u>1.30</u>	<u>2.23</u>
			<u>30</u>	<u>21.37</u>	<u>9.08</u>	<u>2.22</u>	<u>7.53</u>
<u>Fergus Shand Dam</u>	<u>6142400</u>	<u>1950 - 2015</u>	<u>15</u>	<u>17.74</u>	<u>6.66</u>	<u>1.24</u>	<u>2.14</u>
			<u>30</u>	<u>23.42</u>	<u>10.32</u>	<u>1.78</u>	<u>4.04</u>

*Missing values are infilled using observations from the nearest station ID 6144475.¹Kurtosis relative to normal distribution, *i.e.*, kurtosis – 3.

Table 3. Selected stations and their statistical properties for hourly annual maxima rainfall volume for selected durations

<u>Stations</u>	<u>EC-Station ID</u>	<u>Analysis Period</u>	<u>Time Frame (hr)</u>	<u>Mean</u>	<u>Std. Deviation (mm)</u>	<u>Skew</u>	<u>Excess Kurtosis</u>
<u>Toronto Int'l. Airport</u>	<u>6158731</u>	<u>1950 - 2013</u>	<u>1</u>	<u>24.67</u>	<u>11.01</u>	<u>1.98</u>	<u>7.33</u>

Formatted: Space After: 0

Formatted: Font: 11 pt, Bo

Formatted: Font: Bold

Formatted: Font: 11 pt

Formatted: Font: (Default)
11 pt

Formatted Table

Formatted: Font: (Default)
11 pt

Formatted: Space After: 0

			<u>6</u>	<u>38.44</u>	<u>18.09</u>	<u>2.46</u>	<u>7.93</u>
<u>Hamilton Airport</u>	<u>6153194</u>	<u>1960 - 2010</u>	<u>1</u>	<u>25.91</u>	<u>10.30</u>	<u>2.54</u>	<u>10.20</u>
			<u>6</u>	<u>38.72</u>	<u>15.05</u>	<u>2.34</u>	<u>7.07</u>
<u>Oshawa WPCP</u>	<u>6155878</u>	<u>1970 - 2015</u>	<u>1</u>	<u>22.09</u>	<u>8.56</u>	<u>0.54</u>	<u>-0.43</u>
			<u>6</u>	<u>5.94</u>	<u>2.05</u>	<u>1.18</u>	<u>0.99</u>
<u>Windsor Airport</u>	<u>6139525</u>	<u>1950 - 2013</u>	<u>1</u>	<u>29.53</u>	<u>10.43</u>	<u>0.88</u>	<u>0.14</u>
			<u>6</u>	<u>44.36</u>	<u>14.81</u>	<u>1.12</u>	<u>1.81</u>
<u>Kingston P. Station</u>	<u>6104175</u>	<u>1961 - 2007</u>	<u>1</u>	<u>21.21</u>	<u>6.83</u>	<u>0.56</u>	<u>-0.03</u>
			<u>6</u>	<u>37.35</u>	<u>13.61</u>	<u>2.32</u>	<u>8.05</u>
<u>London Airport</u>	<u>6144478/75</u>	<u>1950 - 2015</u>	<u>1</u>	<u>24.26</u>	<u>11.23</u>	<u>2.41</u>	<u>9.79</u>
			<u>6</u>	<u>36.46</u>	<u>12.10</u>	<u>1.73</u>	<u>4.57</u>
<u>Trenton Airport</u>	<u>6158875</u>	<u>1950 - 2015</u>	<u>1</u>	<u>20.36</u>	<u>8.25</u>	<u>1.87</u>	<u>6.03</u>
			<u>6</u>	<u>36.76</u>	<u>12.51</u>	<u>1.19</u>	<u>0.66</u>
<u>Stratford WWTP</u>	<u>6148100</u>	<u>1960 - 2015</u>	<u>1</u>	<u>24.31</u>	<u>11.12</u>	<u>1.71</u>	<u>3.41</u>
			<u>6</u>	<u>41.77</u>	<u>21.61</u>	<u>2.31</u>	<u>6.15</u>
<u>Fergus Shand Dam</u>	<u>6142400</u>	<u>1950 - 2015</u>	<u>1</u>	<u>28.07</u>	<u>13.67</u>	<u>2.02</u>	<u>5.50</u>
			<u>6</u>	<u>39.86</u>	<u>18.59</u>	<u>1.25</u>	<u>2.30</u>

Formatted: Font: (Default)
11 pt

Formatted: Space After: 0

Formatted: Font: (Default)
11 pt

Formatted: Space After: 0

Formatted: Font: (Default)
11 pt

Formatted: Space After: 0

Formatted: Font: (Default)
11 pt

Formatted: Space After: 0

Formatted: Font: (Default)
11 pt

Formatted: Space After: 0

Formatted: Font: (Default)
11 pt

Formatted: Space After: 0

Formatted: Font: (Default)
11 pt

Formatted: Space After: 0

Formatted: Font: (Default)
11 pt

Formatted: Space After: 0

Formatted: Space Before:

Formatted: Font: 12 pt, Bo

Formatted: Space Before:
single

Formatted: Font: Bold

Formatted: Font: 12 pt

Formatted: Line spacing: s

Formatted: Line spacing: s

Table 4. Performance of stationary and nonstationary models for Toronto Pearson International Airport

<u>Time Slice</u>	<u>Model</u>	<u>Location parameter</u>	<u>Scale parameter</u>	<u>Shape parameter</u>	<u>AIC_c</u>	<u>Bayes-factor</u>	<u>LB (100vr)</u>	<u>UB (100vr)</u>	<u>UB/LB</u>
<u>15-min</u>	<u>GEV₁-0</u>	<u>37.02</u>	<u>19.83</u>	<u>-0.073</u>	<u>-465.05</u>	<u>-</u>	<u>78.85</u>	<u>209.56</u>	<u>2.66</u>

	GEV _t -I	$30.11 + 0.194t$	20.86	-0.079	-450.28	4.98	105.3	229.77	2.18
	GEV _t -II	$34.30 + 0.056t$	$\exp(2.68 + 0.0069t)$	-0.11	-383.58	10.47	87.66	119.42	1.36
30-min	GEV _t -0	25.65	13.14	0.019	-442.29	-	58.8	155.85	2.65
	GEV _t -I	$17.32 + 0.21t$	13.44	-0.075	-422.67	1.47	57.9	113.37	1.96
	GEV _t -II	$12.08 + 0.35t$	$\exp(2.77 + 0.0023t)$	-0.20	-351.22	74357.2	58.63	99	1.69
1-hr	GEV _t -0	19.77	7.79	0.07	-477.68	-	45.47	101.33	2.22
	GEV _t -I	$18.27 + 0.022t$	8.59	-0.08	-402.43	78.53	42.27	63.57	1.50
	GEV _t -II	$4.44 + 0.414t$	$\exp(1.71 + 0.015t)$	0.044	-372.2	4.43×10^9	49.65	87.11	1.75
2-hr	GEV _t -0	11.79	4.45	0.11	-477.64	-	28.24	58.52	2.07
	GEV _t -I	$11.0 + 0.02t$	4.74	-0.02	-449.02	13.95	27.24	40.98	1.50
	GEV _t -II	$11.46 - 0.0053t$	$\exp(1.52 - 0.00072t)$	0.28	-421.44	9.08	46.44	61.47	1.32
6-hr	GEV _t -0	4.98	1.50	0.26	-488.39	-	13.71	21	1.53
	GEV _t -I	$5.12 + 0.0005t$	1.57	0.24	-496.92	0.15	12.02	29.77	2.48
	GEV _t -II	$5.44 - 0.0049t$	$\exp(0.77 - 0.0042t)$	0.10	-424.07	52.13	13.71	21.0	1.53
12-hr	GEV _t -0	2.96	0.70	0.36	-503.23	-	6.59	25.72	3.90
	GEV _t -I	$3.02 - 0.0031t$	0.69	0.51	-501.42	1.39	12.4	21.98	1.77
	GEV _t -II	$3.13 - 0.0045t$	$\exp(-0.183 - 0.0032t)$	0.49	-511.69	0.86	12.89	20.58	1.60
24-hr	GEV _t -0	1.71	0.41	0.29	-477.04	-	3.69	11.71	3.17
	GEV _t -I	$1.73 - 0.0006t$	0.41	0.28	-466.25	13.22	3.75	10.41	2.78
	GEV _t -II	$1.66 + 0.00093t$	$\exp(-1.00 + 0.00274t)$	0.30	-460.06	1.30	4.28	8.16	1.91

* GEV_t-0 is stationary model whereas GEV_t-I and GEV_t-II are nonstationary models with time-variant mean, and both time-variant mean and standard deviation respectively. The selected best fitted nonstationary model is marked in bold letters. Bayes factor, $\gamma < 1$ indicates that the nonstationary model fits better than the stationary model. However, in cases $\gamma > 1$, to compare with stationary model, the nonstationary model is selected following minimum AICc criteria. LB and UB indicates lower and upper bound of DSI at 100-year return period.

Table 5. Performance of stationary and nonstationary models for Hamilton Airport

Time Slice	Model	Location parameter	Scale parameter	Shape parameter	AICc	Bayes-factor	LB (100yr)	UB (100yr)	UB/LB
15-min	GEV _t -0	53.84	14.96	0.12	-347.44	-	103.8	221.98	2.14
	GEV _t -I	$56.31 - 0.096t$	14.2	0.14	-338.32	0.67	102.19	223.91	2.19
	GEV _t -II	$55.86 - 0.114t$	$\exp(2.83 - 0.0056t)$	0.19	-351.58	2.07	107.81	285.01	2.64

30-min	GEV _t -0	27.1	7.32	0.20	-369.40	-	56.71	174.66	3.08
	GEV _t -I	28.002-0.06t	7.28	0.11	-346.29	1.73	53.93	99.27	1.84
	GEV _t -II	27.8-0.038t	exp(1.91+0.0009t)	0.25	-365.19	0.28	69.81	110.76	1.59
1-hr	GEV _t -0	21.79	6.41	0.13	-361.35	-	41.92	109.07	2.60
	GEV _t -I	21.33+0.026t	7.06	0.03	-353.4	0.62	45.85	75.54	1.65
	GEV _t -II	20.50+0.046t	exp(1.86+0.0035t)	-0.0039	-350.97	3.09	43.75	68.8	1.57
2-hr	GEV _t -0	12.63	3.68	0.11	-349.70	-	25.81	51.37	1.99
	GEV _t -I	12.15+0.006t	3.76	0.21	-322.00	4.91	32.16	54.78	1.70
	GEV _t -II	11.53+0.042t	exp(1.09+0.0087t)	0.19	-329.09	11.20	32.76	49.51	1.51
6-hr	GEV _t -0	5.32	1.33	0.23	-389.88	-	10.24	32.46	3.17
	GEV _t -I	5.15+0.0037t	1.28	0.29	-396.75	0.94	14.51	21.47	1.48
	GEV _t -II	5.09+0.0052t	exp(0.12+0.0038t)	0.28	-375.03	1.21	14.53	20.47	1.41
12-hr	GEV _t -0	3.11	0.74	0.20	-369.54	-	5.86	15.58	2.66
	GEV _t -I	3.09+0.0022t	0.74	0.27	-366.46	1.73	14.51	21.47	1.48
	GEV _t -II	3.03+0.0023t	exp(-0.305+0.0002t)	0.21	-363.03	1.12	13.97	6.35	2.20
24-hr	GEV _t -0	1.44	0.49	0.16	-338.35	-	3.05	11.47	3.76
	GEV _t -I	1.36+0.0026t	0.48	0.22	-338.33	0.31	3.26	8.42	2.58
	GEV _t -II	1.33+0.0034t	exp(-0.74-0.00019t)	0.20	-326.63	0.99	4.04	6.44	1.59

Table 6. Performance of stationary and nonstationary models for Windsor Airport

Time Slice	Model	Location parameter	Scale parameter	Shape parameter	AIC _c	Baves-factor	LB (100vr)	UB (100vr)	UB/LB
15-min	GEV _t -0	60.04	15.76	0.13	-394.2	-	106.43	300.24	2.82
	GEV _t -I	61.6-0.099t	14.61	0.25	-370.00	0.80	157.9	227.06	1.44
	GEV _t -II	63.33-0.068t	exp(2.67+0.0027t)	0.013	-376.75	5.36	115.47	166.94	1.44
30-min	GEV _t -0	38.92	12.94	0.06	-443.89	-	72.9	179.38	2.46
	GEV _t -I	43.20-0.124t	12.04	0.12	-435.12	0.25	72.6	210.06	2.89

Formatted

Formatted

Formatted

Formatted Table

Formatted

	GEV _t -II	$25.2-0.18t$	$\exp(2.68-0.0194t)$	0.052	-494.92	206.09	40.8	116.9	2.87
2-hr	GEV _t -0	11.93	4.57	0.046	-501.85	-	26.04	56.4	2.17
	GEV _t -I	$13.29-0.044t$	4.49	0.093	-496.55	1.32	30.05	47.1	1.57
	GEV _t -II	$12.60-0.029t$	$\exp(1.42+0.00196t)$	0.20	-462.71	1.27	37.68	54.6	1.45
6-hr	GEV _t -0	5.196	1.47	0.082	-498.12	-	9.79	19.9	2.03
	GEV _t -I	$5.80-0.018t$	1.35	0.058	-501.40	0.05	10.48	14.5	1.38
	GEV _t -II	$5.83-0.018t$	$\exp(0.32-0.0012t)$	0.099	-499.38	0.02	10.14	19.2	1.89
12-hr	GEV _t -0	3.09	0.80	-0.0013	-515.05	-	5.35	10.1	1.89
	GEV _t -I	$3.34-0.008t$	0.80	0.062	-511.60	0.13	6.32	8.7	1.38
	GEV _t -II	$3.49-0.011t$	$\exp(-0.22-0.002t)$	-0.026	-500.35	0.05	5.72	7.5	1.30
24-hr	GEV _t -0	1.72	0.63	-0.051	-473.40	-	3.14	6.3	2.03
	GEV _t -I	$1.98-0.008t$	0.61	-0.054	-450.07	0.17	3.57	5.0	1.41
	GEV _t -II	$2.036-0.008t$	$\exp(-0.45-0.0007t)$	-0.103	-435.8	0.12	3.44	4.9	1.44

Formatted

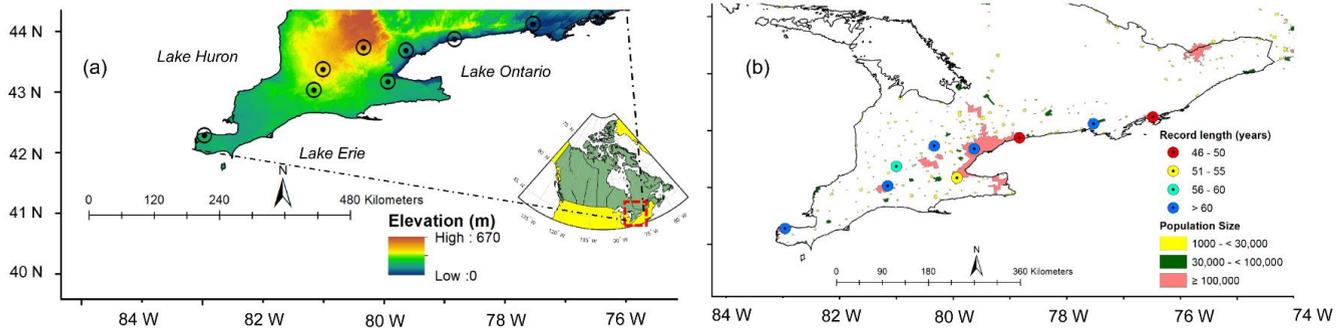
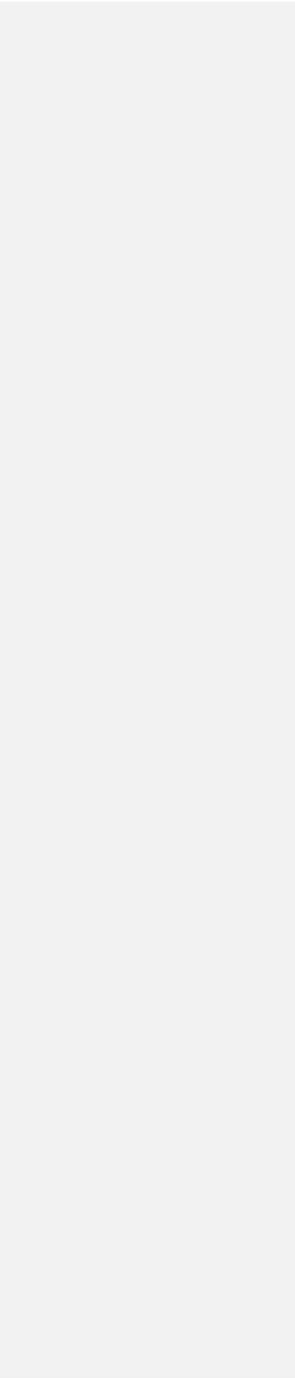
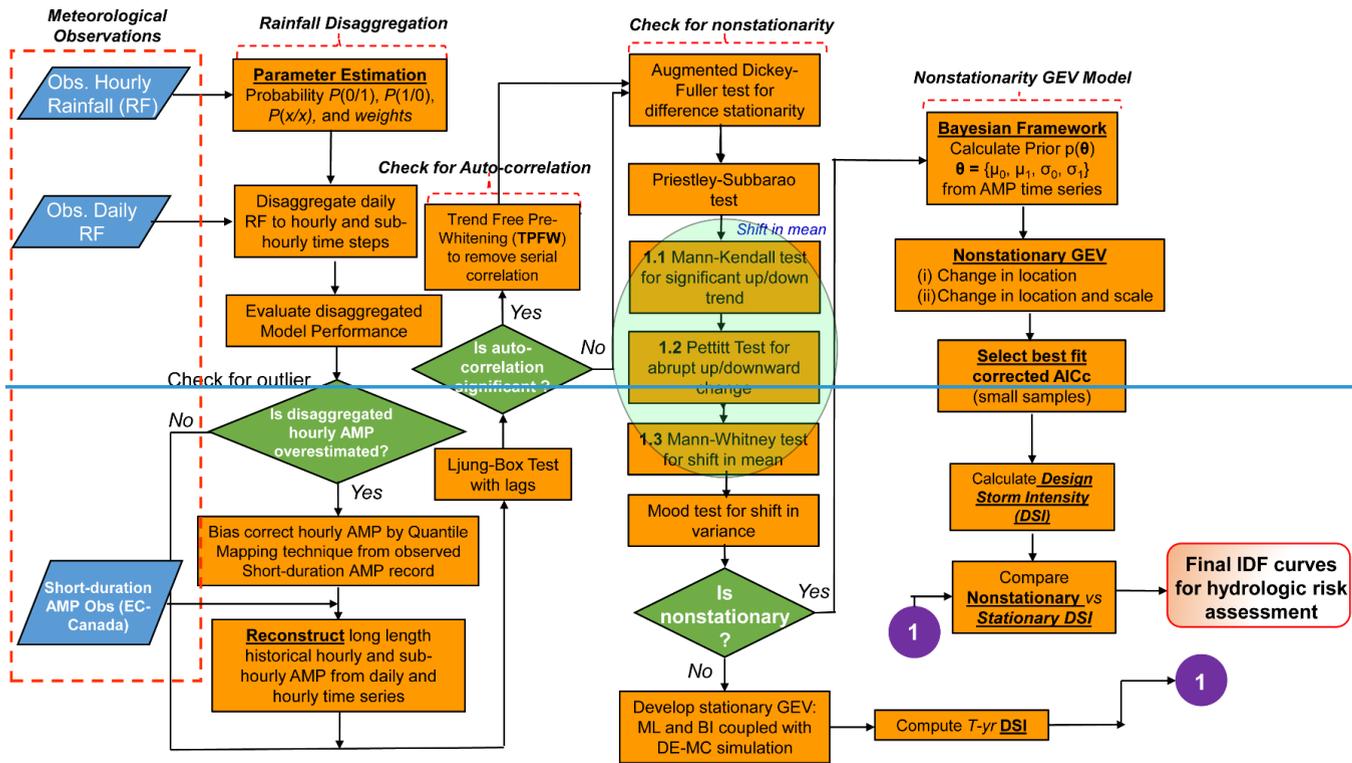


Figure 1. (a) Selected urbanized sites in Southern Ontario. The Southern Ontario (41° - 44° N, 84° - 76° W) is the southernmost region of Canada and is situated on a southwest-northeast transect, bounded by lakes Huron, Erie, and Ontario. The nine locations on the map are (from southwest to northeast corner): Windsor Airport, London International Airport, Stratford Wastewater Treatment Plant (WWTP), Fergus Shand Dam, Hamilton Airport, Toronto International Airport, Oshawa Water Pollution Control Plant (WPCP), Trenton Airport, and Kingston Pumping Station. Topography map indicates the maximum slope of 670 m above mean sea level. (b) The population map shows six the sites: Windsor Airport, London International Airport, Hamilton Airport, Toronto International Airport, Oshawa WPCP, and Kingston P. Station to be located either in or the vicinity of densely populated urbanized area. The remaining three sites are located in the moderately populated area. The short-duration AMP rainfall records in all locations vary between the minimum of 46 and maximum of 66 years.

Formatted: Font color: Cus

|





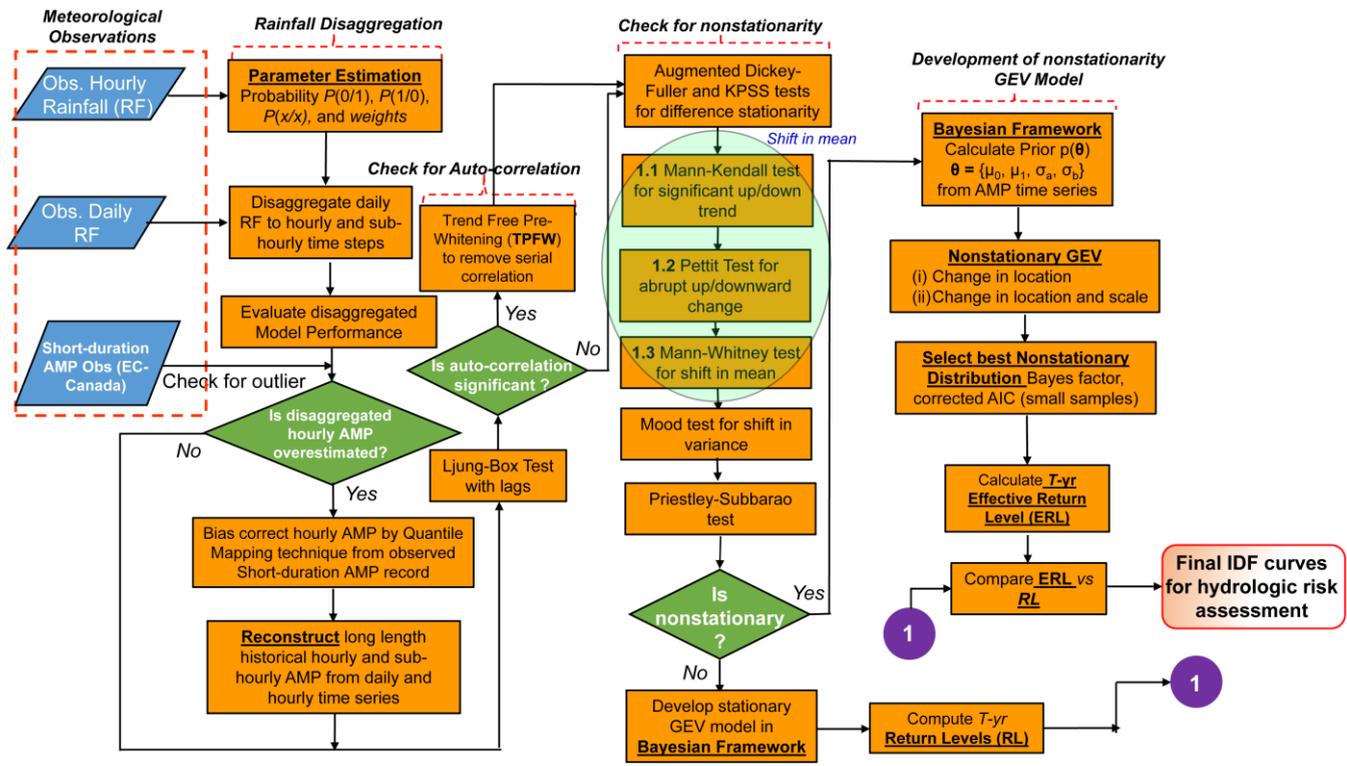
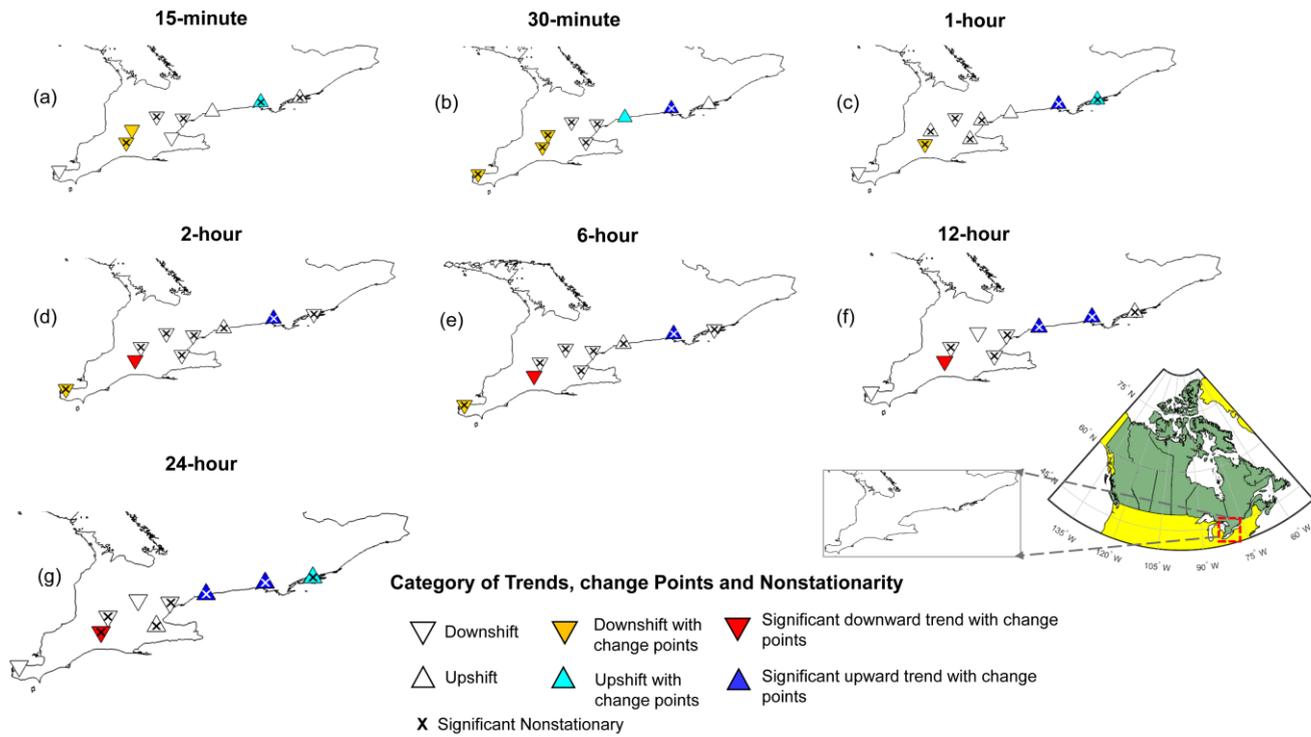


Figure 2. Schematics of the process flow (Blue - input step, orange - process step, and green – decision steps). All three tests – Mann-Kendall, Pettitt’s and Mann-Whitney, check for shifts in the mean. While Mann-Kendall tests for monotonic trends, the other two tests, Pettitt’s and Mann-Whitney check for change point or regime shift in the time series.



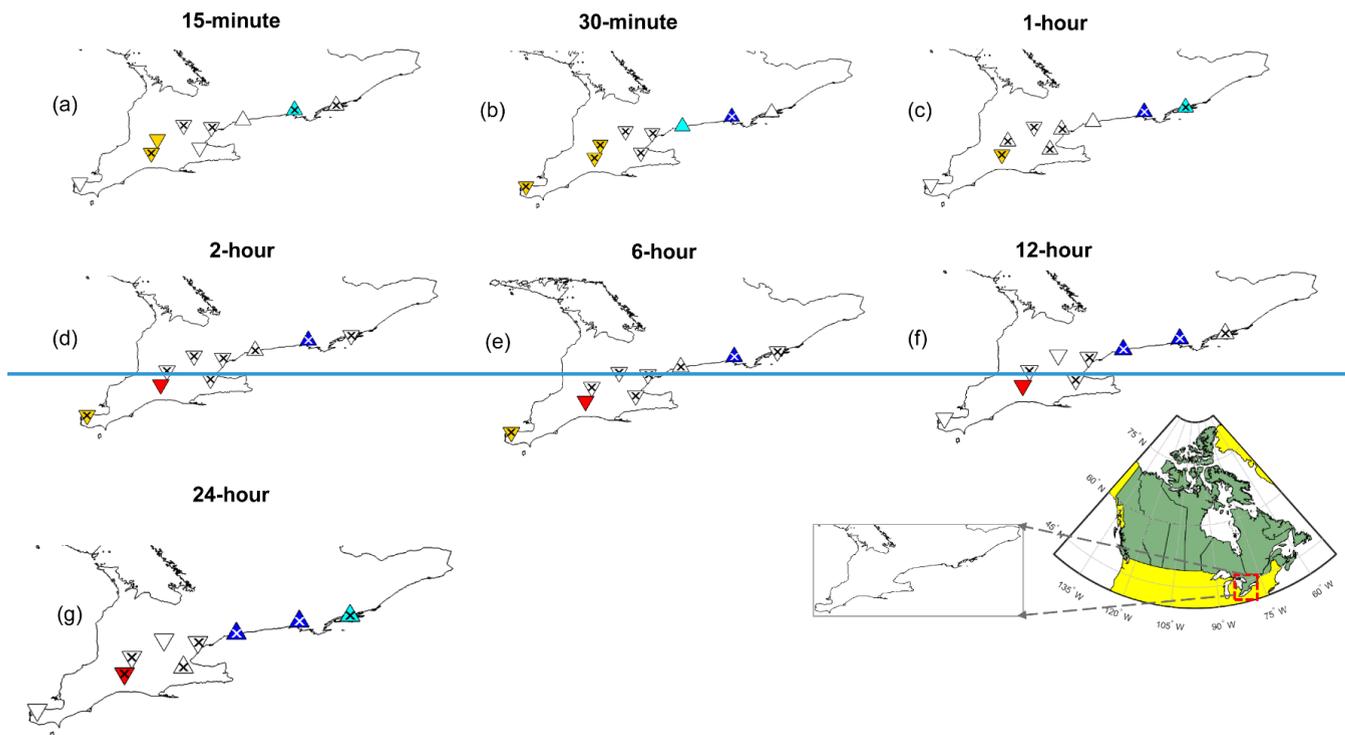
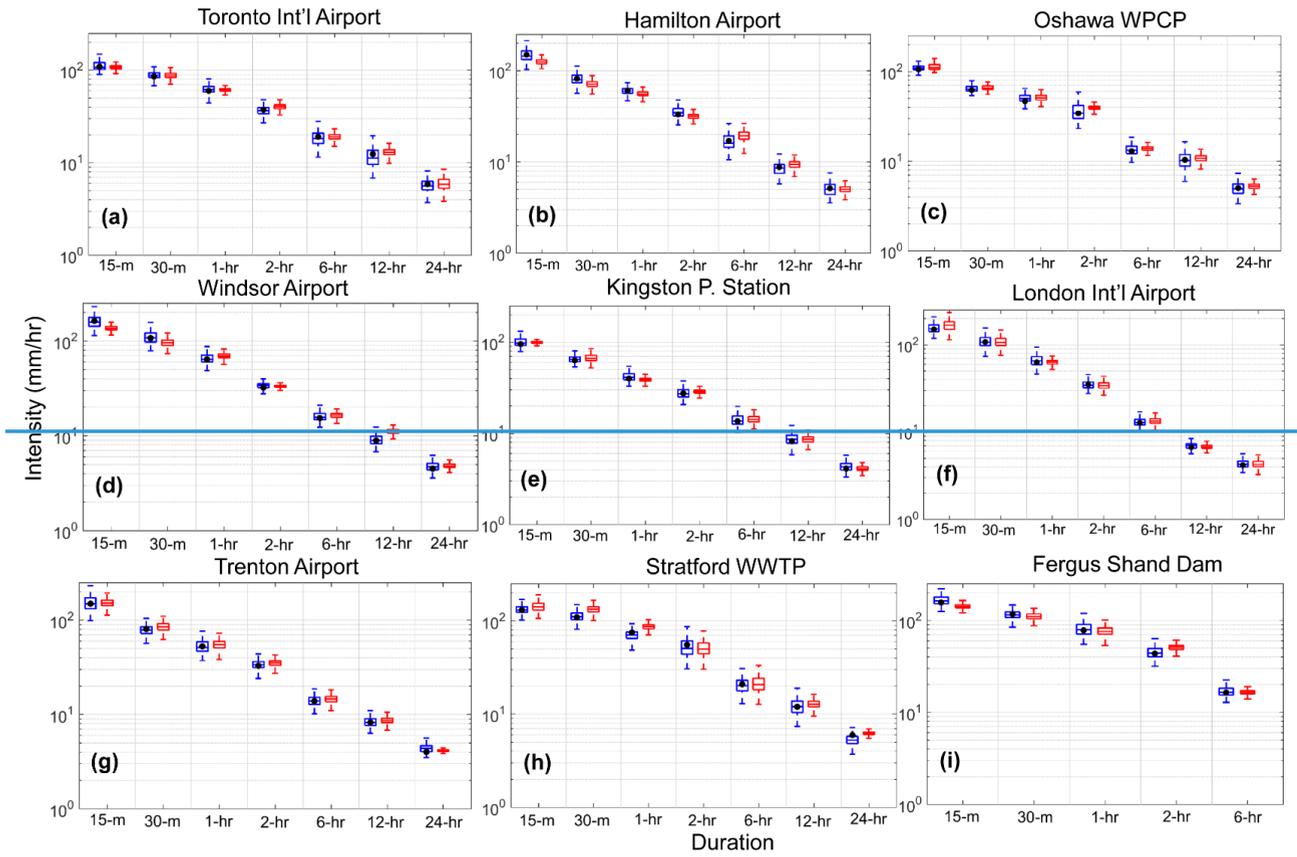


Figure 3. Spatial distribution of trends, change points and nonstationarities in hourly and sub-hourly rainfall extremes of several durations in nine urbanized locations, Southern Ontario (a – g). The population estimates in and the vicinity of urbanized sites varies between 5 Million and 19,000 with highest in Toronto Metropolitan Area and the lowest in Fergus Shand dam area. The up and down triangles in white indicate (statistically insignificant) up and downward shifts; the up and down triangles in cyan and orange indicate shifts with change points only; the up and down triangles in the dark blue and red show presence of (statistically significant) trends including change point(s). Sites with significant nonstationarity are marked with an ‘x’ sign. The trends in short duration AMP extremes are detected using nonparametric Mann-Kendall trend test with correction for ties. The change point in precipitation extremes are identified either as a shift in the mean or the variance in AMP time series. We apply nonparametric Pettitt (Pettitt, 1979) and Mann-

Formatted: Line spacing:

~~Whitney (Ross et al., 2011) tests to identify change point in the mean and Mood's test (Ross et al., 2011) to detect change point in the variance respectively (See section 2 for details). All tests are performed at 5 and 10% significance levels, i.e., $p\text{-value} < 0.10$.~~



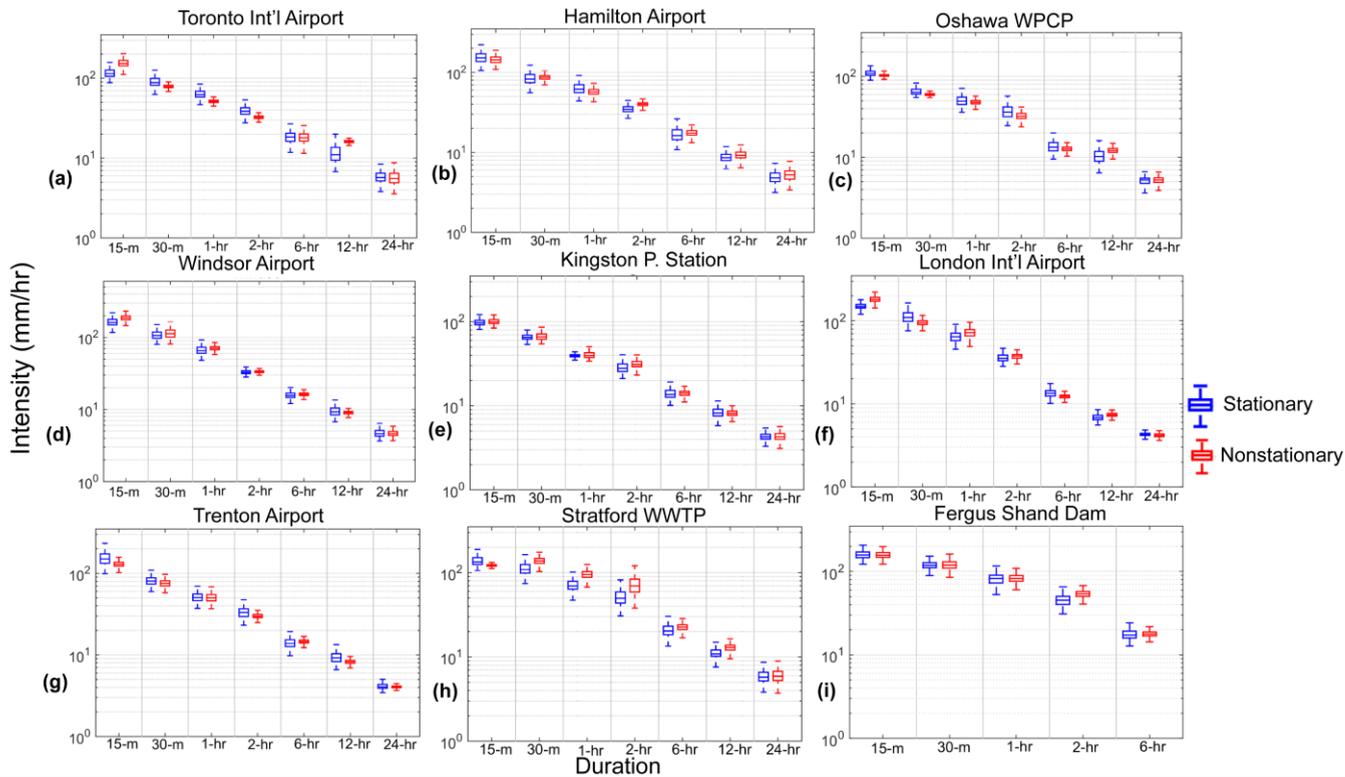


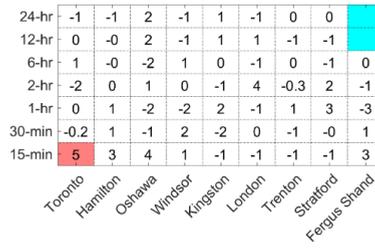
Figure 4. DSI estimates of the median (horizontal line within the box plot) and 95% credible intervals for 100-year return periods of stationary versus nonstationary models. Uncertainty in DSI for 100-year return periods for stationary (blue) versus nonstationary (red) models with durations ranging between 15-min and 24-hr for across nine sites (a - i): Toronto International Airport, Hamilton Airport, Oshawa WPCP, Windsor Airport, Kingston P. station, London International Airport, Trenton Airport, Stratford WWTP, and Fergus Shand Dam). The boxplots indicate the uncertainty in estimated DSI from using Bayesian inference, whereas the DSI obtained from maximum likelihood approach is shown as a black dot in the stationary simulation.

Formatted: Line spacing:

5

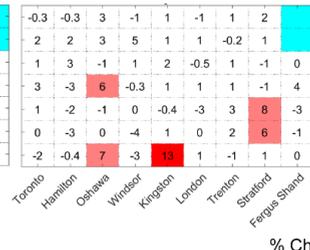
(a)

2-Year



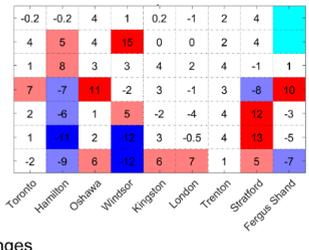
(b)

10-Year



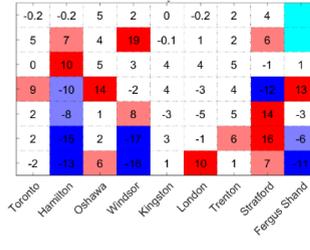
(c)

50-Year

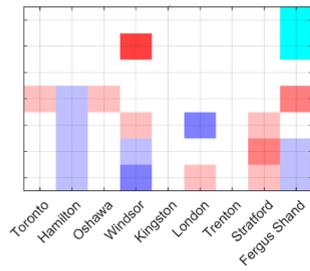
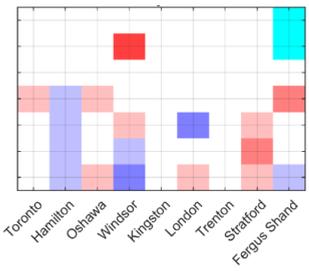
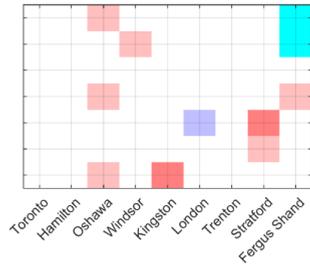
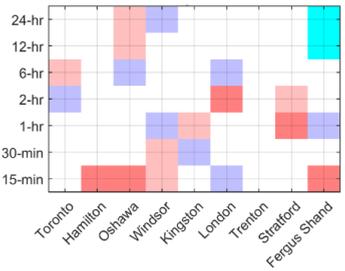
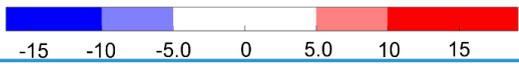


(d)

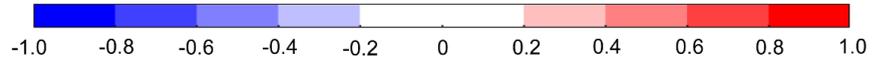
100-Year



% Changes



Z Statistics



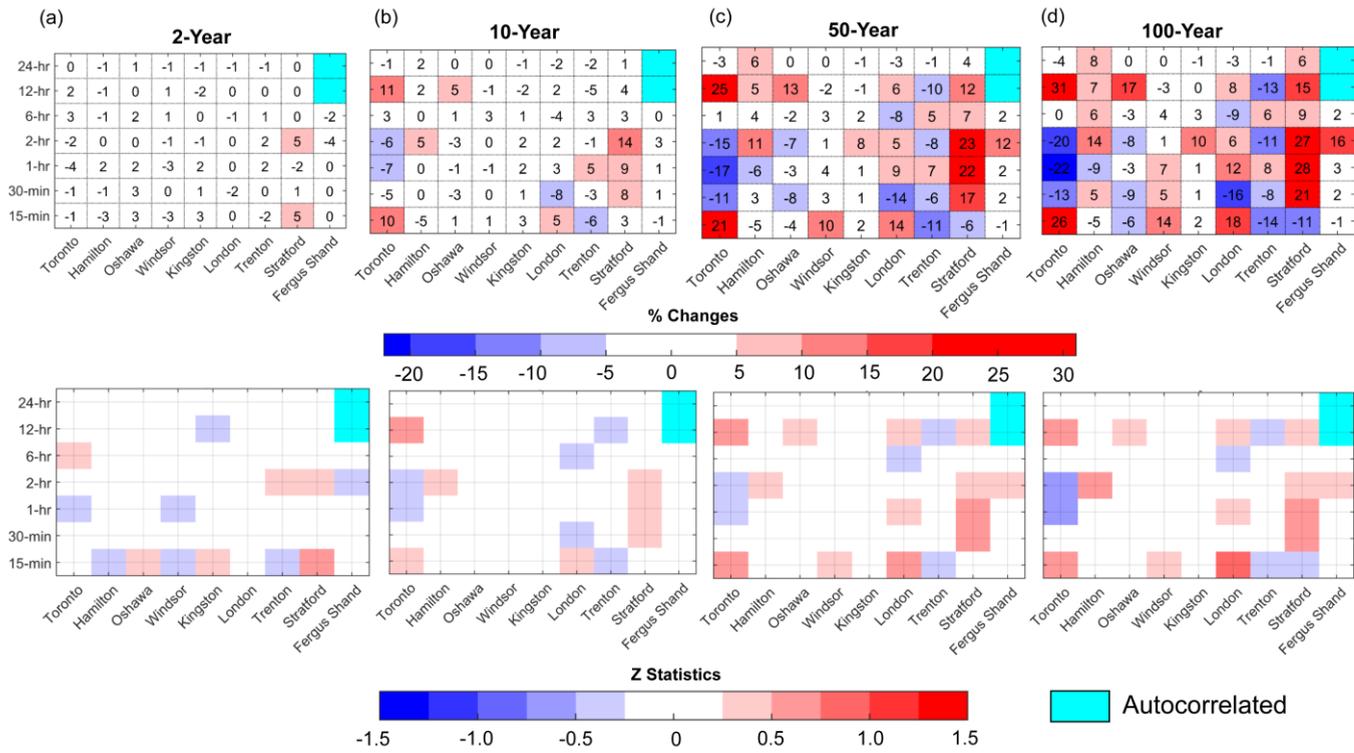
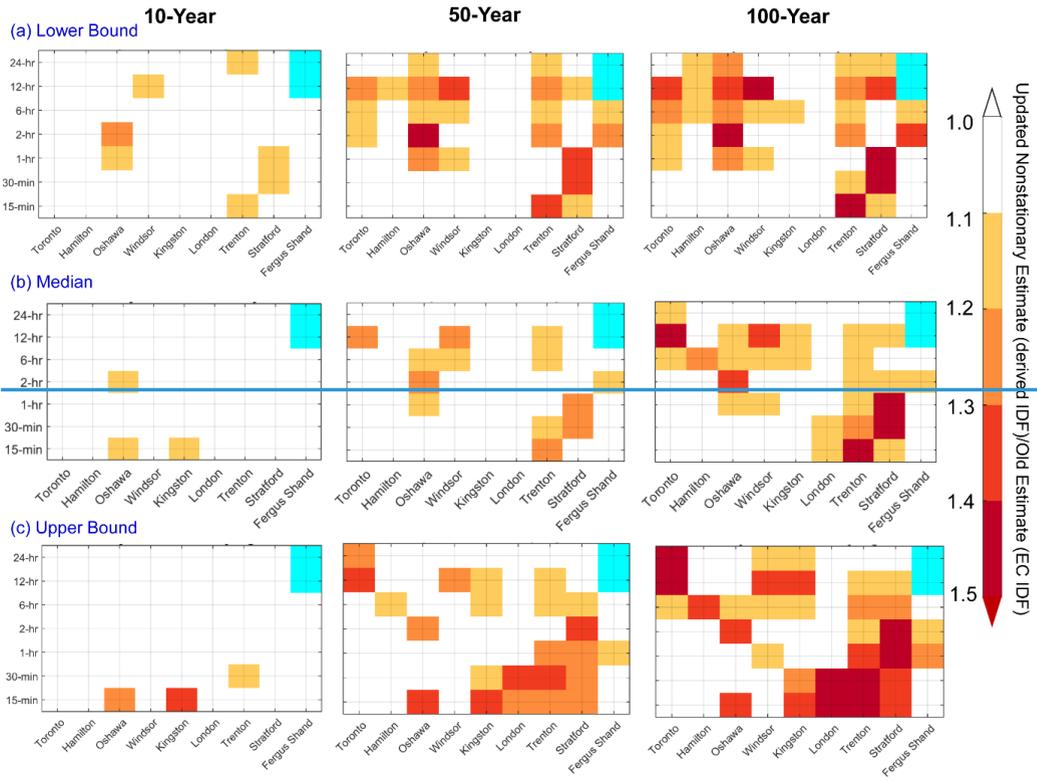
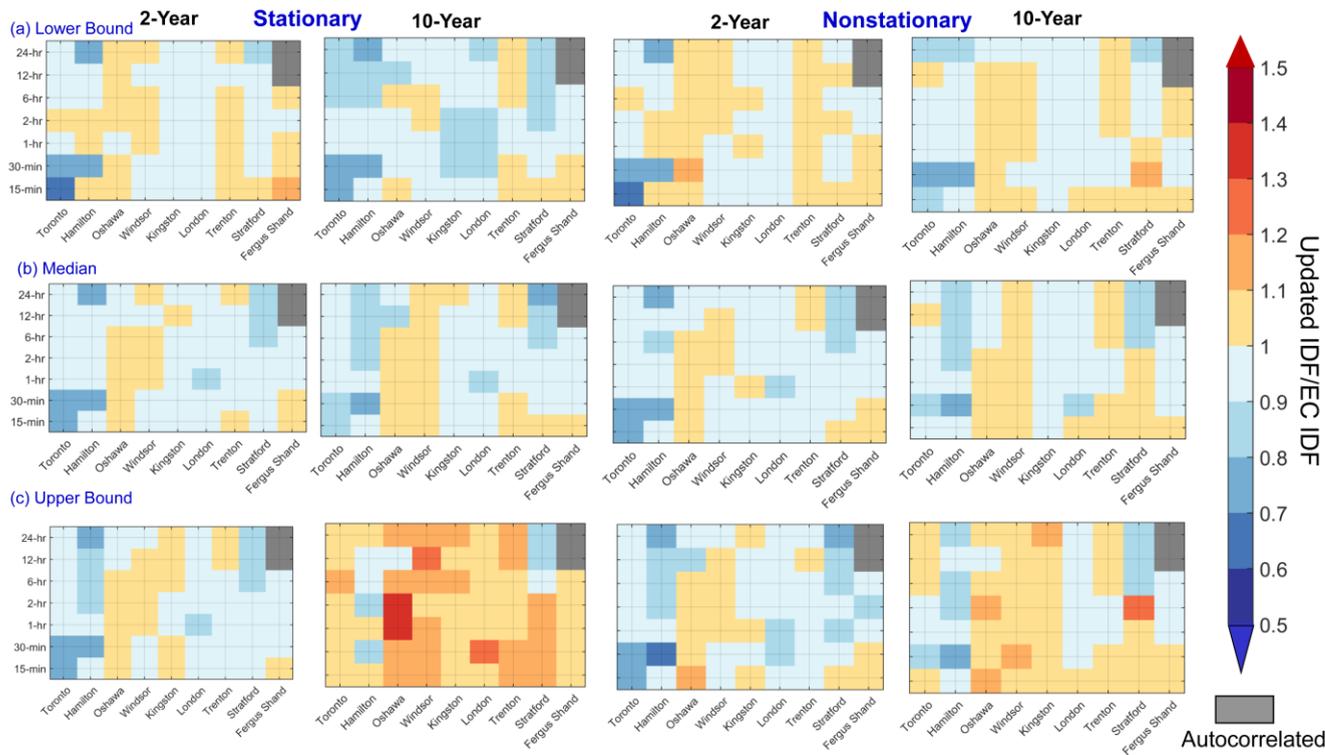


Figure 5. Percentage changes (in top panel) and Z-statistics (bottom panel) of at-site T -year event estimates for $T = 2$ -year to $T = 100$ -year return periods (a – d) with durations between 15-min and 24-hr in nine urbanized locations, Southern Ontario. The Z-statistic represents statistical significance of differences in DSI obtained from the best selected nonstationary versus the stationary model. The Z-statistic is statistically significant when $|Z| > 1.96$ and $|Z| > 1.64$ at 5% and 10% significance levels. The shades in blue and red denote decrease and increase in Z-statistics with the strength of shading represents the magnitude of the test statistics. The cyan shading represents the site durations with significant autocorrelation, which were excluded from the analysis.

Formatted: Line spacing:



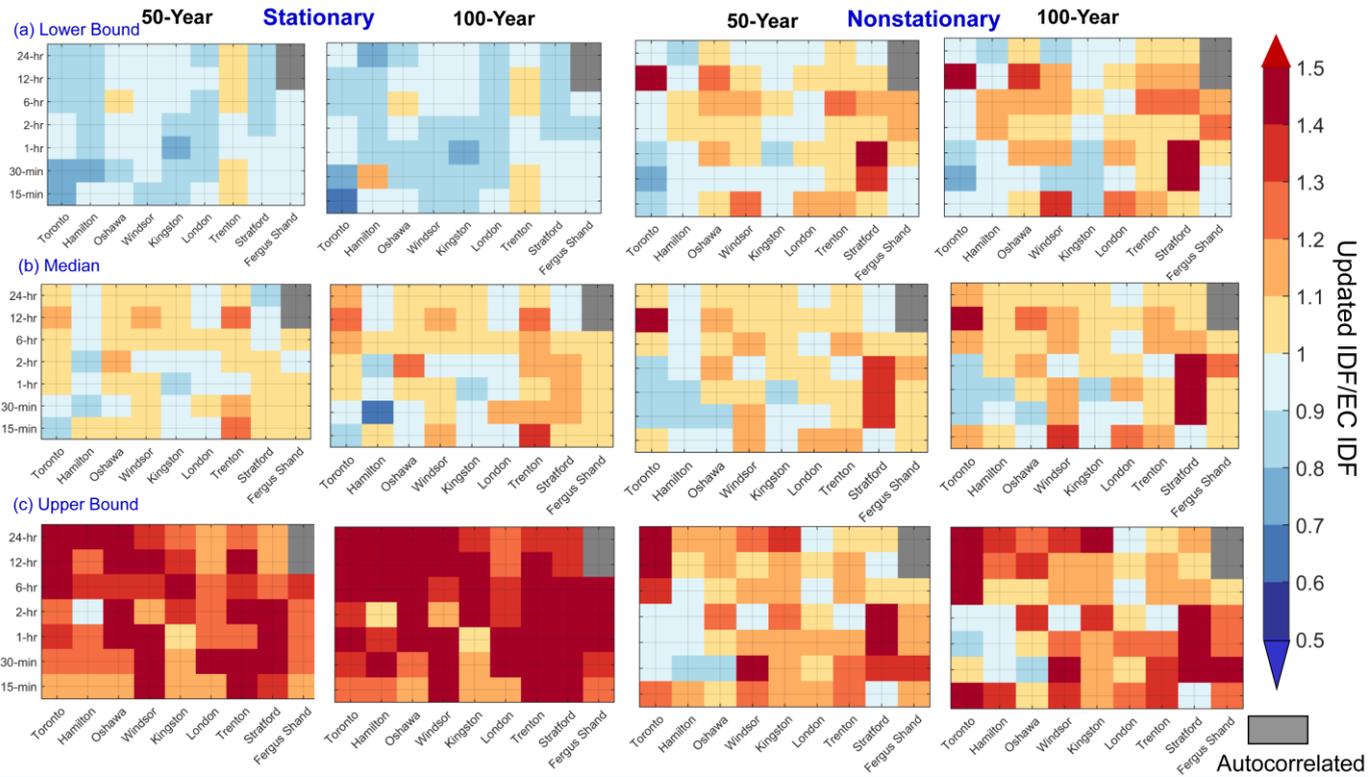


Formatted: Left: 0.93", Right: distance from edge: 0.51"

Figure 6. Central tendency (median, **b**) and the bounds (5th and 95th range credible interval, **a** and **c**) of the updated nonstationary versus EC-generated $T = 2$ - and 10-year event estimates for DSI at selected return periods with durations between 15-min and 24-hr. The minimum, median and maximum T year event estimates of nonstationary models are obtained from time variant GEV parameter(s) by computing the 5th, 50th and 95th percentiles of DE-MC sampled parameters. The DSI and associated 95% confidence limits of EC-generated IDF is obtained from the national archive of Engineering Climate Datasets maintained by the Environment Canada (<http://climate.weather.gc.ca/>). The shades in blue and red denote

decrease and increase in DSI. The strength of shading represents the magnitude of the ratio between updated versus EC-generated DSI. ~~with a deeper shade indicates an increase in the ratio. The cyan shading indicates the site with significant autocorrelation.~~

5



Formatted: Left: 0.8", Right: 0.51"

5 Figure 7. Central tendency (median, b) and the bounds (95% credible interval, a and c) of the updated nonstationary versus EC-generated $T = 50$ - and 100-year event estimates for DSI at selected return periods with durations between 15-min and 24-hr. The DSI and associated 95% confidence limits of EC-generated IDF is obtained from the national archive of Engineering Climate Datasets (<http://climate.weather.gc.ca/>). The shades in blue and red denote decrease and increase in DSI. The strength of shading represents the magnitude of the ratio between updated versus EC-generated DSI.

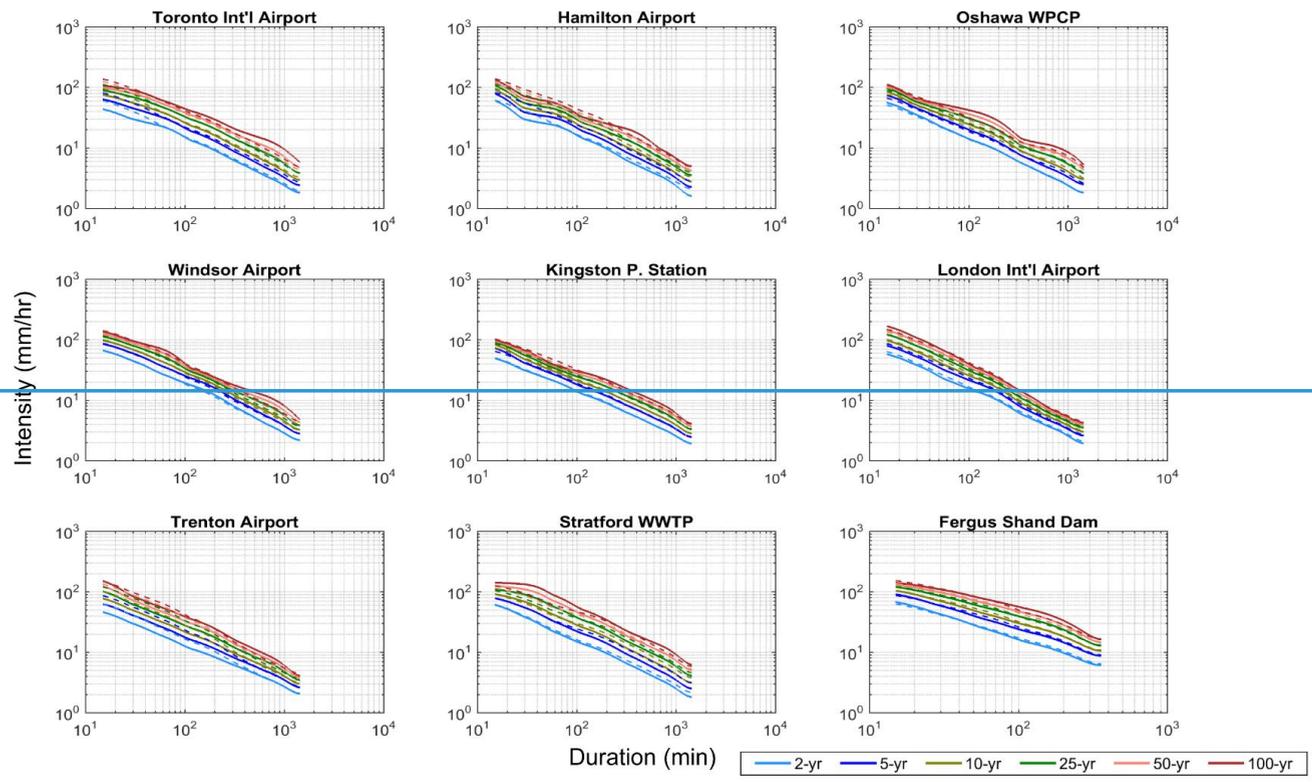
Formatted: Font: Not Bold
 Formatted: Font: Not Bold
 Formatted: Font: Not Bold
 Formatted: Font: Not Bold
 Formatted: Font: Not Bold

|



Formatted: Figure or Table

Formatted: Left: 0.93", Right: 0.3", Footer distance from e



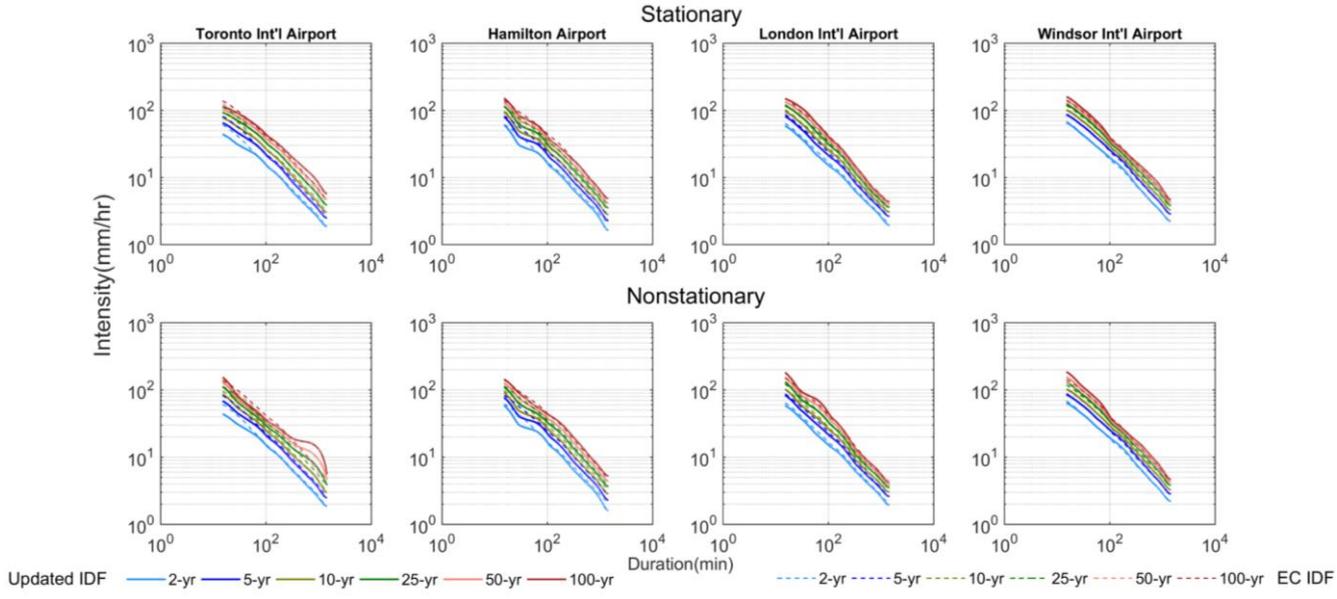


Figure 78. Estimated nonstationary versus EC-generated IDF for ~~return periods~~ $T = 2, 5, 10, 25, 50$ and 100-year ~~return periods~~ for the selected urbanized locations in Southern Ontario, Canada. The updated and EC IDFs are shown using solid and dotted lines respectively. The nonstationary IDFs are shown using solid lines, while EC-generated IDFs are shown using dotted lines.

Formatted: Space Before: