Evaluation of Precipitation Patterns for Designing Climate Change-Resilient Stormwater Management Systems in a Nonstationary Climate

Contestant: Marali Kalra, Penn State University (graduating December 2020)

Advisor: Dr. Cibin Raj, Penn State University

Student Signature: Marali Kalra Date: 5/15/2020

Advisor Signature: Cibin Raj Date: 05/15/2020

This project began as a literature review that I conducted as an honors project for a course taught by Dr. Cibin Raj. During that work, it became clear that there were openings for research into precipitation trends at the local level as well as simple linear models for precipitation change. Over the following year, the project developed into experimental and analytical work using precipitation data made available to the public by NOAA.

Abstract

Conventional stormwater design assumes that the climate is stationary -- that the statistical properties of precipitation, among other variables, remain constant over time. However, climate change-related shifts in the global water cycle over the past century have undermined the stationarity assumption. In Pennsylvania, annual precipitation is expected to increase while extreme storms become more frequent. Stormwater infrastructure designed with an assumption of climate stationarity is at risk for increased flooding. Precipitation projections and revised design protocols are required to build resilient stormwater systems for the changing climate. This study explores the possibility of modeling precipitation nonstationarity as linear change in the parameters of a fitted generalized extreme values (GEV) distribution. Three nonstationary models are applied to annual maximum precipitation time series from seven Pennsylvania precipitation stations, and their accuracy with respect to the observed data is evaluated. The majority of the locations studied experienced upward precipitation trends during the twentieth century. Shifts in the shape, as well as the location, of the probability distribution were common. This type of change was best captured by models that incorporated a time-varying GEV scale parameter. Finally, a case study set in Philadelphia demonstrates an optimization procedure for choosing stormwater design specifications to maintain performance under climate nonstationarity.

1. Introduction

Stormwater engineering design aims to balance two opposing forces: the need to prevent flooding, and the need to minimize costs. Structures large enough to capture or control all flood waters in a catchment typically are not affordable. Risk-based stormwater planning offers a solution: regulators specify an acceptable probability of structural failure, and engineers design structures to handle storms with that probability of occurrence. Flooding does occur, but rarely, so that neither the cost of stormwater infrastructure nor the long-term cost of flood damage is unreasonable.

Stormwater regulations and sizing requirements are defined in terms of design storms, which are hypothetical rainfall events that describe the extreme precipitation distribution at one geographic location. A design storm has three properties: magnitude, expressed as a return level; frequency, expressed as a return period; and duration. In this study, return levels are given as precipitation depths per unit area; they may also be expressed in terms of rainfall intensity. Frequency is a measure of how often a storm is likely to occur. It is typically expressed as a return period, which is the number of years in which a storm of a given magnitude can be expected to occur once. In general, small storms occur frequently and have short return periods; extreme storms occur infrequently and have long return periods. Return period is the inverse of probability; a 100-year storm, for instance, has a 1% probability of occurrence in any given year. Depth-frequency relationships vary for different storm durations. Only storms with 1-day durations are considered in this study. The most recent NOAA precipitation document for the state of Pennsylvania is Atlas 14, Volume 2, released in 2004 and revised in 2006. The web-based Precipitation Frequency Data Server provides return levels for a variety of return periods and durations. A detailed description of the supporting calculations for Atlas 14 may be found in Bonnin et al. 2006.

However, analysis based on historical data is valid only for stationary time series, whose statistical properties do not vary with time (Hosking & Wallis 1997; Khaliq et al. 2006). An analysis appended to Atlas 14, volume 2 determined that observed trends were not significant or consistent enough to undermine depth-frequency results (Bonnin et al. 2006), but others (e.g. Brown 2010; Milly et al. 2008) argue that the stationarity assumption does not apply under climate change. As global temperatures rise, atmospheric moisture content is expected to increase, driving an intensification of the hydrologic cycle (Trenberth 2011; Emori & Brown 2005; Huntington 2005). Though changes will vary geographically (Emori & Brown 2005), the literature predicts overall increases in both quantity of precipitation and proportion of extreme precipitation (Allen & Ingram 2002; Karl & Knight 1998; Shortle et al. 2020). North America is likely to experience broad increases in precipitation (Seneviratne et al. 2012). DeGaetano makes note of a concentration of increasing trends in the Northeast (2009). Furthermore, the Pennsylvania Department of Environmental Protection anticipates that average precipitation in Pennsylvania will increase 15-20%, 10-15%, 0-5%, and 0-5% in winter, spring, summer and fall respectively by the end of the twenty-first century (Shortle et al. 2020).

Retaining the stationarity assumption in a changing climate has consequences for stormwater management. Milly et al. found a significant increase in large-scale floods in the United States during the twentieth century (2002). In regions such as Pennsylvania, where precipitation is increasing over time, conventionally designed stormwater infrastructure will be under-designed, and the risk of failure will be higher. Flood damages contributed to the more than $125 million in infrastructure repairs spent by the Pennsylvania Department of Transportation in 2018 (Governor Wolf's Climate Change Priorities 2019). Guo (2006), investigating a parallel case in Chicago, noted that the city's sewer pipes are under-designed by 36% if precipitation changes during the last century are considered. Lopez-Cantu & Samaras (2018), using precipitation change and state design standards to evaluate the risk of stormwater system failure across the U.S., found that under high-emissions climate scenario RCP 8.5, all but two of the states studied will fall into the highest-priority class for revising stormwater standards.

If stormwater infrastructure is to remain viable under climate change, engineers require a new design protocol that models future precipitation patterns and incorporates time-dependent depth-frequency relationships. To facilitate adoption by stormwater professionals, the method must be as simple as possible, and familiar concepts such as return periods should be retained. The research to date has explored several methods for modeling nonstationary precipitation time series. Several studies (e.g. Cook et al. 2017, Früh et al. 2010, Mailhot et al. 2007) have used climate model outputs to predict future precipitation. In this method, a model or ensemble of models is selected based on emissions scenario, temporal and spatial resolution (Cook et al. 2017), outputs are obtained for the relevant years, and the projected data are used to calculate depth-frequency relationships. This approach takes advantage of the wealth of climate research now available, but it has several drawbacks that make it unsuitable for the development of a simple, local model of precipitation change. Climate models generally output low-resolution data, which must be downscaled to the regional and local scales required for stormwater planning. In areas with complex topography, the downscaled data may not be accurate (Früh et al. 2010). The method can become computationally intensive, especially if multiple models are involved, and uncertainty can be difficult to quantify (Allen & Ingram 2002).

An alternative approach uses historical trends as predictors of future precipitation change. Though this concept disregards the influence of climate variables other than precipitation and time, it is more responsive to local conditions. (Denault et al. 2006) identified linear trends in annual maximum precipitation data and extrapolated them forward to estimate future maxima. Others (DeGaetano 2009, Mailhot & Duchesne 2010) investigated trends in the parameters of the fitted probability distribution. This study explores the use of probability distribution parameters to characterize precipitation trends in Pennsylvania. Additionally, a case study demonstrates a procedure for stormwater design with a nonstationary precipitation model.

2. Methods

Long-term precipitation records were used to compute depth-frequency curves for a selection of NOAA measuring stations in Pennsylvania. Four different curves were examined for each station: one that assumed climate stationarity, and three that modeled nonstationarity as linear change in the values of the probability distribution's parameters. These models were then evaluated on their accuracy with respect to the observed trends in return levels.

2.1 Study Area

The study encompasses seven measuring stations in Pennsylvania (Figure 1). Each of Pennsylvania's major climatic regions is represented: one station (Philadelphia) is located in Pennsylvania's southeastern coastal plain, two stations (Warren and Franklin) are on the Allegheny Plateau, and the remaining four are in the ridge and valley region of central and northern Pennsylvania. The Allegheny Plateau typically experiences frequent low-volume rainfall events, while the coastal plain, which is on the fringes of Atlantic tropical storms, is prone to occasional, extreme storms (Knight 1960). Precipitation patterns vary in the ridge and valley region because of the complex topography (Knight 1960).

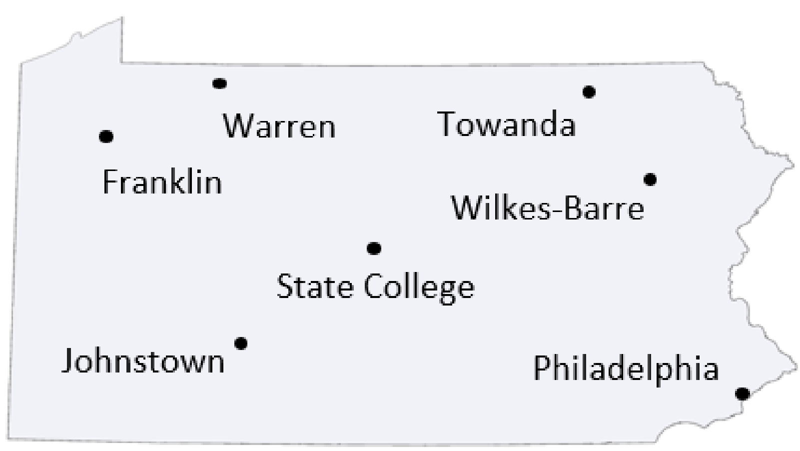


Figure 1. Map of Pennsylvania with measuring stations labeled. Not to scale.

2.2 Data

The data used in this study were 24-hour annual maximum precipitation depths compiled for Atlas 14. These data were extracted from daily observations made at National Climatic Data Center measuring stations as part of the data processing for Atlas 14 (Bonnin et al. 2006). The annual maximum time series are available online on the Precipitation Frequency Data Server maintained by NOAA (https://hdsc.nws.noaa.gov/hdsc/pfds/). To make estimates for low-frequency storms as accurate as possible, analysis was restricted to Pennsylvania stations with at least 100 continuous years of record (Table 1).

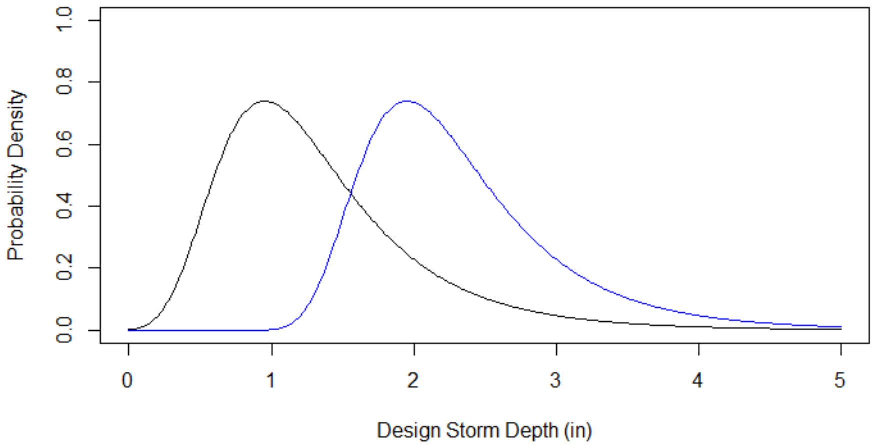
Table 1. Site information for the seven NOAA measurement stations included in the study.

NOAA Station ID	Location	Date Range	Years of Data
36-3028	Franklin	1897-2000	104
36-4385	Johnstown	1890-1992	103
36-6889	Philadelphia	1900-2000	101
36-8449	State College	1888-2000	113
36-8905	Towanda	1895-2000	106
36-9298	Warren	1893-2000	108
36-9702	Wilkes-Barre	1890-1997	108

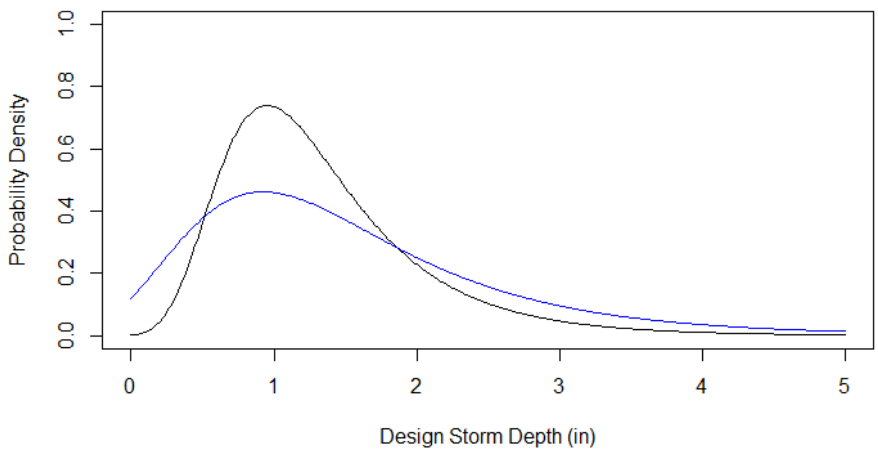
2.3 Model Development

The key to stormwater design is a known, quantitative relationship between storm magnitude and frequency of occurrence. This relationship may be obtained by fitting historical rainfall data to a probability distribution using the principles of extreme value theory as described in Coles (2001). Where the data are block maxima, such as annual maximum precipitation depths, the probability distribution used is the generalized extreme values (GEV) family of distributions, which has three parameters: location, which determines the position of the distribution's peak, which in the case of stormwater design is related to the average magnitude of precipitation. The scale parameter determines the spread of the distribution, and the shape parameter determines tail behavior.

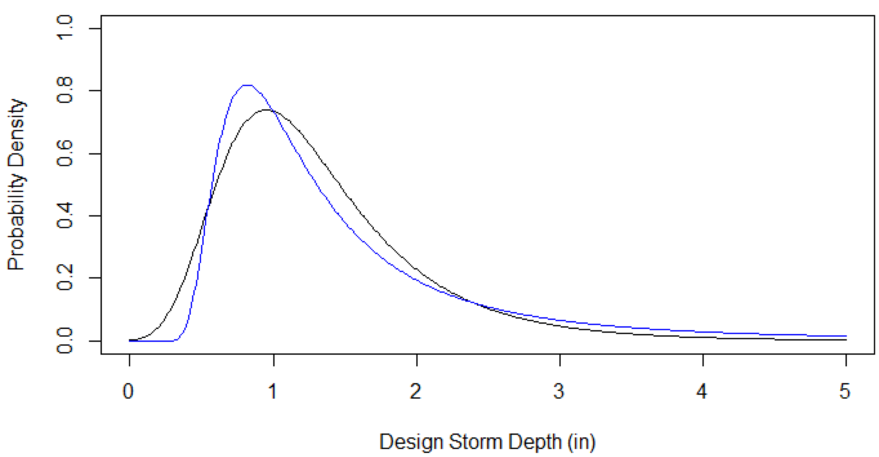
The standard protocol in stormwater design is to compute a stationary depth-frequency curve, where all GEV parameters are held constant over time. This produces a single value for each return level, which is assumed to apply throughout the modeling period and into the future. A nonstationary depth-frequency relationship, in contrast, allows GEV parameters to vary with time, yielding return levels that change from year to year. As shown in Figure 2, variation in each parameter has a different effect on the shape of the distribution. Most modelers have focused on changes to the location parameter, but as Katz & Brown point out, changes in the scale parameter can have more significant effects on return levels (1992).



a)



b)



c)

Figure 2. Effects of increasing each parameter on the GEV distribution. Baseline distributions are shown in black; modified distributions are shown in blue. a) Increased location parameter: the distribution shifts to the right. b) Increased scale parameter: the distribution spreads and flattens,

making extremes more likely. c) Increased shape parameter: the upper tail of the distribution grows heavier, biasing change toward the upper extremes.

The R package extRemes (Gilleland & Katz 2016) was used to fit the Atlas 14 data to four separate models (Table 2): a stationary distribution, used as a baseline for comparison, and three nonstationary distributions in which different combinations of GEV parameters were treated as linear functions of time. Nonstationary model 1 incorporates a time-varying location parameter, making it sensitive to overall increases or decreases in precipitation, but not to changes in the frequency of extreme events. Under this model, return levels increase linearly and at the same rate. Nonstationary model 2, in which both the location and scale parameters of the distribution vary with time, and model 3, in which all three of the GEV parameters vary, capture changes to the distribution's shape as well as rightward or leftward translation. Under model 2, return levels change linearly, but at varying rates. Model 3 describes return level change as a nonlinear curve. The GEV parameters generated by each model were used to calculate 2-, 5-, 10-, 25-, 50-, and 100-year return levels. Additionally, regression lines were fitted to the precipitation data from each station to identify general trends.

Table 2. Summary of linear depth-frequency models.

Model Name	**Time-Varying Parameters**	**Return Level Trend**
stationary	none	none
1	location	linear
2	location, scale	linear, varying with return period
3	location, scale, shape	polynomial, varying with return period

2.4 Model Performance Evaluation

A procedure was developed for evaluating the models described above, with a separate evaluation dataset and specified model performance measures as recommended by ASABE Engineering Practice 621 (2017). Because return levels are derived quantities, it was not possible to compare the models directly to observed data. Instead, an evaluation dataset was derived using a method that captured interannual variability. Annual maximum precipitation data were divided into overlapping 30-year windows, with each window offset by one year from the preceding one. Return levels were estimated by fitting a stationary GEV distribution to each 30-year block of data. This generated a truncated time series of return levels, beginning 29 years after the start of the modeling period, for each measuring station. Root-mean-square error (RMSE), percent bias (PBIAS), and Nash-Sutcliffe efficiency (NSE) were calculated for models 1-3 as well as the stationary model to gauge each model's accuracy with respect to observed precipitation. Pearson coefficients for correlation with time were calculated for 2-, 5-, 10-, 25-, 50-, and 100-year return levels in the evaluation dataset to identify trends.

3. Results & Discussion

3.1 Characterization of Trends in Observed Data and Evaluation Dataset

Figure 3 shows the change in annual maximum precipitation at each measuring station during the modeling period. Extreme precipitation at all stations except Towanda displayed increasing trends, ranging from 4.6% in State College to 48% in Franklin. The average change across the seven stations studied was positive 13%, indicating that precipitation has increased significantly in the past century.

In general, extreme precipitation trends were not uniform across storms of all frequencies at the selected stations. A Pearson correlation test applied to the evaluation dataset revealed that different return levels in State College, Towanda, and Wilkes-Barre had markedly different correlations with respect to time. In Philadelphia and Johnstown, 2-year return levels were weakly correlated with time, with Pearson coefficients of 0.08 for both stations. This contrasted with the correlation coefficients for storms with longer return periods, which exceeded 0.7. Warren and Franklin were the only measuring stations where all return levels showed consistently strong correlations with time, with all values greater than 0.7 and less than 0.85.

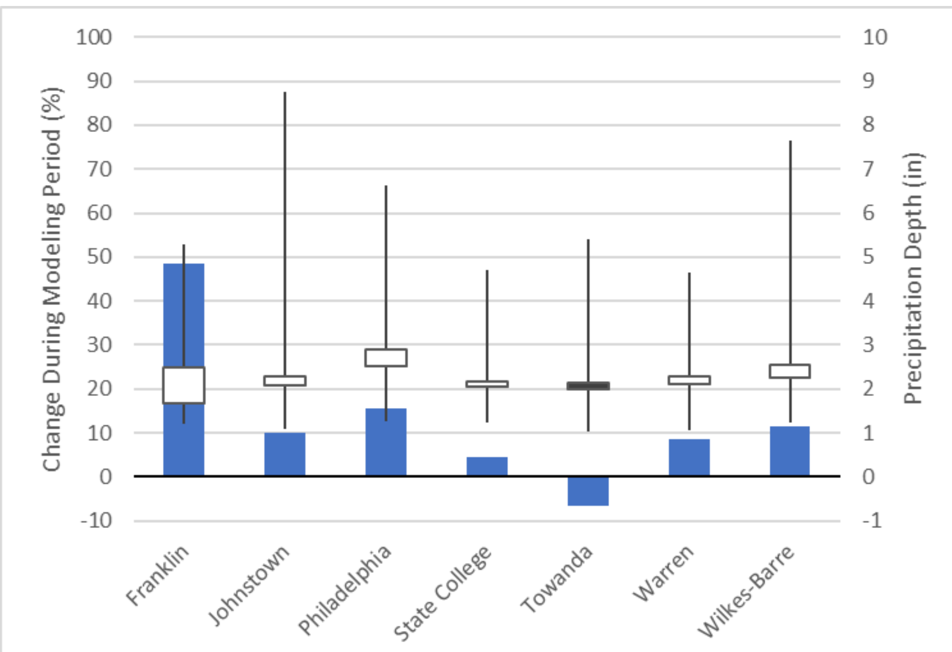


Figure 3. Change in annual maximum precipitation at each measuring station. The black lines show maximum and minimum values over the roughly 100-year modeling period. The boxes outline the absolute change shown when a linear trend was fitted to the data. The blue bars show that change as a percentage of the initial value.

3.2 Nonstationary Modeling Results

Model 1, incorporating a time-varying location parameter, showed less pronounced trends than the direct regression with moving 30-year window discussed above. In Philadelphia, State

College, and Wilkes-Barre, which experienced overall increases in annual maximum precipitation, model 1 showed decreases in design storm depths. This is most likely because the actual precipitation change over the modeling period was not uniform across storms with all return periods. Because model 1 forced low- and high-frequency storms to increase at the same rate, it failed to capture that change.

In general, models with more time-varying parameters predicted greater positive change in return levels. Across all stations, the 10-year storm depth increased on average 1.3% under model 1, 13% under model 2, and 24% under model 3. In Philadelphia and Franklin, the stations with the strongest upward trends, model 3 showed greater than 200% increases in the 100-year storm. The disparity is clear in Figure 4, where the 10-year return level in Philadelphia shows a slight decrease under model 1, but a rapid increase under models 2 and 3. The two exceptions to this pattern were model 2 in Johnstown and model 2 in Towanda, which found that design storms were decreasing over time. In these cases, design storms with longer return periods showed more extreme negative changes over time.

According to models 2 and 3, which did allow trends to vary among storms with different return periods, high-frequency design storms (those with short return periods) tended to remain constant or show negative trends. Low-frequency storms displayed greater positive change. Under model 2, 100-year return levels increased, on average, 11% more than 2-year return levels across all stations. Under model 3, this difference was higher, at 82%. The spread in the magnitude of design storm trends varied considerably from station to station. At three measuring stations, Philadelphia, State College, and Towanda, model 3 found decreases in high-frequency storms while it showed increases in low-frequency storms. Model 2 did not show comparable flexibility; the trends it displayed were either uniformly positive or uniformly negative.

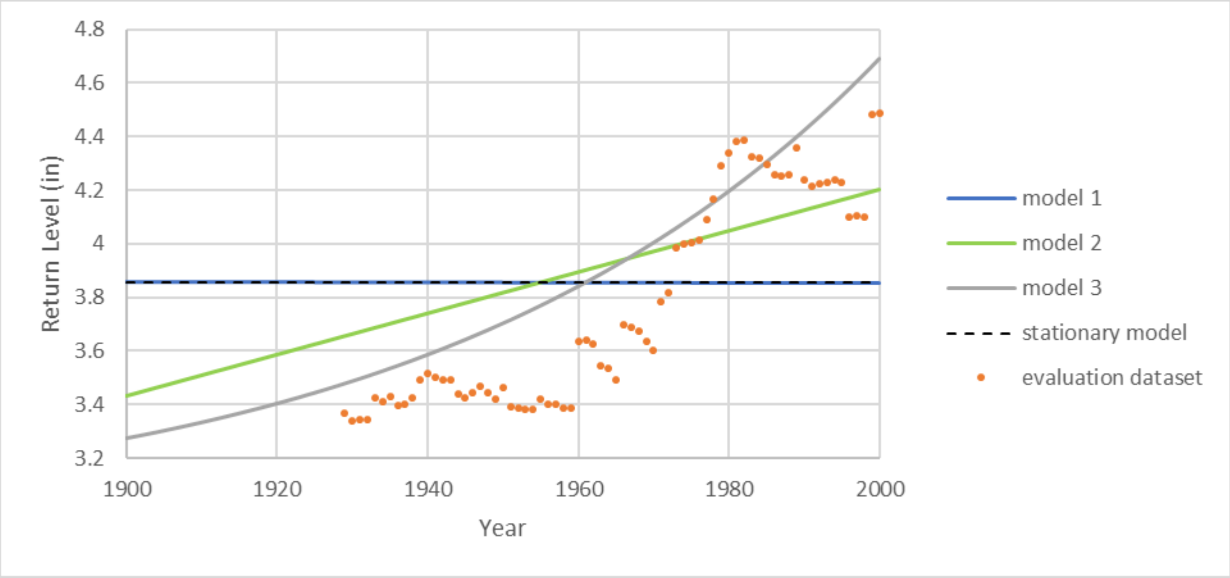


Figure 4. Modeling results for 10-year return levels in Philadelphia. The nonstationary and stationary models, shown as solid lines, were compared to the evaluation dataset, shown as scattered points.

3.3 Selection of Models

No single model applied equally well across all stations studied. Nonstationary models had to be considered on a station-by-station basis to capture local precipitation behavior. Model 3's accelerating rate of change allowed it to fit the observed data more closely, leading to high Nash-Sutcliffe efficiencies and low RMSE. However, its predictions became unreasonably large when extrapolated over long time periods or applied to extreme storms. Furthermore, the lowest PBIAS values were for low-frequency storms under model 3, indicating that it systemically overestimated return levels for extreme storms. The bias was less notable for high-frequency storms. At some stations (Philadelphia, State College, and Towanda) PBIAS was positive for the 2-year return level. Model 3 seemed to be overly inflected -- the rates of change rose to unreliable levels as return period increased. Because of these difficulties, model 3 was removed from consideration as a predictive model.

In general, RMSE increased with return period at all stations. This is most likely the case because the return level values were computed from 30 years of data, which made it difficult to accurately estimate storm depths with longer return periods. In the case of low-frequency design storms in the evaluation dataset, long-term trends were obscured by interdecadal variability. The highest variability was found in Wilkes-Barre and Philadelphia, the two stations in eastern Pennsylvania, where extreme precipitation is connected to tropical storms moving in from the Atlantic coast. RMSE for both stations reached 2.08 inches for the 100-year return level under model 2. In the region affected by storms from the Atlantic, annual maximum precipitation may be extreme in years where a tropical storm makes landfall, while remaining relatively low in other years. This makes it difficult to accurately project return levels for low-frequency storms.

At the State College, Towanda, and Wilkes-Barre stations, the main difficulty in modeling was the large range of design storm trends. In Towanda, for instance, the 2-year return level in the evaluation dataset showed a strong negative correlation with time, with a Pearson coefficient of -0.77; the 100-year return level, however, correlated positively with time, with a coefficient of 0.56. The pattern in State College was similar, though not as extreme. In Wilkes-Barre, the situation was reversed, with low-frequency return levels decreasing more over time. Under all three nonstationary models, PBIAS in Wilkes-Barre was negative for 2-year return periods and positive for 100-year return periods. Separate linear models for each design storm would likely be more effective than a single nonstationary model that attempts to capture the full range of trend behaviors.

In Philadelphia, which demonstrated a moderate, consistent increasing trend throughout the time period, model 2 achieved the highest Nash-Sutcliffe efficiencies across all design storms. However, in Franklin, which underwent a more significant precipitation increase, model 2 was most accurate for storms with return periods of 10 years and greater. Model 1 was a better fit for 2- and 5-year design storms because it assigned them a higher rate of change than model 2. Franklin station data showed a more negative PBIAS for high-frequency storms. In Philadelphia, low-frequency storms experienced a much higher rate of increase than the high-frequency storms. Model 2 predicted a 32% increase in the 100-year storm, but only an 8% increase in the 2-year storm. In Warren, however, the rates of increase were similar for storms of all return periods. Model 2 showed 14% and 9% increases in the 100-year and 2-year storms, respectively. It can be concluded that Warren experienced minimal change in the shape of the probability distribution.

As a result, both model 1 and model 2 produce positive Nash-Sutcliffe efficiencies, with slightly higher values model 2, attributable to the additional degree of freedom.

In Towanda, the design storm/model combinations that showed decreases during the modeling period had a slight positive PBIAS, meaning that the negative trends assigned by all three nonstationary models were accurate but too strong. In State College, PBIAS was consistently negative for both the nonstationary and the stationary models, with values for extreme storms falling below -5%. The implication is that State College is best not modeled with an increasing trend. In Johnstown, the evidence supports an upward trend with faster increases in low-frequency design storms. PBIAS was highly positive under model 2 for all computed return periods longer than 2 years and highly negative under model 3 for all computed return periods, indicating that decreasing models and extreme increasing models did not reflect the observed data. Model 1, with a moderate increasing trend, had lower values for PBIAS.

4. Design Case Study

Once a precipitation model is chosen, its predictions must be incorporated into stormwater design procedures. Regulatory guidelines such as the Pennsylvania Stormwater Best Management Practices Manual set volume and peak rate control standards in terms of return periods (Pennsylvania Department of Environmental Protection 2006). If, for example, a 10% chance of flooding is determined to be the appropriate risk level for a structure, regulations will require that it be sized to contain or convey a 10-year storm.

Problems arise when depth-frequency relationships change over the lifetime of the structure. Structure dimensions must be chosen based on a single design depth, but the risk associated with that depth changes over time, making it difficult to create a design that remains in compliance with regulations throughout its service lifetime. Mailhot and Duchesne (2010) suggest designing for conditions in a “reference year” during the structure’s lifetime. Under conditions of increasing precipitation, such a structure would be overdesigned before the reference year, meaning it would be too large for the amount of rain it actually received. Overdesign results in unnecessarily high construction costs. After the reference year, the structure would be under-designed, meaning it would be too small to convey the design rainfall depth. Risk would rise above the acceptable level and flooding would occur more frequently than intended by the design.

Beginning-of-life overdesign and end-of-life under-design are unavoidable when precipitation is increasing over time. However, the choice of design depth determines the relative length of those periods of over- and under-design. Stormwater design for a nonstationary climate can be treated as an optimization problem in which the goal is to find the design depth for which failure risk remains as close as possible to the prescribed level at both the beginning and the end of the structure’s lifetime.

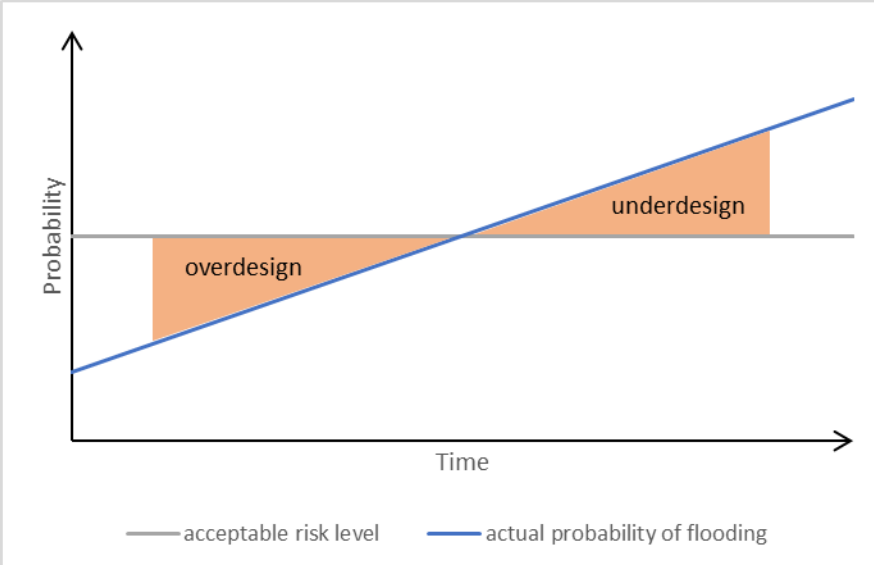


Figure 5. Illustration of over- and underdesign in a stormwater structure built in a region where precipitation increases over time. Because the structure's capacity remains constant over its lifetime, the risk of it failing increases with time, and the return period of the largest storm it can handle decreases.

The Philadelphia measuring station was selected as the site of the case study. Model 2, with time-varying GEV location and scale, was extended forward to the year 2099, then used to compute return levels for the 21st century. According to this model, the 10-year design storm in Philadelphia can be expected to increase from 4.20 inches in 2000 to 4.96 inches in 2099. The model assumes the change will be smoothly linear. These modeling results are applied to a sample stormwater design problem: selecting the optimal design depth for a structure intended to convey a 10-year storm and remain in service from 2000 to 2099. The lower limit on the design depth is the 10-year return level for 2000 (4.36 inches). A structure sized for any smaller depth would be underdesigned for its entire service lifetime. The highest possible design depth is the 10-year return level for 2099 (4.8 inches). Above that level, the structure would be overdesigned over its entire lifetime. The optimal depth falls between those two values. With known depth-frequency relationships for both 2000 and 2099, it is possible to determine beginning- and end-of-life return periods associated with each depth in that range. Depths at the lower end of the range will have beginning-of-life return periods close to 10 years, but their end-of-life return periods will be shorter for storms that can be expected once every 10 years in 2000 will be occurring much more often by 2099. At the upper end of the range, the situation will be reversed: end-of-life return periods will be close to the target value, but beginning-of-life return periods will exceed 10 years.

Equation 1 defines over- and underdesign for a given design depth in any year i as the absolute difference between the target return period, 10 years, and the actual return period for that year, denoted T_i.

$$\Delta T_i = |10 - T_i| \qquad (1)$$

The goal of this nonstationary design problem is to choose a design depth such that ΔT_{2000} and ΔT_{2099} are at a simultaneous minimum. For structures built to last less than 50 years, this depth may be approximated by taking the average of the 10-year design depths for the first and last years of the structures' lifetime. In this case, however, the expected service lifetime is 100 years, which is long enough for the shape of the Philadelphia depth-frequency curve to change appreciably. Return period increases more rapidly with depth in 2099 than in 2000, which shifts the optimal depth toward the lower end of the range.

The optimal depth may be determined graphically by plotting ΔT_{2000} and ΔT_{2099} versus depth, as shown in Figure 6. ΔT_{2000}, representing beginning-of-life overdesign, increases over the depth range. ΔT_{2099}, which corresponds to end-of-life underdesign, decreases with depth. The point where both fall as low as simultaneously possible is the point where the two lines intersect, or where $\Delta T_{2000} = \Delta T_{2099}$. Figure 6 gives the design depth as 4.52 inches, corresponding to a 2000 return period of 11.9 years and a 2099 return period of 8.1 years.

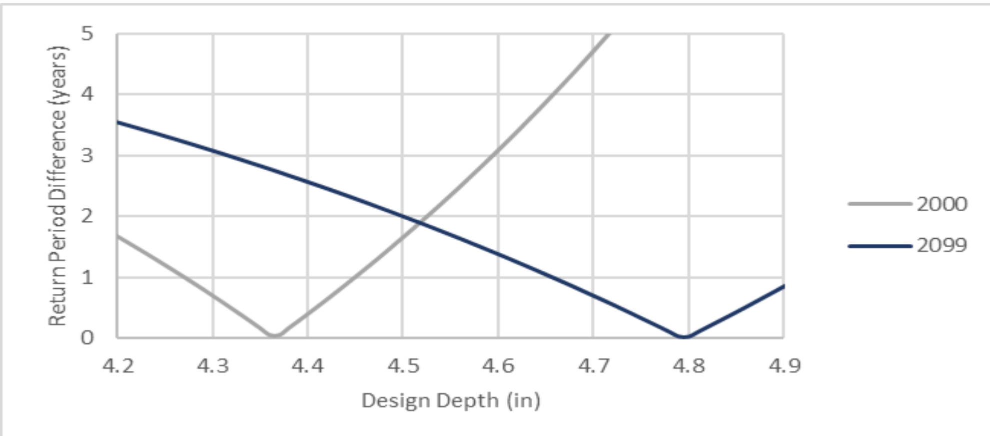


Figure 6. Over- and underdesign represented as the difference, in years, between target and actual return period, computed for 2000 and 2099 for the case study structure over the range of possible design depths.

5. Conclusions

The majority of the measuring stations considered in this study displayed increasing trends in annual maximum precipitation. Generally, storms with longer return periods showed more positive trends. This result is consistent with the consensus in the literature, which predicts that Pennsylvania will experience a rise in total precipitation coupled with more frequent occurrence of extreme storms. Incorporating more time-varying parameters increased a nonstationary model's sensitivity to changes in design storm depths. In the stations in this study, shifts in the shape of the probability distribution were a critical component of precipitation change. This suggests that it is important to consider changes in GEV scale, as well as GEV location, when constructing models with nonstationary distributions. Allowing GEV shape to vary with time, however, generated scenarios of runaway precipitation increase that became unreasonable when extrapolated over long time periods.

Nonstationary modeling was not effective for the four stations in the ridge and valley region. However, a more representative sample of Pennsylvania stations is needed to identify geographical patterns. Analyzing precipitation on a station-by-station basis allows models to be specific to conditions at each region. However, the data available at each station are limited -- it is rare for a single location to have more than 100 years of precipitation records. This makes it difficult to be certain that observed trends are symptomatic of long-term climate-related changes. Land use changes or natural variability may be more likely causes. Nonetheless, methods for characterizing and addressing precipitation change remain valid, no matter the cause of the nonstationarity.

Given projected depth-frequency relationships for each year of a proposed stormwater structure's lifetime, it is possible to adapt design procedures for a nonstationary climate without dispensing with return periods, design storms, or other concepts familiar to stormwater planners. As demonstrated in the Philadelphia case study, climate change considerations can be incorporated at the level of design specifications, altering the way regulatory guidelines are interpreted without complicating the design process itself.

References

Allen, M. R., & W. J. Ingram. 2002. Constraints on future changes in climate and the hydrologic cycle. *Nature* 419: 224-232.

Bonnin, G.M., Martin, D., Lin, B., Parzybok, T., Yekta, M., and D. Riley. 2006. Precipitation-frequency atlas of the United States. *NOAA Atlas 14, 2,* version 3.0. Silver Spring, MD: National Weather Service.

Brown, C. 2010. The end of reliability. *Journal of Water Resources Planning and Management* 136(2): 143-145.

Coles, S. 2001. *An introduction to statistical modeling of extreme values.* London: Springer-Verlag.

Cook, L. M., Anderson, C.J., and C. Samaras. 2017. Framework for incorporating downscaled climate output into existing engineering methods: Application to precipitation frequency curves. *Journal of Infrastructure Systems* 23(4).

DeGaetano, A. T. 2009. Time-dependent changes in extreme-precipitation return-period amounts in the continental United States. *Journal of Applied Meteorology and Climatology* 48: 2086-2099.

Denault, C., Millar, R. G., and B. J. Lence. 2006. Assessment of possible impacts of climate change in an urban catchment. *Journal of the American Water Resources Association* 42(3): 685-697.

Emori, S., and S. J. Brown. 2005. Dynamic and thermodynamic changes in mean and extreme precipitation under changed climate. *Geophysical Research Letters* 32(17).

Früh, B., Feldmann, H., Panitz, H.-J., and G. Schädler. 2010. Determination of precipitation return values in complex terrain and their evaluation. *American Meteorological Society* 23: 2257-2274.

Gilleland, E. & R. W. Katz. 2016. extRemes 2.0: An extreme value analysis package in R. *Journal of Statistical Software* 72(8): 1-39.

Governor Wolf's Climate Change Priorities: Hearing before the Environmental Resources and Energy Committee, Pennsylvania House of Representatives (2019). (Testimony of Secretary Patrick McDonnell).

Guo, Y. 2006. Updating rainfall IDF relationships to maintain urban drainage design standards. *Journal of Hydrologic Engineering*, 11(5): 506-509.

Hosking, J. R. M., & J. R. Wallis. 1997. *Regional frequency analysis: An approach based on L-moments.* Cambridge University Press.

Huntington, T. G. 2005. Evidence for intensification of the global water cycle: Review and synthesis. *Journal of Hydrology* 319: 83-95.

Karl, T. R., and R. W. Knight. 1998. Secular trends of precipitation amount, frequency, and intensity in the United States. *Bulletin of the American Meteorological Society* 79(2): 231-241.

Katz, R. W., & B. G. Brown. 1992. Extreme events in a changing climate: Variability is more important than averages. *Climatic Change* 21: 289-302.

Khaliq, M. N., Ouarda, T. B. M. J., Ondo, J.-C., Gachon, P., and B. Bobée. 2006. Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: A review. *Journal of Hydrology* 329: 534-552.

Knight, P. 1960. Climatography of the United States #60, Cllimate of Pennsylvania. NCDC US Climate Normals.

Lopez-Cantu, T., and C. Samaras. 2018. Temporal and spatial evaluation of stormwater engineering standards reveals risks and priorities across the United States. *Environmental Research Letters,* 13(7).

Mailhot, A., & S. Duchesne. 2010. Design criteria of urban drainage infrastructures under climate change. *Journal of Water Resources Planning and Management* 136(2): 201-208.

Mailhot, A., Duchesne, S., Caya, D., and G. 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: 197-210.

Milly, P. C. D., Wetherald, R. T., and T. L. Delworth. 2002. Increasing risk of great floods in a changing climate. *Nature* 415: 514-517.

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., and R. J. Stouffer. 2008. Stationarity is dead: Whither water management? *Science* 319(5863): 573-574.

Natural Resources and Environmental Systems Hydrology Committee (NRES-21) of the American Society of Agricultural and Biological Engineers. 2017. ASABE engineering practice 621: Guidelines for calibrating, validating, and evaluating hydrologic and water quality (H/WQ) models. St. Joseph, MI: American Society of Agricultural and Biological Engineers.

Seneviratne, S. I., Nicholls, N., Easterling, D., Goodess, C. M., Kanae, S., Kossin, J., Luo, Y., Marengo, J., McInnes, K., Rahimi, M., Reichstein, M., Sorteberg, A., Vera, C., & Zhang, X. (2012). Changes in climate extremes and their impacts on the natural physical environment. In C. B. Field, Barros, V., Stocker, T. F., Qin, D., Dokken, D. J., Ebi, K. L., Mastrandrea, M. D., Mach, K. J., Plattner, G.-K., Allen, S. K., Tignor, M., & Midgley, P. M. (Eds.), *Managing the risks of extreme events and disasters to advance climate change adaptation, a special report of Working Groups I and II of the Intergovernmental Panel on Climate Change* (pg. 109-230). Cambridge, UK, and New York, NY, USA: Cambridge University Press.

Shortle, J., Abler, D., Blumsack, S., Duncan, J., Fernandez, C., Keller, K., Zarekarizi, M., Nassry, M., Nichols, R., Royer, M., and D. Wrenn. 2020. Pennsylvania Climate Impacts. Pennsylvania Department of Environmental Protection.

Trenberth, K. E. 2011. Changes in precipitation with climate change. *Climate Research* 47: 123-138.