

**Developing a Modeling Framework for
Nonstationary Urban Hydraulic Analyses**

A Thesis

Submitted to the Faculty

of

Drexel University

by

Ge Pu

in partial fulfillment of the
requirements for the degree

of

Master of Science in Environmental Engineering

June 2015



Dedications

I dedicated this thesis to my family and extended family. A special feeling of gratitude goes to my loving parents, Desong and Min, who have always been there for me. They have not only supported me through all these years, but also installed in me a tireless work ethic and persistent determination to be whatever I wanted to be in life without limitations.

I also dedicate this work to my aunts, uncles and grandparents for your continued encouragements. Especially, I would like to dedicate this piece to the memory of my dear grandfather, who passed away before I start my Master's Degree. His kindness and believe in my success have been, and will always be my great motivations to achieve the greatest goals.

Acknowledgement

I would like to express my sincere gratitude to the Drexel University, especially **Professor Charles Haas**, for letting me fulfill my dream of being a student here.

I wish to thank both of my Thesis advisors, **Professor Naresh Devineni** for introducing me to this topic as well for the generous support on the way, and **Professor Franco Montalto** for this great opportunity and opening my eyes to the wonderful world of water resource engineering. For these I am sincerely grateful.

I am also grateful to **Professor Patrick Gurian** for agreeing to be on my committee, and your valuable comments and suggestions to this project.

Table of Contents

LIST OF ABBREVIATIONS.....	V
LIST OF FIGURES.....	VI
LIST OF TABLES.....	XI
ABSTRACT.....	XII
INTRODUCTION.....	1
CLIMATE CHANGE AND ITS EFFECTS ON EXTREME PRECIPITATION EVENTS IN THE	
NORTHEAST US.....	1
NONSTATIONARY FRAMEWORK FOR EPE IDF ANALYSIS.....	2
METHODS.....	4
RAINFALL AND CLIMATE DATASETS.....	4
ESTIMATION OF MODEL PARAMETERS AND RESULTS.....	5
RETURN LEVEL INTERPOLATION.....	6
STATIONARY UNIVARIATE AND MULTIVARIATE DURATION EXTREME RAINFALL IDF	
MODELS.....	6
STATIONARY AND NONSTATIONARY EPE MODELS.....	8
<i>Justification of Nonstationarities in PHL AMRD.....</i>	<i>8</i>
<i>Model Basis.....</i>	<i>9</i>
<i>Model Formulation.....</i>	<i>10</i>
<i>Predictor and Data Selections.....</i>	<i>12</i>
<i>Comparison between Models.....</i>	<i>13</i>
RESULTS.....	14
STATIONARY UNIVARIATE AND MULTIVARIATE DURATION EXTREME RAINFALL IDF	
MODELS.....	14
<i>Comparison Between Model 0 and 1.....</i>	<i>14</i>
STATIONARY AND NONSTATIONARY EPE MODELS.....	19
<i>Justification of Nonstationarities in PHL AMRD.....</i>	<i>19</i>
<i>Predictors Selections and Justifications for Model 2 to 4.....</i>	<i>21</i>
<i>Comparisons among Model 2 to 4.....</i>	<i>24</i>
Model 2 Results.....	24
Location Parameter Comparisons.....	25
Scale Parameter Comparisons.....	27
Comparisons of shape parameters.....	29
Comparisons of alpha values.....	30
Comparisons of beta values.....	32
Comparisons between Original JJAS AMRD and Model Outputs.....	34
DISCUSSION AND CONCLUSION.....	43

FUTURE WORK.....	45
BIBLIOGRAPHY.....	47
APPENDIX.....	52
MODEL SCRIPTS.....	52
<i>#model0</i>	52
<i>#model1</i>	58
<i>#model2</i>	63
<i>#model3</i>	64
<i>#model4</i>	66
KNML AND R CORRELATION RESULTS.....	68
.....	69

List of Abbreviations

1. AMRD – Annual Maximum Rainfall Depth
2. ENSO - El Niño Southern Oscillation
3. EPE – Extreme Precipitation Events
4. JJAS – June to September
5. MAM – March to May
6. MDM – Multiduration Model, also see Model 1
7. PHL – Philadelphia International Airport
8. PF – Precipitation Frequency
9. RAI – Rainfall Accumulation Intervals
10. SST – Sea Surface Temperatures

List of Figures

Figure 1. PHL IDF Curves From Model 0.....	15
Figure 2. PHL IDF Curves From Model 1.....	16
Figure 3.PHL AMRD at various RAI vs. Time, also includes information referenced from Teegavarapu (2012) and NOAA (2015) on the years of El Niño, La Niña and natural conditions. Color-coded areas represents time periods of two consecutive years or more El Niño or La Niña events. One needs to note that the ENSO index started from 1950, so there are no available information on El Niño or La Niña before then (NOAA 2015).....	19
Figure 4.Wavelet Plots of PHL AMRD, the bold contour in the graphs means the 95% confidence level, anything falls outside of the contour is not significant. Also, axes of all six charts are exactly the same, and presented on the outside of the six plots. It is important to note that ENSO activities usually fall in the 2 to 5 years region. Multi-decadal oscillations such as Pacific Decadal Oscillation (PDO) and North Atlantic Oscillations (NAO) would often appear above 20 years period. These plots were generated through were generated through KNMI Climate Explorer (Truet and Oldenborgh 2013).....	20
Figure 5. Time of Occurrences Boxplot of PHL AMRD	21
Figure 6. MAM Average NINO3 Index vs. JJAS PHL AMRD, both linear regression and the locally weighted scatterplot smoothing (lowess) are plotted in this chart. Correlations between the two data sets were calculated with spearman’s correlation, because the JJAS AMRD data set is not normally distributed. Directions of El Niño and La Niña events were indicated near the NINO3 index axis to inform how SST changes during these events.	22

Figure 7. 30yr Moving Average of JJAS PHL AMRD vs Original Dataset.....	23
Figure 8. Model 2 Generated Return Level Against Actual JJAS AMRD, Since Model 2 is a stationary model, the return levels remain constant.	24
Figure 9. Comparison of Location Parameter Between Model 2 and 3, it presents the distribution of Model 3 simulated location parameter, “ $a_{t,3}$ ”, compare to the Model 2 location parameter, “ a ”. Due to the fact that a does not change over time, so the stationary Model 2 results are constant, which are shown as dash lines in these plots. Note that the outliers, which are more than 1.3 times of the upper quartile are excluded from this plots. Red rectangles added into this plot is for identifying if the models have successfully picked up increasing trend shown in Figure 3 after year 2000.	25
Figure 10. Comparison of Location Parameter Between Model 2 and 4, it presents the distribution of Model 4 simulated location parameter, “ $a_{t,4}$ ”, compare to the Model 2 location parameter, “ a ”. Due to the fact that a does not change over time, so the stationary Model 2 results are constant, which are shown as dash lines in these plots. Note that the outliers, which are more than 1.3 times of the upper quartile are excluded from this plots. Red rectangles added into this plot is for identifying if the models have successfully picked up increasing trend shown in Figure 3 after year 2000.	26
Figure 11. Comparison of Scale Parameter Between Model 2 and 3, it presents the distribution of Model 3 simulated scale parameter, “ $b_{t,3}$ ”, compare to the Model 2 location parameter, “ b ”. Due to the fact that “ b ” does not change over time, so the stationary Model 2 results are constant, which are shown as dash lines in these plots. Note that the outliers, which are more than 1.3 times of the upper quartile are excluded	

from this plots. Red rectangles added into this plot is for identifying if the models have successfully picked up increasing trend shown in Figure 3 after year 2000.	27
Figure 12. Comparison of Scale Parameter Between Model 2 and 4, it presents the distribution of Model 4 simulated scale parameter, “ $b_{t,4}$ ”, compare to the Model 2 location parameter, “ b ”. Due to the fact that “ b ” does not change over time, so the stationary Model 2 results are constant, which are shown as dash lines in these plots. Note that the outliers, which are more than 1.3 times of the upper quartile are excluded from this plots. Red rectangles added into this plot is for identifying if the models have successfully picked up increasing trend shown in Figure 3 after year 2000.	28
Figure 13. Comparison of Shape Parameter Among Model 2 to 4, this plot shows the distribution of estimated shape parameter from each model.	29
Figure 14. Comparison of α_1 and α_3 , this plot shows the distribution of estimated alpha values in equation 9 and 12.	30
Figure 15. Comparison of α_2 and α_4 , this plot shows the distribution of estimated alpha values in equation 10 and 13.....	31
Figure 16. Comparison of β_1 , β_3 and β_4 , this plot shows the distribution of estimated beta values in equation 9 and 12.....	32
Figure 17. Comparison of β_2 , β_5 and β_6 , this plot shows the distribution of estimated beta values in equation 10 and 13.....	33
Figure 18. Model 3 vs. Original, it presents the distribution of Model 3 simulated AMRD against the original observations. Note that the outliers, which are more than 1.3 times of the upper quartile are excluded from this plots.	34

Figure 19. Model 4 vs. Original, it presents the distribution of Model 4 simulated AMRD against the original observations. Note that the outliers, which are more than 1.3 times of the upper quartile are excluded from this plots.	35
Figure 20. Model 3 Results against Original AMRD Return Levels. This figure matches the Model 3 JJAS AMRD output with original data shown as red dash line. Blue lines were drawn to show the 100yr, 10yr and 2yr return level of the original JJAS AMRD data. Black lines were drawn from the Model 3 output at the same return levels.	36
Figure 21. Model 4 Results against Original AMRD Return Levels. This figure matches the Model 4 JJAS AMRD output with original data shown as red dash line. Blue lines were drawn to show the 100yr, 10yr and 2yr return level of the original JJAS AMRD data. Black lines were drawn from the Model 3 output at the same return levels.	37
Figure 22. Comparison of 10yr Return Level Against Actual Extreme Rainfall, This focus on one single return level is mainly due to the fact that Philadelphia stormwater design is heavily based on 10yr return levels (City of Philadelphia, 2014). One need to note that all return level related information are referred to the right axis, and the nonstationary model that has a similar trend from the original AMRD dataset should have a better modeling performance.....	38
Figure 23. 10 year Return Level Exceedance Comparison, Model 2 has the stationary probability of 0.1 because it was developed in a stationary framework. Unlike Model 2, Model 3 and 4 both have probabilities of exceedance changing over time due to nonstationarity.	39
Figure 24. ENSO Impact on Model 3 Results, this plot was produced to examine how MAM ENSO effect Model 3 JJAS AMRD outcomes, compare that with the original	

dataset. Both blue points and the regressions are plotted in this chart from the 95%, 50% and 5% of all datasets. Red regression lines and points were from the original JJAS AMRD.40

Figure 25. ENSO Impact on Model 4 Results, this plot was produced to examine how MAM ENSO effect Model 4 JJAS AMRD outcomes, compare that with the original dataset. Both blue points and the regressions are plotted in this chart from the 95%, 50% and 5% of all datasets. Red regression lines and points were from the original JJAS AMRD.40

Figure 26. ENSO Impact on Exceeding 10yr Return Levels, This plot represents how ENSO affect the results of exceeding 10yr return level in Model 2, 3 and 4 results. Black line and points represent the results from Model 3, while Model 4 results are shown in blue points and line. Model 2 probability of exceedance remains stationary at 10% in red.41

Figure 27. 3D Plot of Model 4 Results, this was produced through adding the information of 30yr moving average into the chart shown in Figure 26. It not only explains how ENSO affect the probability of exceeding the stationary 10yr level, but also helps to understand how the moving average indicator affects the model outcome dynamically with ENSO.42

List of Tables

1. Table 1. Model 0 Output of Extreme Eainfall at PHL (inch/hr).....	14
2. Table 2. Model 1 Output of Extreme Rainfall At PHL (inch/hr)	15
3. Table 3. Difference Between Model 0 and 1 Results (inch/hr)	16
4. Table 4. Difference Between Model 0 and 1 Results (inch)	17
5. Table 5. DIC results from Model 0 and 1	18
6. Table 6. DIC results from each Model.....	36
7. Table 7. Spearman Correlation Results from R.....	63

Abstract

Developing a Modeling Framework for Nonstationary Urban Hydraulic Analyses

Ge Pu

Naresh Devineni, Ph.D.

Franco Montalto, Ph.D.

Multiple recent national and international climate assessment reports suggest that the frequency, duration and intensity of future extreme precipitation in the North East US will increase. Urban cities in this region like Philadelphia are extremely vulnerable to these impacts due to their combined sewer overflows (CSO). These are mainly due to the fact that the sewer systems receive higher than normal flows during heavy precipitation. Hence, they seek to incorporate the best available information about current and future climate and understand the extent of their impacts on their stormwater network.

This project intends to develop a general modeling framework for integrating multiple sources of information such as rain gauge data and climate scenarios for extreme rainfall events with prescribed duration using a Hierarchical Bayesian modeling technique. It aims to develop a stochastic temporal rainfall field simulator for the analysis of the extreme rainfall precipitation events. The approach here is to minimize uncertainties in stationary hydraulic assessment models with non-stationary ones. Five extreme rainfall models were developed for improvements with both stationary and non-stationary modeling frameworks. Recommendations and evaluations of these model outcomes are evaluated and discussed to provide a critical path towards informing climate adaptation decision on the ground.

This Page Is Left Blank Intentionally

Introduction

Climate Change and its effects on Extreme Precipitation Events in the Northeast US

Climate change is very likely to alter the future precipitation patterns globally; it is expected to see increases in rainfall intensity, duration and frequency in the region of Northeast (NE) US. According to the National Climate Assessment (2014), there was a five inch increase observed between 1895 and 2011 in this region. Meanwhile, NE US has experienced more increase in extreme precipitation events (EPE) than any other regions of US. This increase was documented as much as 70% from 1895 to 2010. IPCC (2014) has also confirmed these findings with various observations in EPE over the USA between the mid-20th and the early 21st century.

Increasing trend of EPE is very likely to continue, because it is among the expected impacts of greater global warming (Wolff et al., 2014). Yet, EPE have historically caused various damages around the globe, in the forms of large scale flooding (Coumou and Stefan, 2012), health related issues (Luber et al., 2014; White House, 2014), agricultures impacts (Freidline, 2011; Hatfield et al., 2011), as well as significant ecological impacts (Darwin, 1978; Knapp et al., 2008). In 2015, EPE (as part of Extreme Weather Events) are now the second most likely to happen global risks according to the World Economic Forum (2015).

Meanwhile, NE US is especially vulnerable to future EPE increase due to its aging storm water infrastructures. Urban cites such as Philadelphia and New York will be significantly impacted due to their combine sewer systems. Not only these systems are hard and expensive to retrofit, but also poses significant hazards during EPE. This is mainly due to the fact that the sewer systems receive higher than normal flows during

heavy precipitation. Treatment plants are unable to handle flows that are more than twice design capacity and when this occurs, a mix of excess stormwater and untreated wastewater discharges directly into the City's waterways at certain outfalls. Cities are concerning about CSOs because of their effect on water quality and recreational uses. Hence, it is vital to incorporate the best available information about current and future climate and understand the extent of their impacts on urban stormwater management in the NE US.

Nonstationary Framework for EPE IDF Analysis

The extent of climate change impacts on EPE is still yet uncertain, because projections of future climate are still being debated. Yet, due to recent years of increases in EPE, the growing need for climate adaptation and mitigation strategies is becoming stronger and stronger (Milly et al., 2007; Craig, 2010; Tarlock, 2012; Sivapalan and Jos 2009). Especially in the field of water resource management, this change from stationary to nonstationary is increasingly looked-for (Milly et al., 2007; Tarlock, 2012; EPA 2014; Cheng et al. 2014). While stationary framework assumes modeling parameters remain constant over time, nonstationary framework does not (Olson et al. 1998). Thus, nonstationary methods of modeling can enable one to directly link EPE with other related information such as climate indicators.

On the other hand, Intensity-Duration-Frequency (IDF) is currently one of the many important tools one could use to predict the future EPE frequencies. IDF analysis outputs have traditionally been utilized in municipalities to set standards for various water related designs regulations. This relationship can be dated back as early as 1932 (Guo 2006). For

detailed description of IDF calculation methods, please refer to Guo (2006) and Sivapalan and Blöschl (1998). There are two outstanding problems when utilizing the current stationary IDF approaches. First, current precipitation frequency (PF) data is very outdated. The most recent national update of precipitation IDF was completed in 1977 (Frederick et al. 1977), while latest updates of IDF varies from state to state. Among the 9 states in the NE US, only three states (PA, NJ, and DL) have updated their IDF data to yet still 2004, whereas the rest have no new records since the 1977 national PF study (NOAA 2006). Second, traditional IDF utilized stationary methods to construct EPE probabilities for informing water resource design regulations. Temporal, spatial and climate relationships are ignored. However, these associations could potentially be significant in predicting future patterns of EPE. More importantly, IDF data plays significant role in conducting hydrologic model simulations. Rainfall data is the primary input of these models. Thus, uncertainties generated from the two problems mentioned above directly affect the outcomes of hydrologic analysis, which in turn will mislead downstream water resource policies and practices. Given this fact and the recent increase in EPE, as well as future predicted patterns of these events (Wolff et al., 2014), more accurate and robust IDF evaluation strategies are ever more needed to prepare for future EPE in a changing climate.

Thanks to the advances in multivariate statistical analysis and hierarchical modeling techniques, nonstationary approaches of IDF analysis are now possible (Cheng et al. 2014). These approaches enable the consideration of multiple indicators in an EPE model, such as climate, temporal and spatial indicators. In theory, dynamic models with hierarchical modeling frameworks can have better precisions and predictions than static

models (Cheng et al. 2014). Olson et al. (1998) was among one of the many first attempts to analyze IDF in nonstationary conditions. Recently, Cheng et al. (2014) utilized nonstationary extreme value analysis on central United States EPE. He addressed the detailed theory behind how to model EPE in a nonstationary framework.

For this study, the intended goals are to 1) develop a multi duration modeling framework for stationary IDF analysis. This model will not only be more efficient when generating IDF plots under various rainfall accumulation intervals (RAI), but also takes in consideration of the inter-correlation of EPE data among various RAI. This correlation is vital to be considered, because univariate IDF modeling has many limitations, such as the fact that RAI are not the actual durations of rainfall events (Kao 2008). 2) Identify the nonstationarity in the existing EPE data, and justify the need for a nonstationary modeling framework 3) Develop and compare various nonstationary EPE models with different indicators, as well as the stationary models. 4) Incorporate predictive capability into developed models 5) Compare and identify the best model to be carried out for future steps.

Methods

Rainfall and Climate Datasets

Rainfall data utilized in this study was obtained from the National Oceanic and Atmospheric Administration (NOAA) Global Hourly dataset at the National Climate Data Center data portal. The duration of data is from 1941/1/1 11:00:00 to 2010/12/01 23:00:00. All rainfall data was recorded at the Philadelphia International Airport (PHL)

on an hourly basis, and the original data was in the unit of 100 of an inch. Climate data utilized in this research includes El Niño-Southern Oscillation (ENSO) related datasets, NINO3, NINO3.4 and NINO4 indexes (NOAA 2015) recorded monthly from 1941/1/1 00:00:00 to 2010/12/01/23:00:00. These indexes document the monthly anomalies of sea surface temperatures (SST) at averaged over three regions off the coast of Ecuador and Colombia. NINO3 covers the area 5°N-5°S latitude, 150°W-90°W longitude. NINO3.4 covers the area 5°N-5°S latitude, 120°W-180°W longitude. NINO4 covers the area 5°N-5°S latitude, 150°W-150°E longitude. During El Niño years, all NINO indexes peak above 0, and the tendency reverse during La Niña years. Other climate indexes such as multidecadial climate oscillations were also considered initially, but then discarded due to the fact that prior research done by Hu and Feng (2012) show ENSO has a significant role in influencing US summertime precipitation variations. Also, this decision was reached because of data unavailability in the other indexes from 1941 to 2010.

Estimation of Model Parameters and Results

All model parameters shown in this study was produced through coupling R (R Core Team 2014) and Openbugs (Thomas et al. 2006). Openbugs utilized the Markov Chain Monte Carlo methods as the basis for estimation, which involves six fundamental steps: 1) constructing a model 2) define prior distribution for model parameters 3) upload model inputs 4) define initials for parameters 5) estimate parameters and results with 10000 iterations 6) check for convergence if “Rhat” is greater than 1.1. Theory behind these steps are explained in detail by Cheng et al. (2014). All coding for this estimation process was done in R and then interface with Openbugs to complete the necessary steps for

estimating model parameters and results. Codes for constructing all models involved with this project are attached in the appendix.

Return Level Interpolation

As mentioned earlier, detailed theory behind IDF interpolation methods are well documented by Guo (2006), and Sivapalan and Blöschl (1998). In general, return levels were produced by the following equation:

$$P_x = \frac{1}{T_x} \quad \text{Equation 5}$$

Where P_x is the probability of exceeding a given rainfall amount, x , T_x is the return period of the same given rainfall amount, x . This given rainfall amount, x , is the return level.

Stationary Univariate and Multivariate Duration Extreme Rainfall IDF Models

Model 0 processes rainfall data independently among various RAI. NOAA PHL climate data was processed through an IDF Curve Generator (IDFCG), and various IDF curve plot based on different RAI were then generated. This IDFCG was developed in R (R Core Team 2014) with the utilization of the *extRemes* (Gilleland 2011), *xts* (Ryan and Ulrich 2014), *data.table* (Dowle et al. 2014), and *lubridate* (Grolemund and Wickham 2011) packages. The purpose of this tool is to transform hourly rainfall data into plots of IDF curves through series of steps, which includes data importing, cleaning, rainfall accumulations, extreme value extractions, GEV statistics analyses (Gilleland and Katz

2011), and IDF curve graph generations. Detailed code involved in developing this model is shown in the appendix.

The purpose of Model 1 is to generate more accurate and efficient IDF curves with considerations of the correlations among different RAI. This model was accomplished through multivariate normal distribution theory. While traditional univariate IDF analysis were conducted independently among various RAI, Multiduration Model (MDM) not only speeds up the IDF curve generation process, but also reflects the correlations among RAI. Theory behind this model is explained in detail by Renard and Lang (2006). General set up of Model 1 is shown in equation 1 and 2.

$$M1: R_{total} \sim \text{Gev}(a^d, b^d, c^d) \quad \text{Equation 5}$$

$$\begin{pmatrix} a_1 \\ \vdots \\ a_{5d} \end{pmatrix} \sim \text{MVN} \left(\begin{matrix} \mu_{a1} & \sigma_{a1}^2 & \cdots & \text{cov}(a_1, a_{5d}) \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{a5d} & \text{cov}(a_{5d}, a_1) & \cdots & \sigma_{a5}^2 \end{matrix} \right) \quad \text{Equation 6}$$

where M1 is model 1, R_{total} is the PHL AMRD, a^d , b^d and c^d are the parameters of model 1, MVN means multivariate normal distribution, cov means covariance of two given values, and σ^2 means the variance of a given value. a_1 to a_{5d} means the location parameters for each RAI, here the RAI are for 1 hr to 5 days. Upon completion of this model, a^d , b^d and c^d were estimated in Openbugs. Reconstruction of each return level based on these estimated parameters was accomplished in R with the extRemes package.

Data selected for Model 0 and 1 input are the PHL AMRD. Since Model 0 does not have any nonstationarity involved, there are no predictors necessary for examination. Model 1 has no predictors involved as well, but it takes in consideration of the covariance between various RAI to ultimate produce the model parameters a_1, b_1 , and c_1 .

Upon selecting data input for Model 0 and 1, simulations of each model was completed. Results from each model output were compared and analyzed. Difference between these two models was interpolated and discussed. Suspected benefits of Model 1 over Model 0 was elaborated as well.

Stationary and Nonstationary EPE Models

Justification of Nonstationarities in PHL AMRD

Initial data exploratory analysis began with data cleaning in R (R Core Team 2014), which was to convert original datasets into more appropriate format for further investigations. Upon completion of cleaning and organizing the original data, the interpolated data was accumulated according to different RAI: 1hr, 6hr, 12hr, 24hr, 3days and 5days. Then, annual maximum values of EPE were extracted from each of the RAI accumulated data. Nonstationarity in the dataset was investigated through timeries and time-frequency (wavelet) plots. Timeseries plot was constructed in R (R Core Team 2014). Productions of wavelet plots were accomplished in the KNMI Climate Explorer (Truet and Oldenborgh 2013). These charts were made through an application developed by Torrence and Compo (1998) based on Wavelet Analysis methodologies. Purposes of these charts are to decompose a time series into time–frequency space, where one is able to determine both the dominant modes of variability and how those modes vary in time. Both timeseres and wavelet plots were constructed to not only to justify the nonstationarities of PHL AMRD over time, but also to identify the potential influences of such nonstationarities.

Model Basis

While various parametric and nonparametric methods are applicable to nonstationary frequency analysis of hydrometeorological data (Khaliq 2006), trends in recent years are shifting toward utilizing the **Generalized Extreme Value** theorem (Lima 2015). Walshaw (2013) documents the definition and assumptions of GEV in further details. Stationary models developed through GEV in this study assume all parameters remain constant over time, while the nonstationary model assumes the opposite. Variations of parameters in the nonstationary model would assumed to be impacted by different indicators such as climate, time, and moving average of extreme rainfall values.

Annual extreme rainfall values, R_e , is mathematically assumed to follow the classical GEV distribution (Sang 2008; see examples in Lima et al. 2015 and Mondal et al. 2015), G_{ev} means a GEV distribution based on stationary parameters μ (location), σ (scale), ξ (shape):

$$R_e \sim Gev(\mu, \sigma, \xi) \quad \text{Equation 1}$$

where the cumulative distribution of G_{ev} is:

$$G_{ev}(x) = \exp\left\{-\left[1 + \xi * \left(\frac{x - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} \quad \text{Equation 2}$$

If we assume the location parameter, μ , vary over time, and then the value of it can be modeled as a function of n indicators:

$$\mu(t) = f(x_1(\omega_1), x_2(\omega_2), x_3(\omega_3), \dots \dots x_n(\omega_n)) \quad \text{Equation 3}$$

where f means a particular function of some parameters and indicators, which can be linear, exponential or others. ω indicates the respective time of indicators, which can be records before or within the time R_e was recorded. Predictable functionality would be

developed if utilizing past ω to inform future Re. These exact same concept shown in Equ. 3 can be applied to the scale parameter, σ :

$$\sigma(t) = f(x_1(\omega_1), x_2(\omega_2), x_3(\omega_3), \dots \dots x_n(\omega_n)) \quad \text{Equation 4}$$

For the shape parameters, ξ , of the GEV distribution, the relationships between ξ and their indicators are assumed to be very complex (Lima et al. 2015). Due to the time constrains, ξ were assumed to remain constant for all of the models in this study. Upon completing the model formulations, each parameter involved were estimated through Bayesian Inference Using Gibbs Sampling methodology (Thomas et al. 2006).

Model Formulation

Model 2 was developed through a traditional univariate GEV framework in Openbugs (Thomas et al. 2006). Equation 7 explains how this model was formulated.

$$M2: Rt \sim Gev(a, b, c) \quad \text{Equation 7}$$

Where M2 is Model 2, Rt is the maximum PHL AMRD during a giving duration. Gev is the GEV distribution. a, b, and c refers to the location, scale and shape parameters of this model.

Model 3 was developed through a Hierarchical GEV framework. Model 3 utilizes a given ENSO index (NINO3, NINO3.4 or NINO4) averaged over a given duration as the climate indicator for modeling PHL EPE. Below, Equation 8, 9 and 10 together explain how this model was formulated:

$$M3: Rt \sim Gev(a_{t,3}, b_{t,3}, c) \quad \text{Equation 8}$$

$$a_{t,3} = \alpha_1 + \beta_1 * enso \quad \text{Equation 9}$$

$$b_{t,3} = abs(alpha_2 + beta_2 * enso) \quad \text{Equation 10}$$

where M3 is Model 3, R_t is the maximum PHL AMRD during a given duration. Gev is the GEV distribution. $a_{t,3}$, $b_{t,3}$, and c refers to the location, scale and shape parameters of this model. $alpha_1$, $alpha_2$, $beta_1$, and $beta_2$ are estimated parameters from Model 3. $enso$ refers to a given ENSO index averaged over a given duration. Unlike model 2, the location and scale parameters of the GEV distribution in this model, a_t , was effected by the $enso$ values and change over time. Model 3 takes in consideration of climate information during the PHL EPE modeling process. One needs to note that it is assumed linear relationship in both Equation 7 and 8.

Model 4 was developed through a Hierarchical GEV framework. It utilizes both a given ENSO index averaged over a given duration and a PHL EPE moving average dataset as indicators for modeling PHL EPE. Equation 10, 11 and 12 explain how this model was formulated:

$$M4: R_t \sim Gev(a_{t,4}, b_{t,4}, c) \quad \text{Equation 11}$$

$$a_{t,4} = alpha_3 + beta_3 * enso + beta_4 * \overline{R_t} \quad \text{Equation 12}$$

$$b_{t,4} = abs(alpha_4 + beta_5 * enso + beta_6 * \overline{R_t}) \quad \text{Equation 13}$$

where M4 is Model 4, R_t is the maximum PHL AMRD during a given duration. Gev is the GEV distribution. $a_{t,4}$, $b_{t,4}$, and c refers to the location, scale and shape parameters of this model. $alpha_3$, $alpha_4$, $beta_3$, $beta_4$, $beta_5$, and $beta_6$ are estimated model parameters. $enso$ refers to a given ENSO index averaged over a given duration. $\overline{R_t}$ is a given trend indicator of the PHL AMRD. One needs to note that it is assumed linear relationship in

both Equation 11 and 12. Both location and scale parameters of this model change over time, and are impacted by both the NINO3 Index and $\overline{R_t}$.

Predictor and Data Selections

Due to the fact this study is a preliminary investigative study on nonstationary framework, only 24hr AMRD was utilized as an example for comparing results among model 2, 3 and 4. The model that had the best result would then be applied to the rest of the RAI cumulated rainfall dataset in future projects.

Selection processes of various indicators began with first looking into which period of year does most of the EPE take place. Duration which cover the 25% to 95% quintiles of EPE time of occurrences was selected as the duration of “Rt” in Model 2, 3 and 4. All Rt values utilized in these models are the AMRD extract from the selected duration.

In order to find the most appropriate “enso” values in Model 3 and 4, initial investigation began with selecting the most appropriate duration. This was done in KNMI Climate Explorer (Truet and Oldenborgh 2013) to correlate averaged HadISST1 1° reconstruction (Rayner 2003) at various periods with the selected “Rt” mentioned above. Reason to use such dataset was because all NINO indexes (NINO3, NINO3.4 and NINO4) were constructed based on the HadISST1 1° data. Among the different periods of comparisons, the one has the strongest correlation with “Rt” was selected as the period of average for “enso” in Model 3 and 4. Then, each of the three NINO index averaged within this period was correlated with the selected “Rt”. Among the three indexes, the one averaged within the selected period, and has the strongest correlation was selected as the “enso” values in

Model 3 and 4. All of the correlation results were generated through KNMI Climate Explorer (Truet and Oldenborgh 2013) and R (R Core Team 2014). Final decision on “enso” was also analyzed and justified.

In addition, temporal trend indicator, $\overline{R_t}$, would need to be selected. Such indicator would help the model in better recognizing the trend with the original PHL AMRD dataset during the same duration as “Rt”. Selection of such indicator was mainly based on literature review. According to Krakauer and Devineni (2015), utilization of 30yr weighted moving average as an indicator for climate normalization generally have good overall performance when compared to other methods. Thus, this study decided to utilized the third indicator for the value of $\overline{R_t}$. Upon selection, analysis of selected $\overline{R_t}$ against the original “Rt” was evaluated and to justify this decision.

Comparison between Models

For Model 2 only, return levels of 100 years and 10 years were reconstructed based on the estimated values of a, b and c. Results of these levels against the original AWRD dataset was produced. Since Model 2 is a stationary model, its return levels remain constant.

On the other hand, estimated parameters and outcomes in Model 3 and 4 should not remain constant overtime, because they are nonstationary. One would also expect that outcomes from the nonstationary models do not deviate from Model 2 by significant amounts. Significant deviations would indicate a bias in the nonstationary models or other possible factors, such as climate anomalies. Comparisons of the various outcomes

were conducted, which includes 1) contrasts of location and shape parameters, 2) value comparison of alphas and betas 3) model fit results against original dataset based on Deviance Information Criterion (DIC), timeseries and ENSO perspectives. On the other hand, detailed comparisons were conducted at the 10yr return level, because of the fact that Philadelphia stormwater design is heavily based on this level (City of Philadelphia, 2014). Both AMRD values and probability of exceedance at the 10yr level was contrasted. These comparisons were also conducted based on timeseries and ENSO perspectives, as well as the selected \overline{R}_t for Model 4 only.

Results

Stationary Univariate and Multivariate Duration Extreme Rainfall IDF Models

Comparison Between Model 0 and 1

	1.5yr	5yr	10yr	20yr	50yr	75yr	100yr
1hr	0.269	0.490	0.630	0.794	1.057	1.195	1.302
6hr	0.079	0.131	0.162	0.196	0.250	0.277	0.298
12hr	0.049	0.079	0.097	0.116	0.144	0.158	0.169
24hr	0.029	0.048	0.060	0.074	0.096	0.107	0.116
3day	0.011	0.018	0.023	0.028	0.036	0.041	0.044
5day	0.007	0.012	0.015	0.018	0.023	0.026	0.028

Table 1. Model 0 Output of Extreme Eainfall at PHL (inch/hr)

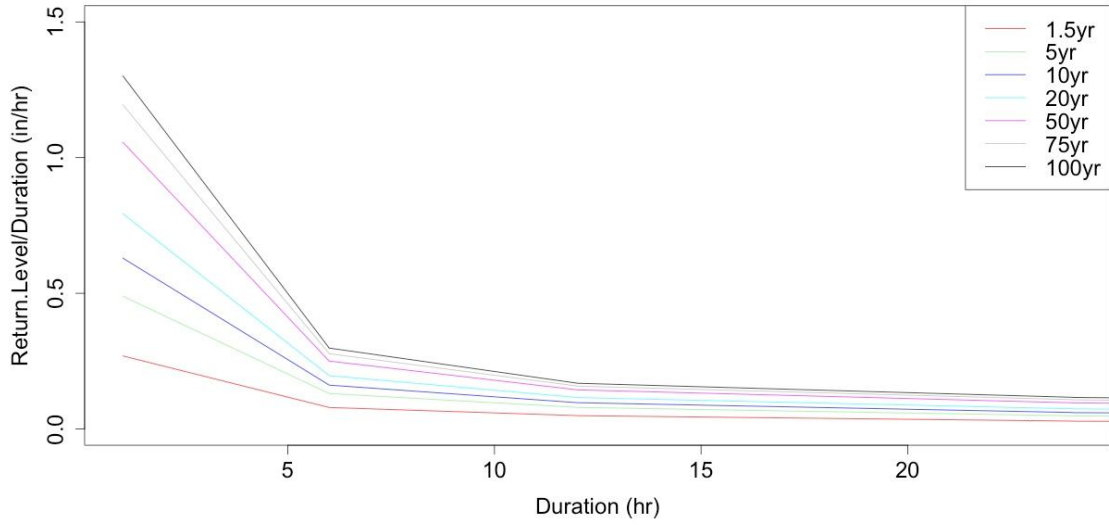


Figure 1. PHL IDF Curves From Model 0

	1.5yr	5yr	10yr	20yr	50yr	75yr	100yr
1hr	0.283	0.524	0.676	0.853	1.136	1.283	1.397
6hr	0.048	0.113	0.147	0.182	0.230	0.252	0.268
12hr	0.023	0.070	0.093	0.116	0.146	0.160	0.169
24hr	0.011	0.042	0.058	0.073	0.094	0.103	0.110
3day	0.003	0.016	0.023	0.029	0.037	0.040	0.043
5day	0.002	0.011	0.015	0.019	0.025	0.027	0.029

Table 2. Model 1 Output of Extreme Rainfall At PHL (inch/hr)

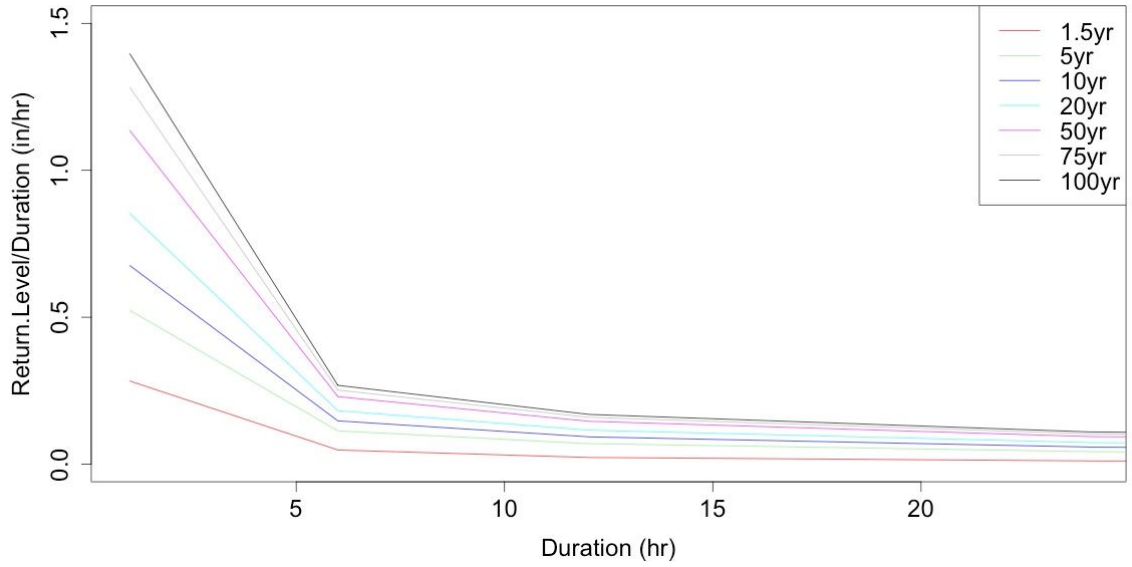


Figure 2. PHL IDF Curves From Model 1

	1.5yr	5yr	10yr	20yr	50yr	75yr	100yr
1hr	-0.014	-0.034	-0.046	-0.059	-0.079	-0.088	-0.014
6hr	0.031	0.017	0.014	0.014	0.020	0.025	0.031
12hr	0.026	0.010	0.004	0.000	-0.002	-0.002	0.026
24hr	0.018	0.006	0.002	0.001	0.002	0.004	0.018
3day	0.008	0.002	0.000	-0.001	-0.001	0.000	0.008
5day	0.005	0.001	0.000	-0.001	-0.001	-0.001	0.005

Table 3. Difference Between Model 0 and 1 Results (inch/hr)

	1.5yr	5yr	10yr	20yr	50yr	75yr	100yr
1hr	-0.014	-0.034	-0.046	-0.059	-0.079	-0.088	-0.014
6hr	0.186	0.102	0.084	0.084	0.12	0.15	0.186
12hr	0.312	0.12	0.048	0	-0.024	-0.024	0.312
24hr	0.432	0.144	0.048	0.024	0.048	0.096	0.432
3day	0.576	0.144	0	-0.072	-0.072	0	0.576
5day	0.6	0.12	0	-0.12	-0.12	-0.12	0.6

Table 4. Difference Between Model 0 and 1 Results (inch)

Range of return levels in Model 0 is from the lowest intensity of 0.028 inch per hr at 5day 100yr rainfall to the highest intensity of 0.269 inch per hour at 1hr 1.5yr rainfall. These intensity amounts could also be represented as volumes, which are 3.36 inch and 0.269 inch, respectively. For Model 1, the range is from 0.029 inch per hour at 5day 100yr rainfall to 0.286 inch per hour at 1hr 1.5yr rainfall. These intensity amounts could also be represented as volumes, which are 3.48 inch and 0.283 inch, respectively. Model 1 results shows uniform increase when compare to Model 0. These results were interpolated from Table 1 and 2, and confirmed in Figure 1 and 2.

When comparing Model 0 and 1 IDF results in details, Table 3 shows deviations as much as 0.03 inches per hour exist in the 1hr, 6hr, 1.5yr and 100yr categories. Table 4 shows difference as much as 0.6 inch at 1.5yr and 5yr return levels. Both tables confirms that Model 0 underestimates in the majority of the return levels, especially at 1.5yr and 100yr. Model 1 is observed to underestimate at 1hr RAI, but the deviations less than 0.08 inches.

	DIC
Model 0	
1hr	-68.45
6hr	-15.74
12hr	5.61
24hr	45.36
3days	54.61
5days	69.89
Model 1	542.5

Table 5. DIC results from Model 0 and 1

DIC results of Model 0 are separated by each RAI. Table 5 shows 1hr in Model 0 has the least difference between model results and the original dataset. This variation increase as the RAI increases. 5 days has the largest DIC, and this shows longer RAI would contribute to more DIC in Model 0. Since Model 1 modeled all of the RAI together, there is only one DIC value for all of the RAI together. Model 1 DIC is significantly larger than all of the DIC results in Model 0. This means the MDM has a much larger deviance from the original dataset than Model 0. Although, this statement might not be appropriate due to the fact that both models utilized different methods of IDF constructions.

Stationary and Nonstationary EPE Models

Justification of Nonstationarities in PHL AMRD

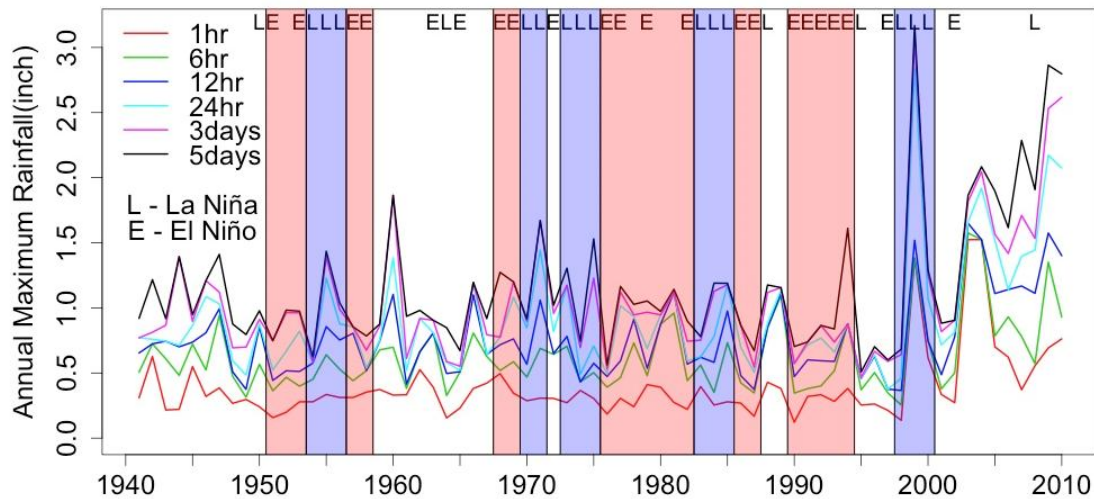


Figure 3. PHL AMRD at various RAI vs. Time, also includes information referenced from Teegavarapu (2012) and NOAA (2015) on the years of El Niño, La Niña and natural conditions. Color-coded areas represent time periods of two consecutive years or more El Niño or La Niña events. One needs to note that the ENSO index started from 1950, so there is no available information on El Niño or La Niña before then (NOAA 2015).

There is a clear upward trend in the AWRD data set at various RAI after 2000 as shown in Figure 3. The general trends among all RAI are fairly consistent with each other. Significant spikes in data are observed during 1955, 1960, 1971, 1975, 1994 and 2000. Four out of six of the spikes occurred during La Niña years. No significant increase was observed during El Niño years except 1994. This graph also shows nonstationarity within the dataset at the occasional spike before 2000 and the upward trend afterwards.

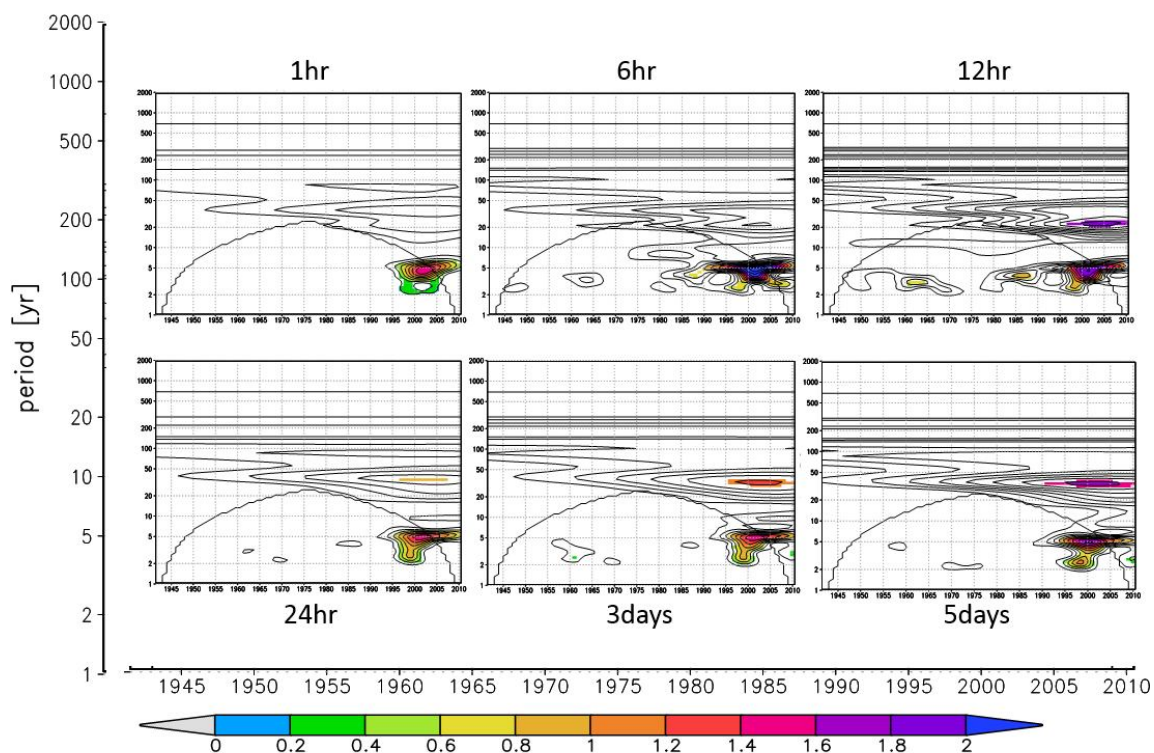


Figure 4. Wavelet Plots of PHL AMRD, the bold contour in the graphs means the 95% confidence level, anything falls outside of the contour is not significant. Also, axes of all six charts are exactly the same, and presented on the outside of the six plots. It is important to note that ENSO activities usually fall in the 2 to 5 years region. Multi-decadal oscillations such as Pacific Decadal Oscillation (PDO) and North Atlantic Oscillations (NAO) would often appear above 20 years period. These plots were generated through were generated through KNMI Climate Explorer (Truet and Oldenborgh 2013).

Regarding to the climate effects on the AMRD data are shown in Figure 4. There is a clear band of influence from 1995 to 2010 and between 2 to 5 years periods, while the 6hr, 12hr and 5 days have the strongest signals within these bands. Especially, when at 6hr and 12hr, these bands are extended into 1980. 12hr, 24hr, 3days and 5days charts all contain a narrow band from 1995 to 2010. These bands are at 30 years for 12hr, and at 40 years for the other RAI.

Predictors Selections and Justifications for Model 2 to 4

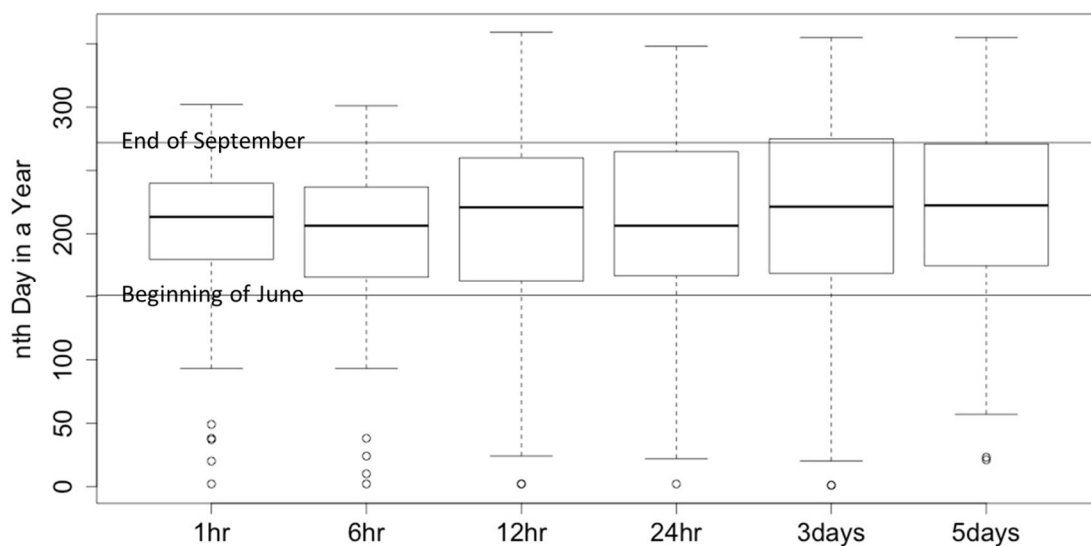


Figure 5. Time of Occurrences Boxplot of PHL AMRD

As shown in Figure 5, 25% to 95% quintile of 1941 to 2010 EPE occurred during JJAS (June to September) based on all PHL AMRD data. Thus, a decision was made to utilize AMRD values from this time period as “Rt” for Model 2, 3 and 4. Correlation results between JJAS AMRD and HadISST1 1° are shown in the appendix. From these results, it was decided to utilize MAM averaged NINO indices for “enso” in Model 3 and 4. Further correlations was investigated between various MAM averaged NINO indices with the PHL JJAS AMRD. Results of this further investigation are shown in the appendix. Based on these results, average MAM (March to May) NINO3 index has the strongest relation with JJAS 24hr annual maximum rainfall values. Once this observation

was drawn from the dataset, this study reached a decision to utilize JJAS AMRD in conjunction with the MAM NINO3 averages as to enhance the predictability of these models. In this way, Model 3 and Model 4 could potentially be used to predict JJAS extreme rainfall depth with MAM NINO3 average. This setup could potentially transform Model 3 or 4 into appropriate warning systems for detecting future JJAS EPE probabilities months in advance.

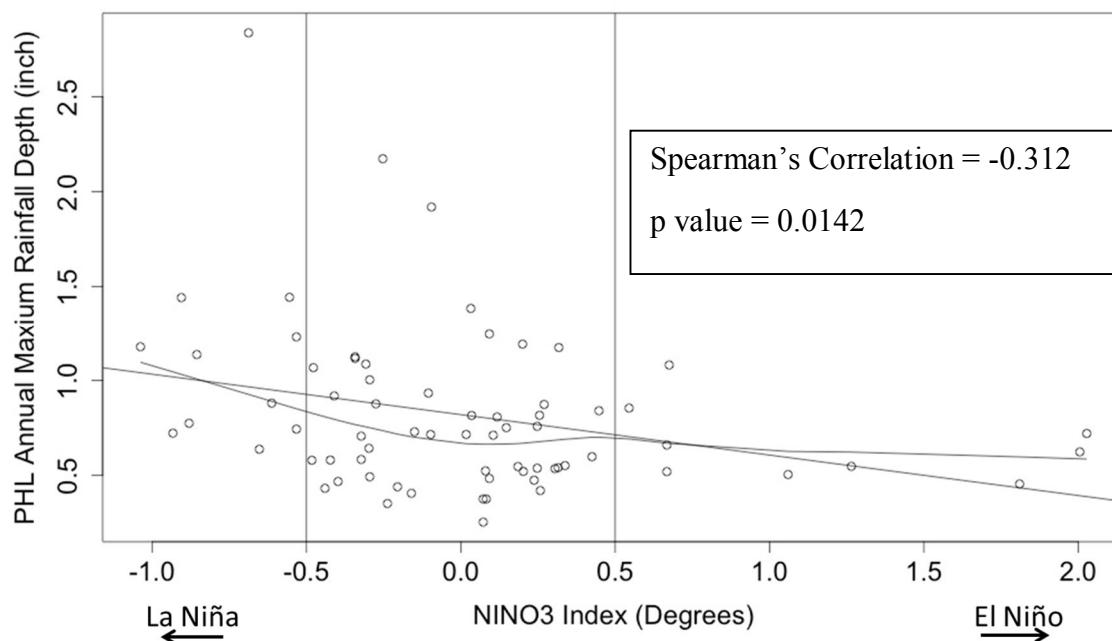


Figure 6. MAM Average NINO3 Index vs. JJAS PHL AMRD, both linear regression and the locally weighted scatterplot smoothing (lowess) are plotted in this chart. Correlations between the two data sets were calculated with spearman's correlation, because the JJAS AMRD data set is not normally distributed. Directions of El Niño and La Niña events were indicated near the NINO3 index axis to inform how SST changes during these events.

Justification of selecting NINO3 as “enso” for Model 3 and 4 was further investigated through Figure 6. The correlation between the two datasets is -0.312, which means the MAM average NINO3 index is moderately correlated with the PHL JJAS annual rainfall extremes. P value also show this correlation to be significant at 0.0142, which is lower than 5% (assumed significant level) and the null hypothesis of no correlation can be safely rejected.

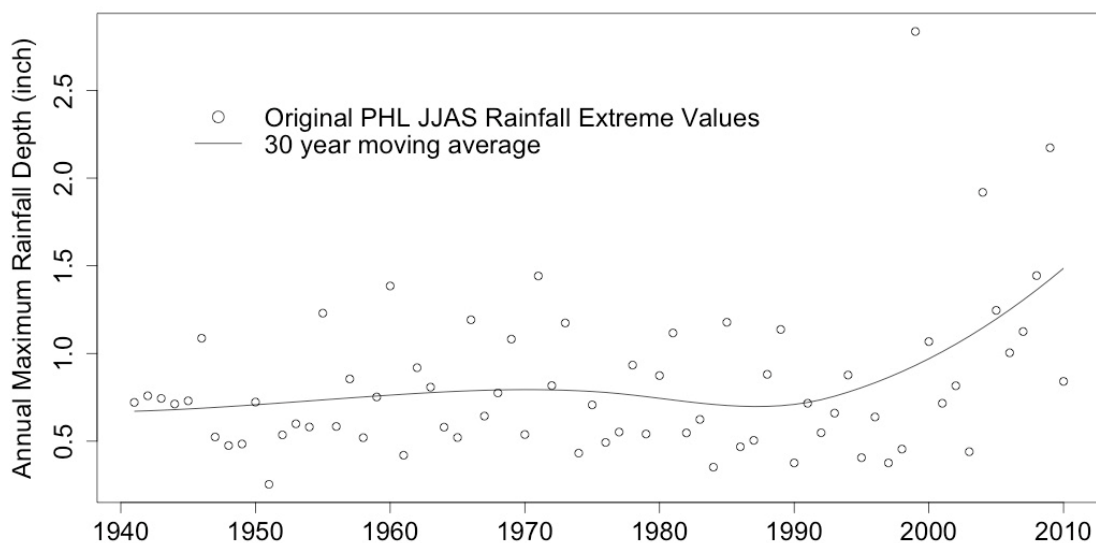


Figure 7. 30yr Moving Average of JJAS PHL AMRD vs Original Dataset

Selection of 30yr moving average is generally appropriate in representing the trend exists in the original PHL AMRD dataset according to Figure 7. This moving average successfully picked up both the slight increase and decrease from 1941 to 1985, and the significant increase after 2000. Ultimately, JJAS PHL AMRD, MAM averaged NINO3

index and 30yr moving average of JJAS PHL AMRD were selected for “ R_t ”, “enso”, and “ \overline{R}_t ”, respectively.

Comparisons among Model 2 to 4

Model 2 Results

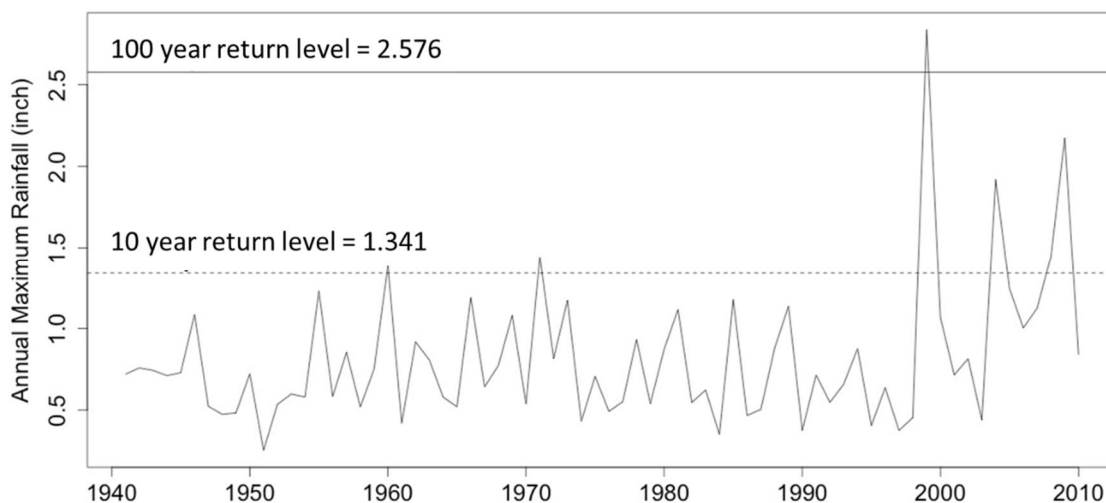


Figure 8. Model 2 Generated Return Level Against Actual JJAS AMRD, Since Model 2 is a stationary model, the return levels remain constant.

Model 2 results show the stationary 10yr and 100yr return level base on the original JJAS AMRD are at 1.341 inch and 2.576 inch, respectively. 5 out of 69 years of AMRD exceed the 10yr return level, which mean the probability of exceeding is 7.24%. For the 100yr

return level, only 1 out of 69 years exceed that. Thus, the probability of exceeding is 1.45%. These results were interpolated from Figure 8.

Location Parameter Comparisons

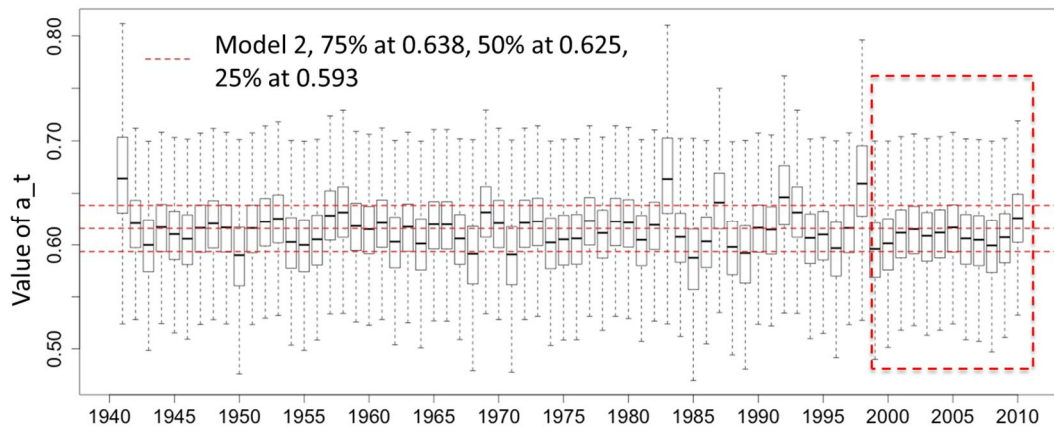


Figure 9. Comparison of Location Parameter Between Model 2 and 3, it presents the distribution of Model 3 simulated location parameter, “ $a_{t,3}$ ”, compare to the Model 2 location parameter, “ a ”. Due to the fact that a does not change over time, so the stationary Model 2 results are constant, which are shown as dash lines in these plots. Note that the outliers, which are more than 1.3 times of the upper quartile are excluded from this plots. Red rectangles added into this plot is for identifying if the models have successfully picked up increasing trend shown in Figure 3 after year 2000.

Majority of the Model 3 location parameter are very similar to the 75%, 50% and 25% quartiles of those in Model 2, as shown in Figure 9. However, there are years where the quintiles are significantly above the Model 2 levels, these years includes 1941, 1983, and 1998. Majorities of maximum and minimum excluding the outliers are fairly consistent at 0.7 and 0.53, respectively. Years which does not fall within this range of maximum and

minimum includes 1941, 1983, 1987, 1992, 1993 and 1998. Area inside the red rectangle shows that Model 3 location parameter does not pick up the increasing trend after 2000 as seen in Figure 3.

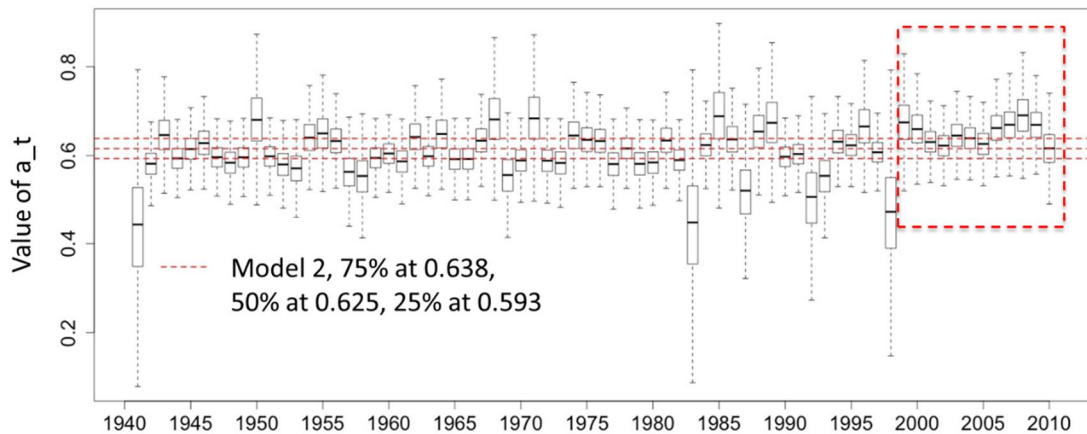


Figure 10. Comparison of Location Parameter Between Model 2 and 4, it presents the distribution of Model 4 simulated location parameter, “ $a_{t,4}$ ”, compare to the Model 2 location parameter, “ a ”. Due to the fact that a does not change over time, so the stationary Model 2 results are constant, which are shown as dash lines in these plots. Note that the outliers, which are more than 1.3 times of the upper quartile are excluded from this plots. Red rectangles added into this plot is for identifying if the models have successfully picked up increasing trend shown in Figure 3 after year 2000.

Meanwhile, majority of location parameters in Model 4 are different from Model 2 results are shown in Figure 10. Variations are very significant during 1941, 1983, 1987, 1992, 1993, and 1998. These significant differences occur at where Model 4 location parameters are lower than Model 2. 1950, 1968, 1971, 1985, 1989, and 1999 are where Model 4 has slightly higher location parameters than the Model 2 results. Area inside the

red rectangle shows that Model 4 location parameter successfully picked up the increasing trend after 2000 as seen in Figure 3.

Scale Parameter Comparisons

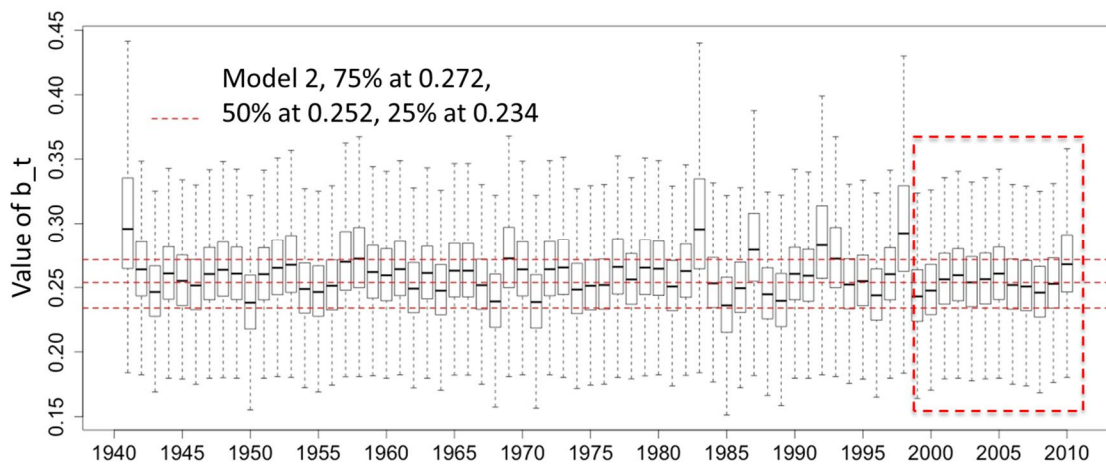


Figure 11. Comparison of Scale Parameter Between Model 2 and 3, it presents the distribution of Model 3 simulated scale parameter, “ $b_{t,3}$ ”, compare to the Model 2 location parameter, “ b ”. Due to the fact that “ b ” does not change over time, so the stationary Model 2 results are constant, which are shown as dash lines in these plots. Note that the outliers, which are more than 1.3 times of the upper quartile are excluded from this plots. Red rectangles added into this plot is for identifying if the models have successfully picked up increasing trend shown in Figure 3 after year 2000.

When comparing Model 3 and 2 scale parameters, as shown in Figure 11, departures from Model 2 are very similar to the location parameter comparison as shown in Figure 9. Positive deviations from Model 2 occur at year 1941, 1983, and 1998. Majority of

maximums and minimums excluding the outliers are fairly consistent at 0.35 and 0.156, respectively. Area inside the red rectangle shows that Model 3 scale parameter did not pick up the increasing trend after 2000 as seen in Figure 3.

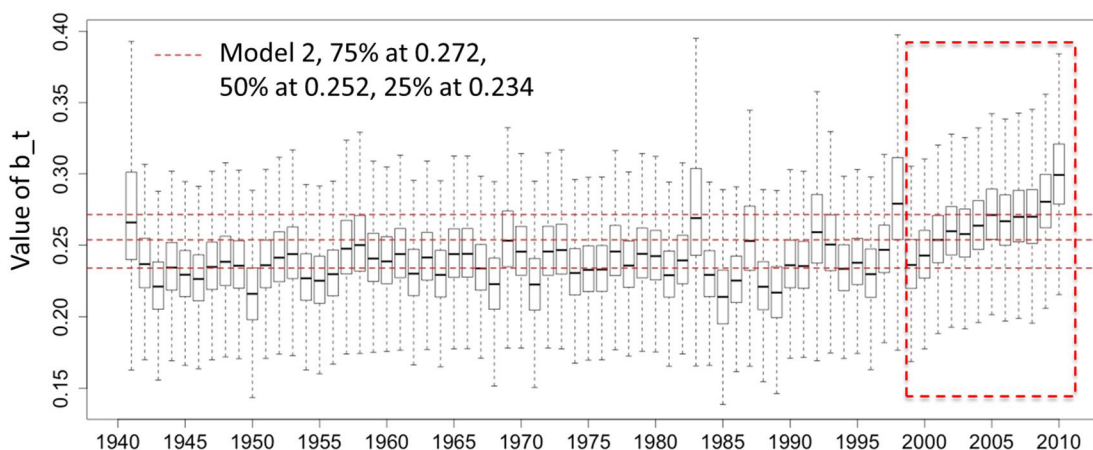


Figure 12. Comparison of Scale Parameter Between Model 2 and 4, it presents the distribution of Model 4 simulated scale parameter, “ $b_{t,4}$ ”, compare to the Model 2 location parameter, “ b ”. Due to the fact that “ b ” does not change over time, so the stationary Model 2 results are constant, which are shown as dash lines in these plots. Note that the outliers, which are more than 1.3 times of the upper quartile are excluded from this plots. Red rectangles added into this plot is for identifying if the models have successfully picked up increasing trend shown in Figure 3 after year 2000.

Scale parameter comparison between Model 4 and 2 reveals that Model 4 is roughly one quartile below the Model 2 results before 2000, as shown in Figure 12. Same spikes in Figure 11 at years of 1941, 1983, and 1998 are reflected here. Area inside the red rectangle shows that Model 4 scale parameter successfully picked up the increasing trend

after 2000 as seen in Figure 3. Majority of maximums and minimums before 2000 excluding the outliers are fairly consistent at 0.30 and 0.155, respectively.

Comparisons of shape parameters

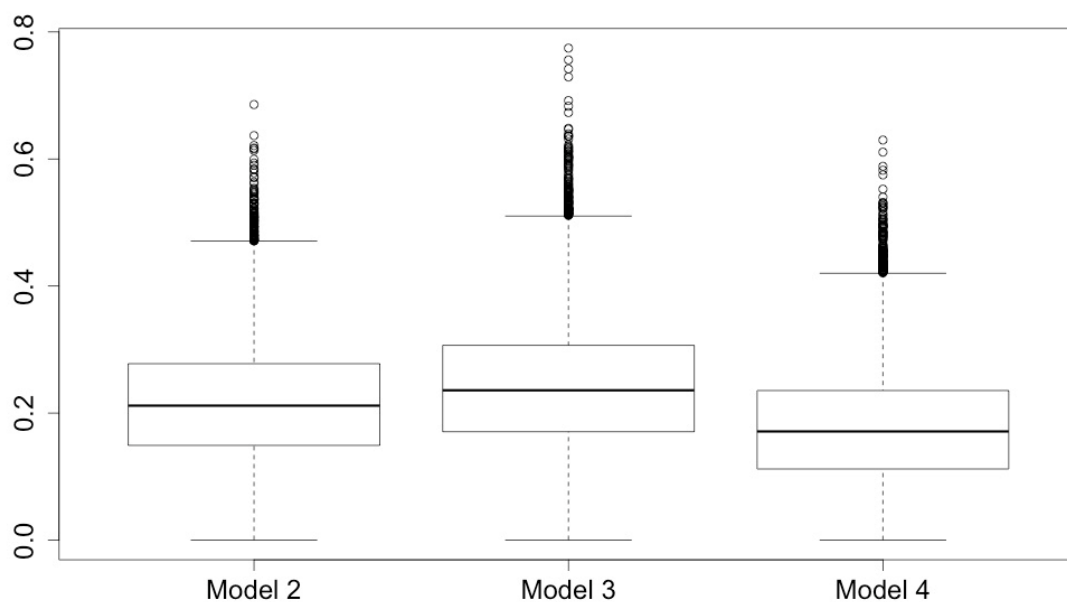


Figure 13. Comparison of Shape Parameter Among Model 2 to 4, this plot shows the distribution of estimated shape parameter from each model.

Among Model 2, 3 and 4, distributions of shape parameter show Model 3 has the highest shape parameter, while Model 4 has the lowest. At the median level, Model 3 shape parameter is 0.236, while Model 2 and 4 has 0.212 and 0.172, respectively. At the

maximum level, Model 3 shape parameter is 0.775, while Model 2 and 4 has 0.686 and 0.630, respectively.

Comparisons of alpha values

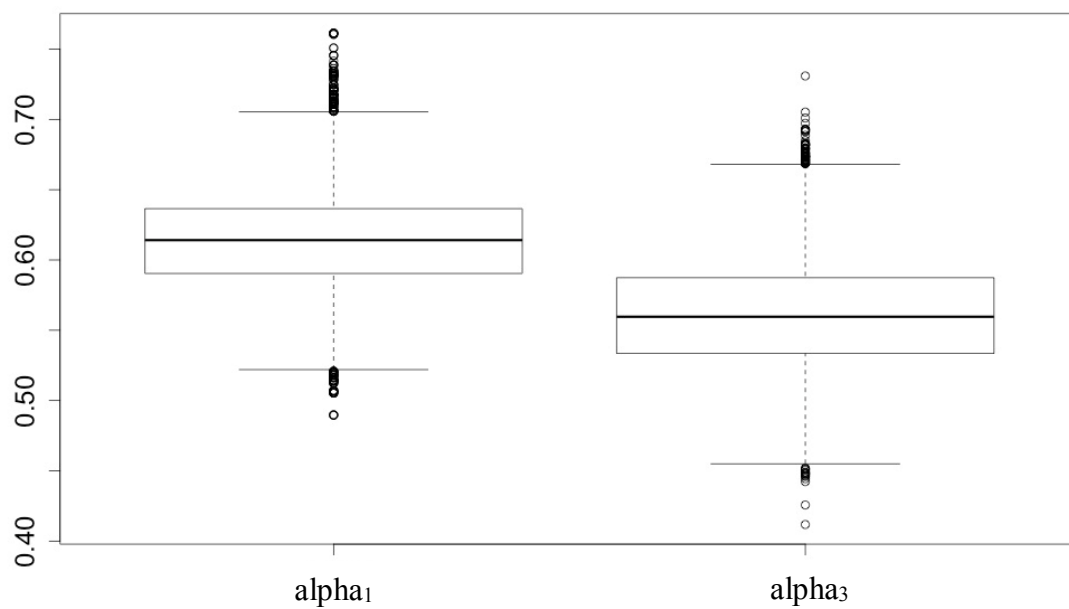


Figure 14. Comparison of α_1 and α_3 , this plot shows the distribution of estimated alpha values in equation 9 and 12.

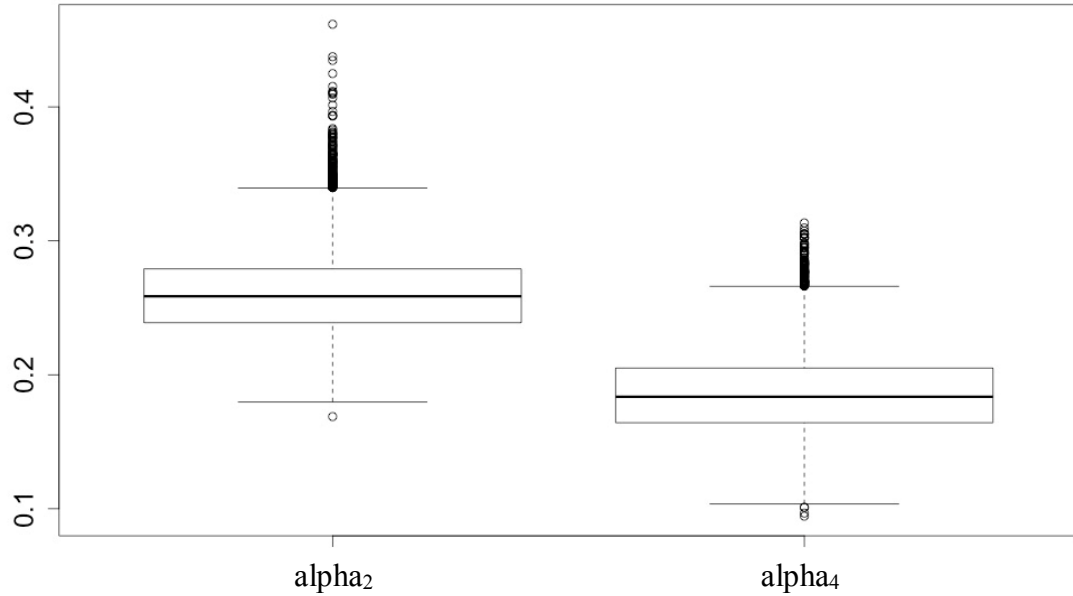


Figure 15. Comparison of α_2 and α_4 , this plot shows the distribution of estimated α values in equation 10 and 13.

$$M3: R_t \sim \text{Gev}(a_{t,3}, b_{t,3}, c) \quad \text{Equation 8}$$

$$a_{t,3} = \alpha_1 + \beta_1 * \text{enso} \quad \text{Equation 9}$$

$$b_{t,3} = \text{abs}(\alpha_2 + \beta_2 * \text{enso}) \quad \text{Equation 10}$$

$$M4: R_t \sim \text{Gev}(a_{t,4}, b_{t,4}, c) \quad \text{Equation 11}$$

$$a_{t,4} = \alpha_3 + \beta_3 * \text{enso} + \beta_4 * \overline{R_t} \quad \text{Equation 12}$$

$$b_{t,4} = \text{abs}(\alpha_4 + \beta_5 * \text{enso} + \beta_6 * \overline{R_t}) \quad \text{Equation 13}$$

Recall equation 8 to 12 above, α_1 and α_2 both contribute to location parameters of Model 3 and 4, respectively. Since α_1 has a higher distribution than α_3 are shown in Figure 14. On the other hand, the higher distribution of α_2 over α_4 was confirmed in Figure 15. When comparing Model 3 against Model 4, α_1 is

significantly higher than α_3 , while same relationship persists between α_2 and α_4 .

Comparisons of beta values

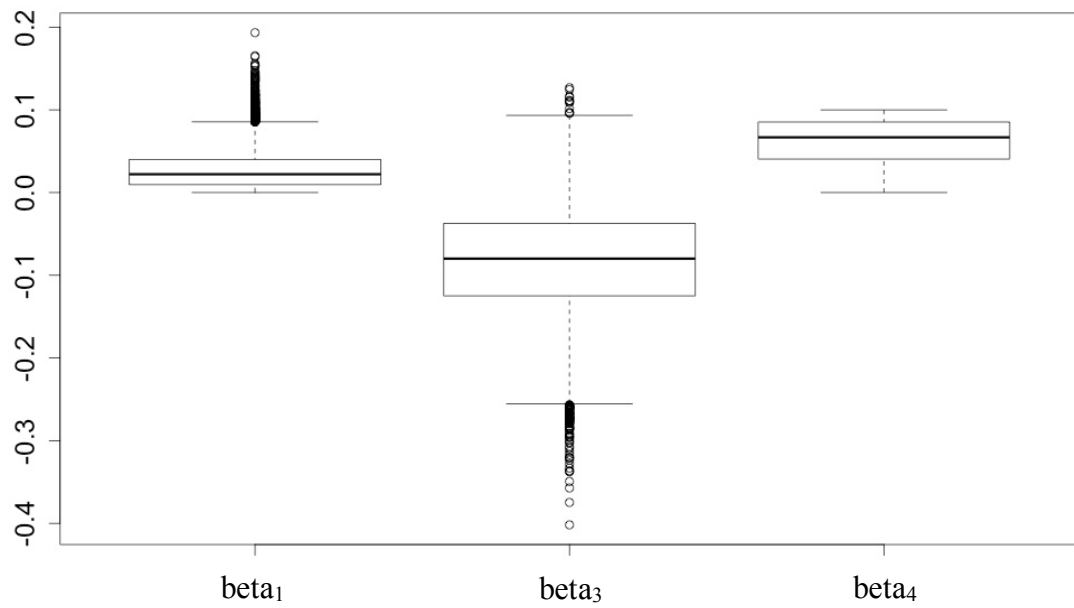


Figure 16. Comparison of β_1 , β_3 and β_4 , this plot shows the distribution of estimated beta values in equation 9 and 12.

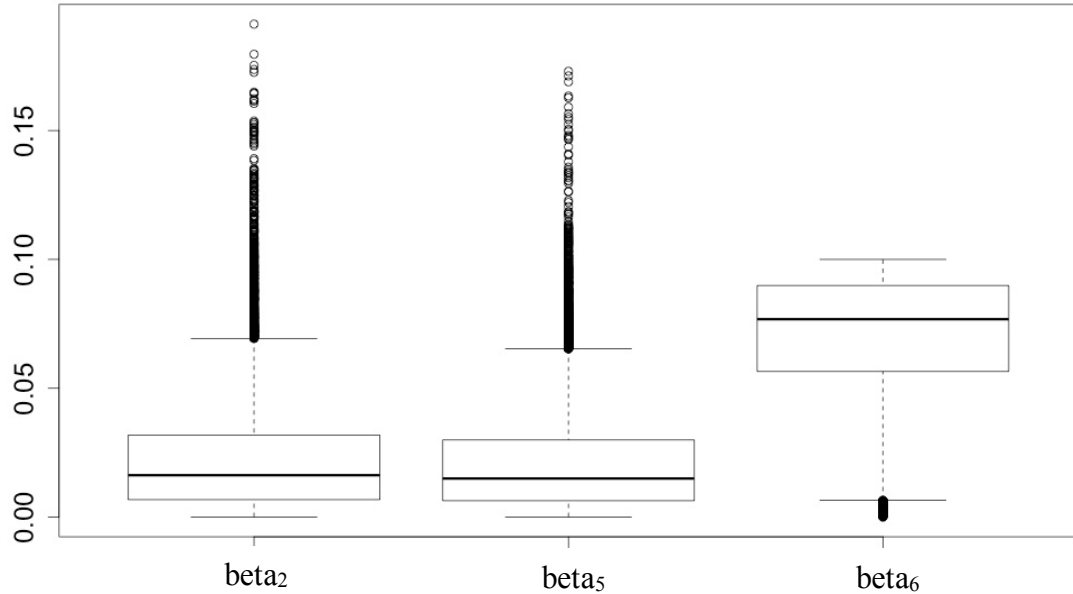


Figure 17. Comparison of beta₂, beta₅ and beta₆, this plot shows the distribution of estimated beta values in equation 10 and 13.

Recall equation 8 to 12 above, beta₁ and beta₃ link the influence of enso with the location parameters of Model 3 and 4, respectively. beta₂ and beta₅ link the influence of enso with the scale parameters of Model 3 and 4, respectively. Figure 16 show that Model 3 location parameter has positive influence from enso, Model 4 has the opposite. Scale parameters of both models has similar amount of influence as shown in Figure 17. beta₄ and beta₆ link the influence of \overline{R}_t with the location and scale parameters of Model 4, respectively. Values of beta₄ and beta₆ show that \overline{R}_t has similar influence on both location and scale parameters of Model 4. It is interesting to notice that Model 4 location parameter has an opposite way of linking enso impact from Model 3. Together, higher alpha and beta values for “enso” in Model 3 contributes to higher location parameters

than Model 4. The negative enso impact on Model 4 location parameter confirms the trend seen in Figure 11 and 12.

Comparisons between Original JJAS AMRD and Model Outputs

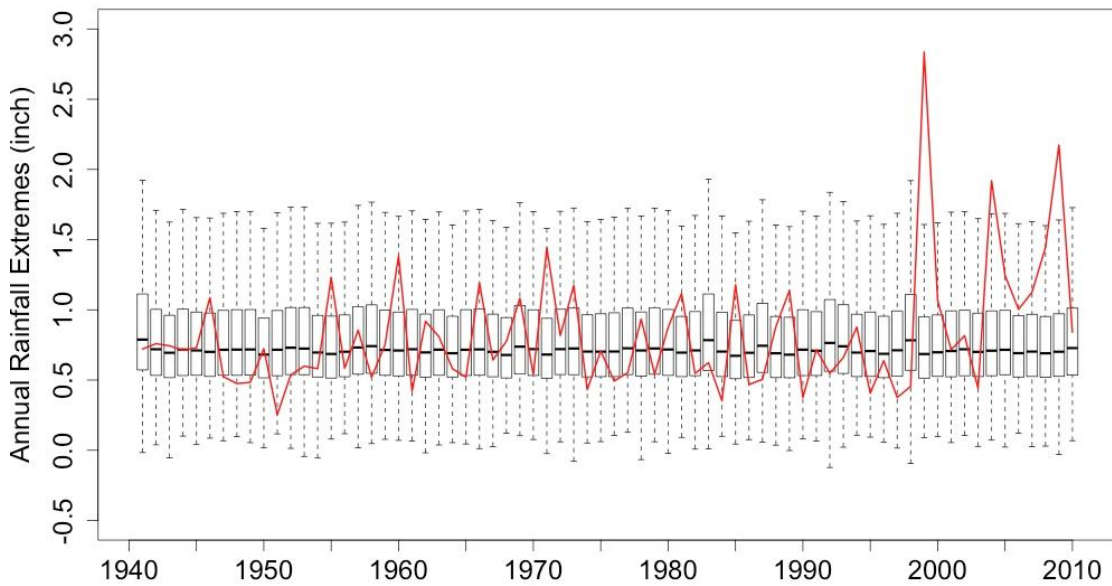


Figure 18. Model 3 vs. Original, it presents the distribution of Model 3 simulated AMRD against the original observations. Note that the outliers, which are more than 1.3 times of the upper quartile are excluded from this plots.

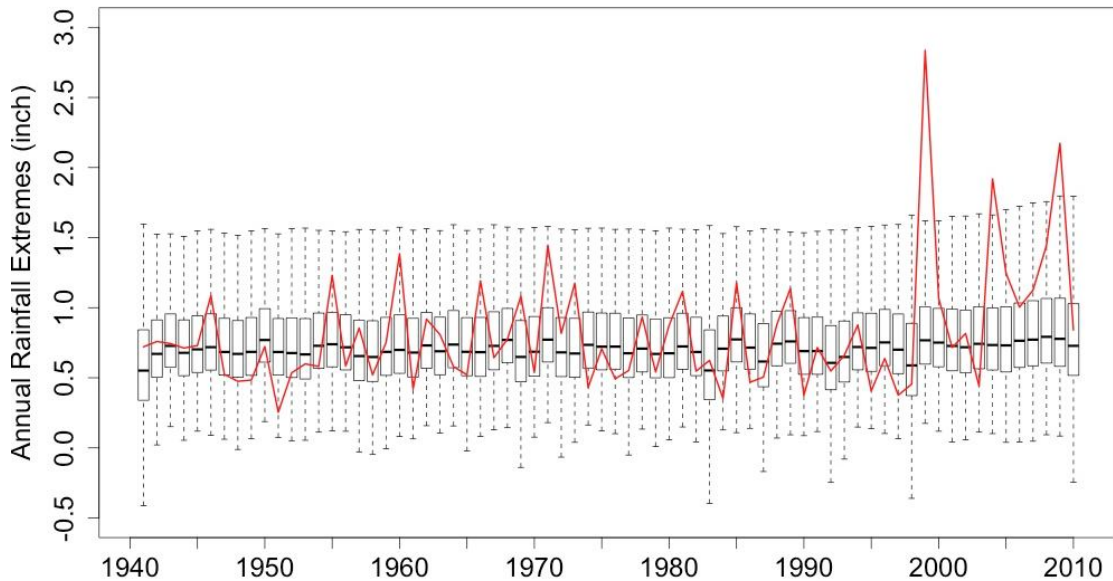


Figure 19. Model 4 vs. Original, it presents the distribution of Model 4 simulated AMRD against the original observations. Note that the outliers, which are more than 1.3 times of the upper quartile are excluded from this plots.

Observations in Figure 18 show Model 3 predicted results have no strong correlation with the trends in the original AMRD dataset. 25%, 50% and 75% quantiles of Model 3 produced AMRD show stationary pattern during the entire study period. On the other hand, Figure 19 show Model 4 AMRD has shown nonstationarity when compare to the original AMRD. Before 2000, trends of increase and decrease in original AMRD was reflected in the Model 4 AMRD results. After 2000, the increasing trend of AMRD was also reflected in Model 4, especially at the upper maximum level excluding outlines.

	DIC
Model 2	45.36
Model 3	49.45
Model 4	44.53

Table 6. DIC results from each Model

Model fit results of Model 2 to 4 were also analyzed under statistical methods. DIC results shown in Table 5 demonstrates that Model 4 has the best overall fit to the original AMRD, because the DIC values is the smallest.

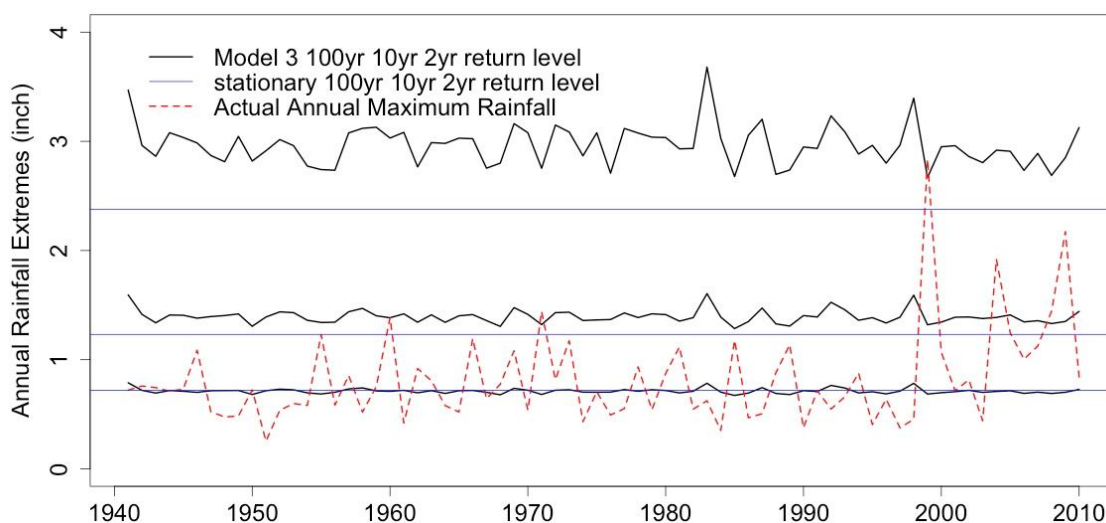


Figure 20. Model 3 Results against Original AMRD Return Levels. This figure matches the Model 3 JJAS AMRD output with original data shown as red dash line. Blue lines were drawn to show the 100yr, 10yr and 2yr return level of the original JJAS AMRD data. Black lines were drawn from the Model 3 output at the same return levels.

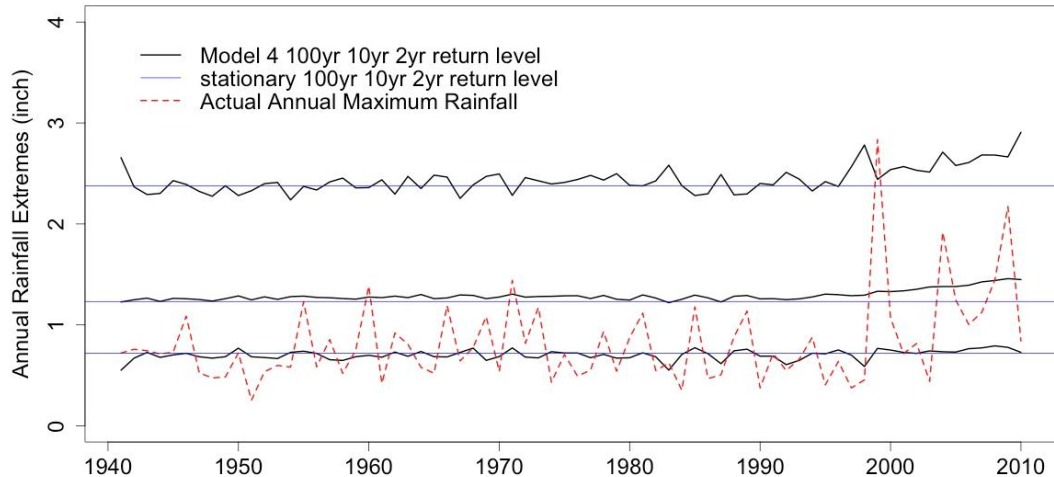


Figure 21. Model 4 Results against Original AMRD Return Levels. This figure matches the Model 4 JJAS AMRD output with original data shown as red dash line. Blue lines were drawn to show the 100yr, 10yr and 2yr return level of the original JJAS AMRD data. Black lines were drawn from the Model 3 output at the same return levels.

From results show in Figure 20 and 21, Model 3 and 4 generated 2yr return period in general is fairly similar as the 2yr return period from Model 2. For 10yr return period, the deviations between Model 2 and Model 4 results are small, while Model 3 increase slightly. This increase is amplified at the 100yr return level when comparing Model 3 to Model 2, while Model 4 does not deviate much from the Model 2 JJAS AMRD 100yr return level. Model 3 did not pick up the increasing trend in JJAS AMRD after 2000, while Model 4 did. This increasing trend amplifies from 2yr to 100yr return levels.

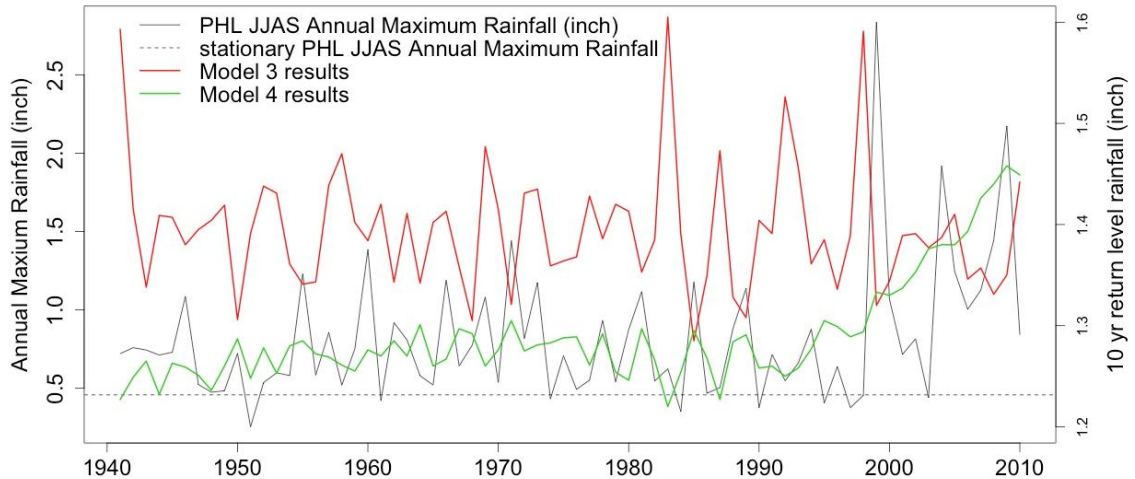


Figure 22. Comparison of 10yr Return Level Against Actual Extreme Rainfall, This focus on one single return level is mainly due to the fact that Philadelphia stormwater design is heavily based on 10yr return levels (City of Philadelphia, 2014). One need to note that all return level related information are referred to the right axis, and the nonstationary model that has a similar trend from the original AMRD dataset should have a better modeling performance.

Detailed trends at the 10yr return level are shown in Figure 22. Model 4 correlates well with the trends in the original JJAS AMRD data set, while Model 3 results do not. Increasing trend after 2000 is much better captured by Model 4 than Model 3. Although, both Model 3 and 4 are above the Model 2 JJAS AMRD 10yr level. Before 2000, the average difference is about 0.005 for Model 4 and 0.16 for Model 3.

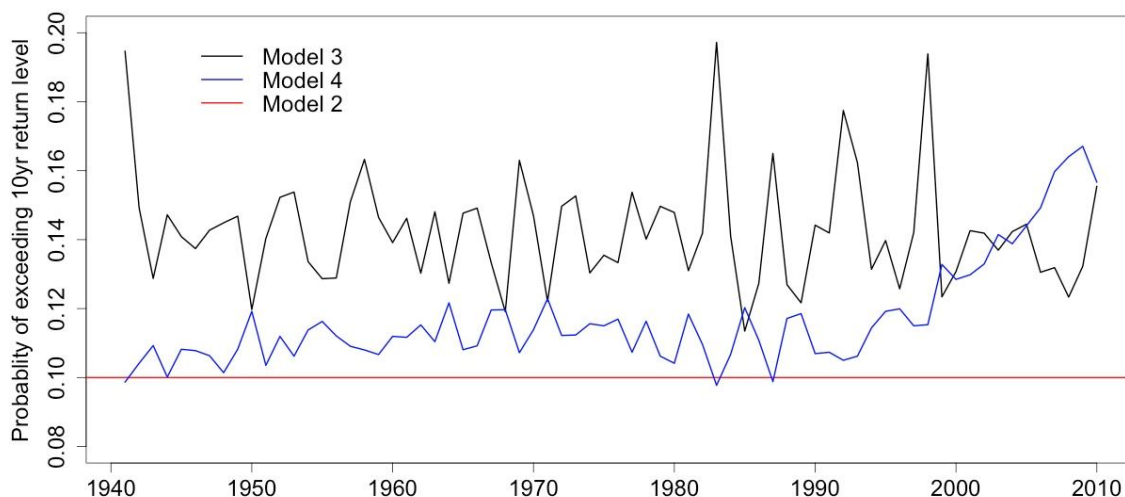


Figure 23. 10 year Return Level Exceedance Comparison, Model 2 has the stationary probability of 0.1 because it was developed in a stationary framework. Unlike Model 2, Model 3 and 4 both have probabilities of exceedance changing over time due to nonstationarity.

Figure 23 shows the average of exceedance at the 10yr return level is roughly around 15% for Model 3. However, some significant spike in Model 3 results was observed, which includes 1941, 1984, 1986, 1993, and 1998. Model 4 has a rough average around 11% exceedance before 1995, then the probability increase towards 17%. Figure 23 reflects the trends in Figure 22, where Model 3 overestimate the Model 2 10yr return level by roughly 5%. Model 4 has a difference from Model 2 at roughly 1% before 1995, and it successfully picked up the increasing trend after 2000 in original JJAS AMRD as shown in Figure 3.

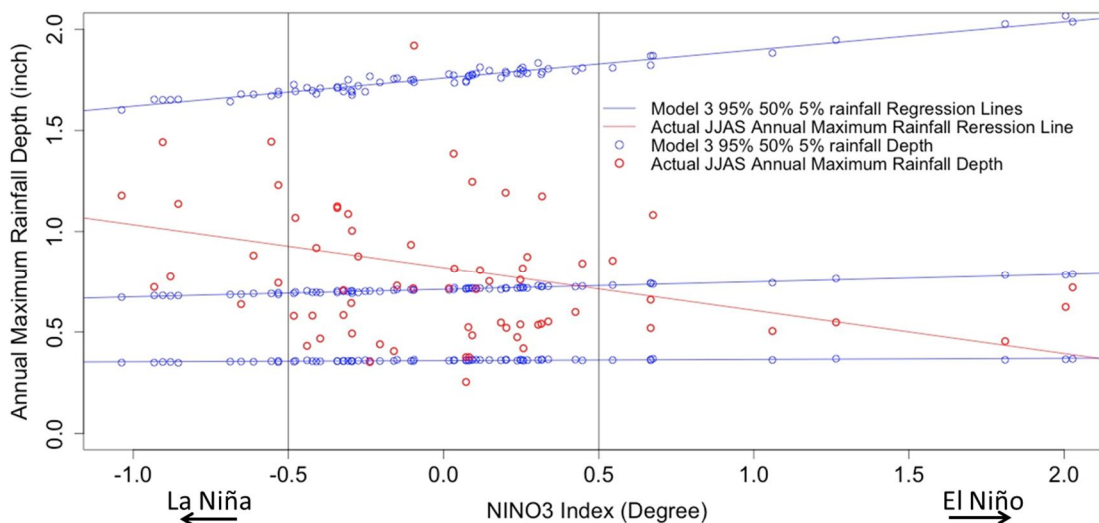


Figure 24. ENSO Impact on Model 3 Results, this plot was produced to examine how MAM ENSO effect Model 3 JJAS AMRD outcomes, compare that with the original dataset. Both blue points and the regressions are plotted in this chart from the 95%, 50% and 5% of all datasets. Red regression lines and points were from the original JJAS AMRD.

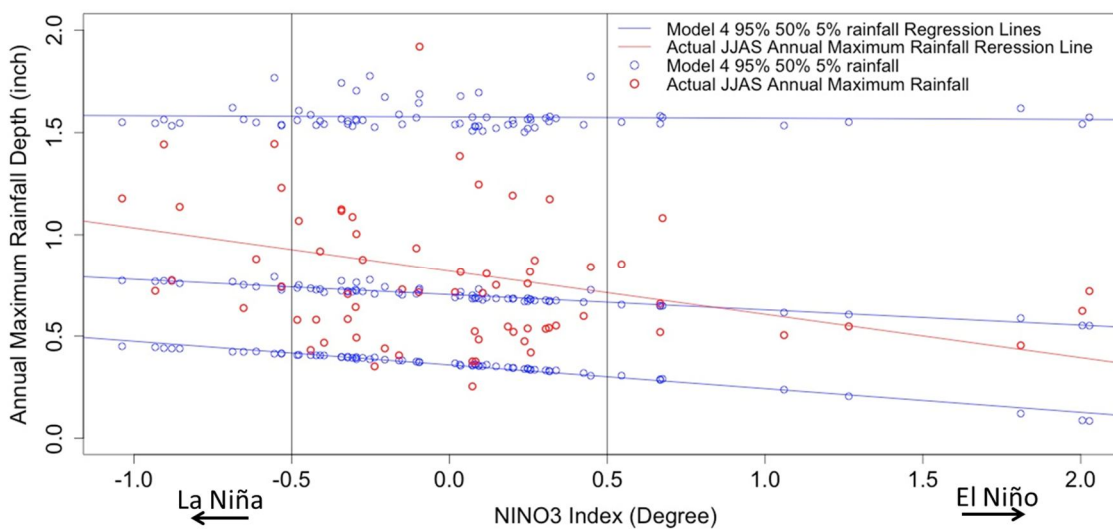


Figure 25. ENSO Impact on Model 4 Results, this plot was produced to examine how MAM ENSO effect Model 4 JJAS AMRD outcomes, compare that with the original dataset. Both blue points and the regressions are plotted in this chart from the 95%, 50% and 5% of all datasets. Red regression lines and points were from the original JJAS AMRD.

ENSO impact on the difference between original JJAS AMRD and Model Outputs, shown in Figure 24 and 25, demonstrated that Model 3 and 4 captures majorities of the original data within their 95%, 50% and 5% percentiles. However, Model 3 deviations from the original at the 95% are greater than Model 4. Slopes of trends in both model outputs are not similar to the original dataset regression line. Although, Model 4 is closer to Model 2 than Model 3, because Model 4 follows the original decreasing trend, shown in Figure 6, where Model 3 did not.

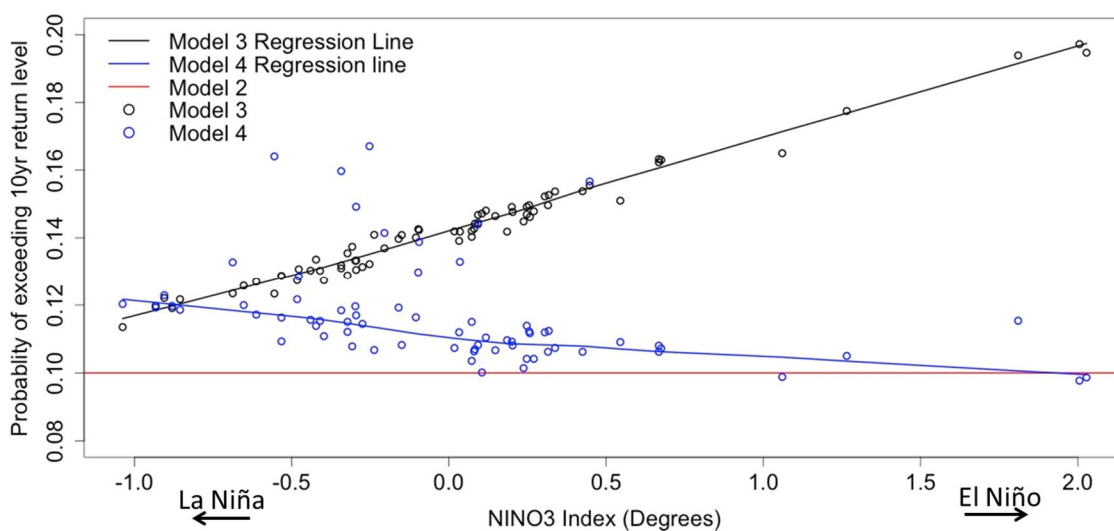


Figure 26. ENSO Impact on Exceeding 10yr Return Levels, This plot represents how ENSO affect the results of exceeding 10yr return level in Model 2, 3 and 4 results. Black line and points represent the results from Model 3, while Model 4 results are shown in blue points and line. Model 2 probability of exceedance remains stationary at 10% in red.

On the other hand, ENSO impact on exceeding 10yr return levels, as shown in Figure 26, demonstrates that Model 3 results has a positive correlation between MAM average NINO3 Index with PHL JJAS AMRD, while Model 4 has the opposite trend. Degrees of departure from Model 2 results are greater in Model 3 than Model 4. Model 3 results show that El Niño will increase PHL JJAS AMRD to up to 20% of probability, while La Niña will decrease it to about 11.5%. On the other hand, Model 4 shows the opposite, La Niña events will increase the chance of exceedance to 12%, while El Niño decrease it to the stationary 10% level.

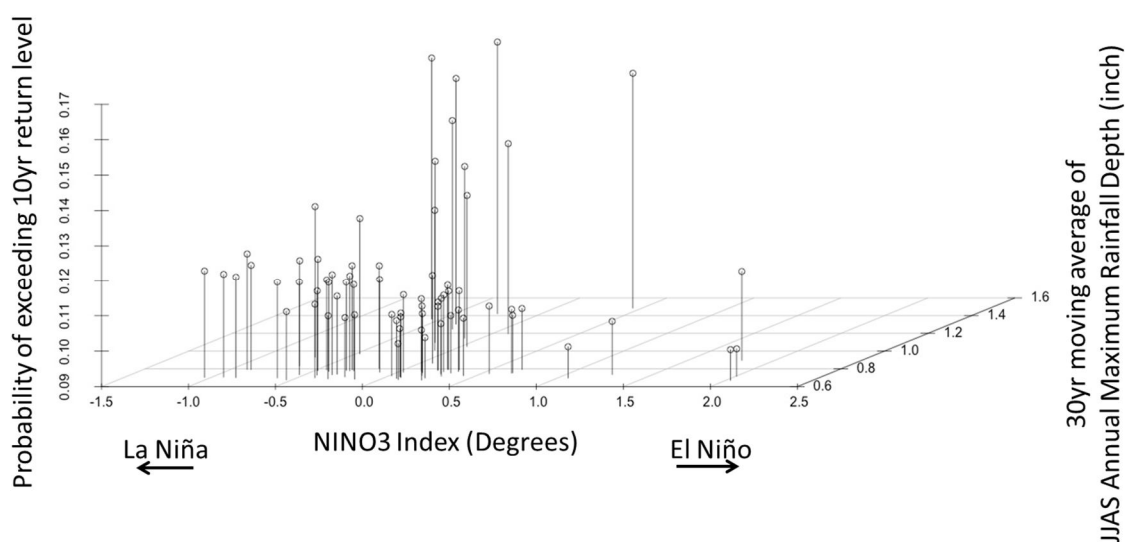


Figure 27. 3D Plot of Model 4 Results, this was produced through adding the information of 30yr moving average into the chart shown in Figure 26. It not only explains how ENSO affect the probability of exceeding the stationary 10yr level, but also helps to understand how the moving average indicator affects the model outcome dynamically with ENSO.

Model 4 3D plot, Figure 27 reflects the same observations in Model 4 results as shown in Figure 26. Also, points of high chances of exceedance between -0.5 to 0.5 degrees (also as known as the “natural” condition in ENSO) are due to corresponding high moving average values. In this range, 30 year moving average has a significant impact on probability of exceeding 10 year in the natural ENSO regions.

Discussion and Conclusion

This paper offers two improvements on the existing EPE modeling practices. One introduces a multivariate model in stationary IDF analysis for faster and more accurate data interpolations. The other completely replaces the traditional framework with non-stationary EPE models, which takes in considerations of climate, temporal and possible spatial indicators. These two improvements are vital to the future water resource planning, especially to improve the resilience and adaptiveness of urban storm water designs in a changing climate.

For the stationary IDF improvement, one could utilize model 1 as a faster and better IDF analysis tool, because it not only can model various RAI together in one simulation, but also can link in the relationship between various RAI. This study demonstrates the difference between analyze IDF independently and dependently among different RAI can have impact on the final IDF results. Results in this project reflect that the deviation can be as much as 0.6 inch as shown in Table 4. Even through this difference is small, such deviation could contribute to large scale damage downstream. For example, if the

impervious area is 10000 ft², 0.6 inch of difference could contribute to a 500 ft³ of additional runoff, which could potentially cause many associated downstream problems.

On the other hand, preliminary analysis in this study confirms the fact that PHL EPE does not follow a stationary trend. The nonstationarity in PHL AWRD is verified by two main observations: 1) Significant increase in AWRD after 2000 are observed in Figure 3. One also need to note this trend was confirmed by NOAA's Climate Extreme Index (Gleason et al. 2008). Reason to this increase is inconclusive. On the other hand, it was observed in the NOAA dataset (2015) that from 2005 to 2010 that there is a consistent transition from El Niño to La Niña or vice versa for each year except 2008. This transition was not observed as frequent as in the years before 2005. From this figure, one could also conclude that there is a strong relationship between increases in annual rainfall extreme depth and La Niña years. Meanwhile, El Niño years suppress or maintain PHL AMRD. 2) Clear ENSO connection to all RAI was conformed in Figure 4, and the connection is strong from 1995 to 2010. Even though associations with Multi-Decadal Oscillations are not very significant, one obvious reason is because the analyzed data only contained 69 years of data. One would assume that if more data was recorded, theses associations would be clearer.

Under these nonstationary conditions, the traditional IDF modeling based on stationary framework is surely not appropriate. This means the need for a nonstationary EPE models is justified. If one utilize the traditional IDF return level modeling, as shown in Model 2, the risk of such decision is significant. This study reveals that with both nonstationary models increase the probability of exceeding the 10yr return level by 6% at maximum,

and 1% at minimum. Most importantly, the chance of exceedance does not follow a stationary trend as shown in Figure 23.

Various comparisons among Model 2 to 4 results show that Model 4 has the best overall performance in modeling the original JJAS AMRD. DIC results in Table 5 have confirmed this fact. Model 4 follows the trend of the original data when comparing the ENSO effects, where Model 2 and 3 does not. On top of that, it successfully picked up the increasing trend of AWRD in Figure 3 after 2000. There is some bias inside Model 3. Not only because it did not show significant nonstationarity, but also it overestimates the AWRD at both 100yr and 10yr return levels. The above findings have proven success in Model 4 to be the best candidate for future modeling work.

Future Work

Future work of this study will address the cross validation of models in this study, which will ultimately address the model fit to the original dataset. This cross validation would utilize the model to predict AMRD based on consecutive or random selected periods from 1941 to 2010, and then validate it against the original dataset. On the other hand, it would also be appropriate to cross validate the results from Model 3 and 4 with consecutive or random time intervals. This would test how well does the predictive function operates.

Meanwhile, there are still vast amount of possible modeling indicators and spatial relationships one could add to the Models in this study. Possible indicators could include the concentration of CO₂ in the atmosphere, AMRD at nearby locations, pressure index

at nearby sites and so on. It is possible that the nonstationary model could improve accuracy when more strong correlated indicators are introduced.

Another possible next step for this study is to plug in the output of this model into hydrologic modeling software, such as SWMM. One could utilize the nonstationary interpolated in models developed by this study to investigate how frequent does one node (ex. drainage pipe or storm basin) will overflow on an annual basis. Future practitioners could utilize this information to either improve the existing stormwater network or better manage the present systems.

Bibliography

1. Kirshen, P., C. Watson, E. Douglas, A. Gontz, J. Lee, and Y. Tian, 2008: Coastal flooding in the northeastern United States due to climate change. *Mitigation and Adaptation Strategies for Global Change*, 13, 437-451, doi:10.1007/s11027-007-9130-5.
2. Horton, R., G. Yohe, W. Easterling, R. Kates, M. Ruth, E. Sussman, A. Whelchel, D. Wolfe, and F. Lipschultz, 2014: Ch. 16: Northeast. *Climate Change Impacts in the United States: The Third National Climate Assessment*, J. M. Melillo, Terese (T.C.) Richmond, and G. W. Yohe, Eds., U.S. Global Change Research Program, 16-1-nn.
3. Coumou, Dim, and Stefan Rahmstorf. "A decade of weather extremes." *Nature Climate Change* 2, no. 7 (2012): 491-496.
4. Luber, G., K. Knowlton, J. Balbus, H. Frumkin, M. Hayden, J. Hess, M. McGeehin, N. Sheats, L. Backer, C. B. Beard, K. L. Ebi, E. Maibach, R. S. Ostfeld, C. Wiedinmyer, E. Zielinski-Gutiérrez, and L. Ziska, 2014: Ch. 9: Human Health. *Climate Change Impacts in the United States: The Third National Climate Assessment*, J. M. Melillo, Terese (T.C.) Richmond, and G. W. Yohe, Eds., U.S. Global Change Research Program, 220–256. doi:10.7930/J0PN93H5.
5. White House Report, cited 2014: *The Health Impacts of Climate Change on Americans*.
6. Darwin, C. M. A. *Origin of species*. DMP, 1978.
7. Freidline, A., 2011: Rain cuts early volume, delays summer harvest. *The Packer*, July 15, 2011. The Packer/Vance Publishing Corp. [Available online at <http://www.thepacker.com/fruit-vegetable-news/shipping-profiles/Rain-cuts-early-volume-delays-summer-harvest-125642283.html>]
8. Hatfield, J. L., K. J. Boote, B. A. Kimball, L. H. Ziska, R. C. Izaurralde, D. Ort, A. M. Thomson, and D. Wolfe, 2011: Climate impacts on agriculture: Implications for crop production. *Agronomy Journal*, **103**, 351-370, doi:10.2134/agronj2010.0303.
9. Romero-Lankao, P., J.B. Smith, D.J. Davidson, N.S. Diffenbaugh, P.L. Kinney, P. Kirshen, P. Kovacs, and L. Villers Ruiz, 2014: North America. In: *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Barros, V.R., C.B. Field, D.J.

- Dokken, M.D. Mastrandrea, K.J. Mach, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L.White (eds.)]. *Cambridge University Press*, Cambridge, United Kingdom and New York, NY, USA, pp. 1439-1498
10. World Economic Forum (2015) Global Risks 2015, 10th Edition. (Geneva: World Economic Forum)
 11. Milly, P. C. D., B. Julio, F. Malin, M. Robert, W. Zbigniew, P. Dennis, and J. S. Ronald. "Stationarity is dead." *Ground Water News & Views* 4, no. 1 (2007): 6-8.
 12. Craig, Robin Kundis. "'Stationarity is Dead'-Long Live Transformation: Five Principles for Climate Change Adaptation Law." *Harvard Environmental Law Review* 34, no. 1 (2010): 9-75.
 13. Tarlock, A. Dan. "United States Flood Control Policy: The Incomplete Transition from the Illusion of Total Protection to Risk Management." (2012).
 14. Sivapalan, Murugesu, and Jos M. Samuel. "Transcending limitations of stationarity and the return period: process-based approach to flood estimation and risk assessment." *Hydrological processes* 23, no. 11 (2009): 1671-1675.
 15. Knapp, Alan K., Claus Beier, David D. Briske, Aimée T. Classen, Yiqi Luo, Markus Reichstein, Melinda D. Smith et al. "Consequences of more extreme precipitation regimes for terrestrial ecosystems." *Bioscience* 58, no. 9 (2008): 811-821.
 16. EPA, Being Prepared for Climate Change, A Workbook for Developing Risk-Based Adaptation Plans. Rep. Washington, DC: EPA Office of Water, 2014. Print.
 17. Olsen, J. Rolf, James H. Lambert, and Yacov Y. Haimen. "Risk of Extreme Events Under Nonstationary Conditions." *Risk Analysis* 18.4 (1998): 497-510. Web.
 18. Guo, Yiping. "Updating Rainfall IDF Relationships to Maintain Urban Drainage Design Standards." *Journal of Hydrologic Engineering J. Hydrol. Eng.* 11.5 (2006): 506-09. Web.
 19. Frederick, R. H., Myers, V. A., and Auciello, E. P. (1977). "Five to 60 minute precipitation frequency for the eastern and central United States." *NWS HYDRO-35*, National Oceanic and Atmospheric Administration Technical Memorandum, U.S. Department of Commerce, Washington, D.C.

20. Bonnin, Geoffrey M., Deborah Martin, Bingzhang Lin, Tye Parzybok, Michael Yekta, and David Riley. "Precipitation-frequency atlas of the United States." *NOAA atlas* 14, no. 2 (2006).
21. Sang, Huiyan. *Extreme value modeling for space-time data with meteorological applications*. ProQuest, 2008.
22. Lima, Carlos HR, Upmanu Lall, Tara J. Troy, and Naresh Devineni. "A climate informed model for nonstationary flood risk prediction: application to Negro River at Manaus, Amazonia." *Journal of Hydrology* (2015).
23. Mondal, Arpita, and P.p. Mujumdar. "Modeling Non-stationarity in Intensity, Duration and Frequency of Extreme Rainfall over India." *Journal of Hydrology* 521 (2015): 217-31. Web.
24. Khaliq, M. N., T. B. M. J. Ouarda, J-C. Ondo, P. Gachon, and B. Bobée. "Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: A review." *Journal of hydrology* 329, no. 3 (2006): 534-552.
25. Walshaw, David. "Generalized extreme value distribution." *Encyclopedia of Environmetrics* (2013).
26. R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
27. Gilleland, E. and Katz, R. W. (2011) New software to analyze how extremes change over time. *Eos*, 92(2), 13—14
28. Jeffrey A. Ryan and Joshua M. Ulrich (2014). xts: eXtensible Time Series. R package version 0.9-7. <http://CRAN.R-project.org/package=xts>
29. M Dowle, T Short, S Lianoglou, A Srinivasan with contributions from R Saporta and E Antonyan (2014). data.table: Extension of data.frame. R package version 1.9.4. <http://CRAN.R-project.org/package=data.table>
30. Garrett Grolemond, Hadley Wickham (2011). Dates and Times Made Easy with lubridate. *Journal of Statistical Software*, 40(3), 1-25. URL <http://www.jstatsoft.org/v40/i03/>.
31. Andrew Thomas, Bob O Hara, Uwe Ligges, and Sibylle Sturtz (2006). Making BUGS Open. *R News* 6: 12-17.

32. NOAA Climate Prediction Center. Eastern tropical Pacific SST, 2015. <http://www.esrl.noaa.gov/psd/data/correlation/nina3.data>
33. NOAA Climate Prediction Center. Central Tropical Pacific SST, 2015. <http://www.esrl.noaa.gov/psd/data/correlation/nina4.data>
34. NOAA Climate Prediction Center. Eastern Central tropical Pacific SST, 2015. <http://www.esrl.noaa.gov/psd/data/correlation/nina34.data>
35. Trouet, V. and G.J. van Oldenborgh, KNMI Climate Explorer: a web-based research tool for high-resolution paleoclimatology, *Tree Ring Research*, 2013, 69, 1, 3-13, doi:10.3959/1536-1098-69.1.3.
36. Catherine Loader (2013). locfit: Local Regression, Likelihood and Density Estimation.. R package version 1.5-9.1. <http://CRAN.R-project.org/package=locfit>
37. Krakauer, Nir Y., and Naresh Devineni. "Up-to-date Probabilistic Temperature Climatologies." *Environ. Res. Lett. Environmental Research Letters* 10.2 (2015): 024014. Web.
38. Hu, Qi, and Song Feng. "AMO- and ENSO-Driven Summertime Circulation and Precipitation Variations in North America." *Journal of Climate* J. Climate 25.19 (2012): 6477-495. Web.
39. Teegavarapu, Ramesh SV. *Floods in a Changing Climate: Extreme Precipitation*. Cambridge University Press, 2012.
40. Gleason, Karin L., Jay H. Lawrimore, David H. Levinson, Thomas R. Karl, and David J. Karoly. "A revised US climate extremes index." *Journal of Climate* 21, no. 10 (2008): 2124-2137.
41. Torrence, Christopher, and Gilbert P. Compo. "A Practical Guide to Wavelet Analysis." *Bulletin of the American Meteorological Society Bull. Amer. Meteor. Soc.* 79.1 (1998): 61-78. Web.
42. Griffiths, Michael L., and Raymond S. Bradley. "Variations of twentieth-century temperature and precipitation extreme indicators in the northeast United States." *Journal of Climate* 20, no. 21 (2007): 5401-5417.
43. City of Philadelphia. *Stormwater Management Guidance Manual. Design Manual*. Philadelphia: Philadelphia Water Department, 2014. Print. Ser. 2.1.

44. Sivapalan, Murugesu, and Günter Blöschl. "Transformation of Point Rainfall to Areal Rainfall: Intensity-duration-frequency Curves." *Journal of Hydrology* 204.1-4 (1998): 150-67. Web.
45. Cheng, Linyin, Amir AghaKouchak, Eric Gilleland, and Richard W. Katz. "Non-stationary extreme value analysis in a changing climate." *Climatic Change* 127, no. 2 (2014): 353-369.
46. Shih-Chieh Kao (2008). Multivariate Frequency Analysis of Extreme Rainfall Events using Copulas [PowerPoint slides]. Retrieved from <http://web.ornl.gov/~5v1/sharefiles/200810Seminar.pdf>
47. Renard, Benjamin, and Michel Lang. "Use of a Gaussian copula for multivariate extreme value analysis: some case studies in hydrology." *Advances in Water Resources* 30, no. 4 (2007): 897-912.
48. Rayner, N.A., D.E. Parker, E.B. Horton, C.K. Folland, L.V. Alexander, D.P. Rowell, E.C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res. (Atmospheres)*, 108(14), 2-1

Appendix

Model Scripts

```

#model0
#Standard IDF curve-----

library("Kendall")
library("extRemes")
library("xts")
library("data.table")
library("lubridate")

##DO NOT SOURCE THIS SCRIPT, JUST RUN IT-----
##data cleaning functions/time conversion-----
clda<-function(input){
  t1 <- read.csv(input, header=FALSE, stringsAsFactors=TRUE)
  t2 <- t1[3:length(t1[,3]),]
  t3 <- data.frame(t2[,3],sprintf( "%04d", as.numeric(t2[,4])),t2[,9])
  t3$year<-substr(t3[,1],1,4)
  t3$month<-substr(t3[,1],5,6)
  t3$date<-substr(t3[,1],7,8)
  t3$Hour<-substr(t3[,2],1,2)
  ###Ignore hour and seconds, POSIXct will fix it
  t4<-paste(paste(t3[,4],t3[,5],t3[,6], sep="/"),t3[,7],sep=" ")
  timef<-as.POSIXct(t4,tz="",format="%Y/%m/%d %H")
  tf<-data.frame(timef,t3[,3])
  return(tf)
}

contab<-function(a,b){
  fl<-clda(a)
  names(fl)<-c("time",b)
}

```

```

return(fl)
}

#Specially made for PHL
##import file##
#PHL.Precip <- read.table("PHL Precip.txt",sep="|",header=TRUE)
#c<-PHL.Precip
#PHL<-data.frame(as.POSIXlt(c[,2],"%Y-%m-%d %H:%M:%S",
tz=""),as.numeric(c[,3]))
#colnames(PHL)<-c("time","PHL")
PHL2<-contab("PHL2.txt","PHL")
#PHL<-rbind(PHL,PHL2[352439:length(PHL2[,1]),])
PHL1<-contab("PHL1.txt","PHL")
PHL<-rbind(PHL1,PHL2)

##Gettiing rid of 999.9-----
CV999<-function(x){
  x[x[,2]==999.9,2]<-NA
  return(x)
}

PHL<-CV999(PHL)

#padding time series-----
-

madetime<-seq(from=as.POSIXct("1900-01-01 00:00:00"), to=as.POSIXct("2014-12-08
00:00:00"), by="hour")
mt<-data.frame(list(time=madetime))

```

```
time.fill<-function(orginal){
  t1<-merge(mt,orginal,all=TRUE)
  return(t1)
}
```

```
PPHL<-time.fill(PHL)
```

```
##*****
```

```
##Remove duplicated timeseries, while keeping maximum recordings-----
-----
```

```
#use next two lines to test which one to drop
```

```
#summary(PALTA[which(as.numeric(diff(PALTA[,1]))!=3600)-1,2])
```

```
#summary(PALTA[which(as.numeric(diff(PALTA[,1]))!=3600),2])
```

```
re.rep<-function(x){
```

```
  t1<-as.xts(x[,2],x[,1])
```

```
  t2 <- t1[ ! duplicated( index(t1), fromLast = TRUE ), ]
```

```
  t3<-data.frame(time=as.POSIXct(index(t2),tz="",format="%Y/%m/%d %H:%M:%S"),
  coredata(t2))
```

```
  return(t3)
```

```
}
```

```
PPHL<-re.rep(PPHL)[1:1007497,]
```

```
##*****
```

```
##cumulate into nhr rain Data###CHANGE "a" to data input rain column name-----
-----
```

```
cumcon<-function(x){
```

```
  cum<-function(i,n){sum(a[(i-n):i],na.rm=TRUE)}
```

```

    mapply(cum,x:length(a),x-1)
  }
hours<-c(1,6,12,24,72,120)

a<-PPHL[,2]
CPHL<-mapply(cumcon,hours)
#####

##add cumulated data-----
--

addc<-function(c,b){
  c$"6hr"<-c(seq(0,0,length=5),b[[2]])
  c$"12hr"<-c(seq(0,0,length=11),b[[3]])
  c$"24hr"<-c(seq(0,0,length=23),b[[4]])
  c$"3days"<-c(seq(0,0,length=71),b[[5]])
  c$"5days"<-c(seq(0,0,length=119),b[[6]])
  return(c)
}

PPHL<-addc(PPHL,CPHL)
#####

rm(a,c,PHL.Precip,PHL1,PHL2,ALTA,ALTM,LEHI,LEHI1,LEHI2,MCGU,MILL,MILL1,
MILL2,MONM,mt,NEPHL,PHL,RDNG,RDNG1,RDNG2,TREN,WILM,CALTA,CALTM,CLEHI,
CMCGU,CMILL,CMONM,CNEPHL,CPHL,CRDNG,CTREN,CWILM,
hours,madetime)

##produce annual max amount and their dates-----
-----

```

```

annu.max.dates<-function(n){
  xd[is.na(xd[,n])==T,2]<-0
  t1<-data.frame(xd[1],xd[,n]/100)
  t3<-as.xts(as.numeric(t1[,2]),t1[,1])
  t4<-apply.yearly(t3,which.max)
  t5<-apply.yearly(t3,max)
  t6<-data.frame(date=index(t4), coredata(t4),coredata(t5))
  t6[t6[,3]==0,2]<-NA
  t6[t6[,3]==0,3]<-NA
  t6[,2]<-t6[,1]+t6[,2]*3600 ##unless you need the exact dates
  t6[,2]<-yday(as.Date(t6[,2]))
  t7<-data.frame(t6[,2],t6[,3])
  return(t7)
}

an.t.v<-function(cn){
  ta1<-lapply(cn,annu.max.dates)
  awvd<-data.frame((1900:2014),ta1)
  colnames(awvd)<-
c("Interval","1_t","1_v","6_t","6_v","12_t","12_v","24_t","24_v","3D_t","3D_v","5D_t"
,"5D_v")
  return(awvd)
}

xd<-PPHL
FPHL<-an.t.v(2:7)
#####
rm(xd)

```

```

####n is column number####
IDF<-function(np){

  ##GevFIT####
  fit<-fevd(FPHL[42:111,np],type = "GEV")

  ##Return Level####
  rl1<-return.level(fit,conf=0.05,return.period=c(1.5,5,10,20,50,75,100,200,250,500))
  #convert results into data frame
  res<-as.data.frame(rl1[1:10])

  return(res)

}

cob<-lapply(c(3,5,7,9,11,13),IDF)
result<-data.frame(cob)
colnames(result)<-c("1hr","6hr","12hr","24hr","3day","5day")
rownames(result)<-
c("1.5yr","5yr","10yr","20yr","50yr","75yr","100yr","200yr","250yr","500yr")

#plot IDF####
result[11,]<-c(1,6,12,24,72,120)
t.r<-as.data.frame(t(result))
par(mar=c(5,5,4,5)+0.1)
plot(t.r[,11],t.r[,1]/t.r[,11],type="l",col=2,ylim=c(0,2),xlim=c(1,24),
      cex.axis=1.5,cex.lab=1.5,
      main="IDF of PHL with Independent RAI",xlab="Duration
(hr)",ylab="Return.Level/Duration (in/hr)")

```

```

lines(t.r[,11],t.r[,2]/t.r[,11],type="l",col="light green")
lines(t.r[,11],t.r[,3]/t.r[,11],type="l",col=4)
lines(t.r[,11],t.r[,4]/t.r[,11],type="l",col=5)
lines(t.r[,11],t.r[,5]/t.r[,11],type="l",col=6)
lines(t.r[,11],t.r[,6]/t.r[,11],type="l",col=8)
lines(t.r[,11],t.r[,7]/t.r[,11],type="l",col=9)
#lines(t.r[,11],t.r[,8]/t.r[,11],type="l",col="dark green")
#lines(t.r[,11],t.r[,9]/t.r[,11],type="l",col=107)
#lines(t.r[,11],t.r[,10]/t.r[,11],type="l",col="dark red")

legend('topright'
      , c("1.5yr", "5yr", "10yr", "20yr", "50yr", "75yr", "100yr")
      , lwd = 1
      , col = c(2,"light green",4:6,8:9)
      , cex = 1.5
      , ncol=1
      )

```

```
write.csv(t.r[,1:10],"Table1.csv")
```

```
#Multiduration Model-----
-----
```

```

#modell
model <- function() {

  for (i in 1:nduration)
  {
    for (t in 1:ntimes)
    {
      predictand[t,i] ~ dgev(a[i],b[i],c[i])
    }
  }
}

```



```

    predictand.new[t,i] ~ dgev(a[i],b[i],c[i])
  }
}

for (i in 1:nduration)
{
  b[i] ~ dunif(0,1)
  c[i] ~ dunif(0,1)
}

for (i in 1:nduration)
{
  a[i] ~ dnorm(mu.a, tau.a) #multi-level modeling
  #shared among different durations
}

mu.a ~ dunif(0,0.3)
tau.a <- pow(sigma.a,-2)
sigma.a ~ dunif(0,0.01)

}

#model to file
library(R2OpenBUGS)
model.file <- file.path(tempdir(),"model.txt")
write.model(model, model.file)

#Data Input

```

```

nduration = 6
ntimes = 70
predictand = cbind(FPHL[42:111,3],FPHL[42:111,5],FPHL[42:111,7],
                   FPHL[42:111,9],FPHL[42:111,11],FPHL[42:111,13])
data <- list("nduration","ntimes","predictand")
#data <- list("y","N")

#parameters
params <- c("a","b","c","mu.a","tau.a","sigma.a","predictand.new")
#params <- c("a","b","c","alpha","y.new")

#initial parameters
inits = NULL
#inits <- function()
# { list(a=rnorm(nduration,1),b=rnorm(nduration,0,1),c=rnorm(nduration,0,1),
#       sigma.a=runif(nduration,0,1))}

#run model, get result
out.model1 <- bugs(data, inits, params,model.file, n.iter=10000,DIC=F,debug=T)

#check Rhat
all(out.model1$summary[,"Rhat"] < 1.1)

#IDF curves

modeltoIDF<-function(n){
  a1<-out.model1$sims.list$a[,n]
  b1<-out.model1$sims.list$b[,n]
  c1<-out.model1$sims.list$c[,n]

```

```

r100<-qevd(1-(1/100),median(a1),median(b1),median(c1))
r75<-qevd(1-(1/75),median(a1),median(b1),median(c1))
r50<-qevd(1-(1/50),median(a1),median(b1),median(c1))
r20<-qevd(1-(1/20),median(a1),median(b1),median(c1))
r10<-qevd(1-(1/10),median(a1),median(b1),median(c1))
r5<-qevd(1-(1/5),median(a1),median(b1),median(c1))
r1.5<-qevd(1-(1/1.5),median(a1),median(b1),median(c1))

out<-list(r1.5,r5,r10,r20,r50,r75,r100)

return(out)
}

IDFfrommodel<-lapply(1:6,modeltoIDF)
IDFfrommodelf<-data.frame(cbind(IDFfrommodel[[1]],
                                IDFfrommodel[[2]],
                                IDFfrommodel[[3]],
                                IDFfrommodel[[4]],
                                IDFfrommodel[[5]],
                                IDFfrommodel[[6]]))

colnames(IDFfrommodelf)<-c("1hr","6hr","12hr","24hr","3day","5day")
rownames(IDFfrommodelf)<-c("1.5yr","5yr","10yr","20yr","50yr","75yr","100yr")
#IDFfrommodelf<-IDFfrommodelf[2:7,1:6]
View(IDFfrommodelf)

t.r2<-data.frame(t(IDFfrommodelf))
t.r2$hour<-c(1,6,12,24,3*34,5*24)

```

```

View(t.r2)
par(mar=c(5,5,4,5)+0.1)
plot(t.r2[,8],as.numeric(t.r2[,1])/t.r2[,8],type="l",col=2,ylim=c(0,2),xlim=c(1,24),
     cex.axis=1.5,cex.lab=1.5,
     main="IDF of PHL with Dependent RAI",xlab="Duration
(hr)",ylab="Return.Level/Duration (in/hr)")
lines(t.r2[,8],as.numeric(t.r2[,2])/t.r2[,8],type="l",col="light green")
lines(t.r2[,8],as.numeric(t.r2[,3])/t.r2[,8],type="l",col=4)
lines(t.r2[,8],as.numeric(t.r2[,4])/t.r2[,8],type="l",col=5)
lines(t.r2[,8],as.numeric(t.r2[,5])/t.r2[,8],type="l",col=6)
lines(t.r2[,8],as.numeric(t.r2[,6])/t.r2[,8],type="l",col=8)
lines(t.r2[,8],as.numeric(t.r2[,7])/t.r2[,8],type="l",col=9)
#lines(t.r[,11],t.r[,8]/t.r[,11],type="l",col="dark green")
#lines(t.r[,11],t.r[,9]/t.r[,11],type="l",col=107)
#lines(t.r[,11],t.r[,10]/t.r[,11],type="l",col="dark red")

legend('topright'
      , c("1.5yr","5yr","10yr","20yr","50yr","75yr","100yr")
      , lwd = 1
      , col = c(2,"light green",4:6,8:9)
      , cex = 1.5
      , ncol=1
      )

t.rf <- cbind(as.numeric(t.r2[,1]),
              as.numeric(t.r2[,2]),
              as.numeric(t.r2[,3]),
              as.numeric(t.r2[,4]),
              as.numeric(t.r2[,5]),
              as.numeric(t.r2[,6]),

```

```
as.numeric(t.r2[,7]))
colnames(t.rf)<-c("1hr","6hr","12hr","24hr","3day","5day")
rownames(t.rf)<-c("1.5yr","5yr","10yr","20yr","50yr","75yr","100yr")

write.csv(t.rf,"Table2.csv")
```

#model2

```
model <- function() {
```

```
  # Prior
```

```
  a ~ dnorm(0, 5)
```

```
  b ~ dunif(0, 1)
```

```
  c ~ dunif(0, 1)
```

```
  # Likelihood
```

```
  for (i in 1:n)
```

```
  {
```

```
    y[i] ~ dgev(a, b, c)
```

```
  }
```

```
}
```

```
#model to file
```

```
library(R2OpenBUGS)
```

```
model.file <- file.path(tempdir(),"model.txt")
```

```
write.model(model, model.file)
```

```
#Data Input
```

```
y = as.numeric(FPHL6to9[42:111,9])
n = length(y)
data <- list("y","n")

#parameters
params <- c("a","b","c")

#initial parameters
#inits = NULL
inits <- function() { list(a=0.5,b=0.1,c=0.1) }

#run model, get result
out.model2 <- bugs(data, inits, params,model.file, n.iter=10000,debug=T,DIC=F)

#model3
model <- function() {

  # Prior
  alpha_a ~ dnorm(0, 10)
  alpha_b ~ dunif(0, 10)
  beta_1_a ~ dunif(0, 1)
  beta_1_b ~ dunif(0, 1)
  c ~ dunif(0, 1)

  # Likelihood
  # Layer 1
  for (t in 1:N)
  {
    a[t] <- alpha_a+beta_1_a*ENSO[t]
```

```
b[t] <- abs(alpha_b+beta_1_b*ENSO[t])
}

for (t in 1:N)
{
  y[t] ~ dgev(a[t], b[t], c)
  y.new[t] ~ dgev(a[t], b[t], c)
}
}

#model to file
library(R2OpenBUGS)
model.file <- file.path(tempdir(),"model.txt")
write.model(model, model.file)

#Data Input
ENSO = ENSO3mean[,2]
y = FPHL6to9[42:111,9]
N = length(y)
data <- list("y", "ENSO", "N")

#parameters
params <- c("a", "b", "c", "alpha_a", "alpha_b", "beta_1_a", "beta_1_b", "y.new")

#initial parameters
inits = NULL

#run model, get result
out.model3 <- bugs(data, inits, params, model.file, n.iter=10000, debug=T)
```

```

#all(out.model4$summary[, "Rhat"] < 1.1)
#model4
model <- function() {

  # Prior
  alpha ~ dnorm(0, 10)
  beta1 ~ dnorm(0, 1)
  beta2 ~ dunif(0, 0.1)
  alpha2 ~ dnorm(0, 100)
  beta3 ~ dunif(0, 10)
  beta4 ~ dunif(0, 0.1)
  #b ~ dunif(0, 1)
  c ~ dunif(0, 1)

  # Likelihood
  # Layer 1
  for (t in 1:N)
  {
    a[t] <- alpha+beta1*ENSO[t]+beta2*mean[t]
    b[t] <- abs(alpha2+beta3*ENSO[t]+beta4*mean[t])
  }

  for (i in 1:N)
  {
    y[i] ~ dgev(a[i], b[i], c)
    y.new[i] ~ dgev(a[i], b[i], c)
  }
}

```



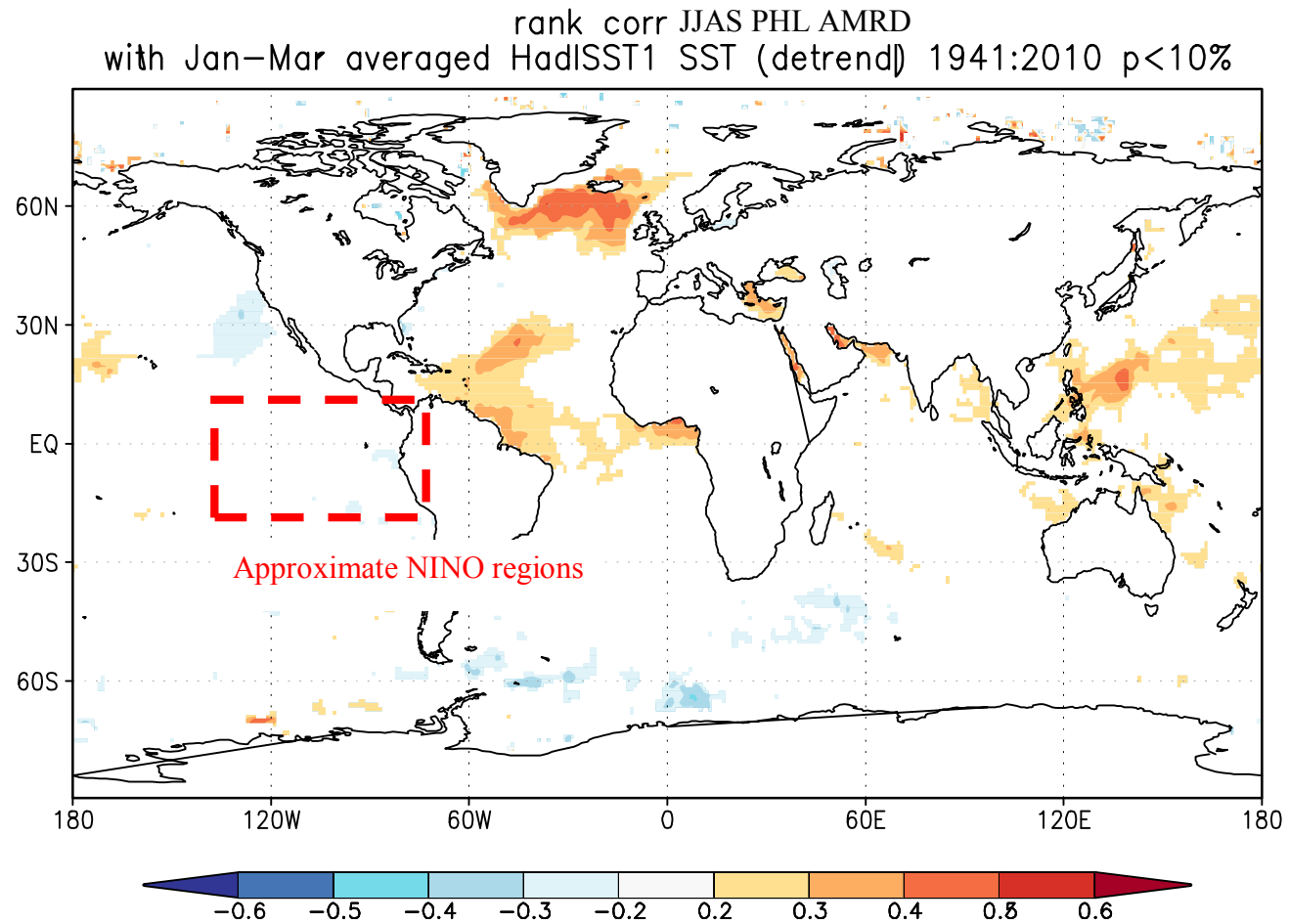
```
#model to file
library(R2OpenBUGS)
model.file <- file.path(tempdir(),"model.txt")
write.model(model, model.file)

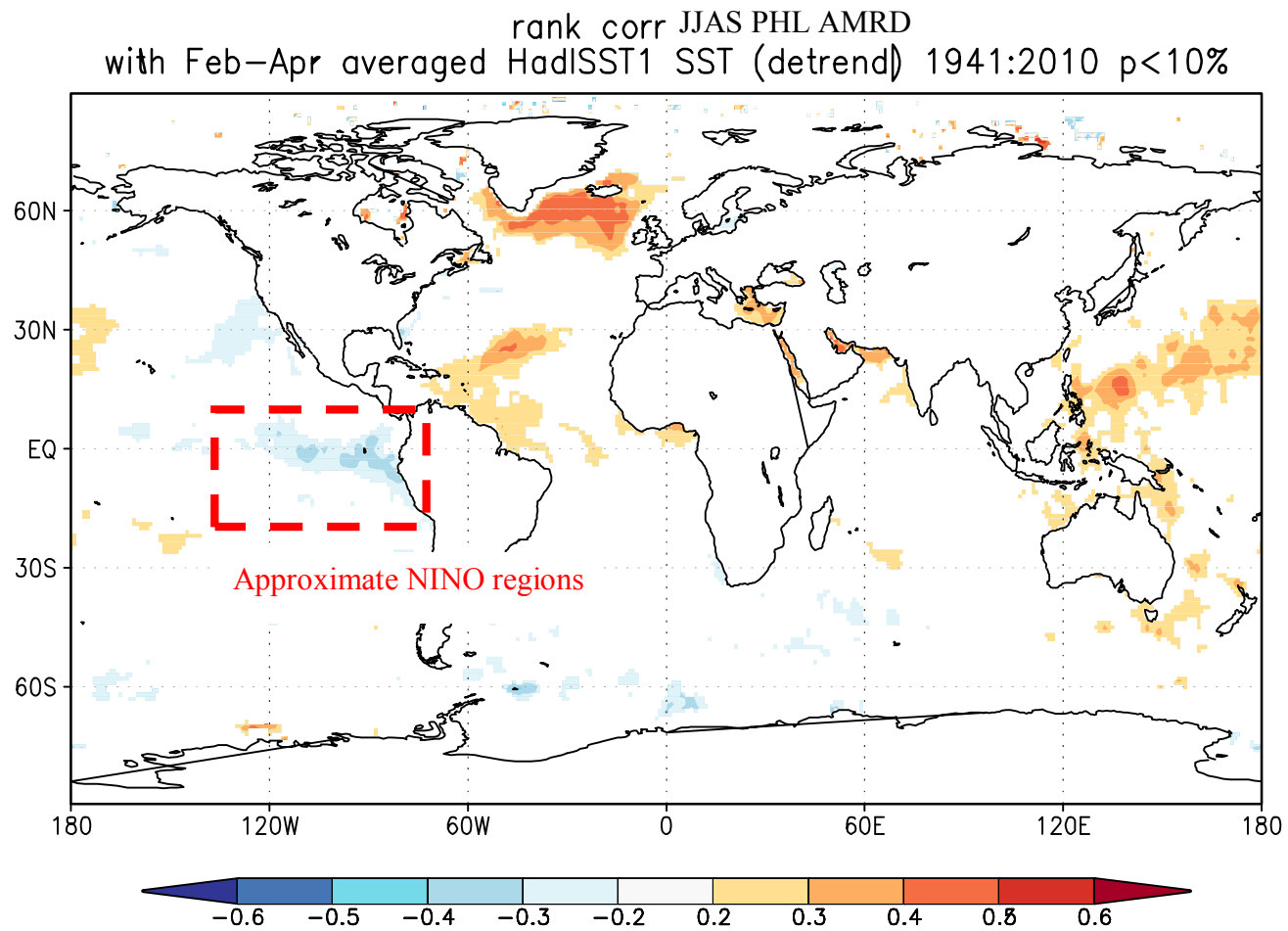
#Data Input
ENSO = ENSO3mean[,2]
mean = f$y
y = FPHL6to9[42:111,9]
N = 70
data <- list("y","ENSO","N","mean")

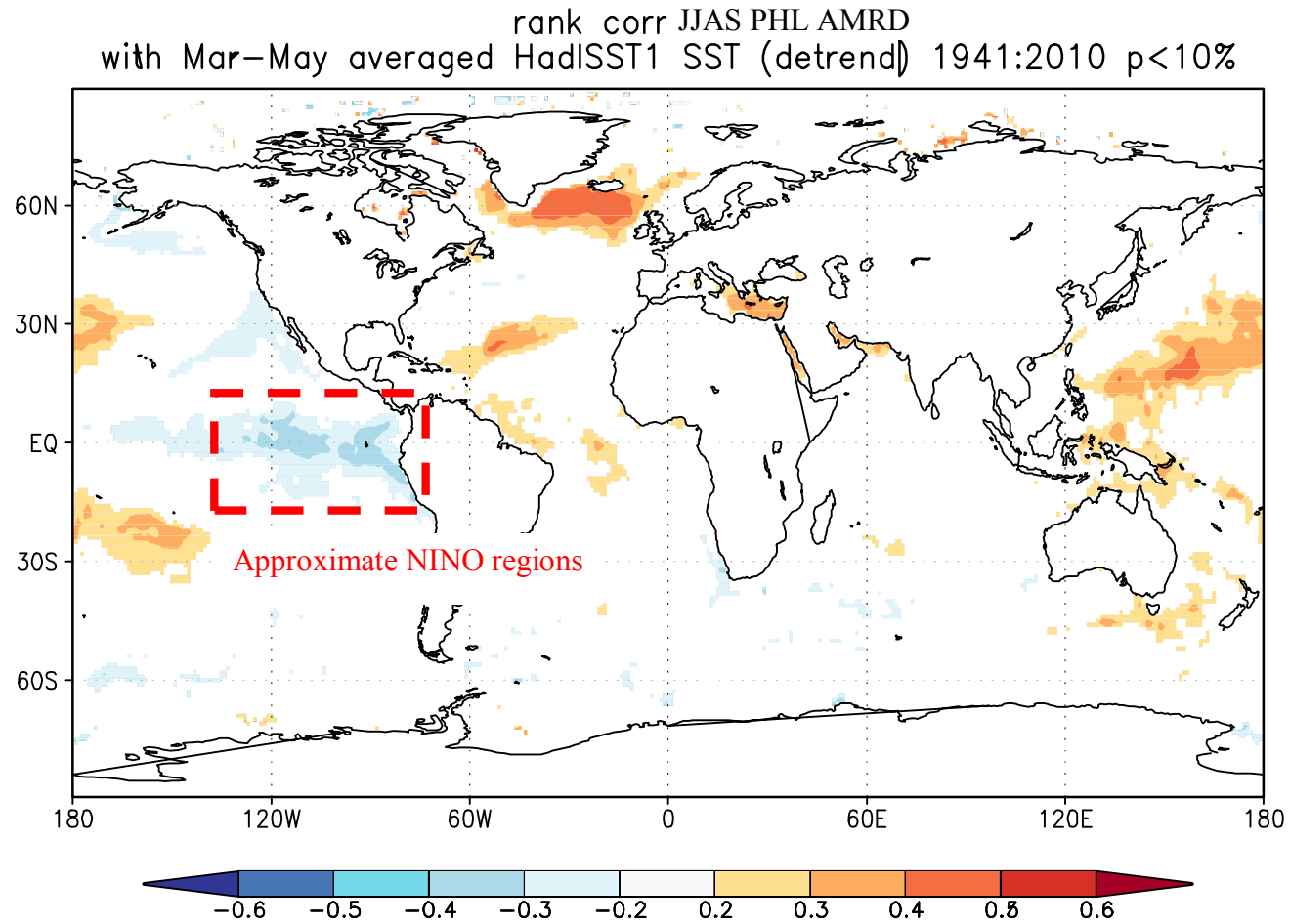
#parameters
params <- c("a","b","c","alpha","beta1","beta2",
           "alpha2","beta3","beta4","y.new")

#initial parameters
inits = NULL
#inits <- function() { list(a=rnorm(N,0,1),b=rnorm(1,0,1),c=rnorm(1,0,1),
#
#           alpha=rnorm(1,0,1),beta=rnorm(1,0,1)) }

#run model, get result
out.model4 <- bugs(data, inits, params,model.file, n.iter=10000,DIC=F,debug=T)
```

KNML and R Correlation Results





	JJAS PHL AMRD	p-values
MAM Averaged NINO 3	-0.312	0.0142
MAM Averaged NINO 3.4	-0.223	0.0163
MAM Averaged NINO 4	-0.172	0.0111

Table 7. Spearman Correlation Results from R

This Page Is Left Blank Intentionally