

The effect of modeling choices on updating intensity duration frequency curves and stormwater infrastructure designs for climate change: Supplementary Information

Lauren M. Cook^{12*}, Seth McGinnis³, and Constantine Samaras¹

¹Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, PA, USA

²Department of Urban Water Management, Swiss Federal Institute for Aquatic Research, Dübendorf, Switzerland

³National Center for Atmospheric Research, Boulder, CO, USA

*Corresponding author: Lmcook88@gmail.com

S1. Data

This section provides additional details about the historical data (Section S1.1) and regional climate model simulations (Section S1.2) that were used to create the updated IDF curves in 34 U.S. cities.

S1.1. Historical data

Table S1 presents the characteristics of the weather station where the data was obtained in each of the 34 cities used in this study, including elevation, latitude, longitude, and Bukovsky climate region. Observed data was available for the period 1950 to 2014 in all cities except Pittsburgh and Wichita, where data records begin in 1953 and 1954, respectively. For consistency, the historical period used in this study begins in 1954 in all cities.

Table S1. Characteristics of cities and weather stations used in this study

City	State	Bukovsky climate region	NOAA station name	Elevation (m)	Latitude	Longitude
Albuquerque	NM	S. Rockies	ALBUQUERQUE INTERNATIONAL AIRPORT NM US	1619	35	-106.6
Amarillo	TX	C. Plains	AMARILLO AIRPORT TX US	1100	35.2	-101.7
Baltimore	MD	Mid Atlantic	BALTIMORE WASHINGTON INTERNATIONAL AIRPORT MD US	13	39.3	-76.5
Benavides	TX	Mezquital	BENAVIDES TX US & BENAVIDES 2 TX US	116	27.6	-98.4
Billings	MT	N. Plains	BILLINGS INTERNATIONAL AIRPORT MT US	1087	45.8	-108.5
Birmingham	AL	Deep South	BIRMINGHAM AIRPORT AL US	188	33.6	-86.7
Boise	ID	Great Basin	BOISE AIR TERMINAL ID US	815	43.6	-116.2

Boston	MA	North Atlantic	BOSTON MA US	4	42.4	-71
Boulder	CO	S. Rockies	BOULDER 2 CO US	1654	40	-105.3
Charlotte	NC	Mid Atlantic	CHARLOTTE DOUGLAS AIRPORT NC US	234	35.2	-80.9
Chicago	IL	Great Lakes	CHICAGO MIDWAY AIRPORT 3 SW IL US	187	41.8	-87.8
Detroit	MI	Great Lakes	DETROIT METROPOLITAN AIRPORT MI US	189	42.4	-83
El Paso	TX	Mezquital	EL PASO INTERNATIONAL AIRPORT TX US	1201	31.8	-106.4
Fargo	ND	N. Plains	FARGO HECTOR INTERNATIONAL AIRPORT ND US	273	46.9	-96.8
Indianapolis	IN	Appalachia	INDIANAPOLIS INTERNATIONAL AIRPORT IN US	247	39.7	-86.3
Jacksonville	FL	Southeast	JACKSONVILLE INTERNATIONAL AIRPORT FL US	12	30.4	-81.7
Las Vegas	NV	Southwest	LAS VEGAS MCCARRAN INTERNATIONAL AIRPORT NV US	659	36.1	-115.2
Los Angeles	CA	Pacific SW	LOS ANGELES INTERNATIONAL AIRPORT CA US	37	33.9	-118.4
Memphis	TN	Deep South	MEMPHIS INTERNATIONAL AIRPORT TN US	79	35.1	-90
Minneapolis	MN	Prairie	MINNEAPOLIS ST PAUL INTERNATIONAL AIRPORT MN US	254	44.9	-93.2
Missoula	MT	N. Rockies	MISSOULA INTERNATIONAL AIRPORT MT US	972	46.9	-114.1
New Orleans	LA	Southeast	NEW ORLEANS AIRPORT LA US	6	29.9	-90.1

New York	NY	North Atlantic	NY CITY CENTRAL PARK NY US	40	40.8	-74
Oklahoma City	OK	S. Plains	OKLAHOMA CITY WILL ROGERS WORLD AIRPORT OK US	390	35.4	-97.6
Phoenix	AZ	Southwest	PHOENIX AIRPORT AZ US	337	33.4	-112
Pittsburgh	PA	Appalachia	PITTSBURGH INTL AIRPORT	367	40.4	-80
Pocatello	ID	N. Rockies	POCATELLO REGIONAL AIRPORT ID US	1356	42.9	-112.6
Portland	OR	Pacific NW	PORTLAND INTERNATIONAL AIRPORT OR US	6	45.6	-122.6
Saint Louis	MO	Prairie	ST LOUIS LAMBERT INTERNATIONAL AIRPORT MO US	176	38.8	-90.4
San Antonio	TX	S. Plains	RANDOLPH AFB TX US	230	29.5	-98.3
San Jose	CA	Pacific SW	SAN JOSE CA US	29	37.3	-121.9
Seattle	WA	Pacific NW	SEATTLE TACOMA INTERNATIONAL AIRPORT WA US	113	47.4	-122.3
Wichita	KS	C. Plains	WICHITA DWIGHT D. EISENHOWER NATIONAL AIRPORT KS US	406	37.7	-97.4
Yakima	WA	Great Basin	YAKIMA AIRPORT WA US	324	46.6	-120.5

24

25 Data gaps existed in four cities: Detroit, Memphis, Seattle, and Birmingham. In this case, data
26 was substituted using data from the closest weather station where that data existed during the
27 missing years. **Table S2** presents the dates where data was supplemented, as well as characteristics
28 of the weather stations that were used to fill data gaps.

Table S2. Characteristics of weather stations used to substitute for missing data in select cities

City	Dates Missing (YYYY-MM-DD)	Station name of data supplement	Elevation (m)	Latitude	Longitude
Detroit	1954-01-01 to 1959-12-31	DETROIT CITY AIRPORT MI US	192.9	42.23333	-83.3333
Memphis	1995-09-01 to 1999-04-04	MUNFORD TN US	136.6	35.45	-89.8
Seattle	1954-01-01 to 1965-01-01	SEATTLE CITY OFFICE WA US	4	47.6	-122.33
Birmingham	1978-10-01 to 1987-07-31	BIRMINGHAM WSFO AL US	226.8	33.46	-86.83
Birmingham	2008-02-20 to 2009-12-31	WARRIOR AL US	158.5	33.79	-86.82

S1.2. Climate model data collection

This analysis uses hourly rainfall time series from RCMs from the North American Coordinated Regional Downscaling Experiment (NA-CORDEX) project (Mearns et al. 2017). For each city, we extracted the rainfall time series from the RCM grid cell whose centroid was closest to the location of the NOAA weather station in the city (coordinates are provided in **Table S1**). Best practices for impacts analyses usually recommend using the 9 or 25 grid cells surrounding each station. However, we chose to use a single grid cell because including values from adjacent grid cells increases uncertainty and makes it difficult to distinguish differences in uncertainty from the modeling assumptions that are the focus of the study. To determine how this choice would alter the future distribution, we analyzed the raw RCM data for the 25 grid cells surrounding each city in **Table S1**. The median daily intensity for each return period of the uncertainty range including each climate model and spatial resolution changes by +/- 10%. **Figures S1(a) to S1(e)** show these results for five selected cities, Albuquerque, Charlotte, El Paso, New York, and Pittsburgh.

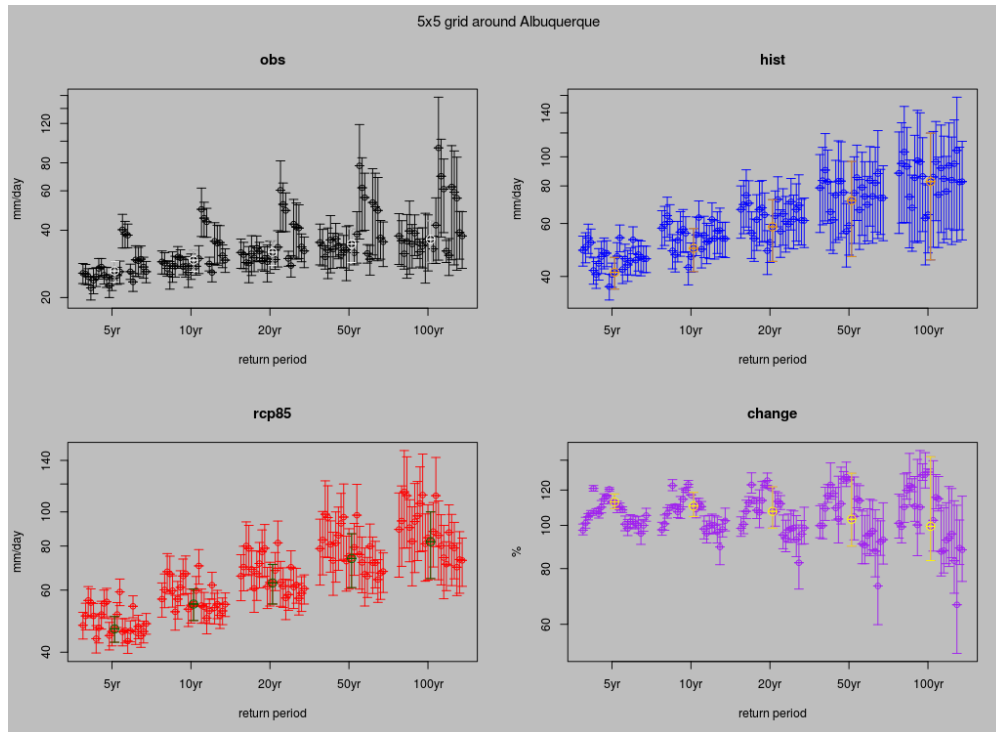


Fig S1(a). Results for the daily return levels for the 25 grid cells surrounding Albuquerque. The change represents the difference between the historical (hist) and future (rcp85) values. The central grid cell is shown in a contrasting color.

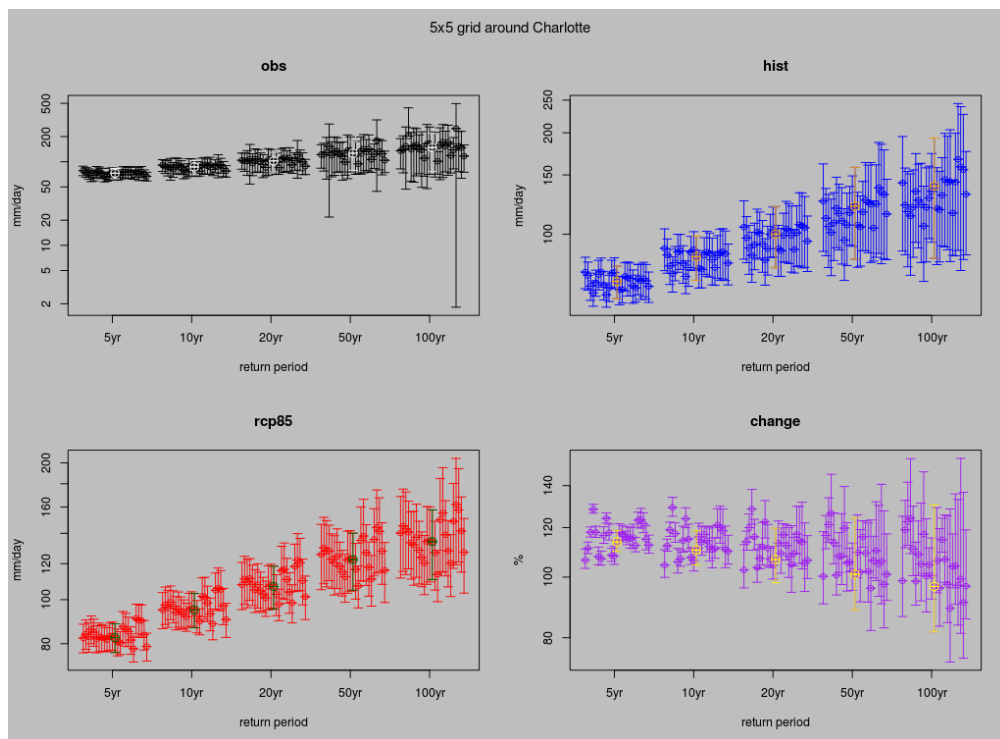


Fig S2(b). Results for the daily return levels for the 25 grid cells surrounding Charlotte. The change represents the difference between the historical (hist) and future (rcp85) values. The central grid cell is shown in a contrasting color.

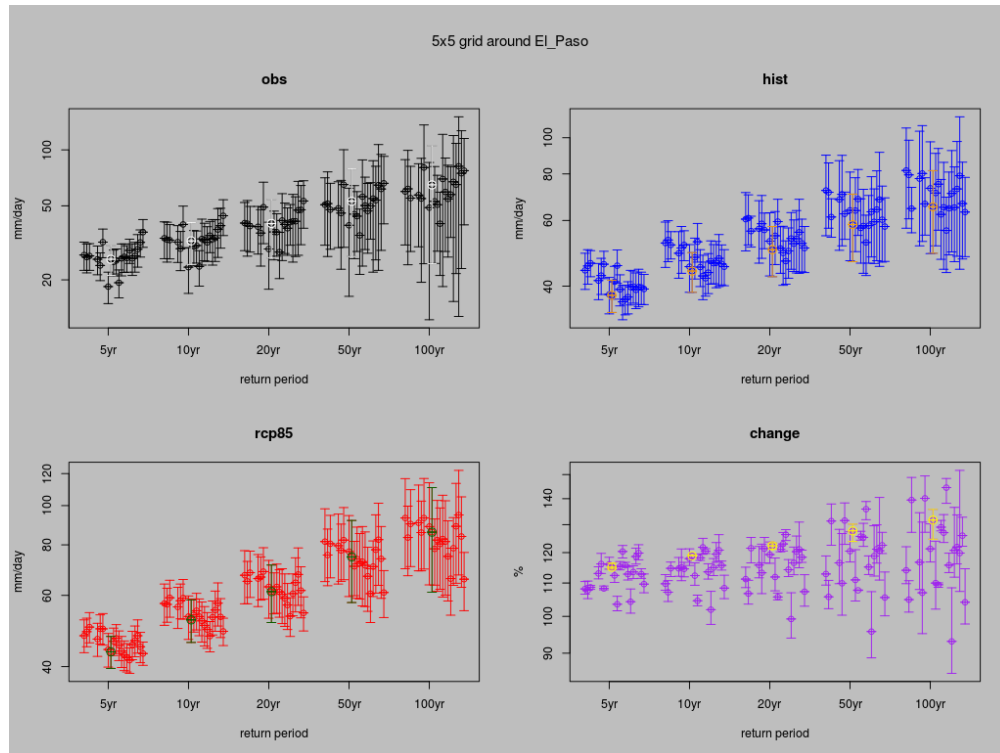


Fig S3(c). Results for the daily return levels for the 25 grid cells surrounding El Paso. The change represents the difference between the historical (hist) and future (rcp85) values. The central grid cell is shown in a contrasting color.

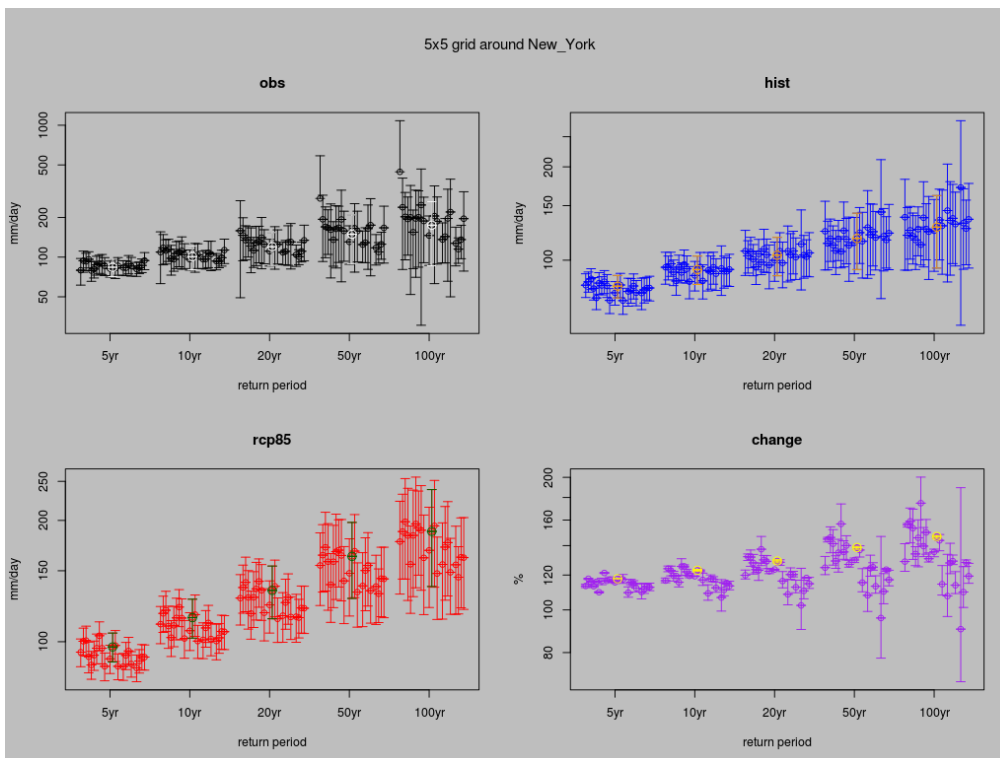


Fig S4(d). Results for the daily return levels for the 25 grid cells surrounding New York city. The change represents the difference between the historical (hist) and future (rcp85) values. The central grid cell is shown in a contrasting color.

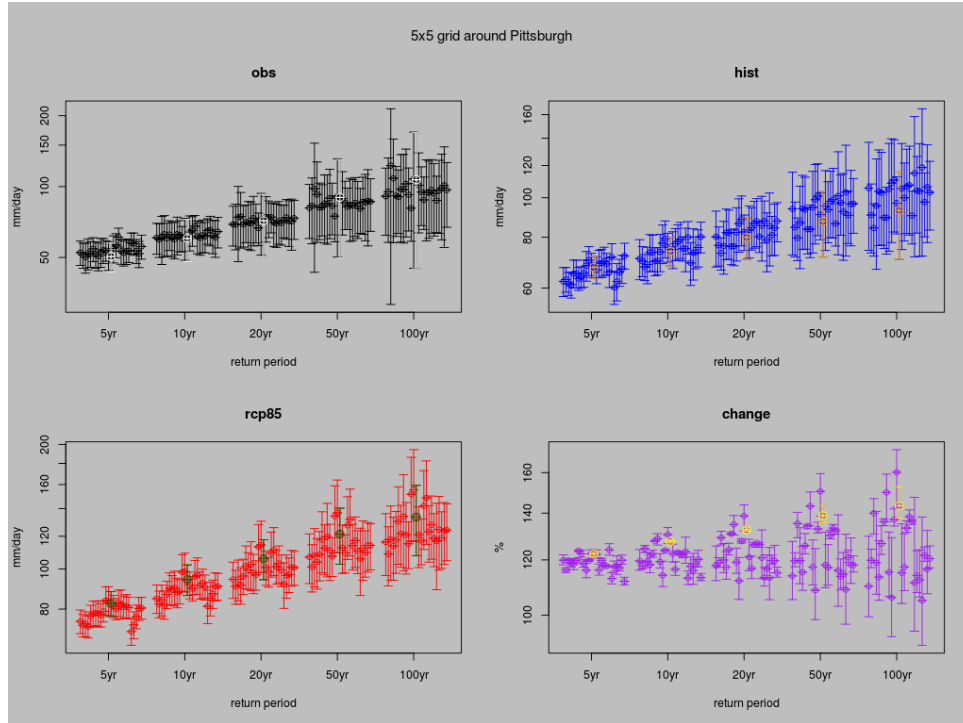


Fig S5(e). Results for the daily return levels for the 25 grid cells surrounding Pittsburgh. The change represents the difference between the historical (hist) and future (rcp85) values. The central grid cell is shown in a contrasting color.

S2. Detailed description of MOS techniques

This analysis uses three MOS techniques to incorporate RCM simulations into the IDF curves. The first method, Kernel Density Distribution Mapping (KDDM) (McGinnis, Nychka, and Mearns 2015), is a form of non-parametric quantile mapping that bias-corrects and spatially adjusts the continuous RCM time-series to match observations. The bias-correction can be completed before or after the time series are aggregated to the desired duration. The second method, Parametric AMS Mapping (PAM), is similar to the Equidistance Quantile Matching Method technique in Solaiman and Simonovic (2011), and uses parametric quantile mapping to adjust the annual maximum series (AMS) of the RCMs. The gridded AMS are spatially adjusted by finding their equivalent return level within the inverse GEV CDF of the observed AMS. The final method, the Change Factor (CF) method, is a “delta” type method that uses areal reduction factors to adjust the

observed rainfall depth up or down depending on the change between the historical and future climate model simulations, as seen in Cook et al. (2017) and Mailhot et al. (2007). **Figure S2** presents an overview of the sequence of steps required for each technique. Method (1a) and (1b) refer to the KDDM method used before and after aggregation, respectively. Additional details about each technique are provided in the following subsections.

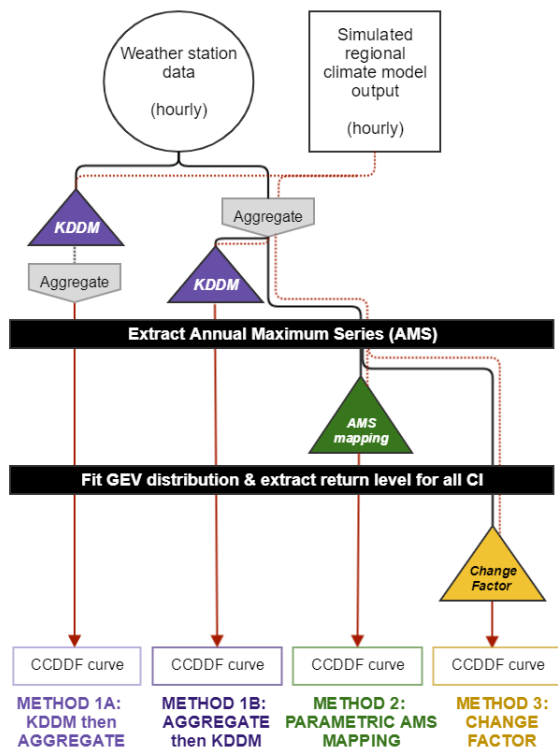


Figure S2. Sequence of steps carried out in each of the IDF adjustment techniques used in this study. This process is repeated for each climate model and location. Red, dashed lines represent inputs or outputs at the grid scale, while solid lines represent inputs (black) or outputs (red) at the station scale.

S2.1. Kernel Density Distribution Mapping

S2.1.1.1. Detailed description

The first method used for updating the IDF curves uses a type of non-parametric bias-correction, called *Kernel Density Distribution Mapping (KDDM)* (McGinnis, Nychka, and Mearns

2015) to adjust the underlying time series from the climate model. KDDM defines a relationship between the observed rainfall time series and the gridded climate model time series for the historical time period (1954–2013). Using this relationship, the method bias-corrects the entire gridded climate model time series, including zero values, to the station scale by adjusting the model values so that their statistical distribution in the historical time period matches that of the observations. Once the time series is bias-corrected, we extract the AMS and obtain the return levels and confidence intervals for each duration, climate model, and city using the general techniques in **Section S2.1**.

The method captures the relationship between the observed rainfall time series and the gridded climate model time series by fitting a transfer function between their empirical CDFs. First, it computes empirical PDFs using kernel density estimation, then calculates CDFs by integrating these PDFs using the trapezoid rule. The methods then maps equal points of probability from the CDFs against each other and fits the resultant mapping with a spline. The equation of this spline is the transfer function between the observed data and the historical climate model simulation output, which is then applied to the 150-year time series of the climate model simulations to obtain bias-corrected values at the station scale.

The time series can be bias-corrected with KDDM before or after it is aggregated to the desired duration (e.g., 24-hour). The former, referred to as method (1a), bias-corrects the 1-hour time series and then aggregates using convolution after bias-correction. The latter, referred to as method (1b), aggregates using convolution and then bias-corrects the convolved time series. Although the convolved time series contains many values that are not independent of one another, this does not present a problem for the KDDM technique, because it makes no attempt to fit a parametric

distribution. As long as the two datasets have similar dependency structures, the method remains applicable.

S2.1.1.2. Comparison of KDDM before or after aggregation

For IDF curves obtained at the underlying temporal resolution of the observations and climate model simulations (1 hour in this study), the KDDM technique is straightforward. It can be applied directly to the 1-hour time series from both the observations and the climate model simulations. We then extract the AMS from the bias-corrected time series and fit a GEV distribution to develop 1-hour duration IDF curves at the station scale. For IDF curves at durations greater than the underlying temporal resolution of the observations (i.e., greater than 1 hour), we tested two methods. The first method uses KDDM at the 1-hour time step and then aggregates the bias-corrected 1-hour time series to the desired duration using sliding maxima (van Montfort 1990).

The second method applies KDDM to the time series after aggregating the observations and climate model simulations to the desired duration through convolution. We aggregate both the observed and the climate model time series using the moving window approach: if the desired duration is 3 hours, the first timestep in the aggregated timeseries is the sum of the first three 1-hour timesteps of the original data, the second aggregated timestep is the sum of the second, third, and fourth timesteps from the original time series, and so on.

To determine whether one of these methods is preferable, we compare the AMS of the observed data (1950–2013) to the AMS of the bias-corrected time series for the same time period at a given duration using two metrics: (1) the mean absolute error (MAE) of 100 evenly spaced quantiles, and (2) a Kolmogorov–Smirnov (K-S) test of each AMS series. The former analyzes the difference in the empirical distributions, while the latter tests for equivalence in the PDFs.

We chose to use MAE instead of root mean squared error (RMSE) because the latter weights larger values more heavily, which can be misleading when analyzing non-Gaussian distributions such as annual maximum rainfall. Given the heavy-tailed distribution of rainfall, the MAE is more appropriate to test the model performance. MAE is calculated as follows:

$$MAE(d) = \frac{\sum_{q=1}^n |y_q - x_q|}{n} \quad [S1]$$

where: $MAE(d)$ is the mean absolute error of the empirical CDF for a given duration, d , y_q is the value of annual maximum series for a quantile, q , of the historical climate model simulation, and x_q is the value of the AMS series for the same quantile in the observation series, and n is the number of quantiles.

The Kolmogorov–Smirnov (K-S) test compares the probability distribution of one sample to another sample. If the null hypothesis is accepted ($h_0 = 0$), then the two distributions are not from different distributions. Since the KDDM method used the observed time series for bias-correction, the MAE is expected to be approximately 0, and the K-S test null hypothesis should be accepted.

Figure S3 reports results for the MAE and KS-test for methods (1a) and (1b) at the two climate model resolutions (25-km and 50-km) and various durations for 6 of the 34 cities. Durations are shown along the x-axis, and cities are shown in each subplot. The two methods are shown as different-colored error bars. The median, minimum, and maximum MAE of the climate model ensemble are shown as a square and the lower and upper bounds of the error bar, respectively. Filled squares represent the 25-km resolution climate model ensemble, while the open square represents the 50-km resolution. The number of stars at the top of the bar signifies the number of climate models where $h_0 = 1$ (rejected) for the K-S test, meaning that the historical climate model AMS is *not* from the same continuous distribution as the observed AMS.

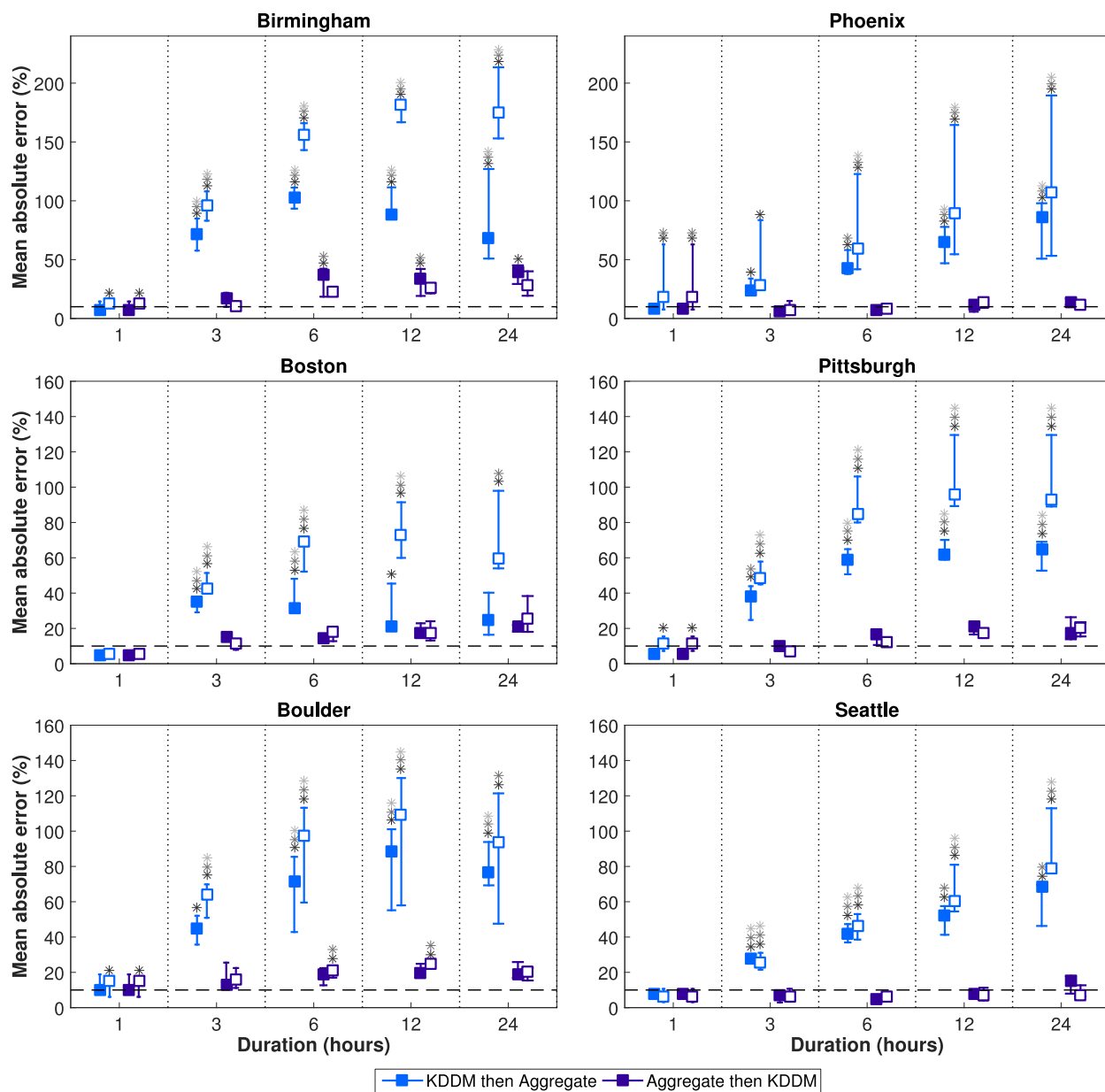


Figure S3. Mean absolute error (MAE) of the AMS from the bias-corrected CORDEX ensemble compared to observed AMS for the historical time period (1950–2013) in 6 cities. The different colors represent the different bias correction methods: method (1a), KDDM then aggregation (light blue), and method (1b), aggregation then KDDM (dark blue). The median MAE from the 25-km climate model ensemble is represented by a solid, square marker, while the median of the 50-km ensemble is an open square. The maximum of the MAEs from the climate model ensemble is the upper bound of the error bar and the minimum of the MAEs is the lower bound of the error bar. The number of stars above the error bar shows the number of climate models that are *not* from the same continuous distribution as the observed AMS ($h_0=1$ from K-S test). The horizontal, dashed line represents a MAE of 10%.

This comparison shows that method (1a), bias-correcting then aggregating, has a considerably larger error than method (1b), aggregating then bias-correcting, for both climate model resolutions in all cities. The error using this method (1a) tends to increase as the duration (interval of aggregation) increases. The reason that the error grows as the duration increases could be an artifact of the large grid cell of the climate model. For example, as a large storm moves across the grid cell in the model, several instances of heavy rainfall could be recorded in consecutive time steps while the storm remains in the grid. However, at the station scale, only one instance of heavy rainfall would be recorded, since the storm would pass over the station quickly. When the model values are bias-corrected to the station scale at the 1-hour duration, these wet time steps remain in the bias-corrected time series. If the 1-hour duration maximum rainfall is extracted, it will be similar to the observations, since only one maximum value is extracted. However, if the values are aggregated and then the maximum is extracted, the consecutive time steps with large values of rainfall from the climate model will aggregate to a higher value than would be the case in the observations. Although the evidence here supports this hypothesis, further research is still needed to validate it.

For method (1a), the K-S test rejects the null hypothesis for nearly all climate models and cities. For the 50-km resolution (open square), when the time series is bias-corrected and then aggregated to higher durations (3 – 24 hours), nearly all of the resulting AMS from climate models end up not being from the same continuous distribution as the historical AMS. A larger number of bias-corrected models are from the same continuous distribution as the historical AMS for the 25-km resolution model ensemble. This is the case for the 3-hour duration in Phoenix, Pittsburgh, and Boulder, and the 12 and 24 hour durations in Boston and Seattle. The results for the 25-km ensemble for method (1a) in Boston are closer to results for method (1b) than any other city.

These results show that method (1b), aggregation before bias-correction, outperforms method (1a), bias-correction before aggregation for all durations. Method (1a) considerably overestimates the AMS compared to method (1b), and has a larger number of bias-corrected AMS that are not from the same continuous distribution as the observed AMS. Therefore, we only use method (1b), aggregation then bias-correction, for comparison to other techniques.

S2.2. Parametric AMS Mapping

S2.2.1. Detailed description of technique

Parametric AMS Mapping, method (2), is also a bias-correction method like the KDDM method. However, instead of bias-correcting the entire time series, PAM adjusts the annual maximum series of the climate model by estimating the GEV distributions of the climate model and observed AMS, and then mapping between them. This method is similar to the Equidistance Quantile Matching Method developed by Solaiman and Simonovic (2011) as part of an effort to update Canadian IDF curves to reflect future extremes (Solaiman and Simonovic 2011; Simonovic et al. 2016). That method, which makes use of daily output from GCMs, spatially downscales from the grid scale to the station scale, and then temporally downscales from the daily GCM to the sub-daily level. For the present work, the temporal downscaling step is unnecessary, because RCM output from NA-CORDEX is available at the sub-daily level. Instead we apply separate mapping functions for each duration.

The steps for developing the adjusted AMS using PAM are as follows:

1. Extract AMS from the observed data (X_d) and the climate model output (Y_d) using sliding maxima.
2. Fit a GEV distribution, GEV_{X_d} , to the observed AMS, X_d .

3. Fit a GEV distribution, $GEV_{Y_{hd}}$, to the AMS of the climate model, Y_d , for the historical period, Y_{hd} . The number of years does not have to equal the number of years in the observed time period, but they should cover similar time periods.
4. Calculate $F_{GEV,Y_{hd}}(Y_d)$, the CDF of $GEV_{Y_{hd}}$ at the values of Y_{hd} , in order to get the probability for each value of Y_{hd} .
5. Evaluate $F^{-1}_{GEV,X_d}[F_{GEV,Y_{hd}}(Y_d)]$, the inverse CDF of GEV_{X_d} at $F_{GEV,Y_{hd}}(Y_d)$. The result is the spatially adjusted AMS, X_{fd} .
6. Repeat step 1 for each duration and city, and steps 2–4 for each climate model, duration, and city.

S2.2.2. Comparison of PAM and KDDM techniques

We evaluated the performance of PAM and KDDM by comparing the AMS of the historical climate simulation to the observed AMS time series for all aggregated durations (3, 6, 12, 24, and 48 hours) using the mean absolute error (MAE) and the K-S test. **Figure S4** presents these results for each climate model resolution (25-km and 50-km), duration, and 6 of 34 cities. Durations are shown along the x-axis, and cities are shown in each subplot. The two methods are shown as different colored error bars. The median, minimum, and maximum MAE of the climate model ensemble are shown as a square and the lower and upper bounds of the error bar, respectively. Filled squares represent the 25-km resolution climate model ensemble, while the open square represents the 50-km resolution. The number of stars at the top of the bar signifies the number of climate models where $h_0 = 1$ (rejected) for the K-S test, meaning that the historical climate model AMS is *not* from the same continuous distribution as the observed AMS.

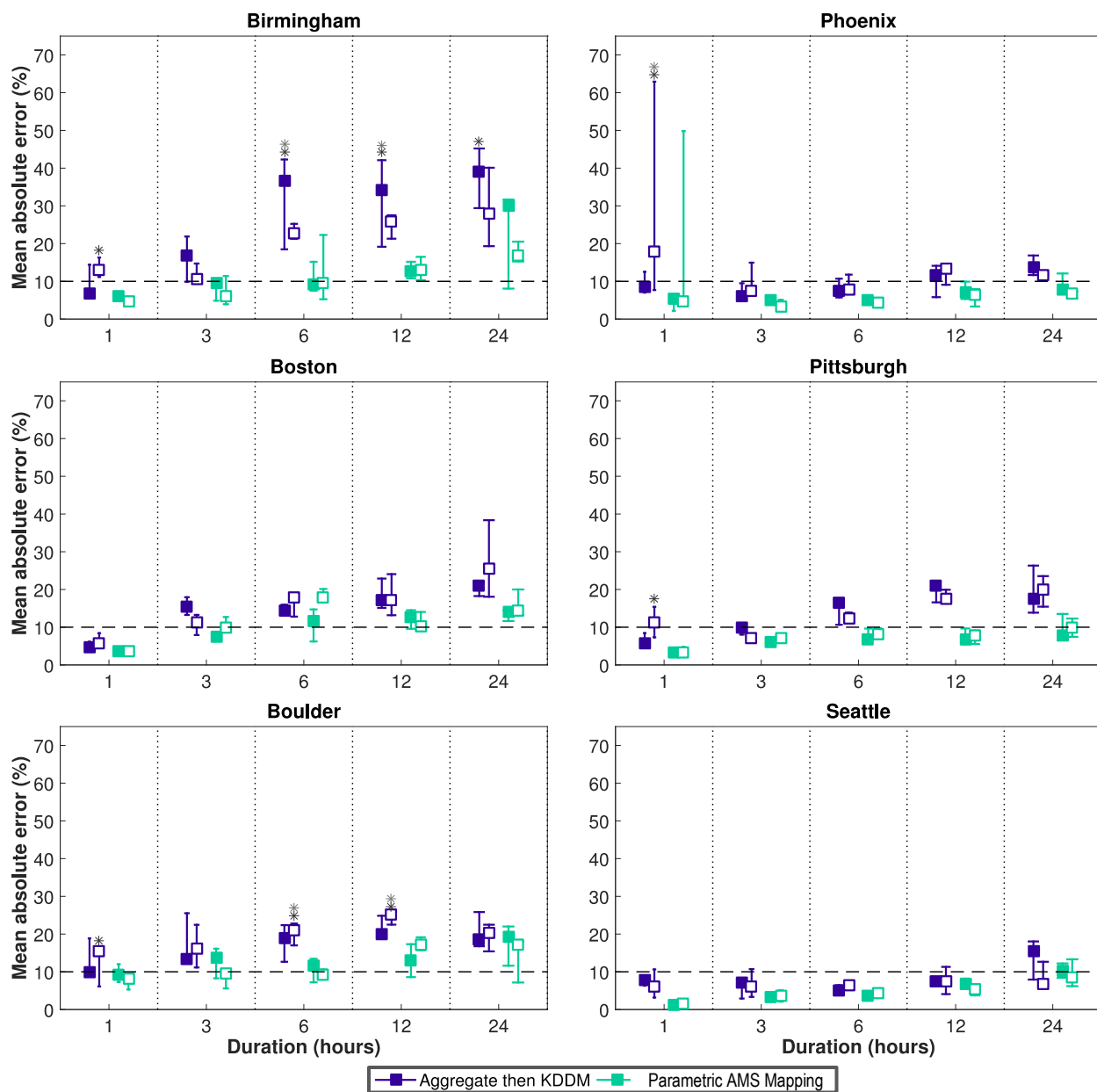


Figure S4. Mean absolute error (MAE) of the AMS from the bias-corrected NA-CORDEX ensemble compared to observed AMS for the historical time period (1950–2013) in 6 cities. The different colored error bars represent the different bias correction methods: method (1b), aggregation then KDDM (dark blue), and method (2), parametric AMS mapping. The median MAE from the 25-km climate model ensemble is represented by a solid, square marker, while the median of the 50-km ensemble is an open square. The maximum of the MAEs from the climate model ensemble is the upper bound of the error bar and the minimum of the MAEs is the lower bound of the error bar. The number of stars above the error bar shows the number of climate models that are not from the same continuous distribution as the observed AMS ($h_0=1$ from K-S test). The horizontal, dashed line represents a MAE of 10%.

225
226

Overall, PAM exhibits better performance than KDDM because it has a smaller MAE and never results in a rejection of the K-S test. Despite the better performance of PAM, we still use KDDM in our analyses in order to understand how these differences in performance affect the resulting IDF curves. Conceptually, the two methods are entirely analogous: in both cases, model values are transformed into probabilities according to CDF of the model data, then transformed back into values using the CDF of the observations. The difference is in where in the workflow (Figure S2) that transformation is applied. Since the distribution of block maxima such as the AMS is known to converge to the GEV distribution, it is not surprising that applying the transformation to the extreme values directly using the correct parametric distribution exhibits better performance than applying it implicitly using a non-parametric estimate.

S2.3. Change Factor

Method 3, the Change Factor technique, adjusts the observed depth of rain based on the ratio of change between the historical and future climate model simulations. This method is also referred to as the areal reduction factor method (Allen and DeGaetano 2005; Zhu, Stone, and Forsee 2012; Cook, Anderson, and Samaras 2017). Future depths at the station scale are obtained with the following equation:

$$D_F^{(s)}(p, d) = D_H^{(s)}(p, d) \frac{D_F^{(g)}(p, d)}{D_H^{(g)}(p, d)} \quad [2]$$

where D denotes the depth of rainfall, or return level, for a given probability of occurrence (p), duration (d), and CI. This depth is either at the station scale (s) or grid scale (g), for future (F) or historical (H) time periods. Historical depths at the station scale refer to depths obtained using historical observations obtained using the general techniques in **Section S2.1**. Depth is converted to intensity by dividing by the duration (in hours).

Depths at the grid scale for the historical or future period refer to return level depths obtained directly from the gridded regional climate model simulations. These depths are obtained by following the steps in **Section S2.1** for each of the six climate model simulations for historical and future periods. For this analysis, we apply the change factor to observations of the time period 1954–2013. The historical climate model baseline is also from 1954–2013, and selection of the future period is discussed in the following section (**Section 2.3.4**).

The CF method estimates the median future depth using the median observed return level, the median historical climate return level, and the median future return level. Confidence intervals for the future depth are obtained in the same manner, i.e., the 90% CI uses the 90% CI from the observed data, historical climate model, and future climate model.

S3. Comparison of MOS techniques for multiple durations and return periods

Figure S5 presents the results in 6 cities of the future 50-year rainfall depth estimated using the KDDM, PAM, and CF techniques for the 1-hour (left) and 24-hour (right) durations. The observed period is shown as a black dashed line and the future period is shown as different colored lines for each technique. The sloped line passes through the median of the climate model ensemble (all resolutions) and the range shows the upper and lower bounds defined in Section 2.2.4. The observed range shows the 10% and 90% CI for the observed period. The upper bound on the y-axis (highlighted in red) varies in each subplot.

For the smaller return periods, the mean of the ensemble for each method is relatively consistent for most cities; however, as the return period increases, the differences in each method are larger. This is consistent with results from Sarr et al. 2015, who found that the choice of downscaling technique matters more for return periods of 10 years or longer. In this analysis, the

mean and bounds of the three methods can vary by several orders of magnitude for the larger return periods. We find that the PAM and KDDM methods can generate problematically large values (see **Figure S6**). For instance, in Birmingham, the PAM method predicts the mean ensemble value for the 24-hour 100-year storm to be about 50 inches (120 cm) of rain. The historical value is around 11 inches (30 cm) —a value that is five times lower than the future prediction. This unrealistic future value is the result of several consecutive large rainstorms in the 25-km WRF-GFDL model run. After PAM is applied to the model to bias-correct to the station scale, these large rainstorms are exaggerated, and the adjusted future rainfall comes out to be unrealistically large, which distorts the entire subsequent chain of analysis. It is beyond the scope of this paper to evaluate the nature of this event and whether it is realistic, but the PAM and KDDM methods have problems when there is a significant mismatch in the presence of the very largest rare events (e.g., hurricanes, atmospheric rivers, etc.) in the input datasets. To guard against this problem, we remove events from the model datasets that are more than five times larger than the next largest event.

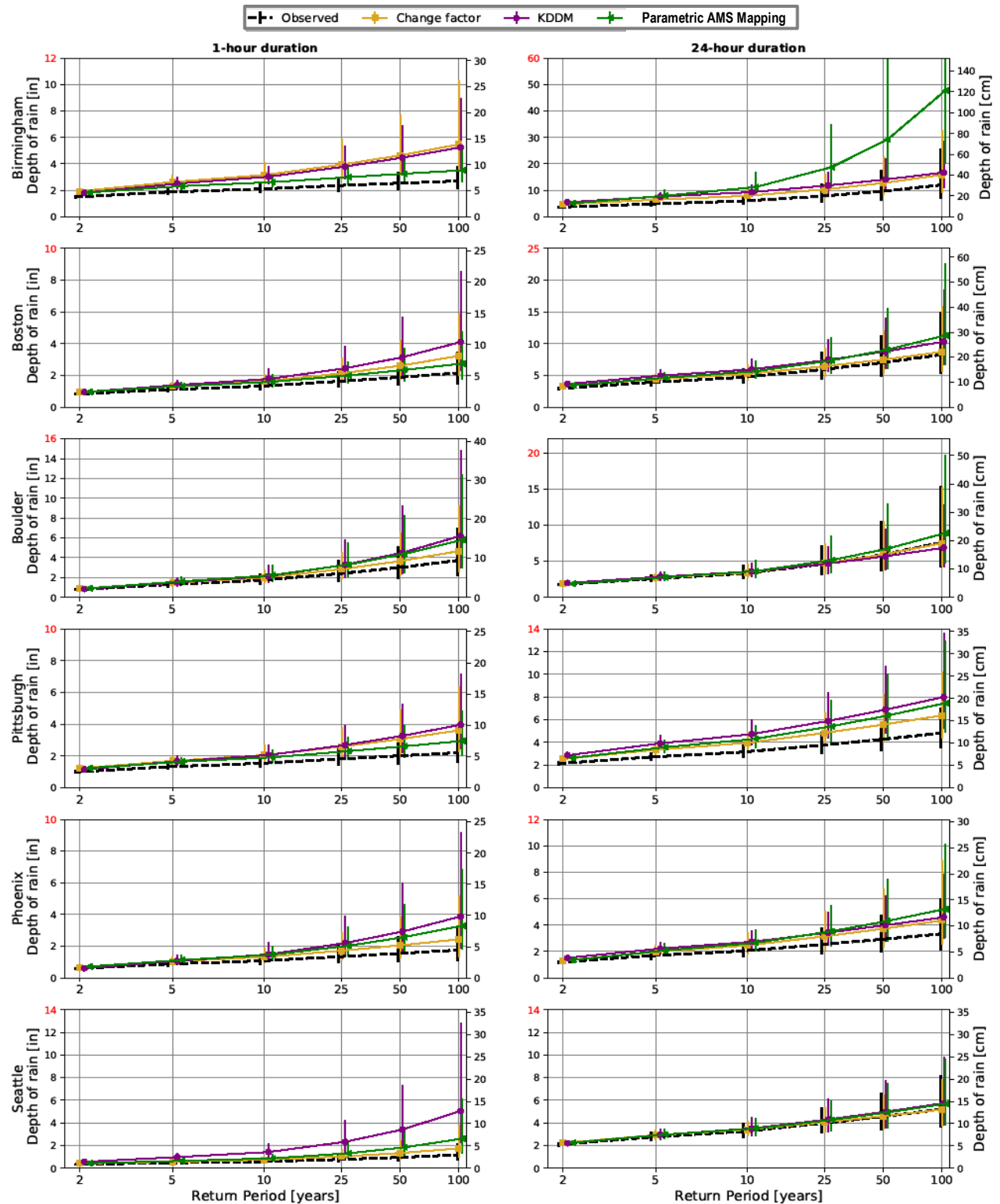


Figure S5. Effect of different correction methods on future rainfall depth for 6 cities and two durations: 1-hour (left) and 24-hour (right). Observed depth is in black; CF method in yellow (square markers); KDDM method in purple (circle markers); and PAM method in green (triangle markers). The sloped line passes through the average of the climate model ensemble (all resolutions); error bars represent the upper and lower bounds of this ensemble.

S4. Additional methods

S4.1. Selection of length of future time period

To evaluate the length of the future time period in the analysis, we represent uncertainty as the distance between the 10% CI and 90% CI of the GEV fit and normalize the results for all durations, return periods, climate models, and cities by taking the percent difference between this range as $(90\%CI - 10\%CI)/10\%CI$. **Figure S6** presents the range of these values as a box plot. The red line in the box represents the median of these values, the top and bottom of the box plot represent the 25th and 75th quantiles, and the whiskers extend to the 90th quantile. Red plus signs represent outliers above the 90th quantile. The x-axis shows the length of the time period in the future for the end of the 21st century. For instance, the 30-year length refers to the period 2070–2099, the 40-year refers to 2060–2099, etc.

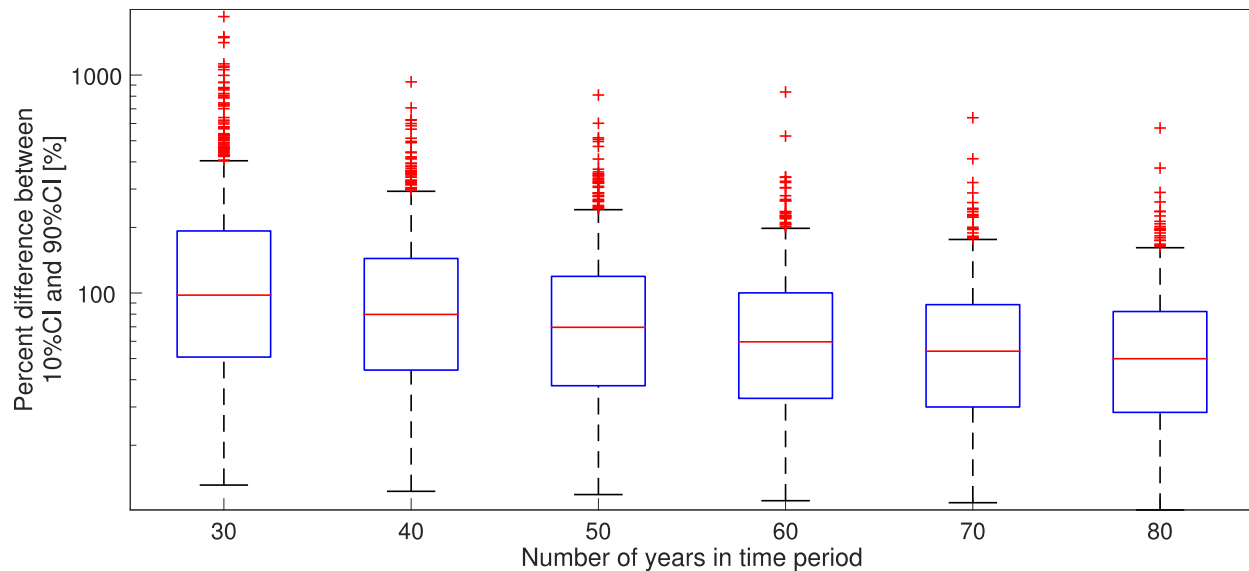


Figure S6. Percent difference between the 10% and 90% confidence interval of the GEV distribution. The range of values for all durations, return periods, climate models, and cities is represented as a box plot. The red line in the box represents the median of these values, the top and bottom of the box plot represent the 25th and 75th quantiles, and the whiskers extend to the 90th quantile. Red plus signs represent outliers above the 90th quantile.

Results show that the range between the 10% and 90% CI of the GEV fit decreases as the number of years in the sample increases. This is expected, as it is well-known that uncertainty decreases with a larger sample size. We selected the longest, 80-year period for use because the

uncertainty range is the smallest, and the additional years only marginally change the median value (red line).

S4.2. Illustrative stormwater infrastructure design example

We calculate peak discharge for the design example using the rational method: $Q_p = ciA$, where Q_p is the peak flow for the watershed, c is the runoff coefficient, i is the rainfall intensity, and A is the watershed area. A runoff coefficient of 0.65 is assumed for a moderately urbanized neighborhood (McCuen 2005).

The rational method assumes that the duration of storm to select from the IDF curve is equal to the time of concentration of the watershed, or the time it takes to travel from the most distant part of the watershed to the outlet point (McCuen 2005). This analysis uses a small 4-hectare (10-acre) watershed with an assumed time of concentration of 1 hour, and thus uses the 1-hour design storm.

Using the peak discharge, the pipe diameter is calculated with Manning's equation:

$$Dr = \left(\frac{nQ_p}{0.31k_n\sqrt{S_0}} \right)^{\frac{3}{8}} \quad [3]$$

where Dr is the pipe diameter, n is the roughness coefficient, S_0 is the channel slope, and k_n is a coefficient of the velocity versus slope relationship. The roughness coefficient is assumed to be 0.013 for ordinary concrete lining, kn equal to 1, and S_0 equal to 0.005 (or 0.5%).

We rounded the resulting diameters from Manning's equation up to the nearest standard U.S. pipe size to determine the diameter of the pipe that would be installed. The minimum storm sewer diameter recommended by the Pennsylvania Department of Transportation is 450 mm (18 inches). Pipes are available in 75 mm (3 inch) increments until the 750 mm (30 inch) diameter, after which they are available in 150 mm (6 inch) increments (PennDOT 2015). The pipe sizes used in this

333 analysis are: 18, 21, 24, 27, 30, 36, and 42 inches (equivalent to 450, 525, 600, 675, 750, 900, and
334 1050 mm).

335

S5. References

- Allen, Robert J, and Arthur T DeGaetano. 2005. "Areal Reduction Factors for Two Eastern United States Regions with High Rain-Gauge Density." *Journal of Hydrologic Engineering* 10 (4): 327–35.
- Cook, Lauren M, Christopher J. Anderson, and Constantine Samaras. 2017. "Framework for Incorporating Downscaled Climate Output into Existing Engineering Methods: Application to Precipitation Frequency Curves." *Journal of Infrastructure Systems* 23 (4). [https://doi.org/10.1061/\(ASCE\)IS.1943-555X.0000382](https://doi.org/10.1061/(ASCE)IS.1943-555X.0000382).
- Mailhot, Alain, Sophie Duchesne, Daniel Caya, and Guillaume Talbot. 2007. "Assessment of Future Change in Intensity–Duration–Frequency (IDF) Curves for Southern Quebec Using the Canadian Regional Climate Model (CRCM)." *Journal of Hydrology* 347 (1–2): 197–210.
- McCuen, Richard H. 2005. *Hydrologic Analysis and Design*. 3rd ed. Pearson Prentice Hall.
- McGinnis, Seth, Doug Nychka, and Linda O Mearns. 2015. "A New Distribution Mapping Technique for Climate Model Bias Correction." In *Machine Learning and Data Mining Approaches to Climate Science*, edited by Valliappa Lakshmanan, Eric Gilleland, Amy McGovern, and Martin Tingley, 91–99. Springer.
- Mearns, L. O., Seth McGinnis, Daniel Korytina, Raymond Arritt, Sebastien Biner, Melissa Bukovsky, Hsin-I Chang, et al. 2017. "The NA-CORDEX Dataset, Version 1.0." Boulder, CO: NCAR Climate Data Gateway. <https://doi.org/10.5065/D6SJ1JCH>.
- Montfort, Martin AJ van. 1990. "Sliding Maxima." *Journal of Hydrology* 118 (1–4): 77–85.
- PennDOT. 2015. "Chapter 13, Storm Drainage Systems." In *PennDOT Drainage Manual 2015 Edition*, 13–25. Harrisburg, PA: Pennsylvania Department of Transportation. <https://www.dot.state.pa.us/public/pubsforms/Publications/PUB%20584.pdf>.
- Sarr, M.A., O. Seidou, Y. Trambly, and S. El Adlouni. 2015. "Comparison of Downscaling Methods for Mean and Extreme Precipitation in Senegal." *Journal of Hydrology: Regional Studies* 4 (September): 369–85. <https://doi.org/10.1016/j.ejrh.2015.06.005>.
- Simonovic, Slobodan P., Andre Schardong, Dan Sandink, and Roshan Srivastav. 2016. "A Web-Based Tool for the Development of Intensity Duration Frequency Curves under Changing Climate." *Environmental Modelling & Software* 81 (July): 136–53. <https://doi.org/10.1016/j.envsoft.2016.03.016>.
- Solaiman, Tarana A, and Slobodan P Simonovic. 2011. "Development of Probability Based Intensity-Duration-Frequency Curves under Climate Change." *Water Resources Research Report* 34: 1–93.
- Zhu, Jianting, Mark C. Stone, and William Forsee. 2012. "Analysis of Potential Impacts of Climate Change on Intensity–Duration–Frequency (IDF) Relationships for Six Regions in the United States." *Journal of Water and Climate Change* 3 (3): 185–96. <https://doi.org/10.2166/wcc.2012.045>.