STUDY ON POISSON CLUSTER STOCHASTIC RAINFALL GENERATORS

A Dissertation

by

DONG KYUN KIM

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2009

Major Subject: Civil Engineering

STUDY ON POISSON CLUSTER STOCHASTIC RAINFALL GENERATORS

A Dissertation

by

DONG KYUN KIM

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Francisco Olivera
Committee Members,	Ralph Wurbs
	Scott Socolofsky
	Ming-Han Li
Head of Department,	John Niedzwecki

December 2009

Major Subject: Civil Engineering

ABSTRACT

Study on Poisson Cluster Stochastic Rainfall Generators. (December 2009)

Dong Kyun Kim, B.S., Hanyang University, South Korea; M.S., Stanford University

Chair of Advisory Committee: Dr. Francisco Olivera

The purpose of this dissertation is to enhance the applicability and the accuracy of the Poisson cluster stochastic rainfall generators.

Firstly, the 6 parameters of the Modified Bartlett-Lewis Rectangular Pulse (MBLRP) stochastic rainfall simulation model were regionalized across the contiguous United States. Each of the parameters of MBLRP model estimated at 3,444 National Climate Data Center (NCDC) rain gages was spatially interpolated based on the Ordinary Kriging technique to produce the parameter surface map for each of the 12 months of the year. Cross-validation was used to assess the validity of the parameter maps. The results indicate that the suggested maps reproduce well the statistics of the observed rainfall for different accumulation intervals, except for the lag-1 autocorrelation coefficient. The estimated parameter values were also used to produce the maps of storm and rain cell characteristics.

Secondly, the relative importance of the rainfall statistics in the generation of watershed response characteristics was estimated based on regression analyses using the rainfall time series observed at 1099 NCDC rain gages. The result of the analyses was used to weigh the rainfall statistics differently in the parameter calibration process of

MBLRP model. It was observed that synthetic rainfall time series generated weighing the precipitation statistics according to their relative importance outperformed those generated weighing all statistics equally in predicting watershed runoff depths and peak flows. When all statistics were given the same weight, runoff depths and peak flows were underestimated by 20% and 14%, respectively; while, when the statistics were weighed proportionally to their relative importance, the underestimation was reduced to 4% and 3%, which confirms the advantage of weighing the statistics differently. In general, the value of the weights depends on the hydrologic process being modeled.

Lastly, a stochastic rainfall generation model that can integrate year-to-year variability of rainfall statistics is suggested. The new framework consists of two parts. The first part generates the short-term rainfall statistics based on the correlation between the observed rainfall statistics. The second part generates the rainfall time series using the modified Bartlett-Lewis rectangular pulse model based on the simulated rainfall statistics. The new approach was validated at 104 NCDC gages across the United States in its ability to reproduce rainfall and watershed response characteristics. The result indicates that the new framework outperformed the traditional approach in reproducing the distribution of monthly maximum rainfall depths, monthly runoff volumes and monthly peak flows.

DEDICATION

This dissertation is dedicated to God, the creator of the universe and the provider of love through Jesus Christ.

This dissertation is also dedicated to my family, especially to my parents

ACKNOWLEDGEMENTS

Firstly, I give my greatest thanks to God, who created this amazing universe. Everything He created is wonder and miracle. Especially, I sincerely appreciate Him sending Jesus Christ to us and let this world realize the greatness of love. My heart will never stop loving Him.

Secondly, I would like to thank my parents, Jooyul and Hwasun; and my brother, Namkyun. They sacrificed so much for my education. Due to their love and sacrifice, my sons and daughters, and their sons and daughters too, will have bigger dreams than we have and will thrive in the environment where their dreams do not end up with dreams.

I also give great thanks to my wife, Eugene and my daughter, Minso. Eugene reshaped my angular character with her tender love. Through Minso, I realize a portion of God's love. Also, I would like to thank my in-laws, Bongchan and Kyungsook, who love me just like their own son.

I would like to thank my academic advisor, Dr. Francisco Olivera. His brilliant guidance throughout my doctoral study played the most significant role in shaping me into a better scholar and engineer. Especially, I would like to thank him for his thoughtfulness, patience, and encouragement. I truly appreciate the freedom of research that he granted me for the last year of my doctoral study. His trust on me was the greatest motivation of my research. In fact, I have considered him as a second father in the U.S.

I would like to thank Dr. Jean-Louis Briaud. Through his distinguished leadership, thorough insight, and kind-yet-strong encouragement, I accomplished more than I ever imagined.

I would like to thank my doctoral committee members. Dr. Ralph Wurbs has been my role model especially in terms of work ethic. Dr. Scott Socolofsky provided many pieces of advice which dramatically improved the quality of this dissertation. Dr. Ming-Han Li, provided invaluable advice on the ARF study, which resulted in my first official journal publication.

I also would like to thank Dr. Yongsik Cho, professor of Hanyang University, who introduced me to the field of hydrology and hydraulics. His encouragement and support led me into the path of life that is full of scholastic adventures and excitement.

Dr. Janghwoan Choi and Dr. Huidae Cho are simply the answer for my prayer. The word “guidance angel” might be the best word to describe these great friends of mine. I truly missed the casual conversations and scholastic debates that I used to have with Dr. Choi and Dr. Cho after they left.

I also would like to thank Dr. Anand Govnidasamy. His excellent project management skills played a crucial role in the successful completion of the bridge-scour project.

Lastly, I would like to thank many members of Vision Mission Church at College Station, including Pastor Youngsik Ahn. Not many people that I will meet later in my life will come to church at 6 o’clock every day morning and pray for my daughter’s health. I will pay it forward.

TABLE OF CONTENTS

Page

Page

LIST OF FIGURES

Page

Page

Page

LIST OF TABLES

1. INTRODUCTION

Stochastic rainfall generators provide synthetic rainfall input to hydrologic simulation models whenever the observed data are not available. Because they enable the Monte-Carlo simulation approach by providing infinitely long length of rainfall time series to hydrologic simulation models, they are extensively utilized to assess the risks associated with hydrologic systems such as floods [*Arnaud and Lavabre*, 1999; 2002; *Leander et al.*, 2005; *Brath et al.*, 2006; *Aronica and Candela*, 2007; *Haberlandt et al.*, 2008; *McMillan and Brasington*, 2008], draughts [*Fowler et al.*, 2003; *Harrold et al.*, 2003; *Mehrotra and Sharma*, 2007; *Mohan and Sahoo*, 2008a; b], contaminant transport [*Fernandez-Galvez et al.*, 2007; *McGrath et al.*, 2008; *Nolan et al.*, 2008], landslides [*Benda and Dunne*, 1997; *Iida*, 2004], water availability [*Lettenmaier and Sheer*, 1991; *Fowler et al.*, 2000; *Bouvier et al.*, 2003; *Shamir et al.*, 2007; *Bae et al.*, 2008; *Nunes et al.*, 2008], climate changes [*Wilby*, 1994; *Semenov and Barrow*, 1997; *Fowler et al.*, 2003; *Dibike and Coulibaly*, 2005; *Hreiche et al.*, 2007; *Kang and Ramirez*, 2007; *Kilsby et al.*, 2007; *Nunes et al.*, 2008] and ecosystems [*Porporato et al.*, 2004; *Muneepeerakul et al.*, 2008].

Poisson cluster rainfall models are one of the most widely utilized stochastic rainfall generators due to their superior structural robustness and applicability. The structure of the model was originally suggested by Rodriguez-Iturbe et al. [1987] and

---

This dissertation follows the style of *Water Resources Research.*

has been constantly improved with an indication that the models over-simplify the complex nature of physical rainfall processes. For example, Rodriguez-Iturbe [1988] introduced an additional parameter to the Bartlett-Lewis rectangular pulse model [*Rodriguez-Iturbe et al.*, 1987] to better model the randomness of cell duration that varies from one storm to another. The modified model reproduces dry periods in rainfall time series better than the original version. Cowpertwait [1994] developed a model which uses distinct set of parameters modeling cell durations and depths corresponding to different storm types (e.g., frontal and convective, etc.), thus shows an improved performance in reproducing hourly statistics of rainfall time series. Kakou [1998] developed a method in which an explicit dependence between the depth and duration of the rain cells can be considered. Cowpertwait et al. [2007] developed a model in which rain depth varies within a rain cell, thus is capable of capturing the properties of rainfall time series at the time scales as fine as 5 minutes.

Some other studies, instead of modifying the structure of models, suggest that the models need more information about the observed rainfall time series to have better performance. Cowpertwait [1991] derived an equation representing the probability of an arbitrary interval of any chosen length of rainfall time series being dry. Cowpertwait [1998] also derived the equation representing the skewness of rainfall depth distribution, thus outperforms its predecessors in reproducing extreme values.

Even with these efforts to improve the performance of the models, Onof et al. [2000] indicates that the model have rooms for improvement concerning regionalization;

model fitting (parameter calibration); and the reproduction of extreme values and watershed response characteristics.

This dissertation presents the results of the additional endeavors to enhance the applicability and performance of the Poisson cluster stochastic rainfall generators, particularly focusing on the following three topics: regionalization of parameters, parameter calibration method, and further model development.

In Section 2, 6 parameters of the Modified Bartlett-Lewis rectangular pulse model are regionalized across the contiguous United States. Using the parameter surface maps provided by this study, synthetic rainfall time series can be generated at any location of the country without having to calibrate and validate the model each time. Additionally, rainfall characteristics such as average number of rain cells per storm; average rate of storm and cell arrival; average storm duration; and average rain depth per storm were derived from the estimated model parameters and were spatially interpolated to produce maps. These maps can be used to understand regional rainfall characteristics.

Studies on the regionalization of point-process based stochastic rainfall simulation models have been conducted in the past decades [*Econopouly et al.*, 1989; *Hawk and Eagleson*, 1992; *Cowpertwait et al.*, 1996b]. An early attempt at regionalizing the parameters of the Hershenhorn and Woolhiser's [1987] point-process rainfall model was conducted by Econopouly et al. [1989]. They showed that parameter sets could be transferred up to 470 km away within the same climatological regime. Their study used data from five rain gages and its spatial extent included the states of Nebraska, Missouri, Wisconsin, Illinois and Iowa. Hwak and Eagleson [1992] also regionalized the

parameters of the modified Bartlett-Lewis rectangular pulse (MBLRP) simulation model across the country and generated contour maps of each parameter for each month of the year. Their maps were generated based on data from 40 rain gages and four additional gages were used for validation. Cowpertwait et al. [1996b] suggest regression equations that relate the parameters of the Neyman-Scott rectangular pulse (NSRP) model [*Rodriguez-Iturbe et al.*, 1987] to regional properties (e.g., altitude and distance to the coast). Their analysis was performed using rainfall data from 112 sites in the United Kingdom (with an area of 243,820 km^2). The regionalization of the parameters of the MBLRP method presented here has similar goals as previous study [*Hawk and Eagleson*, 1992] had, but it is fundamentally different in the robustness of its results because of the much larger number of rain gages used for map generation and validation, and the approach used in calibration for handling the multi-modality of the MBLRP model structure.

In Section 3, an improved calibration approach for stochastic rainfall generation models is presented. In previous applications available in the literature, these models were calibrated to match long-term statistics of rainfall records without distinguishing them based on their relative importance. In this new approach, the relative importance of each statistics is used to weigh them differently.

The calibration of the Poisson cluster stochastic rainfall generators typically consists of matching the statistics of the observed rainfall time series. However, finding the parameters that accurately reproduce observed rainfall statistics is difficult because of the models' complex mathematical structure [*Chandler*, 1997], and it is not

uncommon to find synthetic rainfall time series with systematic statistical biases (i.e., over- or under-estimation of some statistics). A number of articles discuss and/or suggest solutions to address these biases. Chandler [1997] presented a methodology for parameter estimation based on approximate likelihood functions of collections of sample Fourier coefficients. Likewise, Calenda and Napolitano [1999] proposed an approach based on the scale of fluctuation of the observed rainfall process. Calenda and Napolitano [1999] also observed that the parameters estimated using the method of moments are highly sensitive to the choice of statistics considered in the calibration, and that the method often requires the use of complex heuristic optimization algorithms. Favre et al. [2004] recommend a method for calibrating the Neyman-Scott rectangular pulse model [*Rodriguez-Iturbe et al.*, 1987] – a Poisson cluster stochastic rainfall generator – in which two of the five model parameters are determined by optimization and the remaining three, analytically. Burton et al. [2008] indicate the difficulty in matching the statistics of both time series and suggest a method to exact-fit the mean rainfall. Their method consists of modifying the parameters corresponding to the mean rain cell intensity to account for the difference between the observed and synthetic mean rainfall. While these efforts help improving the calibration of Poisson cluster stochastic rainfall generators from the rainfall statistics perspective, they do not address the validity of the synthetic rainfall time series from the watershed point of view. In this study, the rainfall statistics that affect the most the watershed response, as indicated by its runoff depth and peak flow, have been identified and weighted in the calibration of a stochastic rainfall generation model according to their relative importance. It was observed that the

use of these different weights helped generate rainfall time series that better reproduced the watershed processes.

In Section 4, a novel approach of Poisson cluster rainfall model is introduced to address the problem of the current framework of the Poisson cluster rainfall models that cannot account for the year-to-year variability of the rainfall statistics.

The parameters of the Poisson cluster rainfall models determine the random behavior of storms and rain cells and are calculated such that the differences between the statistics of the observed and the synthetic rainfall time series are minimized [*Rodriguez-Iturbe et al.*, 1987]. The statistics considered in the model calibration include mean, variance, lag-1 autocorrelation coefficient and probability of zero rainfall at various temporal accumulation levels. These rainfall statistics are calculated for each calendar month referring to the entire length of the rainfall time series. Here, it is worth noting that each of these statistics has year-to-year variability. For example, a mean rainfall of the month of January of a given precipitation gage containing the records from 1951 to 2000 is the lump average of the mean rainfall of January, 1951; January, 1952; …; January 2000. This study suggests that this year-to-year variability of the statistics is significant, and that lumping all this information in a single value is a rough simplification that causes problems on Poisson cluster rainfall modeling. These problems include systematic error in reproducing important rainfall and watershed response characteristics. The novel approach presented by this study incorporates this variability by simulating the short-term rainfall statistics (e.g., January, 1951; January, 1952; …; January 2000 in the former example) based on the correlation between the observed

rainfall statistics. Then, the model generates the rainfall time series using the modified Bartlett-Lewis rectangular pulse model (MBLRPM) based on the simulated short-term rainfall statistics. The validation of the model based on 104 gages across the United States showed that the newly suggested approach significantly outperformed the traditional ones in reproducing of the distribution of monthly maximum rainfall depths, monthly peak flows and monthly runoff volumes.

2. REGIONALIZATION OF THE MODIFIED BARTLETT-LEWIS RECTANGULAR PULSE STOCHASTIC RAINFALL MODEL

2.1 Introduction

Precipitation partitioning consists of subdividing rainfall into evapotranspiration, infiltration, storage as soil moisture and surface runoff. Among the numerous factors that affect the partitioning of precipitation, its temporal intermittence plays a significant role [*Marani et al.*, 1997]. *Marani et al.* [1997], for example, discuss the case of hypothetical storm events with different temporal intermittence, yet with the same accumulated depth, that generate runoff depths that differ by an order of magnitude. *Marani and Zanetti* [2007] also indicate that precipitation measurements must have a sub-daily temporal resolution to accurately estimate runoff depths. However, according to the *National Climate Data Center* [2008], there are 25,396 rain gages in the contiguous United States (9,826,630 km^2) but only 25% of them keep sub-daily (i.e., hourly) records; that is, on average, one gage each 1,550 km^2, but it can be even sparser in some areas of the Midwest. This study addresses the regionalization of the parameters of the Modified Bartlett-Lewis Rectangular Pulse (MBLRP) stochastic point process rainfall model [*Rodriguez-Iturbe et al.*, 1987; 1988] to support the generation of synthetic rainfall time series of sub-daily temporal resolution at any point of the contiguous United States.

The MBLRP model has been widely accepted for its firm statistical and mathematical basis [*Rodriguez-Iturbe et al.*, 1988], and can be used to generate synthetic precipitation time series that reproduce observed rainfall statistics at various

accumulation intervals or to disaggregate precipitation time series to a finer temporal resolution [*Koutsoyiannis and Onof*, 2001]. Using the parameter maps developed in the present study, simulation of sub-daily rainfall and disaggregation of daily into sub-daily rainfall at any point of the country is possible without having to calibrate and validate the model each time.

Rodriguez-Iturbe et al. [1987] present the Bartlett-Lewis Rectangular Pulse (BLRP) model and derived equations of selected statistics of rainfall time series synthetically generated with the model. *Rodriguez-Iturbe et al.* [1988] further modified the model by introducing an additional parameter to enhance its ability to capture dry periods, and called it the MBLRP model. Among others, *Islam et al.* [1990], *Bo et al.* [1994], *Onof and Wheater* [1994a] and *Chandler* [1997] discuss the complexity of the model calibration; and *Isham et al.* [1990], *Bo et al.* [1994], *Onof and Wheater* [1994b], *Glasbey et al.* [1995], *Khaliq and Cunnane* [1996], *Onof et al.* [1996a], *Cowpertwait et al.* [1996a], and *Verhoest et al.* [1997] assess the overall correctness of the fundamental assumptions of the model by comparing statistics and extreme values of observed and synthetic precipitation time series.

Studies on the regionalization of point-process based stochastic rainfall simulation models (including the MBLRP model) have been conducted in the past decades. An early attempt at regionalizing the parameters of the *Hershenhorn and Woolhiser* [1987] point-process rainfall model was conducted by *Econopouly et al.* [1989]. They showed that parameter sets could be transferred up to 470 km away within the same climatological regime. Their study used data from five rain gages and its spatial

extent included the states of Illinois, Iowa, Missouri, Nebraska and Wisconsin. *Hawk and Eagleson* [1992] regionalized the parameters of the MBLRP model across the country and generated contour maps of each parameter for each month of the year. Their maps were generated based on data from 40 rain gages and four additional gages were used for validation. *Cowpertwait et al.* [1996b] suggest regression equations that relate the parameters of the Neyman-Scott Rectangular Pulse (NSRP) model [*Rodriguez-Iturbe et al.*, 1987] to regional properties (e.g., altitude and distance to the coast). Their analysis was performed using rainfall data from 112 sites in the United Kingdom (243,820 km^2). Other regionalization studies address rainfall disaggregation models in which coarse temporal precipitation time series are disaggregated into fine temporal precipitation time series. *Gyasi-Agyei* [1999] suggests a method for estimating the parameters of the *Gyasi-Agyei and Willgoose* [1997] disaggregation model for regions where only coarse temporal data are available. His analysis was performed with 13 rain gages over a region of several hundred square kilometers in Australia. Gyasi-*Agyei's* [1999] approach generated rainfall time series that matched well with observed precipitation depth statistics such as mean, variance, probability of zero and auto-correlation coefficient at different accumulation intervals ranging from six minutes to 24 hours. A more recent study on parameter regionalization of rainfall disaggregation is described in *Gyasi-Agyei and Mahbub* [2007]. They implemented the approach of *Gyasi-Agyei* [1999] on the entire Australian continent (7,686,850 km^2) using rainfall observations from 43 six-minute rain gages. As in the previous case, the generated rainfall time series matched well with observed precipitation depth statistics at different accumulation intervals. *Choi*

et al. [2008] applied *Socolofsky et al.'s* [2001] single-parameter rainfall disaggregation method to Texas (670,200 km^2). The method uses hourly precipitation records from one or more stations to simulate hourly precipitation at another station where only daily data are available. They used a total of 532 hourly rain gages that constituted a database of historical precipitation from which rainfall intermittence patterns were derived. The single parameter represented the smallest expected one-hour event, and it was found to have a strong seasonal but a weak regional variability. Resulting simulated precipitation time series matched well with observed precipitation statistics but had the exception of high variances and high lag-one hour autocorrelation coefficients, which were systematically underestimated.

In this study, maps of each of the six MBLRP model parameters for each month of the year – for a total of 72 maps – have been developed for the contiguous United States. These maps illustrate the regional and seasonal variability of the parameters. Additionally, rainfall characteristics such as the average rainfall depth per storm, storm duration, number of rain cells per storm, or rain cell duration were derived from the estimated model parameters and spatially interpolated to produce maps. These maps can be used to understand regional rainfall characteristics. The method for regionalizing the parameters of the MBLRP model presented here has similar goals previous studies had, but it is fundamentally different in the robustness of the results because of the much larger number of rain gages used for map generation and validation, and the approach used in calibration for handling the multi-modality of the MBLRP model structure.

2.2 Modified Bartlett-Lewis Rectangular Pulse (MBLRP) Model

In the MBLRP model, rainfall time series are represented as sequences of storms comprised of rain cells (see Figure 2-1).

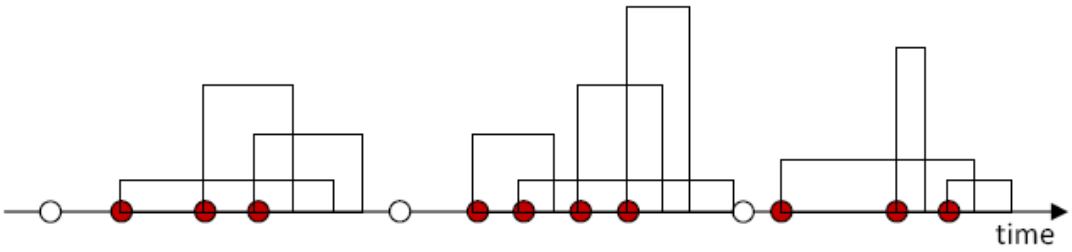


Figure 2-1 Schematic of the MBLRP model. White and gray circles represent the arrival time of storms and rain cells, respectively. Each rain cell is represented by a rectangle whose width and height represent its duration and rainfall intensity.

In the model, X_1 [T] is a random variable that represents the storm arrival time, which is governed by a Poisson process with parameter λ; X_2 [T] is a random variable that represents the duration of storm activity (i.e., the time window after the beginning of the storm within which rain cells can arrive), which varies according to an exponential distribution with parameter γ; X_3 [T] is a random variable that represents the rain cell arrival time within the duration of storm activity, which is governed by a Poisson process with parameter β; X_4 [T] is a random variable that represents the duration of the rain cells, which varies according to an exponential distribution with parameter η that, in turn, is a random variable represented by a gamma distribution with parameters ν and α; and X_5 [L/T] is a random variable that represents the rain cell intensity, which varies according to an exponential distribution with parameter $1/\mu$. From the physical

viewpoint, λ [1/T] is the expected number of storms that arrive in a given period; γ [1/T] is the inverse of the expected duration of storm activity; β [1/T] is the expected number of rain cells that arrive within the duration of storm activity; η [1/T] is the inverse of the expected duration of the rain cells; and μ [L/T] is the average rain cell intensity. Parameters ν [T] and α [dimensionless] do not have a clear physical meaning, but the expected value and variance of η can be expressed as α/ν and α/ν^2. Therefore, the model has six parameters: λ, γ, β, ν, α and μ; however, it is customary to use the dimensionless ratios $\varphi = \gamma/\eta$ and $\kappa = \beta/\eta$ as parameters instead of γ and β.

The estimation of the model parameters is accomplished by matching statistics of the simulated and observed rainfall time series. Some commonly used statistics are the precipitation depth mean, variance, probability of zero rainfall and lag-s covariance at various time scales [*Khaliq and Cunnane*, 1996]. According to *Rodriguez-Iturbe et al.* [1988], the statistics of the synthetically generated rainfall time series at an accumulation interval of T are:

$$E[Y_t^{(t)}] = \lambda \mu \mu_c \frac{\nu}{\alpha - 1} T \tag{2-1}$$

$$\begin{aligned} Var[Y_t^{(T)}] = {} & \frac{2\,\nu^{2-\alpha} T}{\alpha - 2}\left(k_1 - \frac{k_2}{\phi}\right) - \frac{2\,\nu^{3-\alpha}}{(\alpha-2)(\alpha-3)}\left(k_1 - \frac{k_2}{\phi^2}\right) \\ & + \frac{2}{(\alpha-2)(\alpha-3)}\left[k_1 (T+\nu)^{3-\alpha} - \frac{k^2}{\phi^2}(\phi T+\nu)^{3-\alpha}\right] \end{aligned} \tag{2-2}$$

$$\mathrm{Cov}[Y_t^{(T)}, Y_{t+s}^{(T)}] = \frac{k_1}{(\alpha-2)(\alpha-3)}\left\{\begin{array}{l}[T(s-1)+\nu]^{3-\alpha} \\ +[T(s+1)+\nu]^{3-\alpha} - 2(Ts+\nu)^{3-\alpha}\end{array}\right\}$$
$$+\frac{k_2}{\phi^2(\alpha-2)(\alpha-3)}\left\{\begin{array}{l}2(\phi Ts+\nu)^{3-\alpha} \\ -[\phi T(s-1)+\nu]^{3-\alpha} - [\phi T(s+1)+\nu]^{3-\alpha}\end{array}\right\} \quad (2\text{-}3)$$

$$\mathrm{P(zero\ rainfall)} = \exp\left\{-\lambda T - \frac{\lambda\nu}{\phi(\alpha-1)}\left[1+\phi(\kappa+\phi) - \frac{1}{4}\phi(\kappa+\phi)(\kappa+4\phi)\right.\right.$$
$$\left.+\frac{\phi(\kappa+\phi)(4\kappa^2+27\kappa\phi+72\phi^2)}{72}\right] + \frac{\lambda\nu}{(\alpha-1)(\kappa+4\phi)}\left(1-\kappa-\phi+\frac{3}{2}\kappa\phi+\phi^2+\frac{\kappa^2}{2}\right) \quad (2\text{-}4)$$
$$\left.+\frac{\lambda\nu}{(\alpha-1)(\kappa+\phi)}\left[\frac{\nu}{\nu+(\kappa+\phi)T}\right]^{\alpha-1}\frac{\kappa}{\phi}\left(1-\kappa-\phi+\frac{3}{2}\kappa\phi+\phi^2+\frac{\kappa^2}{2}\right)\right\}$$

where

$$k_1 = \left(2\lambda\mu_c\mu^2 + \frac{\lambda\mu_c\kappa\phi\mu^2}{\phi^2-1}\right)\left(\frac{\nu^\alpha}{\alpha-1}\right) \quad (2\text{-}5)$$

$$k_2 = \left(\frac{\lambda\mu_c\kappa\phi\mu^2}{\phi^2-1}\right)\left(\frac{\nu^\alpha}{\alpha-1}\right) \quad (2\text{-}6)$$

$$\mu_c = 1+\frac{\kappa}{\phi} \quad (2\text{-}7)$$

s is the lag time in number of accumulation intervals, and $Y_t^{(T)}$ is the rainfall time series at an accumulation interval T.

2.3 Development of the MBLRP Model Parameter Maps

There are 6,356 hourly rain gages in the contiguous United States, 3,444 of which had a period of record longer than 20 years as of 2006 [*NCDC*, 2008] (see Figure 2-2). These gages were used for model calibration and validation.

Figure 2-2 Location of the 3,444 NCDC rain gages with hourly recording and period of record longer than 20 years.

Model parameters at each rain gage for each month were determined such that the difference between the statistics of the simulated and observed rainfall time series was minimized. Once the parameters were determined at each rain gage, they were spatially interpolated generating maps from which parameter values for any point can be obtained.

2.3.1 Rain Gage Statistics

For each of the 3,444 gages and month of the year (i.e., January, February, March, …), the precipitation depth mean, variance and lag-1 autocorrelation coefficient and the probability of zero rainfall were calculated at 1-, 3-, 12- and 24-hour accumulation levels, for a total of 13 statistics per month (note that the mean is not affected by the accumulation level).

Different accumulation intervals and months were considered in order to capture the sub-daily and seasonal variability of the precipitation. As a reference, maps of the four statistics for an accumulation interval of one hour for the month of May are shown in Figure 2-3. The Ordinary Kriging interpolation technique was used to generate the maps. For all months and accumulation intervals, the mean, variance and probability of zero rainfall showed a strong regional tendency. The regional tendency of the lag-1 autocorrelation, though, was not as strong as that of the other three statistics. Detailed discussion on the weak regional tendency of the lag-1 autocorrelation and its effect on the parameter regionalization is presented later in this article.

2.3.2 Objective Function

Because it is not possible to analytically solve equations (1) through (4) for the MBLRP model parameters of each rain gage and month [*Onof et al.*, 2000], the parameters were obtained by minimizing the following objective function, which represents the discrepancy between the statistics of the observed and simulated rainfall time series:

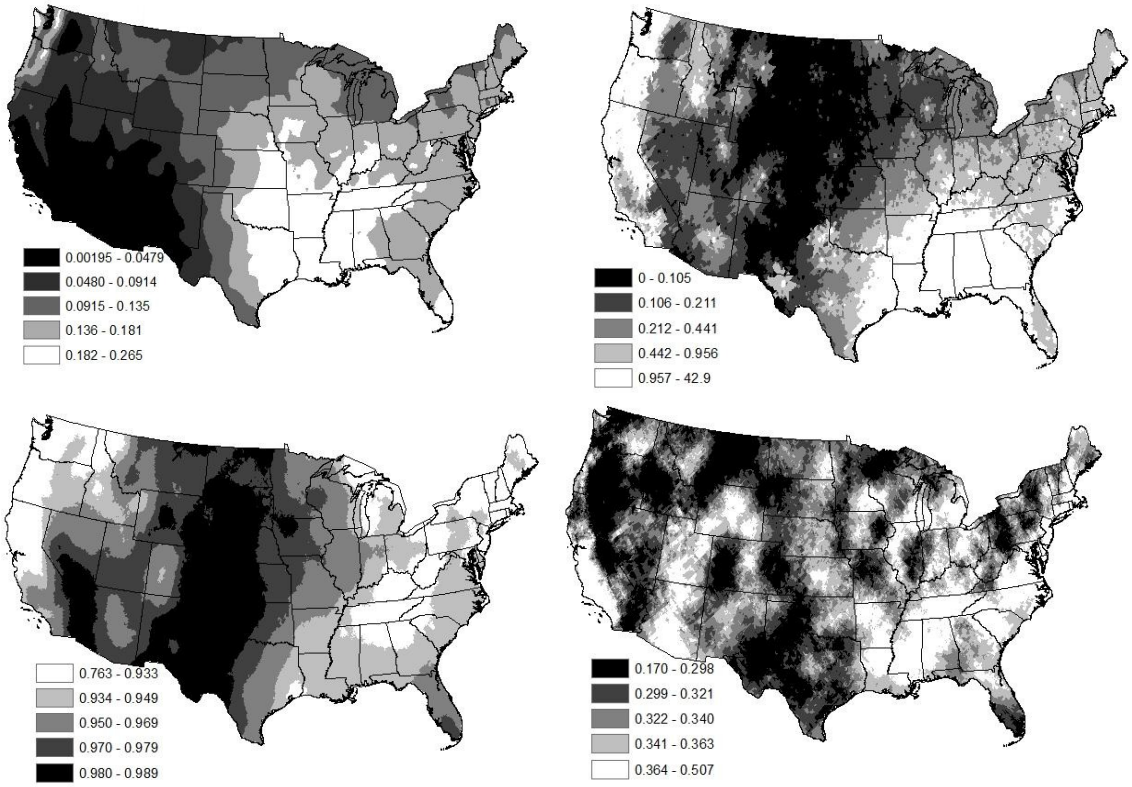


Figure 2-3 Rainfall statistics for one hour accumulation interval for the month of May. From top to bottom and left to right, mean (mm/hr), variance (mm^2/hr^2), probability of zero rainfall and lag-1 autocorrelation coefficient.

$$\sum_{k=1}^{n} w_k \left(1 - \frac{F_k(\vec{\theta})}{f_k} \right) \tag{2-8}$$

where $\vec{\theta}$ is the parameter set (λ, ν, α, μ, φ, κ), n is the number of statistics being matched, $F_k\left(\vec{\theta}\right)$ is the k^{th} statistic of the simulated rainfall time series, f_k is the k^{th} statistics of the observed rainfall time series and w_k is a weight factor given to the k^{th} statistic. The statistics used in the model calibration were the same 13 statistics calculated for the gages. According to *Khaliq and Cunnane* [1996], models calibrated

based on these statistics at various accumulation levels produce rainfall time series that resemble historical observations. Higher-order moments were not included in the objective function despite the fact that matching them would have helped capture better low-frequency events [*Cowpertwait*, 1998]; however, at the expense of losing accuracy for higher-frequency events. All statistics were given the same weight throughout the study despite they might not all have the same relative importance. Addressing this issue was considered beyond the scope of this article, but is currently a matter of research by the authors.

2.3.3 Multi-Modality of the Objective Function

In order to generate rainfall time series at any given point, parameter maps were obtained by interpolation of the parameter values estimated for the rain gages. Obtaining reliable results from this interpolation, however, is not straightforward as explained below. Let rain gages A and B be located close to each other. Despite the gages are at different locations, for the sake of simplicity, it is assumed that both are affected identically by the storm depicted in Figure 2-4.

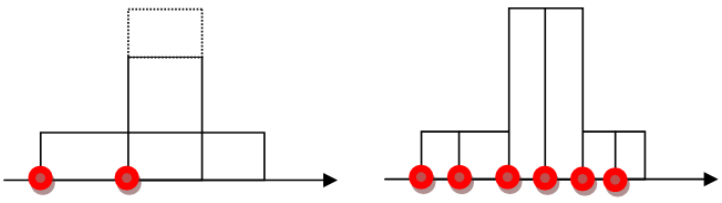


Figure 2-4 Rainfall event modeled with two different parameter sets.

In the figure, the storm in gage A is represented by two rain cells while the one in gage B by six rain cells, where the rain cells in A have longer duration and greater rainfall depth than those in B, but the overall result is the same. That is, a storm, or a precipitation pattern (characterized by its rainfall statistics) if we refer to a longer period, can be represented by different parameter sets. The existence of different and equally-correct solutions for the parameter set is represented as multiple minima of the objective function in the parameter space, and is referred to as *multi-modality* [*Gyasi-Agyei*, 1999; *Onof et al.*, 2000]. Because of this multi-modality, spatially interpolating parameter values between gages A and B might yield incorrect estimates.

Solving this multi-modality problem consisted of identifying as many minima of the objective function at each gage as possible, one of which would be assumed to be the correct parameter vector. The Isolated Speciation-Based Particle Swarm Optimization (ISPSO) method [*Cho et al.*, In Review], which identifies as many minima as possible within a given parameter space, was used for this purpose. A gage in which the global minimum is clearly lower than the other minima is chosen as an anchor gage for estimating the parameter sets in other gages, and the parameter set that generates the global minimum is taken as the correct one at the gage. To select the parameter set at a neighboring gage that best captures the regional tendency, normalized Euclidean distances in the six-dimension parameter space between the parameter vector of the anchor gage and the many parameter vectors of the new gage are calculated. The parameter vector of the new gage that minimizes the normalized Euclidean distance constitutes the best parameter set. Mathematically, let $\overrightarrow{A_1}$ be the already-determined

parameter vector at gage A, and let $\overrightarrow{B_1}, \overrightarrow{B_2}, ...\overrightarrow{B_j}, ...\overrightarrow{B_n}$ be the estimated parameter vectors at neighboring gage B, each of which is associated to a local minimum. The normalized Euclidean distance Δ between parameter sets $\overrightarrow{A_1}$ and $\overrightarrow{B_j}$ is defined as follows:

$$\Delta_j = \sqrt{\sum_{i=1}^{6}\left[\frac{P_i^A - P_{ij}^B}{\max(P_i) - \min(P_i)}\right]^2} \qquad (2\text{-}9)$$

where P_i^A represents the i^{th} parameter at gage A, P_{ij}^B represents i^{th} parameter of the j^{th} parameter vector at gage B, and $\max(P_i)$ and $\min(P_i)$ represent the maximum and minimum values of parameter P_i estimated at all gages, respectively. The parameter vector $\overrightarrow{B_j}$ that minimizes the normalized Euclidean distance is selected as the parameter set at gage B. The procedure continues one gage at a time targeting the gages that are closer to the ones at which parameter sets have already been determined. By doing so, the variability between the parameter values in neighboring gages is significantly reduced and the robustness of the interpolation between them increased.

2.3.4 Cross Validation of the Estimated Parameter Values

Cross-validation was used to determine the relation between the parameter values at each gage and at its surrounding gages. In cross-validation, observed/estimated values at given points in space are compared to the ones obtained by spatial interpolation of their neighboring points. As a reference, Figure 2-5 illustrates the result of cross-validating the six parameters for the month of May. The parameter space for the IPSO

runs was determined based on the results of *Bo et al*. [1994], who estimated the parameters of MBLRP model at two different climate regions located in the United States and Italy, and was redefined iteratively based on cross-validation results. The range of the parameter space that was used in the final iteration of the ISPSO run is given in **Table 2-1**. Plots of other months show similar patterns. Note that, if the estimated parameter values and those calculated by interpolation were equal, the points in the graph would coincide with the 1:1 line. However, the scatter around the 1:1 line is not necessarily an indication of incorrect predictions, but an estimate of the variability of the parameter in the vicinity of the point.

Table 2-1 Parameter range for model calibration

	Parameter Range	
	Minimum	Maximum
λ (1/hr)	0.00001	0.025
ν (hr)	5.5	8.8
α	20	140
μ (mm/hr)	1	40
ϕ	0.01	0.06
κ	0.03	0.44

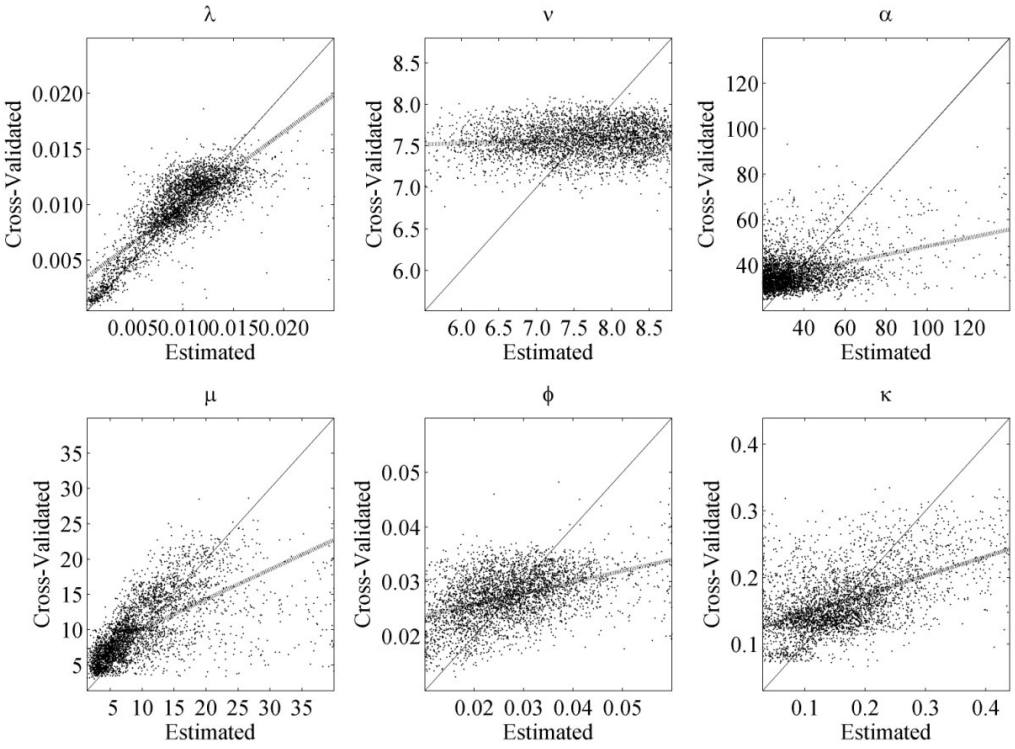


Figure 2-5 Results of cross-validation of the parameters for the month of May. Parameter values estimated by optimization are shown on the x-axis and by cross-validation (interpolation of neighboring gages) on the y-axis. From the top to bottom and left to right, λ (1/hr), ν (hr), α, μ (mm/hr), ϕ and κ.

2.3.5 Spatial Interpolation of the Parameter Values

After the parameter sets were estimated at each of the 3,444 rain gages, the Ordinary Kriging geostatistical interpolation method [*Journel and Huijbregts*, 1978] was applied to obtain surface maps of each parameter and month. A total of 72 monthly parameter maps (i.e., 6 parameters × 12 months) were generated. The variograms used in the interpolation of the model parameters were assumed to have a spherical shape, and their properties were determined such that they minimized the sum of the square of the

residuals between the model and sample variogram. The reader is referred to *Journel and Huijbregts* [1978] for detailed discussion of the Kriging interpolation method.

2.4 Results

As observed by *Isham et al.* [1990], the sensitivity of each rainfall statistic to each parameter is different. Specifically, they indicate that the lag-1 autocorrelation coefficient is strongly sensitive to κ, ν, and α, and insensitive to λ, μ and φ; while the probability of zero rainfall is strongly sensitive to λ, somewhat sensitive to ν, α and φ, and insensitive to μ. In fact, statistics sensitive to parameters with strong regional and/or seasonal patterns also tend to show strong regional and/or seasonal variability, and vice versa.

2.4.1 Regional and Seasonal Patterns of the Model Parameters

In monthly statistic maps (four of which are shown in Figure 2-3), the mean, variance and probability of zero rainfall show clear regional patterns, while the lag-1 autocorrelation coefficient not to the same extent. Consequently, λ is expected to show a strong regional variability, while κ, ν, and α not necessarily.

In Figure 2-5, note a stronger concentration of points around the 1:1 line for λ and μ than for φ, κ, ν, and α. Correlation coefficients between the estimated and cross-validated parameters are 0.81 and 0.63 for λ and μ, respectively, which could be interpreted as an indicator of a strong spatial pattern; while, for φ, κ, ν and α, the correlation coefficients are 0.43, 0.52, 0.11 and 0.40, respectively, indicating a much weaker one. Likewise, t-tests were performed on the slope of the regression lines with

the null hypotheses that they were equal to zero; that is, the null hypotheses implied that the parameter values at a gage are not related to those of their neighbor gages and that no spatial pattern existed. The t-tests resulted in p-values that lead to the rejection of the null hypotheses with a 95% of significance, implying that all slopes were statistically different from zero. In other words, the value of all parameters at a given location can be predicted, to some extent, by the value of the parameters at its neighboring gages.

Figure 2-6 shows maps of interpolated parameter values for the month of May. As observed above, a spatial pattern is clearer for λ and μ than for the other parameters; however, based on the maps, φ, κ, ν and α also show a discernible spatial pattern. Overall, it was observed that those parameters associated to observable physical variables, such as the number of storms or the rainfall intensity, and that are, consequently, less affected by the multi-modality of calibration showed a stronger regional pattern.

Figure 2-7 shows the seasonal variability of the parameters and rainfall statistics of two randomly-selected gages in the states of New York (NCDC 3851 located at 75.5208°W, 43.5753°N) and Washington (NCDC 6858 located at 124.5550°W, 47.9375°N). In the figure, it can be seen that seasonality is not apparent in gage NCDC 3851, while it is reflected in the parameter λ and the statistics of NCDC 6858. Based on the temporal variability of parameter values and statistics in a large number of rain gages, it was concluded that seasonality of the parameters depended on the seasonality of the statistics. In general, λ showed the strongest seasonal variability and ν the weakest.

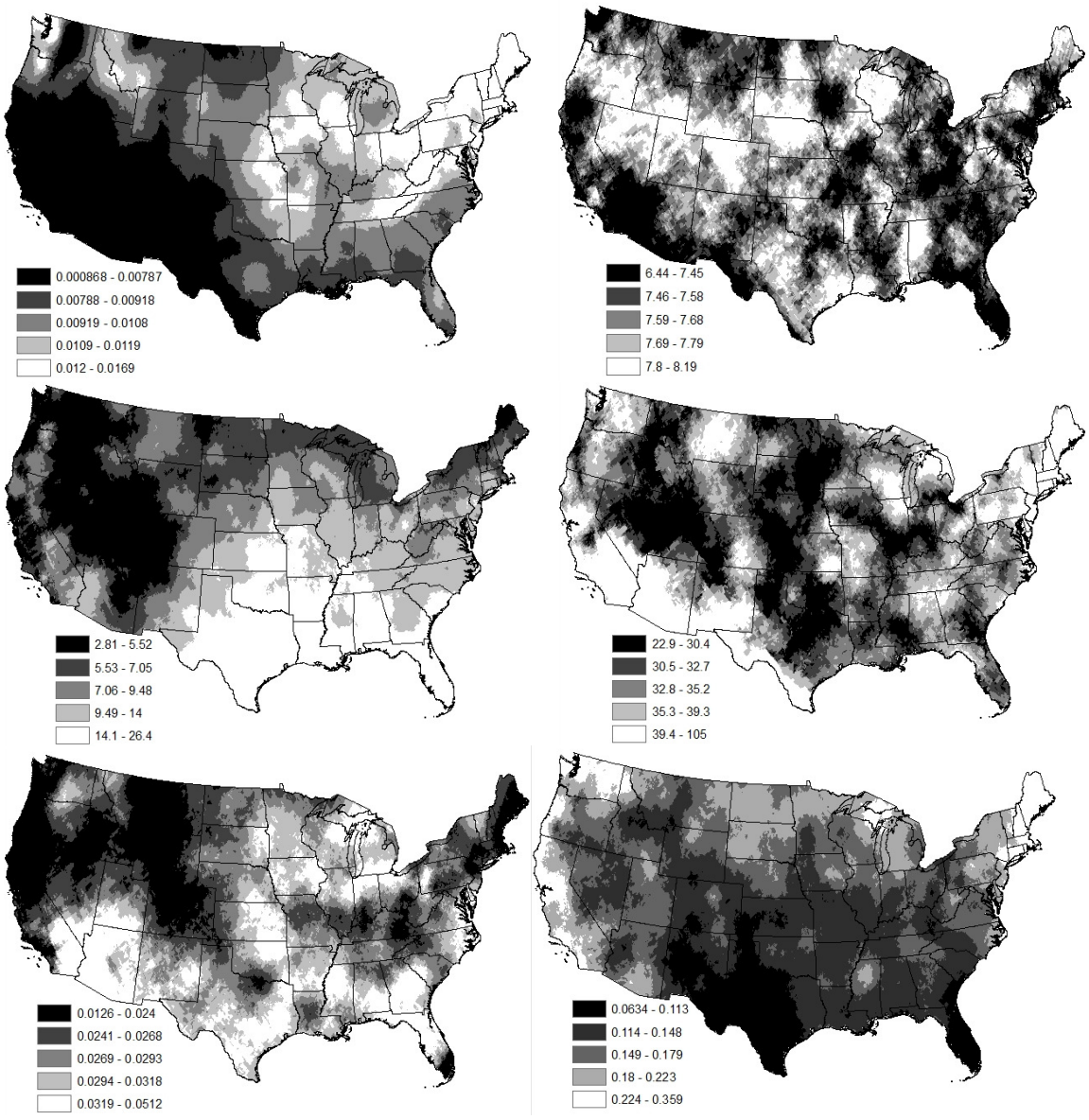


Figure 2-6 MBLRP model parameters for the month of May. From the top to bottom and left to right, λ (1/hr), ν (hr), μ (mm/hr), α, ϕ and κ.

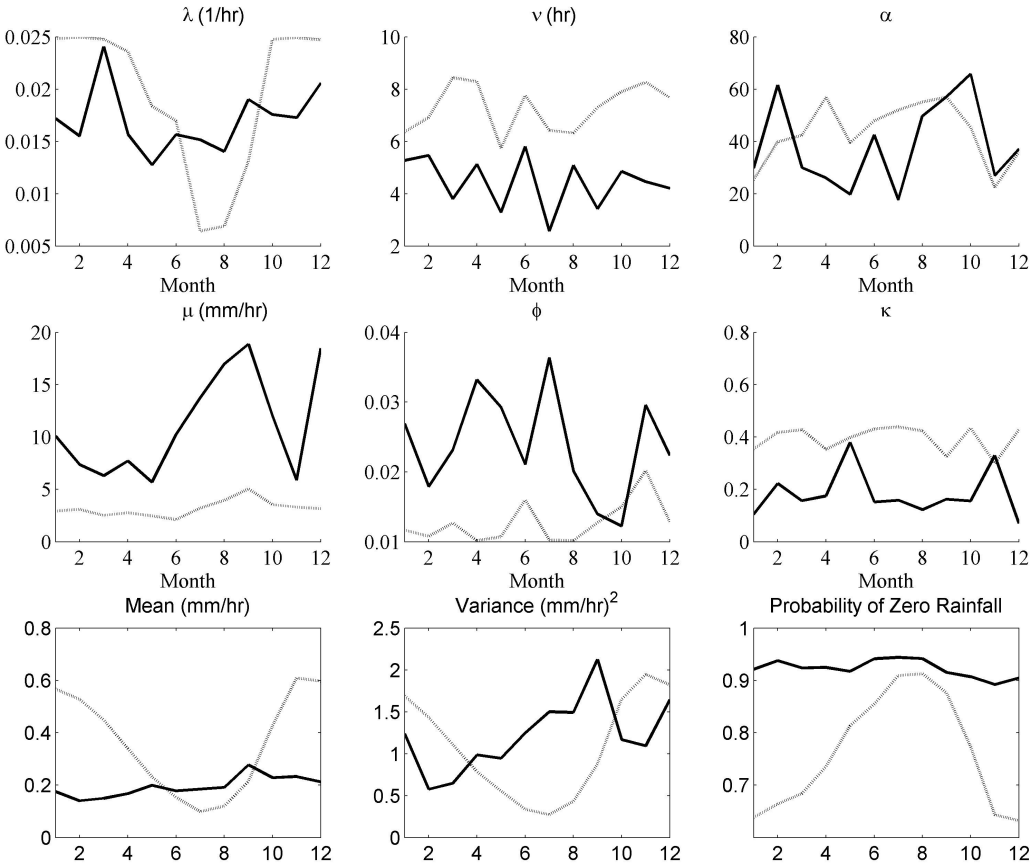


Figure 2-7 Monthly variation of the parameters and statistics of gages NCDC 3851 (black line) located at 75.5208°W, 43.5753°N in the state of New York, and NCDC 6858 (gray line) located at 124.555°W, 47.9375°N in the state of Washington.

2.4.2 Rainfall Characteristics Based on the Model Parameter Values

Rainfall characteristics, such as the average rainfall depth per storm, the average storm duration, the average number of rain cells per storm, the average rain cell arrival rate or the average rain cell duration can be estimated based on the MBLRP parameters. According to *Hawk and Eagleson* [1992], these rainfall characteristics can be expressed as:

Average rainfall depth per storm [L] = $(1 + \kappa / \varphi)\ (\nu / \alpha)\ \mu$ (2-10)

Average storm duration [T] =

$$\frac{1}{\phi\frac{\alpha}{\nu}\left[1+\phi(\kappa+\phi)-\frac{1}{4}\phi(\kappa+\phi)(\kappa+4\phi)+\frac{1}{72}\phi(\kappa+\phi)(4\kappa^2+27\kappa\phi+72\phi^2)\right]} \quad (2\text{-}11)$$

Average number of rain cells per storm = $1 + \kappa / \varphi$ (2-12)

Average rain cell arrival rate [1/T] = $\kappa\ \alpha / \nu$ (2-13)

Average rain cell duration [T] = ν / α (2-14)

These rainfall characteristics were calculated at all gages and interpolated with the Ordinary Kriging technique. Figure 2-8 illustrates these interpolated surfaces for the month of May. Other storm and rain cell characteristics such as the average number of storms arriving during a given period (λ) and average precipitation rate of each rain cell (μ) were already presented in Figure 2-6. As a reference,

Table 2-2 compares the rainfall characteristic for the month of May at four locations of the United States with the same mean monthly rainfall (141 mm). In this case, it was observed that the rainfall depth per storm, which is an observable physical variable and, therefore, less affected by the multi-modality of calibration, showed a stronger regional pattern. The other storm and rain cell characteristics show less well defined spatial patterns, most likely because an error in any of them can be compensated by an error in other.

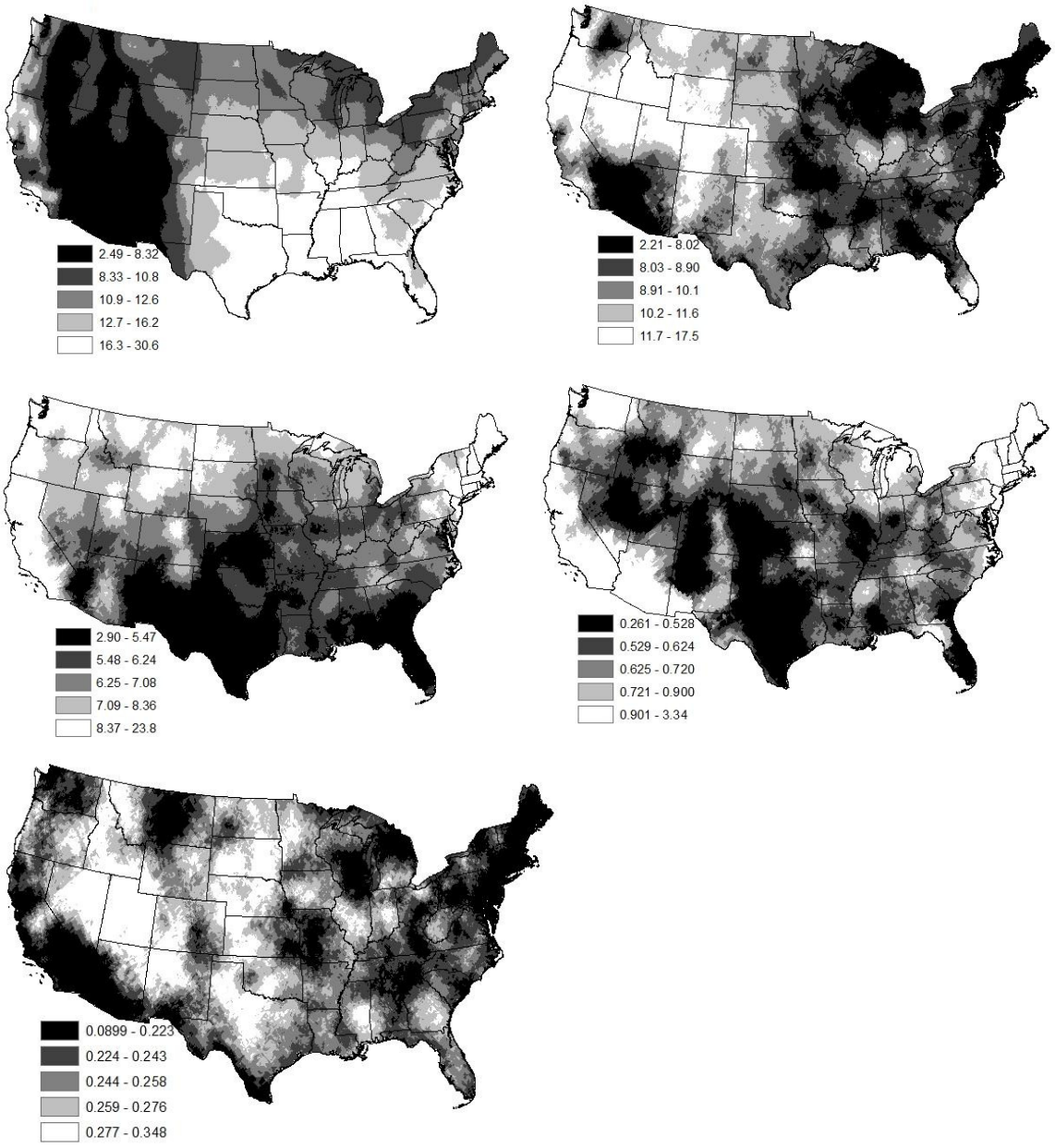


Figure 2-8 Storm and rain cell characteristics for the month of May according to the MBLRP model parameters. From top to bottom and left to right, average rainfall depth per storm (mm), average storm duration (hr), average number of rain cells per storm, average rain cell arrival rate (1/hr) and average rain cell duration (hr).

Table 2-2 Average **r**ainfall characteristics for the month of May for selected locations with mean monthly rainfall depth of 141 mm.

State	Texas	Iowa	Washington	Florida
Longitude	-95.592°	-93.591°	-123.975°	-80.913°
Latitude	30.349°	41.966°	47.862°	27.050°
Rainfall depth per storm (mm)	22.2	13.5	14.2	19.2
Storm duration (hr)	7.9	8.7	12.1	11.7
Number of rain cells per storm	4.7	6.0	16.4	4.0
Rain cell arrival rate (1/hr)	0.57	0.61	1.30	0.38
Rain cell Duration (hr)	0.24	0.28	0.21	0.24

2.4.3 Validation of the Parameter Maps

Validation of the parameter maps consisted of comparing the statistics of the observed rainfall and of the simulated rainfall obtained using the cross-validated parameter sets. Cross-validated parameter values were used because they have the level of accuracy of the parameters estimated by interpolation at points where no gages are located. The statistics used for validation were precipitation depth mean, variance and lag-1 autocorrelation coefficient, and probability of zero rainfall for accumulation intervals of 1, 3, 6, 12 and 24 hours (note that the 6-hour accumulation interval was not used in calibration). This comparison was performed for all 3,444 NCDC hourly precipitation gages used to generate the maps. The results for the month of May are shown in Figure 2-9. In the figure, the two lines on each side of the 1:1 line bind the prediction interval with 95% confidence. The width of the range represents the

uncertainty in the rainfall statistics of the simulated rainfall. Qualitatively speaking, the predictions can be considered satisfactory for the mean, variance and probability of zero rainfall for different levels of accumulation, but not to the same extent for the lag-1 autocorrelation coefficient. However, despite the statistics of the observed rainfall time series were successfully reproduced for the great majority of rain gages, they were not for a number of them. For an accumulation interval of one hour, the 172 gages (i.e., 5% of the gages) with the greatest residuals between the statistics of the observed and synthetic precipitation time series showed the following characteristics: (1) for the mean, the average annual rainfall was 1,650 mm, and the gages were located mostly along the northeast and northwest coast although a number of them were evenly distributed east of the 100°W meridian; (2) for the variance, the average annual rainfall was 1,750 mm, and the gages were located mostly evenly distributed east of the 100°W meridian although a small number along the northwest coast; (3) for the probability of zero rainfall, the average annual rainfall was 1,250 mm, and the gages were located in the northeast and northwest parts of the country; and (4) for the lag-1 autocorrelation coefficient, the average annual rainfall was 750 mm, and the gages were located mostly west of the 100°W meridian with a higher concentration along the coast although a small number of them in the northeast. Note that the average annual precipitation in the country is 1,250 mm. Even though mismatches in all statistics were found, based on the results presented in Figure 2-9, the lag-1 autocorrelation coefficient was of concern. For this statistic, the

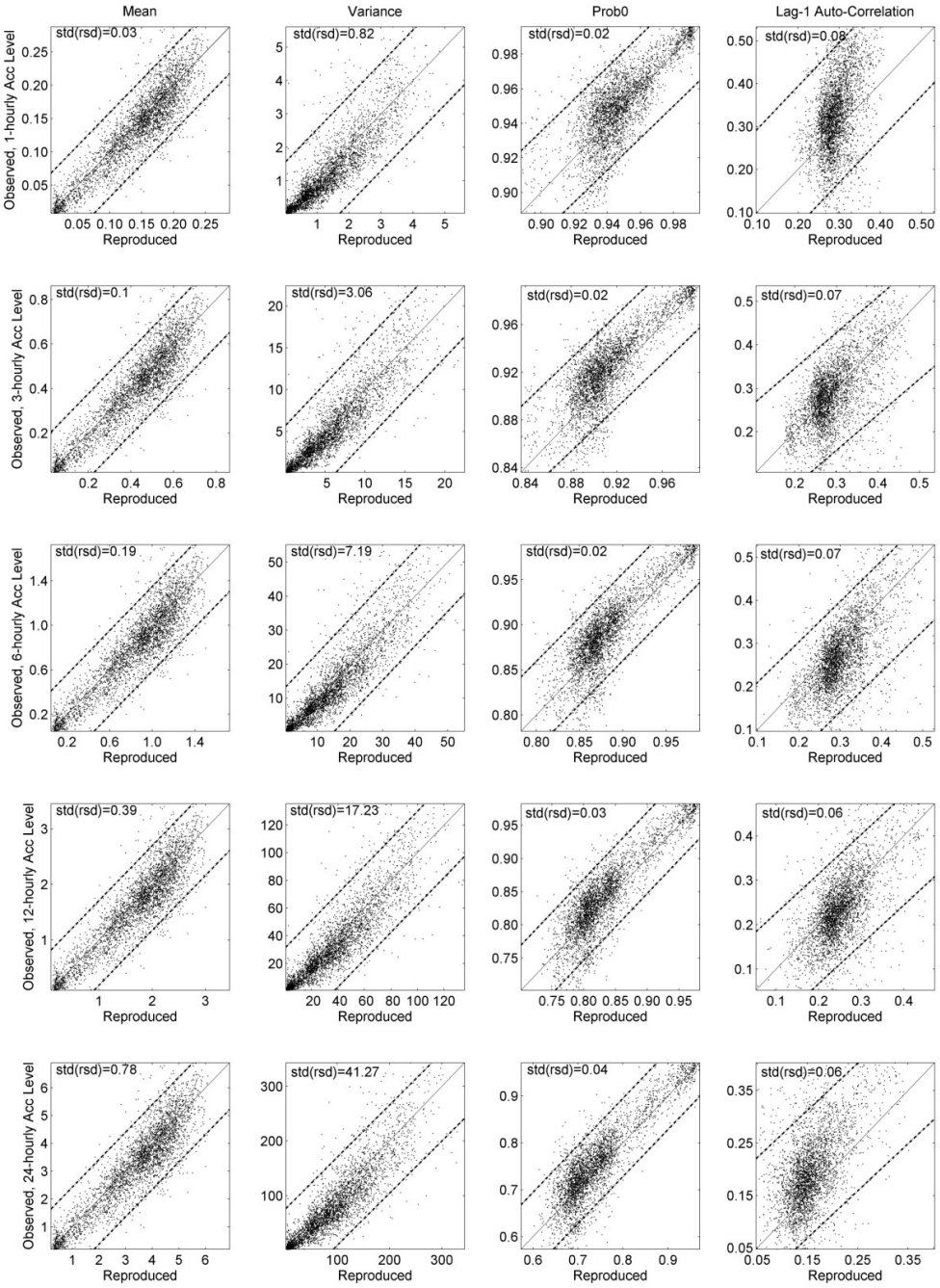


Figure 2-9 Mean (mm/hr), variance (mm^2/hr^2), probability of zero rainfall and lag-1 autocorrelation coefficient of the observed precipitation (y-axis) vs. those calculated from the cross-validated parameters (x-axis) at various accumulation intervals for the month of May. The dotted lines correspond to the prediction interval with 95% confidence.

location of the gages where the mismatch was more pronounced coincided in part with the driest parts of the country. Likewise, the fact that gages in the northwest coast have some of the greatest residuals for all statistics seems to respond to high spatial variability in the area and the limitations of the interpolation technique to capture it with the existing gage network.

2.5 Summary and Conclusions

72 maps of the six parameters of the MBLRP stochastic rainfall generation model for each of the 12 months of the year were generated. These maps allow the implementation of the MBLRP model at any location in the contiguous United States without having to calibrate and validate the model each time. The parameters were estimated at 3,444 NCDC rain gages by matching rainfall statistics (i.e., mean, variance, probability of zero rainfall and lag-1 autocorrelation coefficient) at different rainfall accumulation levels (i.e., 1, 3, 12 and 24 hours). The parameters obtained at the gages were then interpolated using the Ordinary Kriging technique to generate surface maps. Parameters λ and μ were identified to have clear regional patterns, while φ, κ, α had less clear yet discernible spatial tendencies. On the contrary, parameter ν did not show any clear pattern over space because the lag-1 autocorrelation coefficient, to which it is highly sensitive, does not have clear regional tendencies either. Seasonality of the parameters appeared to depend on location, and some parameters showed strong time dependence in some regions and no dependence at all in others.

The parameter maps were cross-validated to assess the correctness of the Ordinary Kriging interpolation. It was observed that the statistics of synthetic rainfall

time series generated with parameters obtained from the maps presented here matched well the means, variances and probabilities of zero rainfall (except for a small number of low probabilities of zero rainfall) of observed rainfall time series, but not the lag-1 autocorrelation coefficients (whose higher values were underestimated and lower values, overestimated). A study of the relevance of matching the lag-1 autocorrelation coefficient, when using synthetic rainfall in complex hydrologic applications, is beyond the scope of this article but a matter currently being researched by the authors. Even though mismatches in all statistics were found, according to the value of the standard deviation of the residuals, only the lag-1 autocorrelation coefficient was of concern. For this statistic, the location of the gages where the mismatch was more pronounced coincided (in part) with the driest parts of the country. Likewise, the fact that gages in the northwest coast have some of the greatest residuals for all statistics seems to respond to high spatial variability in the area and the limitations of the interpolation technique to capture it with the existing gage network.

Overall, it was observed that model parameters and storm and rain cell characteristics associated to observable physical variables and that are, consequently, less affected by the multi-modality of calibration showed stronger regional patterns. Other model parameters and storm and rain cell characteristics show less well defined spatial patterns, most likely because the error in any of them can be compensated by the error in other.

3. ON THE RELATIVE IMPORTANCE OF THE DIFFERENT RAINFALL STATISTICS IN THE CALIBRATION OF STOCHASTIC RAINFALL GENERATION MODELS

3.1 Introduction

Stochastic rainfall generators are widely used in hydrologic analysis because they can provide precipitation input to models whenever data are not available. Not being able to reproduce actual rainfall records, stochastic models aim at generating synthetic precipitation time series whose statistics match those of observed ones [*Rodriguez-Iturbe et al.*, 1987]. That is, the model calibration consists of finding the parameter values that minimize the discrepancy between the statistics of the observed and synthetic time series. Typically, these statistics are the mean, variance, probability of zero rainfall, covariance and extreme values [*Rodriguez-Iturbe et al.*, 1987], but can change depending on the needs of the specific study. For this reason, the performance of stochastic rainfall generators is often measured by comparing the statistics of the observed and synthetic precipitation time series [*Rodriguez-Iturbe et al.*, 1987; 1988; *Isham et al.*, 1990; *Onof and Wheater*, 1994a; *Cowpertwait*, 1995; *Cowpertwait et al.*, 2007, among others]. However, while a measure of the discrepancy between the statistics gives an estimate of the model's performance, it is a rather indirect metric from a watershed-hydrology perspective. In fact, from a watershed viewpoint, realistic synthetic rainfall should reproduce well the catchment response, including runoff depths and flows at the outlet. In this study, the sensitivity of the watershed long-term average

monthly runoff depth and long-term average monthly peak flow – *runoff depth* and *peak flow* in the following – to each rainfall statistic is analyzed, with the understanding that the statistics that most affect the catchment response should be given a greater weight in the calibration of rainfall generation models.

Poisson cluster stochastic rainfall generators [*Rodriguez-Iturbe et al.*, 1987; 1988; *Cowpertwait*, 1995; *Cowpertwait and O'Connell*, 1997; *Cowpertwait et al.*, 2007] are regarded as a robust and practical approach to simulate continuous rainfall time-series [*Olsson and Burlando*, 2002], and their calibration has traditionally consisted of matching the statistics of the observed rainfall time series. However, finding the parameters that accurately reproduce observed rainfall statistics is difficult because of the models' complex mathematical structure [*Chandler*, 1997], and it is not uncommon to find synthetic rainfall time series with systematic statistical biases (i.e., over- or under-estimation of some statistics). A number of articles discuss and/or suggest solutions to address these biases. Chandler [1997] presents a methodology for parameter estimation based on approximate likelihood functions of collections of sample Fourier coefficients. Likewise, Calenda and Napolitano [1999] propose an approach based on the scale of fluctuation of the observed rainfall process. Calenda and Napolitano [1999] also observed that the parameters estimated using the method of moments are highly sensitive to the choice of statistics considered in the calibration, and that the method often requires the use of complex heuristic optimization algorithms. Favre et al. [2004] recommend a method for calibrating the Neyman-Scott rectangular pulse model [*Rodriguez-Iturbe et al.*, 1987] – a Poisson cluster stochastic rainfall generator – in

which two of the five model parameters are determined by optimization and the remaining three, analytically. Burton et al. [2008] indicate the difficulty in matching the statistics of both time series and suggest a method to exact-fit the mean rainfall. Their method consists of modifying the parameters corresponding to the mean rain cell intensity to account for the difference between the observed and synthetic mean rainfall. While these efforts help improve the calibration of Poisson cluster stochastic rainfall generators from the rainfall statistics perspective, they do not address the validity of the synthetic rainfall time series from the watershed point of view. Other studies have addressed the relationship between the rainfall characteristics and the catchment response. Marani et al. [1997], for example, analyzed the sensitivity of the hydrologic partitioning to the temporal resolution and intermittency of rainfall; Gabellani et al. [2007] investigated how the spectral slope of the rainfall time series affects the peak discharge of the watershed hydrograph; and Van Werkhoven et al. [2008] analyzed the sensitivity of streamflow to the spatio-temporal structure of precipitation events of equal accumulation depth.

In this study, the rainfall statistics that affect the most the watershed response, as indicated by its runoff depth and peak flow, have been identified and weighted in the calibration of a stochastic rainfall generation model according to their relative importance. It was observed that the use of these different weights helped generate rainfall time series that better reproduced the watershed processes.

3.2 Effect of the Rainfall Statistics on the Watershed Response

A multiple linear regression analysis was used to quantify the sensitivity of the watershed (long-term average monthly) runoff depth R [L] and peak flow Q_p [L^3/T] to the rainfall statistics. In this multiple linear regression analysis, the predictor (or independent) variables were the rainfall statistics, and the response (or dependent) variables, the runoff depth and peak flow. Because, in the absence of multi-collinearity among the predictors, each regression coefficient reflects the rate of change of the response caused by a unit change in its corresponding predictor, they can be interpreted as a measure of the sensitivity of the catchment response to the different rainfall characteristics.

1,249 NCDC [*NCDC*, 2008] precipitation gages across the coterminous United States, with period of record of 50 years or longer, were considered in the study (Figure 3-1). 1,099 of these gages were chosen at random for the regression analysis and the remaining 150, for the validation of the approach. The data at each station was subdivided per month to capture the seasonality of the precipitation processes. That is, for each gage and month, a time series was created by piecing together the values corresponding to the same month over the years (e.g., Jan 1, 1950; … Jan 31, 1950; Jan 1, 1951; … Jan 31, 1951; Jan 1, 1952; … Jan 31, 1952; and so on). As a result, each of the 1,249 gages had 12 merged rainfall time series, for a total of 14,988 merged time series (i.e., 1,249 gages × 12 months). Specifically, there were 13,188 time series for the regression analysis and 1,800 for the validation of the approach. Next, for each of the

merged time series, the mean, variance, skewness, probability of zero rainfall and lag-1 auto-correlation coefficient were calculated at an hourly accumulation level.

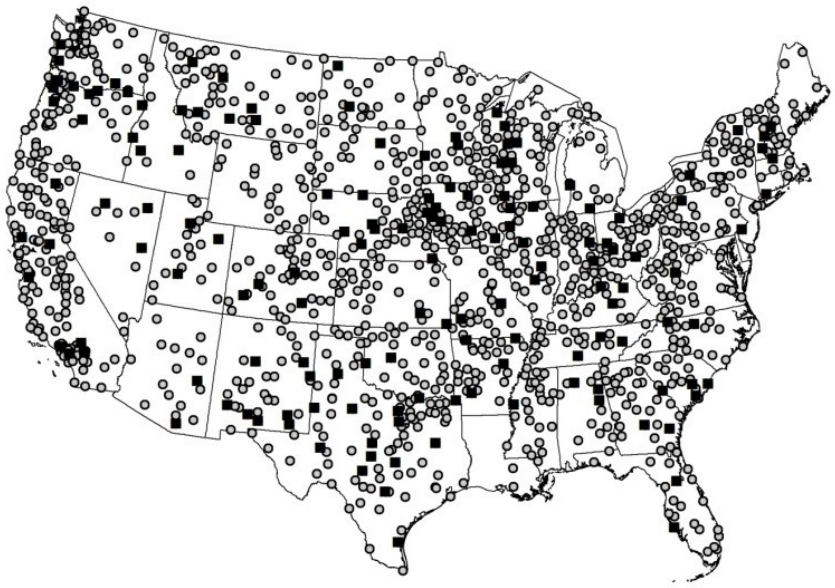


Figure 3-1 1,099 NCDC precipitation gages were used in the regression analysis (gray circles), and 150 in the approach validation (black circles).

Runoff depths and peak flows were estimated for a 7.50-km^2 virtual watershed with lag time of 120 minutes and a range of average-condition curve numbers varying from 50 to 90. A somewhat small size watershed was adopted to minimize the effect of the spatial variability of the precipitation and land use/land cover, which were assumed to be uniform. The Soil Conservation Service (SCS) curve number method [*SCS*, 1972] was used to calculate runoff depth time series for each gage and month. Even though there are well known physically based models suitable for continuous hydrologic simulations, such as the numerical solution of Richard's equation [*Richards*, 1931] or

the Green-Ampt method [*Green and Ampt*, 1911], the curve number method was used mostly because of its simplicity and because performance comparisons have not clearly proven superiority of the other models, especially when exact hydrologic parameters are not available [*Vanmullem*, 1991; *James et al.*, 1992; *Nearing et al.*, 1996; *Ponce and Hawkins*, 1996; *Michel et al.*, 2005]. The antecedent soil moisture condition (AMC), which plays an important role in runoff generation, was taken into account by modifying the curve number values according to the precipitation depth in the previous five days. That is, if the conditions are considered dry based on the precipitation depth in the previous five days, then $CN = 4.2\,CN_0/(10 - 0.058\,CN_0)$; if, on the contrary, the conditions are considered wet, then $CN = 23\,CN_0/(10 + 0.13\,CN_0)$; otherwise, $CN = CN_0$; where CN is the actual curve number and CN_0 is the average-condition curve number [*SCS*, 1972]. Likewise, a one-hour SCS unit hydrograph [*SCS*, 1972] was used to estimate the hydrograph at the watershed outlet generated by the runoff time series. The (long-term average monthly) runoff depth R was calculated by dividing the runoff volume over the entire period of simulation by the duration of the period; and the (long-term average monthly) peak flow Q_p was calculated as the mean of the maximum flows in each month-year. Although different watershed sizes were not considered in the analysis, the approach presented here would apply as long as the watershed is small enough to qualify as uniform and to have a lag-time short enough not to filter out the rainfall variability in the routing process.

Table 3-1 Mean and standard deviation of the rainfall statistics used to calculate z-scores

	MEAN (cm/hr)	VAR $(cm/hr)^2$	SKEW	PROB0	AC
Mean	0.0117	0.0106	25.4	0.951	0.330
Standard Deviation	0.0074	0.0300	18.5	0.032	0.097

In order to make the rainfall statistics comparable in magnitude, they were transformed into z-scores (i.e., value minus the mean divided by the standard deviation) before using them in the regression analysis. Table 3-1 shows the mean and standard deviation of each statistic used in these transformations. Data records with z-scores greater than 10 were assumed to be outliers and were excluded from the analysis. The threshold value of 10 is somewhat arbitrary, but was implemented consistently throughout the study. Similarly, runoff depth R and peak flow Q_p were transformed into residuals (i.e., value minus the mean) before being used in the regression analysis.

Table 3-2 shows the mean values used in these transformations. It is noteworthy that there is a large discrepancy between the mean rainfall (Table 3-1) and mean runoff depth (Table 3-2). In fact, it was observed that, on average, only 20% of the rainfall becomes runoff even when the curve number is as high as 90. This low fraction of precipitation that becomes runoff is explained by the large number of short duration rainfall events (i.e., less than 3 hours) that follow dry periods (in the order of days), which causes significant abstractions.

Table 3-2 Mean of the long-term average monthly runoff depths and peak flows used to calculate residuals.

Curve Number	50	60	70	80	90
Runoff Depth (cm)	0.15	0.29	0.53	0.99	1.76
Peak flow (m^3/sec)	0.39	0.72	1.23	2.04	3.09

The following two multiple linear regression equations were used to quantify the sensitivity of the watershed response to the rainfall statistics:

$$\Delta R = A_1\ \mathrm{MEAN} + A_2\ \mathrm{VAR} + A_3\ \mathrm{SKEW} + A_4\ \mathrm{PROB0} + A_5\ \mathrm{AC} \qquad (3\text{-}1)$$

$$\Delta Q_p = B_1\ \mathrm{MEAN} + B_2\ \mathrm{VAR} + B_3\ \mathrm{SKEW} + B_4\ \mathrm{PROB0} + B_5\ \mathrm{AC} \qquad (3\text{-}2)$$

where ΔR [L] is the runoff depth residual and ΔQ_p [L^3/T] is the peak flow residual; MEAN, VAR, SKEW, PROB0 and AC represent the dimensionless z-scores of the mean, variance, skewness, probability of zero rainfall and lag-1 autocorrelation coefficient of the rainfall time series at an hourly accumulation level, respectively; and $A_{1,\ldots,}$ A_5, $B_{1,\ldots}$ and B_5 are the regression coefficients. Because all predictor and response variables were transformed such that their means were zero, the constant terms of the regression equations were also zero.

The pairwise scatter plots of the z-scores of the statistics are shown in Figure 3-2,

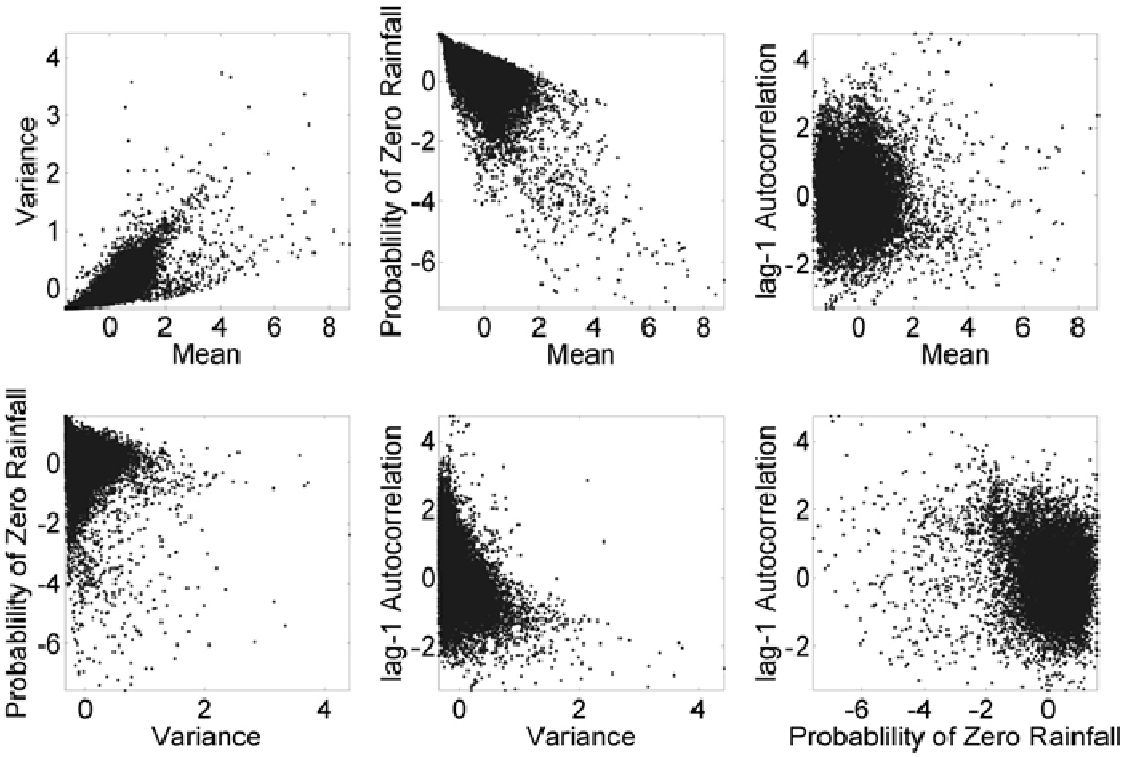


Figure 3-2 Pairwise scatter plots of the z-scores of the rainfall statistics.

where it can be seen that there is a level of correlation between the mean and the variance, and between the mean and the probability of zero rainfall. In case of multi-collinearity among the predictors, the interpretation that the regression coefficients are a measure of the sensitivity of the response to the predictors is not straightforward, because some predictors contain information of the other predictors with which they are correlated. However, even when multi-collinearity exists, the measures of sensitivity still apply provided all the predictors considered in the regression equations are included when using the results. The Principal Component Analysis (PCA) – first envisioned by

Pearson [1901] – was conducted to determine whether the multi-collinearity between these statistics needed to be addressed. According to the PCA, condition numbers of the correlation matrix less than 15 and sum of the reciprocals of the eigenvalues less than five times the number of predictors correspond to cases in which the multi-collinearity is not strong enough and can be ignored [*Chatterjee et al.*, 2006]. In our case, however, the PCA resulted in a condition number of the correlation matrix of 8.4 and a sum of the reciprocals of the eigenvalues of 45, which indicated that the multi-collinearity was strong. The reader is referred to Chatterjee et al. [2006] for a detailed description of the PCA.

Because of the multi-collinearity between the rainfall statistics, ridge regression [*Hoerl and Kennard*, 1970] was implemented to develop a model with lower uncertainty in the regression coefficients than what would have been obtained with standard multiple regression. In ridge regression, rather than calculating the regression coefficients by minimizing the sum of the square of the residuals (SSR), they are obtained by minimizing the sum of the SSR plus a regularization term. That is, by minimizing

$$\| \mathbf{A}\,\mathbf{x} - \mathbf{b} \|^2 + \| \boldsymbol{\Gamma}\,\mathbf{x} \|^2 \qquad (3\text{-}3)$$

where $\mathbf{A}$ is an m × n array in which each of the m rows stores the z-scores of the n rainfall statistics of each gage, $\mathbf{x}$ is an n × 1 column array that stores the regression coefficients of the n z-scores of the rainfall statistics, $\mathbf{b}$ is an m × 1 column array that stores the residuals of the runoff depth or peak flow at the m gages, and $\boldsymbol{\Gamma}$ is an n × n array used to reduce the singularity of matrix $\mathbf{A}$. $\boldsymbol{\Gamma}$ is often the identity matrix but can also be a different operator depending on the specific problem being studied. In this

case, $\boldsymbol{\Gamma}$ was taken as the identity matrix. Given that m = 1,099 (the number of gages for calibration) and n = 5 (the number of statistics), it follows

$$\mathbf{A} = \begin{bmatrix} MEAN_1 & VAR_1 & SKEW_1 & PROB0_1 & AC_1 \\ \dots & \dots & \dots & \dots & \dots \\ MEAN_i & VAR_i & SKEW_i & PROB0_i & AC_i \\ \dots & \dots & \dots & \dots & \dots \\ MEAN_{13188} & VAR_{13188} & SKEW_{13188} & PROB0_{13188} & AC_{13188} \end{bmatrix} \tag{3-4}$$

$$\mathbf{x} = \begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \end{bmatrix} \text{ or } \begin{bmatrix} B_1 \\ B_2 \\ B_3 \\ B_4 \\ B_5 \end{bmatrix} \tag{3-5}$$

$$\mathbf{b} = \begin{bmatrix} \Delta R_1 \\ \dots \\ \Delta R_i \\ \dots \\ \Delta R_{13188} \end{bmatrix} \text{ or } \begin{bmatrix} \Delta Q_{p1} \\ \dots \\ \Delta Q_{pi} \\ \dots \\ \Delta Q_{p13188} \end{bmatrix} \tag{3-6}$$

where the first array in equation (3-5) corresponds to the first array in equation (3-6), and the second in (3-5) to the second in (3-6). It is important to stress that, although ridge regression helps calculate regression coefficients with a lower level of uncertainty, it does not eliminate the multi-collinearity among the independent variables.

The regression coefficients and the adjusted R^2 values are shown in Table 3-3 and Table 3-4. It can be seen that the adjusted R^2 values ranged between 0.59 and 0.87 for the runoff depth and 0.77 and 0.93 for the peak flow indicating that the rainfall statistics explain a significant portion of the catchment response variability. Note that the adjusted R^2 decreases as the curve number decreases (with the exception of the curve

number equal to 90), which is caused by the fact that the runoff generation process is more complex for watersheds with low curve number than with high curve numbers; that is, more dependent on antecedent moisture condition and less on rainfall variability. The fact that the adjusted R^2 values are higher for the peak flow than for the runoff depth is somewhat counterintuitive because one would have expected the runoff depth, which results from all rainfall events and not only from the greatest annual events, to be better represented by the rainfall statistics than the peak flow.

Table 3-3 Regression coefficients of the regression analysis for runoff depth R

Curve Number	A_1	A_2	A_3	A_4	A_5	Adjusted R^2
50	0.201	0.096	0.080	0.029	0.090	0.591
60	0.376	0.154	0.114	0.070	0.132	0.685
70	0.552	0.276	0.126	0.112	0.168	0.786
80	0.911	0.395	0.131	0.217	0.199	0.866
90	1.230	0.554	0.046	0.251	0.168	0.829

Table 3-4 Regression coefficients of the regression analysis for peak flow Q_p

Curve Number	B_1	B_2	B_3	B_4	B_5	Adjusted R^2
50	0.272	0.378	0.091	0.087	0.161	0.766
60	0.503	0.586	0.095	0.187	0.208	0.851
70	0.766	0.845	0.067	0.310	0.238	0.903
80	1.190	1.090	0.020	0.490	0.246	0.929
90	1.370	1.370	-0.151	0.522	0.145	0.844

The sensitivity of R or Q_p to each statistic was estimated as the absolute value of the coefficient of the statistic divided by the sum of the absolute values of the coefficients. For example, the sensitivity of R to the mean was calculated as $|A_1|/(|A_1|+|A_2|+|A_3|+|A_4|+|A_5|)$ and the sensitivity of Q_p to the skewness as $|B_3|/(|B_1|+|B_2|+|B_3|+|B_4|+|B_5|)$. These sensitivity values, which are also a measure of the relative importance of the statistics, were then used to weight the rainfall statistics differently when calibrating rainfall generation models.

Figure 3-3 shows the sensitivity of the runoff depth to the different statistics. Based on the plot, the mean rainfall depth has a 0.40 to 0.55 relative importance and the variance 0.20 to 0.25. These results indicate that the total amount and variability of the

rainfall depth play a very important role in the runoff depth generation. It was also observed that the rainfall intermittency, as indicated by the sum of the relative importance of the probability of zero rainfall and autocorrelation coefficient, varied in a narrow range around 0.20 for the different curve numbers. The positive value of the regression coefficients of the probability of zero rainfall and autocorrelation coefficient confirms that the temporal concentration of rainfall causes greater runoff depths. The effect of extreme events, as reflected by the skewness, ranged between 0.02 and 0.16 with higher values for lower curve numbers. That is, in watersheds with low curve numbers, the skewness is of a greater importance because extreme events generate most of the runoff, while with high curve numbers extreme values are just some of the many events that contribute to the total runoff.

Similarly, Figure 3-4 shows the sensitivity of the peak flows to the different statistics. More than in the previous case, the mean and variance govern the variability of the peak flow, now with a relative importance of 0.28 to 0.39 for the mean and 0.36 to 0.39 for the variance. These high values indicate that the mean and variance of the rainfall time series capture well the rainfall events that cause the peak flows. Likewise, the rainfall intermittency, as indicated by the sum of the relative importance of the probability of zero rainfall and autocorrelation coefficient, varied between 0.20 and 0.25. Again, in this case, the positive value of the regression coefficients of the probability of zero rainfall and autocorrelation coefficient confirms that the temporal concentration of rainfall causes greater peak flows.

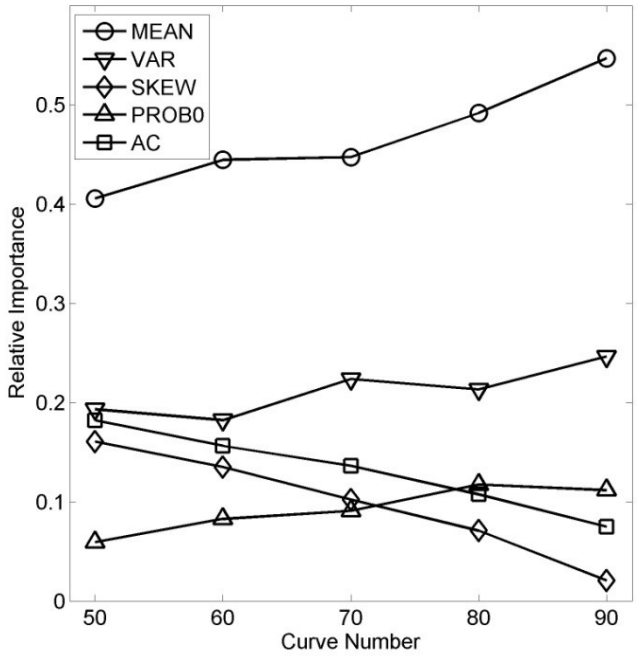


Figure 3-3 Relative importance of the different rainfall statistics on the runoff depth.

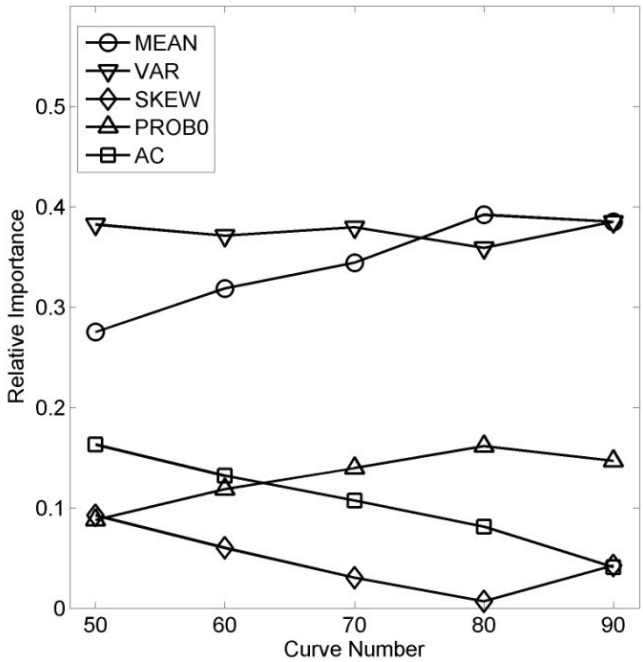


Figure 3-4 Relative importance of the different rainfall statistics on the peak flow.

The effect of extreme events, as reflected by the skewness, ranged between 0.04 and 0.09 with higher values for lower curve numbers for the same reasons mentioned above for runoff depth. The negative regression coefficient for the skewness and curve number of 90 has been considered a mathematical artifact that does not reflect any physical process. Again, the greater skewness for lower curve number values indicates the greater importance of extreme events in watersheds with greater abstractions.

3.3 Effect of the Statistic Weights on Synthetic Rainfall Time Series

The modified Bartlett-Lewis Rectangular Pulse model (MBLRP) [*Rodriguez-Iturbe et al.*, 1988] – also a Poisson cluster stochastic rainfall generator – was used to test our approach. The MBLRP model characterizes rainfall time series under the assumption that rain storms arrive in time according to a given Poisson process, and that each rain storm has a random duration within which a cluster of rain cells, with random depth and duration, arrive in time according to another Poisson process. The parameters of MBLRP models define the probability distributions that govern the random nature of rainfall and are typically determined by minimizing the following objective function:

$$\sum_{k=1}^{n} w_k \left[1 - \frac{F_k(\vec{\theta})}{f_k} \right]^2 \qquad (3\text{-}7)$$

where $\vec{\theta}$ is a vector that contains the model parameters, n is the number of rainfall statistics being matched (i.e., five in this case), F_k is the k^{th} statistic of the synthetic rainfall time series, f_k is the k^{th} statistic of the observed rainfall time series, and w_k is a weight factor of the k^{th} statistic.

The validation of the approach consisted of comparing runoff depth and peak flow calculated using synthetic rainfall time series to those calculated using observed precipitation at the 150 gages selected for this purpose. Ideally, the runoff depths and peak flows obtained with synthetic time series should be equal to those obtained with observed precipitation but, in reality, that might not be the case. However, it is expected that, if the statistics are weighed proportionally to their relative importance for estimating the runoff depth or peak flow, the resulting rainfall time series will reproduce better the catchment response than one based on equally weighed statistics.

Therefore, at each of the 150 gages and for each of the 12 months of the year, three 50-year synthetic rainfall time series were generated using three different parameter sets. These parameter sets were obtained by calibrating the MBLRP model with different weight factors in equation (3-7); that is, (1) weights w_k equal to one; (2) weights w_k taken from Figure 3-3 (which optimize runoff depths R); and (3) weights w_k taken from Figure 3-4 (which optimize peak flows Q_p). In the following, "Figure 3 (which optimize runoff depths R)" will be referred to as "(R)", and "Figure 3-4 (which optimize peak flows Q_p)" as "Figure 3-4 (Q_p)".

Figure 3-5 presents scatter plots of the statistics of the synthetic rainfall time series obtained with the weight factors of Figure 3-3 (R) and observed precipitation. Each of the four plots of Figure 5 present 1,800 points (i.e., 150 gages × 12 months). It was observed that the mean was underestimated on average by 2%, the variance by 1%

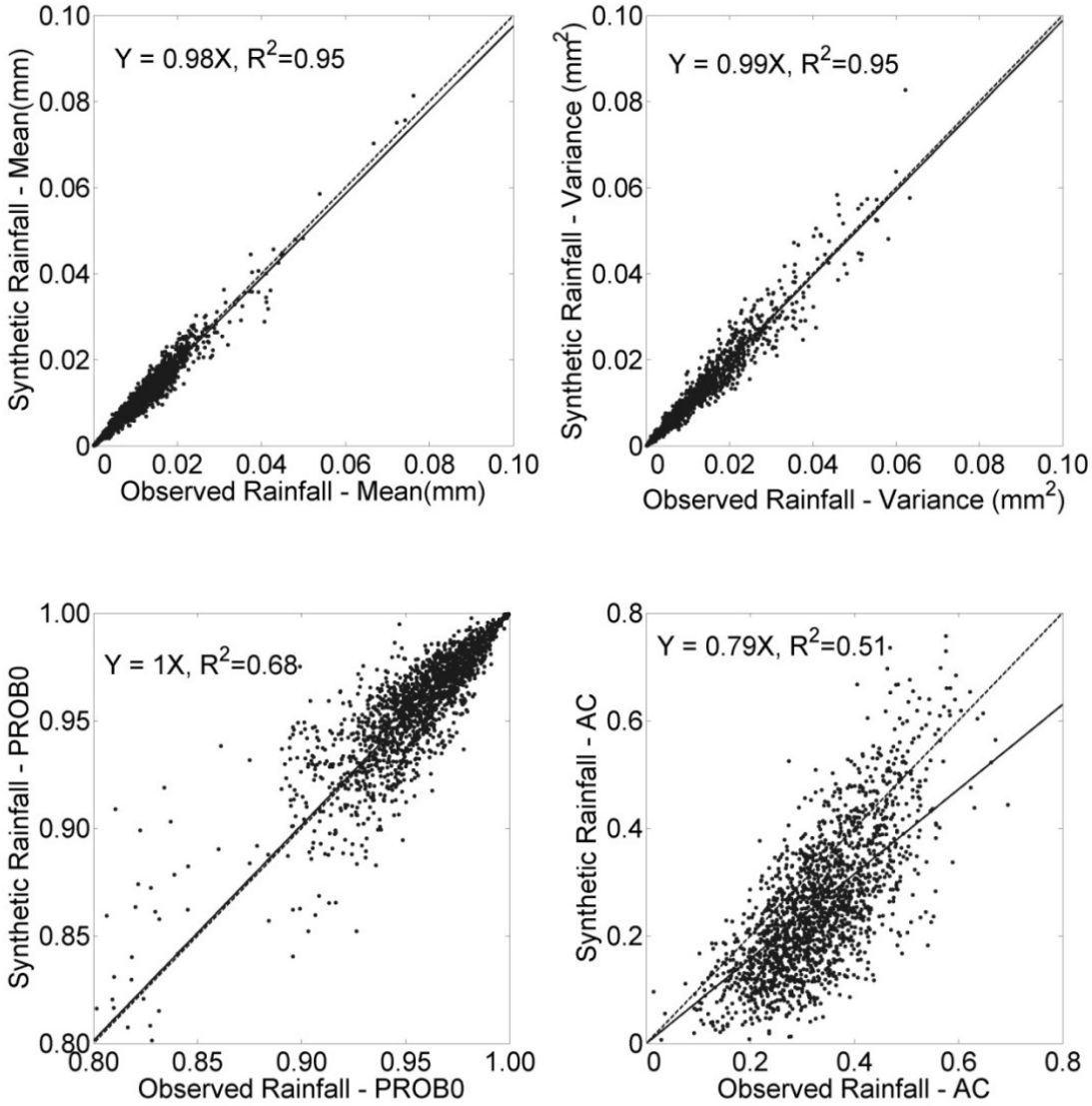


Figure 3-5 Rainfall statistics. The vertical axis corresponds to the statistics of the synthetic rainfall obtained with the weights of Figure 3 (R), and the horizontal axis to those of the observed rainfall.

and the probability of zero rainfall was neither over- or underestimated; while their scatter around the 1:1 line was given by the adjusted R^2 of 0.95, 0.95 and 0.68, respectively. The lag-1 auto-correlation coefficient, however, was underestimated most times and showed no clear pattern. The skewness was not included in Figure 3-5 because

of its relatively low importance compared to the other statistics. Similar behavior was observed when the statistics were calculated for synthetic rainfall time series obtained with the weight factors of Figure 4 (Q_p). Overall, it can be considered that the match of the statistics of the observed rainfall was satisfactory except for the case of the lag-1 autocorrelation coefficient.

Additionally, it is important to indicate that the internal structure of the synthetic rainfall time series is significantly different depending on the weight factors used. Figure 3-6 and Figure 3-7 present characteristics of the storms and rain cells, respectively. In each of these seven plots, the number of points was 1,800 (i.e., 150 gages × 12 months). Figure 3-6 shows scatter plots of average number of storms per 30-day period, average storm duration and average rainfall depth per storm calculated with weight factors obtained from Figure 3-3 (R) and Figure 3-4 (Q_p). Although not insignificant scatter of points around the 1:1 line was observed, no obvious preference to over- or underpredict based on the chosen weight factors was noticed, except for average storm duration, which was mostly over-predicted. Similarly, Figure 3-7 shows scatter plots of average number of cells per storm, average cell arrival rate, average rainfall depth per cell and average cell duration calculated with both sets of weight factors. In this case, almost no correlation was found, which could have been caused by the multi-modality of the MBLRP model calibration process [*Gyasi-Agyei*, 1999; *Onof et al.*, 2000].

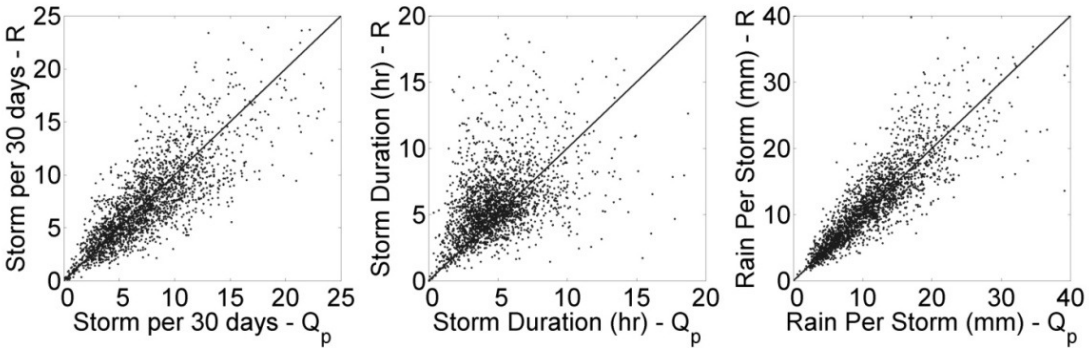


Figure 3-6 Average storm characteristics based on the MBLRP model parameters. The vertical axis corresponds to the parameters obtained with the weights from Figure 3-3 (R) and the horizontal axis to those obtained with the weights from Figure 3-4 (Q_p).

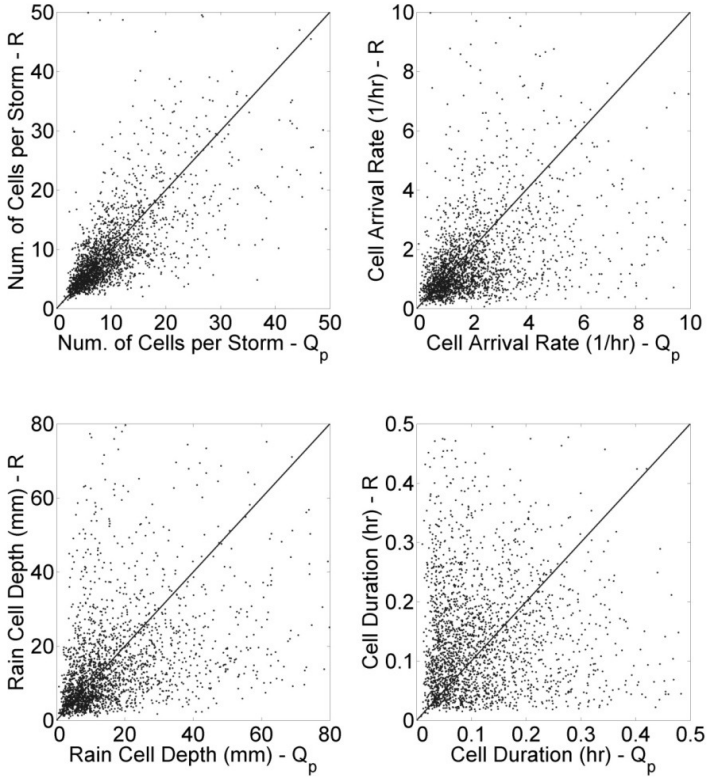


Figure 3-7 Average rain cell characteristics based on the MBLRP model parameters. The vertical axis corresponds to the parameters obtained with the weights from Figure 3-3 (R) and the horizontal axis to those obtained with the weights from Figure 3-4 (Q_p).

3.4 Effect of the Statistic Weights on Runoff Depths and Peak Flows

The synthetic rainfall time series were then routed through one of the virtual watersheds mentioned above – area of 7.5 km^2, lag time of 120 minutes and curve number of 80 – to determine the hydrograph at its outlet.

Figure 3-8 shows that, when all statistics are weighed equally, predicted runoff depths are underestimated on average by 20%; while, when the weight factors of Figure 3-3 (R) are used, the values are underestimated on average by only 4%. The adjusted R^2 were 0.85 and 0.88, and the standard deviation of the residuals 0.51 cm and 0.50 cm, respectively. Similarly, Figure 3-9 shows that, when all statistics are weighed equally, predicted peak flows are underestimated on average by 14%; while, when the weight factors of Figure 3-4 (Q_p) are used, the values are underestimated on average by only 3%. The adjusted R^2 were 0.86 and 0.88 and the standard deviation of the residuals 0.75 m^3/s and 0.73 m^3/s, respectively. Note that, in all plots of Figure 3-8 and Figure 3-9, the number of points is 1,800 (i.e., 150 gages × 12 months). Similar results were obtained for average-condition curve numbers of 50, 60, 70 and 90 but the results are not included here. Overall, when using the weights of Figure 3-3 (R) or Figure 3-4 (Q_p) instead of weighing equally the statistics, the systematic underestimation of runoff depths and peak flows, the scatter around the regression line and the standard deviation of the residuals decrease. This observation validates the initial concept that, when calibrating stochastic rainfall generators, not all statistics are equally important and some of them should be given a greater weight in the calibration process.

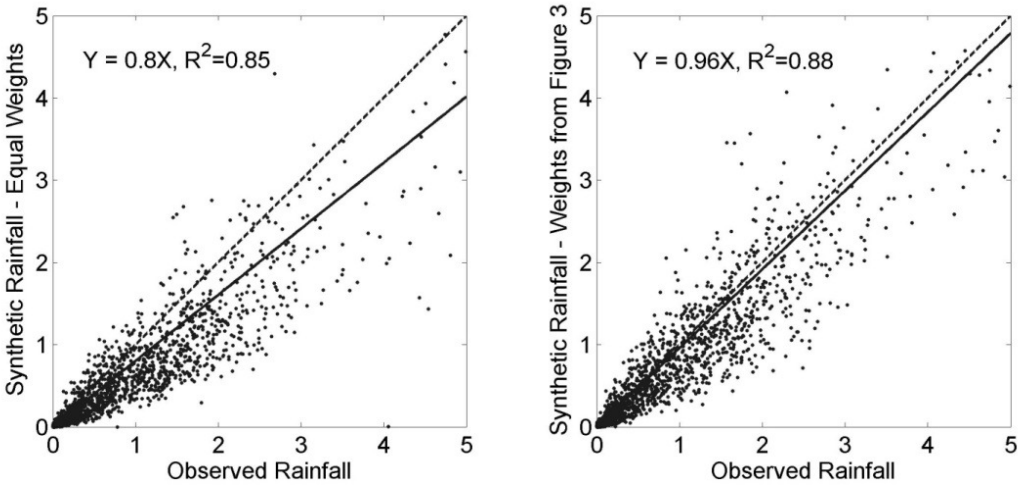


Figure 3-8 (a) Runoff depth calculated with synthetic rainfall obtained with weights equal to one versus the same but calculated with observed rainfall; and (b) runoff depth calculated with synthetic rainfall obtained with the weights of Figure 3-3 (R) versus the same but calculated with observed rainfall.

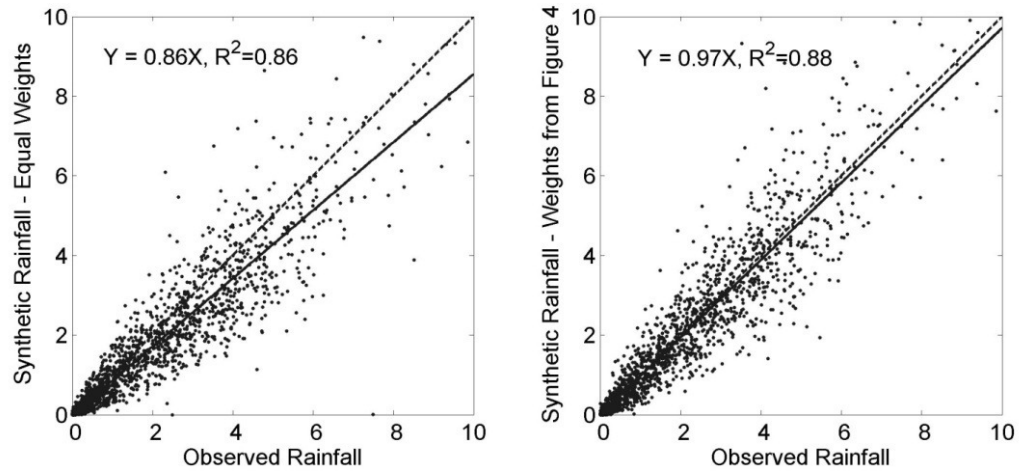


Figure 3-9 (a) Peak flow calculated with synthetic rainfall obtained with weights equal to one versus the same but calculated with observed rainfall; and (b) peak flow calculated with synthetic rainfall obtained with the weights of Figure 3-4 (Q_p) versus the same but calculated with observed rainfall.

Finally, an estimate of the expected error at a given gage could be estimated as follows. Equations (3-1) and (3-2) and the regression coefficients of Table 3-3 and Table 3-4 have been used above to relate the statistics of the observed precipitation to runoff depth and peak flow calculated also from observed precipitation. If it is assumed that the same regression coefficients relate the statistics of the synthetic precipitation time series to runoff depth and peak flow calculated from synthetic precipitation, it would follow:

$$\begin{aligned}\Delta R_{obs} - \Delta R_{syn} &= A_1\left(MEAN_{obs} - MEAN_{syn}\right) \\ &+A_2\left(VAR_{obs} - VAR_{syn}\right) + A_3\left(SKEW_{obs} - SKEW_{syn}\right) \\ &+A_4\left(PROB0_{obs} - PROB0_{syn}\right) + A_5\left(AC_{obs} - AC_{syn}\right)\end{aligned} \tag{3-8}$$

$$\begin{aligned}\Delta Q_{p\,obs} - \Delta Q_{p\,syn} &= B_1\left(MEAN_{obs} - MEAN_{syn}\right) \\ &+B_2\left(VAR_{obs} - VAR_{syn}\right) + B_3\left(SKEW_{obs} - SKEW_{syn}\right) \\ &+B_4\left(PROB0_{obs} - PROB0_{syn}\right) + B_5\left(AC_{obs} - AC_{syn}\right)\end{aligned} \tag{3-9}$$

where the subscript *obs* stands for observed and *syn* for synthetic. In these two equations, the left hand side and the terms in parenthesis on the right hand side constitute expected discrepancies, or expected errors, in predicted runoff depth residuals, peak flow residuals and rainfall statistic z-scores, and can be re-written as:

$$\varepsilon_R = A_1\,\varepsilon_{MEAN} + A_2\,\varepsilon_{VAR} + A_3\,\varepsilon_{SKEW} + A_4\,\varepsilon_{PROB0} + A_5\,\varepsilon_{AC} \tag{3-10}$$

$$\varepsilon_{Q_p} = B_1\,\varepsilon_{MEAN} + B_2\,\varepsilon_{VAR} + B_3\,\varepsilon_{SKEW} + B_4\,\varepsilon_{PROB0} + B_5\,\varepsilon_{AC} \tag{3-11}$$

where ε stands for expected error. Ideally, from the watershed hydrology perspective, the synthetic rainfall time series is such that the terms ε_R and ε_{Qp} are equal to zero. However, the fact that the synthetic time series does not exactly match the statistics of the observed rainfall (i.e., ε_{MEAN} , ε_{VAR}, ε_{SKEW} and ε_{Qp} are different from zero) causes the terms ε_R and ε_{Qp} to be greater than zero as well. However, even if the rainfall statistics were matched perfectly, there might still be an error in the runoff depth and peak flow values caused by the intrinsic error in regression equations (3-1) and (3-2) and the assumption that the regression coefficients also apply to the synthetic time series. Figure 3-10 (a) shows the runoff depth calculated from the synthetic rainfall assuming the weights from Figure 3-3 (R) in the vertical axis, and from the observed rainfall in the horizontal axis; and Figure 3-10 (b) shows the same as (a) but after adding the expected error of equation (3-10) to the runoff depth in the vertical axis. The axes in both charts of Figure 3-10 have been limited to 3 cm to better display the scatter of the data points; however, that has been done at the expense of not including 4% of the points. It can be seen that the values in the right hand side chart lie closer to the 1:1 line indicating the estimated expected error resembles the actual error.

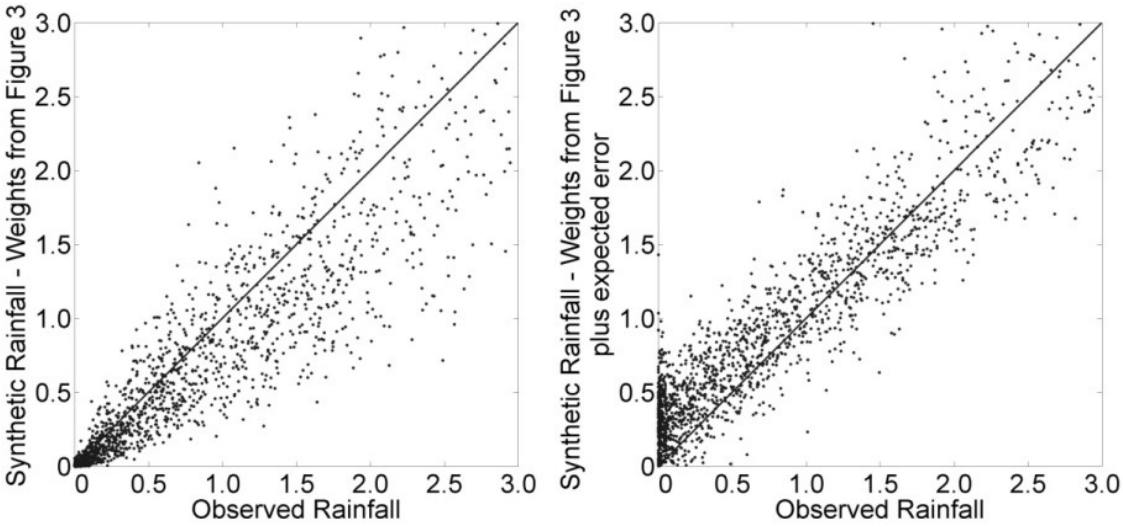


Figure 3-10 (a) Runoff depth calculated with the synthetic rainfall obtained with weights from Figure 3-3 (R) versus the same but calculated with the observed statistics; and (b) same as (a) but after adding the expected error of equation (3-10).

3.5 Conclusions

An improved calibration approach for stochastic rainfall generation models has been presented. In previous applications available in the literature, these models were calibrated to match long-term statistics of rainfall records without differentiating them based on their relative importance. In this new approach, the relative importance of the different statistics is quantified proportionally to the regression coefficients of linear regression equations that relate catchment response characteristics to the rainfall statistics. In this specific application, the selected catchment response characteristics were the long-term average monthly runoff depth and the long-term average monthly runoff peak flow, which were considered representative of the hydrograph volume and

variability. 1099 and 150 NCDC rain gages were used for the regression analysis and validation of the approach, respectively.

Based on the coefficients obtained in the regression analysis, it was concluded that the rainfall mean explained between 40% and 55% of the runoff depth variability; the rainfall variance, around 20%; the probability of zero rainfall and lag-1 autocorrelation coefficient, around 20% combined; and the skewness, less than 15%. Similarly, the rainfall mean explained between 30% and 40% of the peak flow variability; the rainfall variance, between 35% and 40%; the probability of zero rainfall and lag-1 autocorrelation coefficient, around 25% combined; and the skewness, less than 10%.

It was noted that synthetic rainfall time series generated weighing the precipitation statistics according to these figures reproduced well the statistics of the recorded rainfall, with the exception of the lag-1 autocorrelation coefficient. Furthermore, it was observed that these synthetic rainfall time series outperformed those generated weighing all statistics equally at predicting watershed runoff depths and peak flows. In this comparison, the benchmark for assessing the performance of the two models consisted of the runoff depths and peak flows generated by the observed precipitation. In fact, when all statistics were given the same weight, runoff depths and peak flows were underestimated by 20% and 14%, respectively; while, when the statistics were weighed proportionally to their relative importance, the underprediction was reduced to 4% and 3%, which confirms the advantage of weighing the statistics differently.

Finally, it is important to stress that the specific weigh values used in this study should not be considered recommended values for other applications. The value of the weights depends on the hydrologic process being modeled.

4. EFFECT OF THE TEMPORAL VARIABILITY OF RAINFALL STATISTICS ON STOCHASTICALLY GENERATED RAINFALL TIME SERIES

4.1 Introduction

Stochastic rainfall generators are used in hydrologic analysis whenever rainfall data are not available; that is, areas with no rain gages and periods with no records. The Poisson cluster rainfall models [*Rodriguez-Iturbe et al.*, 1987; 1988], a type of stochastic rainfall generators, represent rainfall as sequences of storms composed of clusters of rain cells [*Kavvas and Delleur*, 1975]. This approach to simulate the storm physical processes makes them a robust and practical option for modeling continuous rainfall time series [*Olsson and Burlando*, 2002]. The applicability of the Poisson cluster rainfall models has been validated over various geographic locations with different rainfall characteristics [*Isham et al.*, 1990; *Bo et al.*, 1994; *Onof and Wheater*, 1994b; *Glasbey et al.*, 1995; *Khaliq and Cunnane*, 1996; *Onof et al.*, 1996b; *Cowpertwait et al.*, 1996a; *Verhoest et al.*, 1997] and the models themselves have been used in a wide range of studies dealing with flooding [e.g. *Wheater et al.*, 2005], drought [e.g. *Yoo et al.*, 2008], contaminant transport [e.g. *Botter et al.*, 2006], and ecosystem behavior [e.g. *Laio et al.*, 2009], among others. Important research on the Poisson cluster rainfall models include the modification of the Bartlett-Lewis rectangular pulse model [*Rodriguez-Iturbe et al.*, 1987] to better represent the randomness of the cell durations, which varies from storm to storm, and the presence and duration of dry periods [*Rodriguez-Iturbe et al.*, 1988];

Cowpertwait [1991] derived an equation that represented the probability that an arbitrary time interval of any chosen duration would be dry; Cowpertwait [1994] also developed a model that uses different sets of parameters for representing rainfall cell durations and depths corresponding to different storm types (e.g. frontal or convective), which shows an improved performance in reproducing hourly statistics of rainfall time series; Kakou [1998] developed a method in which an explicit dependence between the depth and duration of the rain cells is considered; Cowpertwait [1998] derived an equation that represented the skewness of the rainfall depth distribution, which allowed it to reproduce better extreme values; Cowpertwait et al. [2007] present a model in which the rainfall intensity varies within the rain-cell duration, which makes it capable of capturing the rainfall variability at time scales as fine as 5 minutes.

In this study, a novel approach based on Modified Bartlett-Lewis Rectangular Pulse (MBLRP) model [*Rodriguez-Iturbe et al.*, 1988] that can account for the year-to-year variability of the rainfall statistics is suggested. This approach has the similarity with those of Cowpertwait [1991] and Cowpertwait [1998] in that it improved performance of the model by incorporating more information about the observed rainfall time series than modifying the model structure.

4.2 Modified Bartlett-Lewis Rectangular Pulse (MBLRP) Model

In the MBLRP model, rainfall time series are represented as sequences of storms comprised of rain cells (Figure 2-1). In the model, X1 [T] is a random variable that represents the storm arrival time, which is governed by a Poisson process with parameter λ; X2 [T] is a random variable that represents the duration of storm activity (i.e., the time

window after the beginning of the storm within which rain cells can arrive), which varies according to an exponential distribution with parameter γ; X3 [T] is a random variable that represents the rain cell arrival time within the duration of storm activity, which is governed by a Poisson process with parameter β; X4 [T] is a random variable that represents the duration of the rain cells, which varies according to an exponential distribution with parameter η that, in turn, is a random variable represented by a gamma distribution with parameters ν and α; and X5 [L/T] is a random variable that represents the rain cell intensity, which varies according to an exponential distribution with parameter $1/\mu$. From the physical viewpoint, λ [1/T] is the expected number of storms that arrive in a given period, γ [1/T] is the inverse of the expected duration of storm activity, β [1/T] is the expected number of rain cells that arrive within the duration of storm activity, η [1/T] is the inverse of the expected duration of the rain cells, and μ [L/T] is the average rain cell intensity. Parameters ν [T] and α [dimensionless] do not have a clear physical meaning, but the expected value and variance of η can be expressed as α/ν and $\alpha/\nu 2$. Therefore, the model has six parameters: λ, γ, β, ν, α and μ; however, it is customary to use the dimensionless ratios $\varphi = \gamma/\eta$ and $\kappa = \beta/\eta$ as parameters instead of γ and β.

The estimation of the model parameters is accomplished by matching statistics of the simulated and observed rainfall time series. Some commonly used statistics are the precipitation depth mean, variance, probability of zero rainfall and lag-s covariance at various time scales [*Khaliq and Cunnane*, 1996].

4.3 Problem Definition

The parameters of the Poisson cluster rainfall models determine the random behavior of storms and rain cells and are calculated such that the differences between the statistics of the observed and the synthetic rainfall time series are minimized. The following equation is typically used as an objective function in the calibration of Poisson cluster rainfall models [*Bo et al.*, 1994].

$$\sum_{k=1}^{n} w_k \left[1 - \frac{F_k(\vec{\theta})}{f_k} \right]^2 \qquad (4\text{-}1)$$

where $\vec{\theta}$ is the vector containing the model parameters, n is the number of rainfall statistics being matched (i.e., five in this case), F_k is the k^{th} statistic of the synthetic rainfall time series, f_k is the k^{th} statistic of the observed rainfall time series, and w_k is the weight factor of the k^{th} statistic.

The statistics considered in this study included mean, variance, lag-1 autocorrelation coefficient and probability of zero rainfall. To capture the rainfall seasonality, different models were developed for each month of the year (i.e., January, February, … December) for a total of 12 models. In this objective function, the statistics of the observed rainfall f_k are typically calculated for the entire period of record, overlooking the fact that they vary from year to year. For example, for the case of NCDC rain gage FL-9148 (see Figure 4-1), which has a period of record of 61 years, Figure 4-2 illustrates both the long-term as well as the individual-month statistics. It can be observed in the figure that the variability of the statistics from year to year is

significant, and that lumping all this information in a single value is a rough simplification. Considering that in previous studies found in the literature only the long term rainfall statistics have been taken into account in the model calibration, it is to wonder how well the year-to-year variability of the statistics is reproduced and what problems are caused by using only long-term statistics in the parameter calibration of Poisson cluster rainfall models.

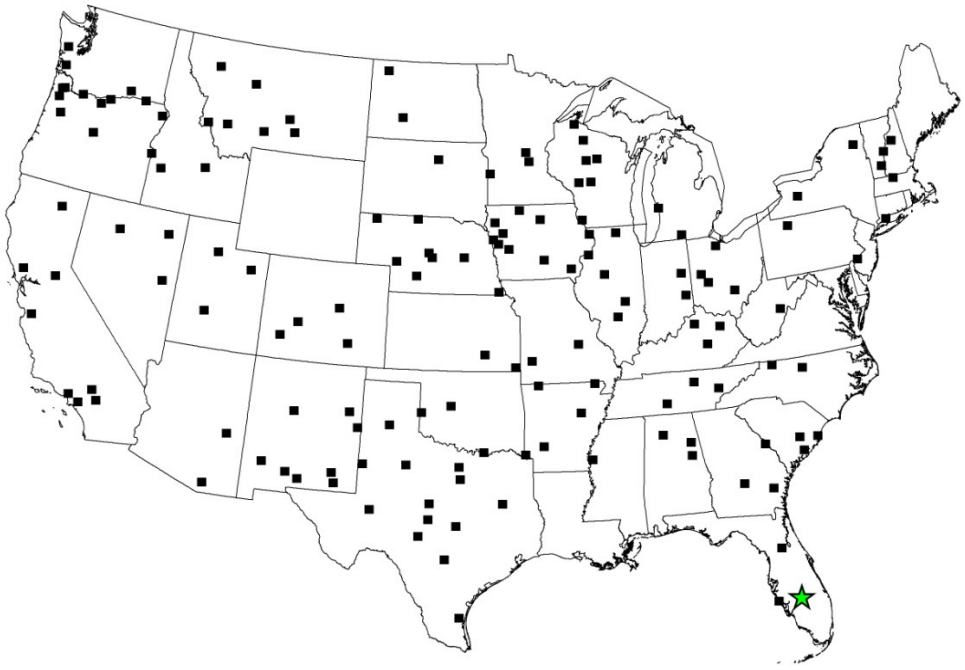


Figure 4-1 Location of the 150 NCDC Precipitation Gages used to identify the problem of the traditional approach of Poisson cluster rainfall modeling (black squares) and the location of the NCDC gage FL-9148 (green star) at which the new approach was developed and validated.

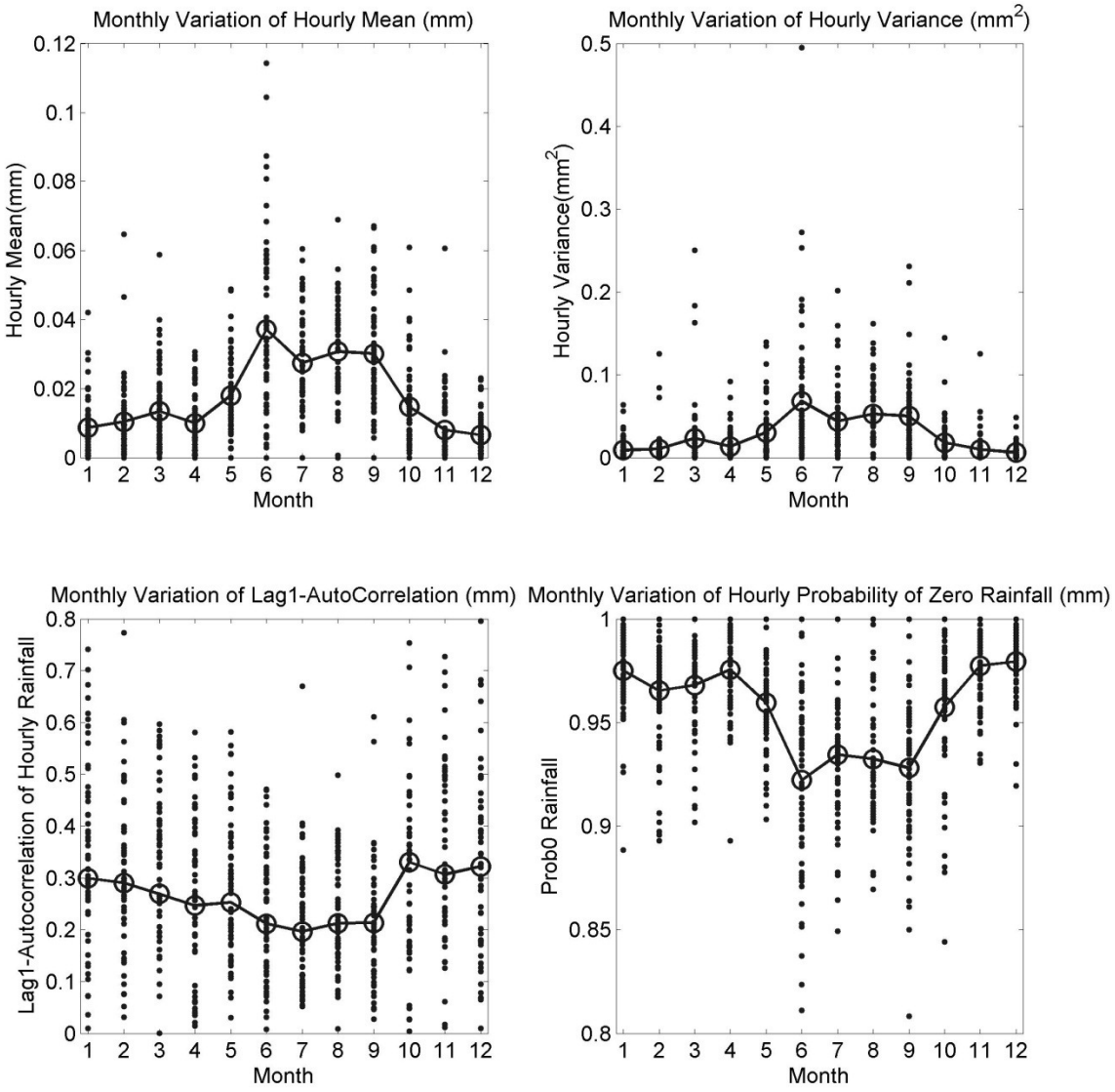


Figure 4-2 Monthly variation of the rainfall statistics observed at the gage FL-9148. The statistics referring to the month of the entire length of the rainfall time series are shown in circle along with the lines and the ones referring to the month of a specific year are shown as dots.

Figure 4-3 suggests an answer to the first question. The top plots of Figure 4-3 are the distribution of the statistics corresponding to June short-term statistics of the 61 years of the rainfall time series observed at the gage FL-9148. The bottom plots of

Figure 4-3 in the second rows are the distribution of the statistics corresponding to short-term statistics of the 100 years of the synthetic rainfall time series at the same gage. Here, the synthetic rainfall time series were generated using the Modified Bartlett-Lewis rectangular pulse model – MBLRPM [*Rodriguez-Iturbe et al.*, 1988] with the parameters identified by optimizing equation (1) based on the following f_ks: mean rainfall at 1-hourly accumulation level; variance of rainfall at 1-, 3-, 12-, and 24-hourly accumulation level; lag-1 autocorrelation at 1-, 3-, 12-, and 24-hourly accumulation level; probability of zero rainfall at 1-, 3-, 12-, and 24-hourly accumulation level. It can be noted that the spread of the statistics of the observed rainfall time series is greater than those of the synthetic rainfall time series. Especially, the discrepancy of the variability of mean and probability of zero rainfall is significant. The mean of the distribution, however, is similar for the upper and lower histogram. This is because the model is designed to match the long term statistics. This problem was identified to some extent by former studies that discuss the inherent characteristic time scale of the Poisson cluster rainfall models. [*Olsson and Burlando*, 2002; *Marani*, 2003]. These studies indicate that the Poisson cluster models can reproduce the variance of rainfall depth only within a given temporal accumulation level (e.g., from 20 minute to 1 week). The finding of the present study is distinct from the former ones in that it identified the mismatch of the variation of not only the mean but also the general statistics at the accumulation level outside of the inherent characteristic time scale of Poisson cluster rainfall models.

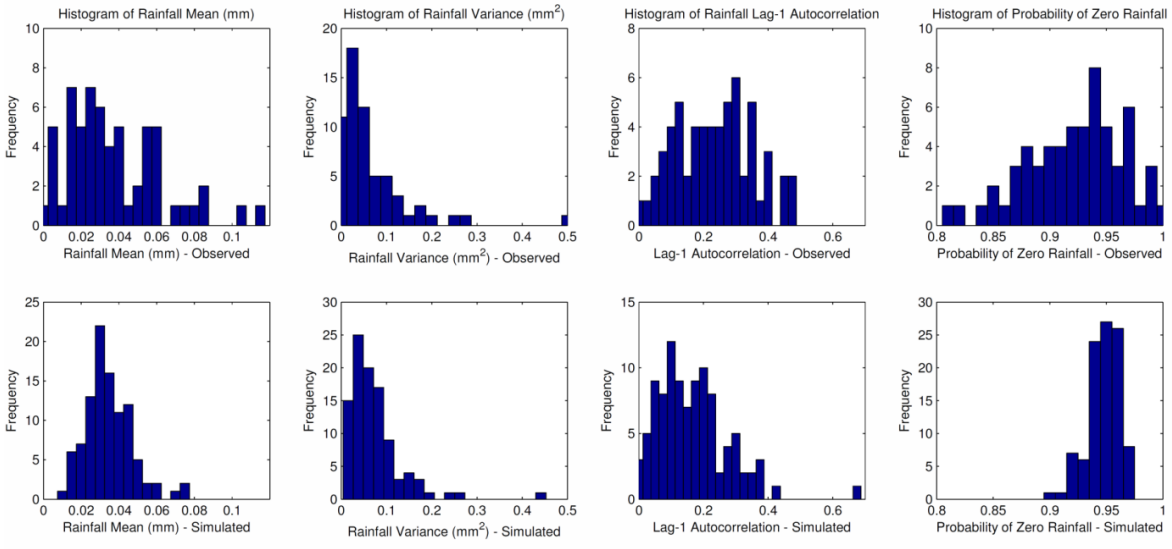


Figure 4-3 Histograms of the statistics of the rainfall time series observed in June at the gage FL-9148 (top) and the ones corresponding to the synthetic rainfall time series generated by the traditional approach of MBLRPM (bottom).

Figure 4-4 shows what problems can be caused due to the systematic underestimation of the year-to-year variability of the rainfall statistics made by the current framework of MBLRPM. The figure compares the standard deviation of the monthly runoff depth and monthly peak flow derived from the observed and synthetic rainfall time series. Instead of applying the typical approach of measuring the model performance based on rainfall statistics, watershed response variables (i.e., runoff depth and peak flow) were used as metric to measure the performance of the model. This is because, from a watershed viewpoint, realistic synthetic rainfall should reproduce well the catchment response, including runoff depths and flows at the outlet. The comparison was made for 150 randomly selected gages across the coterminous United States (Figure 4-1). The random selection was performed out of the pool of rain gages that has at least

50 years of rainfall record. The threshold of 50 years was used to guarantee the representativeness of the rainfall characteristics at the location of the gage.

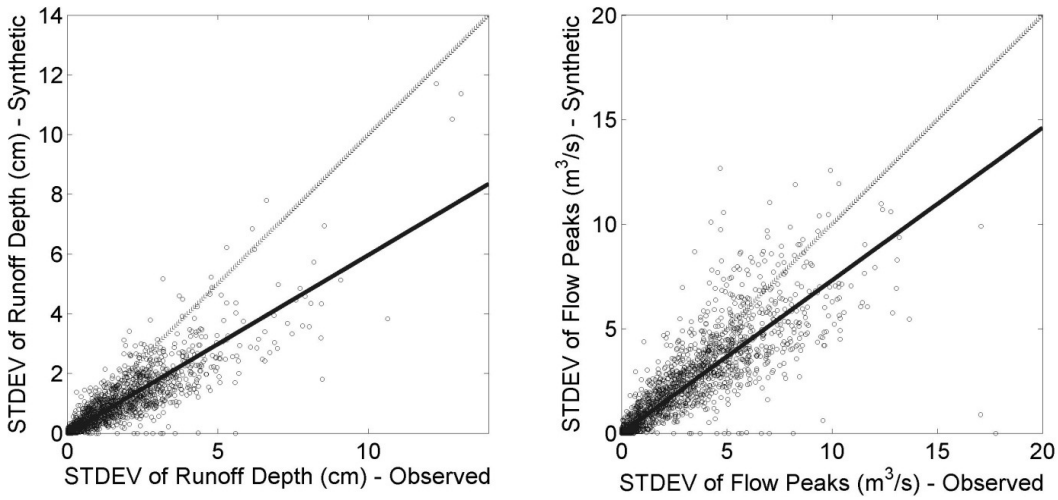


Figure 4-4 Scatter plots of the standard deviations of the monthly runoff depth (left) and monthly peak flow (right) derived from the observed (x) and synthetic rainfall time series (y). The plot compares the values calculated at 150 NCDC rain gages across the coterminous United States (Figure 4-1).

The x-coordinate of each of the points in the plots is calculated according to the following steps:

1. For each of the 150 selected gages, separate the observed rainfall time series by months. (e.g. January 1953, February 1953, March 1953, … , December 2006)
2. Apply each of the separated rainfall time series on a virtual watershed with a drainage area of 7.50 km^2, lag time of 2.00 hours, and curve number of 80. Here, The size and the lag time of the virtual watershed were determined referring to the findings of McCuen et al. [1984] which estimated the average drainage area

(6.47 km^2) and the average time of concentration (1.49 hour) of the 48 major urban watersheds across the United States. The values adopted by this study are slightly greater than these values, but the discrepancy is small enough to assume that the selected values can represent a "typical" urban watershed.

3. Calculate the monthly runoff depth and peak flows using SCS curve number and SCS curvilinear unit hydrograph method [*SCS*, 1972]. In the process of modeling runoff depth, the antecedent soil moisture condition (AMC), which plays an important role in runoff generation, was taken into account by modifying the curve number values according to the precipitation depth in the previous five days (63 when dry, 90 when wet).
4. Calculate the standard deviation of the monthly runoff depths and peak flows of each month (e.g., standard deviation of monthly runoff of January 1953, 1954, … , 2006)

The y-coordinate of each of the points in the plots is calculated according to the following steps:

1. For each of the 150 selected gages, estimate the parameters of the BLRP model by optimizing Equation (1). The f_ks that are used in the parameterizations are: mean rainfall at 1-hourly accumulation level; variance of rainfall at 1-, 3-, 12-, and 24-hourly accumulation level; lag-1 autocorrelation at 1-, 3-, 12-, and 24-hourly accumulation level; probability of zero rainfall at 1-, 3-, 12-, and 24-hourly accumulation level. Based on the estimated parameters, generate synthetic rainfall time series for the length of 50 years.

2. Calculate the standard deviation of the monthly runoff depths and peak flows of each month (e.g., standard deviation of monthly runoff of January 1953, 1954, … , 2006) using the exactly same steps applied to calculate the observed runoff depths and peak flows.

Least square fit regression lines with the intercept being the center of origin (solid line) and 45 degree lines (dotted line) are shown along with the scatter of points. The slope of the regression lines are 0.60 and 0.73 for the plots of runoff depths and peak flows, respectively. This suggests that the variability of runoff depth and peak flows of a given month is systematically underestimated when they are derived from synthetic rainfall time series. In other words, the runoffs and peak flows of a given month derived from the synthetic rainfall time series can miss the upper and lower extreme that are observed in reality, which typically leads to water resources related problems in practice. This study makes an hypothesis that this systematic underestimation of variability of catchment responses are caused by underestimation of the year-to-year variation in rainfall statistics. It is expected that a model that accounts for the year-to-year variability of the rainfall statistics will outperform the traditional Poisson cluster models in reproducing the variability of catchment responses.

The model presented by this study incorporates this variability by simulating the short-term rainfall statistics based on the correlation between the observed rainfall statistics. Then, the model generates the rainfall time series using the MBLRPM based on the short-term rainfall statistics simulated in the first part of the model. The model is named “The Hybrid Model – THM” because it combines the process of generating

rainfall statistics and generating rainfall time series. The procedure of the model development is explained in detail based on the precipitation data observed at one gage in Florida (FL-9148, Figure 4-1). Then, the model is validated using the rainfall data observed at 104 rain gages located across the United States (Figure 4-8).

4.4 The Hybrid Model (THM)

4.4.1 Stochastic Generation of the Short-term Rainfall Statistics

The first step of the THM is to stochastically generate the monthly rainfall statistics. The model generates the following monthly rainfall statistics based on correlation between them; mean at hourly accumulation level (MEAN1), variance at hourly, 3-hourly, 12-hourly, and 24-hourly accumulation level (VAR1, VAR3, VAR12, and VAR24); probability of zero rainfall at hourly, 3-hourly, 12-hourly, and 24-hourly accumulation level (PROB0-1, PROB0-3, PROB0-12, and PROB0-24); lag-1 autocorrelation at hourly, 3-hourly, 12-hourly, and 24-hourly accumulation level (AC1, AC3, AC12, and AC24).

4.4.2 Correlation between Rainfall Statistics

Figure 4-5 shows scatter plots of the variance, probability of zero rainfall and lag-1 autocorrelation coefficient vs. mean, all at one hour accumulation level, of the rainfall time series observed at gage FL-9148 for the month of June.

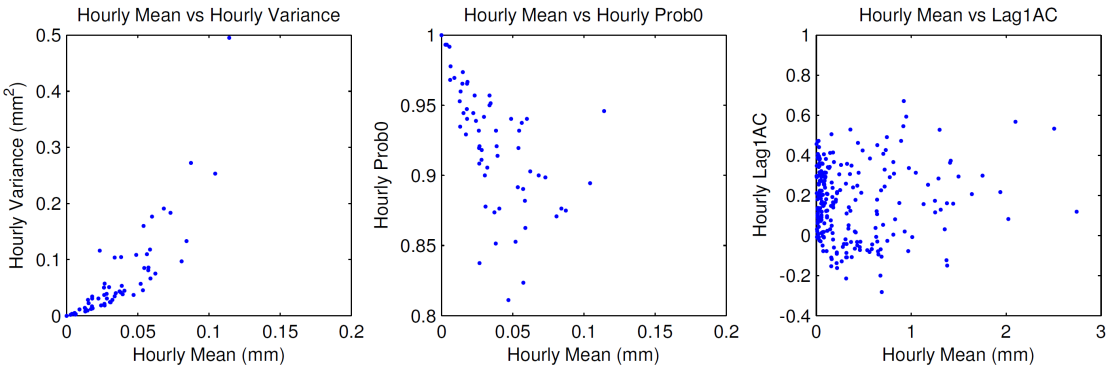


Figure 4-5 Correlation between MEAN1 and VAR1; MEAN1 and PROB0-1; and MEAN1 and AC1 at gage FL-9148.

In the plots of Figure 4-5, there are as many points as years of record in the gage (i.e., 61 points). Regression analyses were performed to identify the relationship between the variables. Table 1 shows the regression coefficients between the variables. Only AC1 did not show a clear relationship with MEAN1 compared to the other two statistics. THM generates MEAN1 first, and then generates the VAR1 and PROB0-1 based on the generated MEAN1 according to the relationship identified through the regression analyses. AC1 was generated independently since it did not show a significant correlation with the other statistics. A strong linear correlation was identified between the rainfall statistics at different accumulation levels as can be seen in Figure 4-6. The statistics at 3, 12, and 24 hours of accumulation were generated based on the relationship identified through this regression analysis. The result of the analysis is presented in Table 4-1.

Table 4-1 Results of the regression analysis between the rainfall statistics at the gage NCDC-FL9148

The linear regression equation y = ax + b was assumed.

Var(Error) represents the variance of the residuals. $\left[\frac{\sum (y_{observed} - y_{regressed})^2}{n-1}\right]$

Independent Variable (x)	Dependent (y)	a	b	Var(Error)	R^2
MEAN1	VAR1	2.7558	-0.0348	0.002	0.7108
MEAN1	PROB0-1	-0.9718	0.9583	0.0013	0.3236
MEAN1	AC1	-0.2059	0.2422	0.0142	0.0019
VAR1	VAR3	3.4488	0.0287	0.0017	0.9799
VAR3	VAR12	5.8397	0.0011	0.2463	0.9202
VAR12	VAR24	1.9743	0.0664	0.5789	0.9541
PROB0-1	PROB0-3	1.3929	-0.4056	0.0002	0.9592
PROB0-3	PROB0-12	1.6505	-0.7018	0.0013	0.8896
PROB0-12	PROB0-24	1.4899	-0.5256	0.0017	0.9368
AC1	AC3	0.4605	0.0748	0.0149	0.1692
AC3	AC12	0.5399	-0.0547	0.0255	0.17
AC12	AC24	0.5698	0.1757	0.0405	0.1975

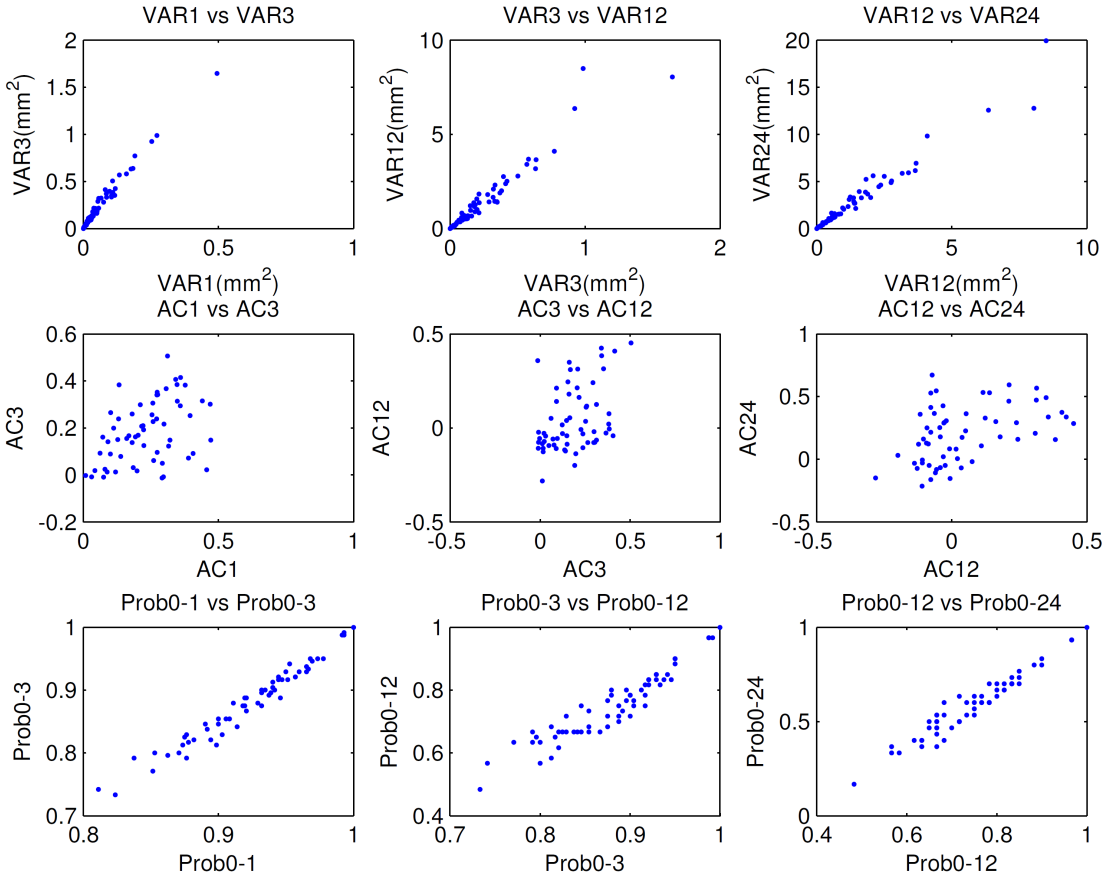


Figure 4-6 Correlation between the statistics measured at different levels of accumulation time at gage FL-9148.

4.4.2.1 Generation of MEAN1

The histogram of MEAN1 for the month of June for gage FL-9148 and its corresponding fitted gamma distribution are shown in Figure 4-7. A gamma distribution was used because MEAN1 cannot adopt negative values and because of its positive skew [*Ozturk*, 1981]. The two parameters of the distribution were estimated using the method of maximum likelihood. For generating synthetic precipitation time series, a different MEAN1 value is randomly drawn from its distribution for each year of simulation.

4.4.2.2 Generation of AC1

Histogram of AC1 observed at the gage FL-9148 during the month of June is shown in Figure 4-8. Because of its near-symmetrical shape, a normal distribution was used to fit the distribution of AC1. AC1 is randomly drawn from this fitted normal distribution.

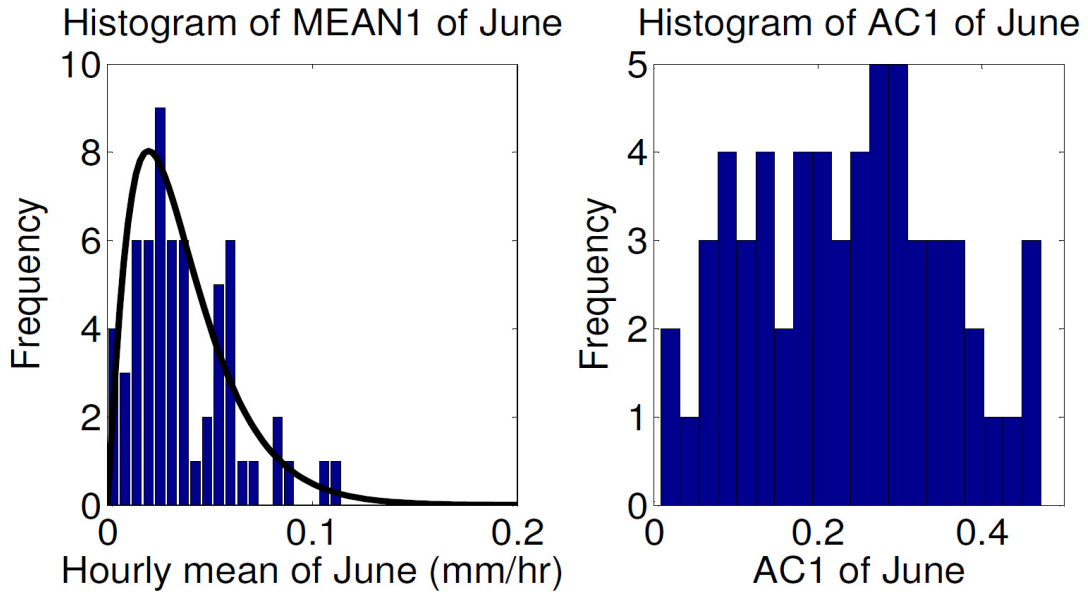


Figure 4-7 Histogram of MEAN1 and AC1. The fitted curve of gamma distribution for MEAN1 is shown along with the histogram (left side).

4.4.2.3 Generation of VAR1 and PROB0-1

Because VAR1 and PROB0-1 are highly correlated with MEAN1 (Figure 5), VAR1 and PROB0-1 are generated based on MEAN1 using the result of the regression analysis (Table 1). The following two equations specifically address the relationships between these variables for the gage FL-9148:

$$VAR1 = 2.76\ Mean1 - 0.0348 + \varepsilon_1 \quad (4\text{-}2)$$

$$PROB0\text{-}1 = -0.9718\ Mean1 + 0.9583 + \varepsilon_2 \quad (4\text{-}3)$$

where ε_1 and ε_2 represent the residual error expected when using the regression equation. ε_1 and ε_2 in the equations were randomly drawn from a normal distribution with mean 0 and variance equal to the error variance of the regression equation (Table 1).

4.4.2.4 Generation of the statistics at 3-, 12-, and 24-hourly accumulation levels

Good linear relationships were identified between the statistics measured at different levels of accumulation time through the regression analysis. The regression coefficients between the statistics are shown in Table 1. This relationship was used in the generation of the statistics at 3, 12 and 24 hours of accumulation level. The methodology is exactly same as the one used for the generation of VAR1 and PROB0-1. VAR1 is used for the generation of VAR3, which is used for the generation of VAR6. VAR6, then again, is used as the basis of the generation of VAR12, which is used for the generation of VAR24. The same principle was applied for the generation of PROB0-3, PROB0-6, PROB0-12, PROB0-24, AC3, AC6, AC12, and AC24.

4.4.3 Generation of Rainfall Time Series Using MBLRPM

THM generates synthetic rainfall time series using MBLRPM. Before the generation of the synthetic rainfall time series, the parameters of the model should be estimated. The 6 parameters of the MBLRPM were estimated based on the rainfall statistics generated in the first part of the model. ISPSO [*Cho et al.*, In Review] was used

to find the set of parameters ($\vec{\theta}$) that minimizes the objective function (OF) given in equation (1)

THM first generates 20 months of synthetic rainfall time series based on one estimated parameter set. Then, it calculates the difference between the statistics of the simulated rainfall time series of the j^{th} month and the target statistics based on the following equation:

$$R_j = \sum_{i=1}^{n} \left[(Z_i^{simul} - Z_i^{target})^2 \right] \quad (4\text{-}4)$$

where Z_i^{simul} is the Z-score of the ith statistics of simulated rainfall time series and Z_i^{target} is the Z-score of the ith target statistics. Here, Z-score was calculated based on the mean and the standard deviation of the rainfall statistics that were estimated using the rainfall time series observed at 1099 rain gages across the United States (Table 3-1)

This study used MEAN1, VAR1, AC1, and PROB0-1 to calculate the value of R (i = 4). Then, the synthetic rainfall time series of the j^{th} month with the lowest value of R_j is chosen among the 20 months of generated rainfall. This selection process prevents THM from incorporating the redundant stochasticity. An alternative approach in which only a month of synthetic rainfall time series is produced for a given parameter set as opposed to the proposed selection procedure incorporates the stochasticity from the process of modeling statistics (first part of the model) and the process of generating rainfall time series.

4.5 Model Validation

The performance of THM was tested based on its ability to reproduce the distribution of the monthly maximum rainfall depths, monthly peak flows, and monthly runoff volumes. A total of 104 months of precipitation data observed at 104 NCDC precipitation gages (one month per one gage) across the coterminous United States (Figure 4-8) were used for this validation procedure. These gages were randomly drawn from the gages that contain at least 50 years of records. For each of the 104 gage-months, 100 months of synthetic rainfall time series were generated using both THM and the traditional approach of Modified Bartlett-Lewis Rectangular Pulse (MBLRP) Model. Thus, each of the gage-months has three different types of rainfall time series including the observed ones. Then, the following values are calculated based on all three types of rainfall time series: monthly maximum rainfall depths with the duration of 1, 3, 6, 12, and 24 hours; monthly runoff depth; and monthly peak flow. SCS curve number method and SCS curvilinear unit hydrograph method [*SCS*, 1972] were used to calculate the last two. The watershed characteristics assumed are the lag time of 2 hours; drainage area of 7.5 km^2; and the curve number of 50, 60, 70, 80, and 90.

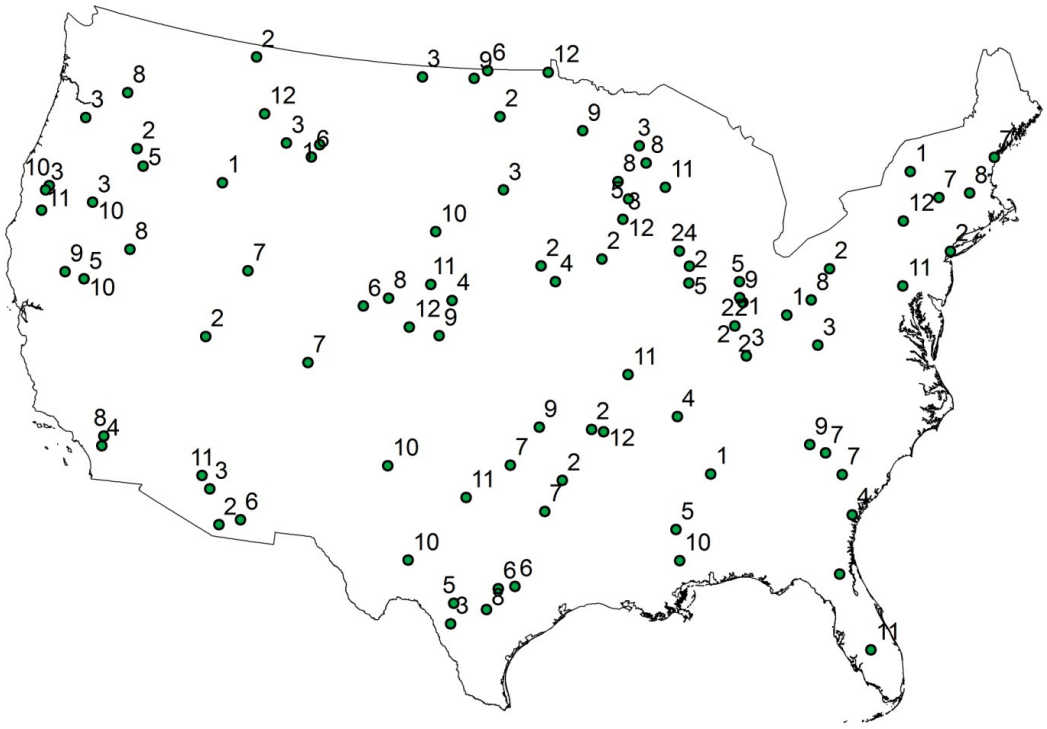


Figure 4-8 Location of the 104 NCDC precipitation stations used for the validation of THM. The numbers next to the dots represent the calendar month on which the validation was performed.

As a result, each gage-month is associated with 500 monthly maximum rainfall depths accounting for statistic variability (i.e., 5 rainfall durations × 100 years of simulation) ($P^{1}_{THM}, P^{3}_{THM}, P^{6}_{THM}, P^{12}_{THM}, P^{24}_{THM}$); 500 monthly maximum rainfall depths not accounting for statistic variability (i.e., 5 rainfall durations × 100 years of simulation) ($P^{1}_{Trad}, P^{3}_{Trad}, P^{6}_{Trad}, P^{12}_{Trad}, P^{24}_{Trad}$); a number of monthly maximum observed rainfall depths (i.e., 5 rainfall durations × years of record) ($P^{1}_{Obs}, P^{3}_{Obs}, P^{6}_{Obs}, P^{12}_{Obs}, P^{24}_{Obs}$); 5 (different curve numbers) sets of 100 monthly runoff depths based on THM ($R^{50}_{THM}, R^{60}_{THM}, R^{70}_{THM}, R^{80}_{THM}, R^{90}_{THM}$); 5 sets of 100 monthly runoff depths based on the

traditional approach ($R^{50}_{Trad}, R^{60}_{Trad}, R^{70}_{Trad}, R^{80}_{Trad}, R^{90}_{Trad}$); 5 sets of +50 monthly runoff depths based on observed rainfall time series ($R^{50}_{Obs}, R^{60}_{Obs}, R^{70}_{Obs}, R^{80}_{Obs}, R^{90}_{Obs}$); 5 (different curve numbers) sets of 100 monthly peak flows on THM ($F^{50}_{THM}, F^{60}_{THM}, F^{70}_{THM}, F^{80}_{THM}, F^{90}_{THM}$); 5 sets of 100 monthly peak flows based on the traditional approach ($F^{50}_{Trad}, F^{60}_{Trad}, F^{70}_{Trad}, F^{80}_{Trad}, F^{90}_{Trad}$); 5 sets of +50 monthly peak flows based on observed rainfall time series ($F^{50}_{Obs}, F^{60}_{Obs}, F^{70}_{Obs}, F^{80}_{Obs}, F^{90}_{Obs}$).

The two-sample Kolmogorov-Smirnov test was used to compare the distributions of the variables calculated from the observed rainfall time series and the ones calculated from the synthetic rainfall time series. The test statistic of the two-sample Kolmogorov-Smirnov test which compares the distributions of the data set x_1 and x_2 is as follows:

$$\max \left| F_1(x) - F_2(x) \right| \tag{4-5}$$

where $F_1(x)$ is the proportion of a set x_1 less than or equal to x. The null hypothesis of the test is that the sets x_1 and x_2 are from the same continuous distribution. Therefore, if the result of the test indicates that the null hypothesis is not rejected, one can say that the set x_1 and set x_2 are from the same continuous distribution with a given significance level that is specified in the test. In this study, the significance level of 5% was used.

In this study, a set of two tests should be performed to tell if THM outperforms the traditional approach. For example, if the test comparing P^1_{Obs} and P^1_{THM} indicates that both variables are from the same continuous distributions and the test comparing P^1_{Obs} and P^1_{Trad} indicates that they are from the different distributions, the advantage of using

THM over the traditional approach to predict the maximum precipitation depth at hourly duration is proved. This set of tests was repeated for the 15 variables ($P^1, P^3, P^6, P^{12}, P^{24}$, $R^{50}, R^{60}, R^{70}, R^{80}, R^{90}$, $F^{50}, F^{60}, F^{70}, F^{80}, F^{90}$) to see how the performance of THM compares to the one of the traditional approaches based on long-term statistics.

4.6 Results

4.6.1 Reproduction of Monthly Maximum Rainfall Depth

Table 4-2 shows the proportion, among 104 gage-months, that THM and the traditional approach succeeded / failed in reproducing the distribution of the observed monthly maximum rainfall depths.

Table 4-2 Proportion that THM and the traditional approach succeeded / failed in reproducing the distribution of the monthly maximum rainfall depths.

Rainfall Duration	1hr	3hr	6hr	12hr	24hr
Proportion both approaches succeeded	0.125	0.2212	0.2308	0.2308	0.2115
Proportion only THM succeeded	0.3462	0.5	0.6058	0.625	0.5481
Proportion only traditional approach succeeded	0.0096	0	0.0192	0.0481	0.0769
Proportion both approach failed	0.5192	0.2788	0.1442	0.0962	0.1635

Overall, THM outperforms the traditional long-term statistics based approach in reproducing the distribution of monthly maxima rainfall depths. For only 3.1% of the

entire comparisons, the traditional approach succeeded in reproducing the distribution of the monthly maximum rainfall depths that THM cannot reproduce. Conversely, THM succeeded in reproducing the distribution of the monthly maximum rainfall depths for 73% of the entire gage-months. At 1, 3, 6, 12, and 24 hours of rainfall durations, THM succeeded 47%, 72%, 84%, 86%, and 76% of the times in reproducing the distribution of the monthly maximum rainfall depths. The same values for the traditional approach was 13%, 22%, 25%, 27%, and 29%, respectively. By comparing the success proportion of THM and the traditional approach, it can be noted that incorporating more statistical information about rainfall time series while keeping the fundamental model structure of Poisson cluster models can significantly enhance the performance in reproducing the extreme rainfall characteristics. However, the fact that extreme rainfall characteristics at fine time scales were not well reproduced even when THM approach was adopted indicates that more refining and tailoring are required on the fundamental structures of Poisson cluster models.

4.6.2 Reproduction of Peak Flows

Table 4-3 shows the proportion, among 104 gage-months, that THM and the traditional approach succeeded / failed in reproducing the distribution of the peak flows derived from the observed rainfall time series. Overall, THM shows the satisfactory performance with the success proportion of 98%, 97%, 93%, 92%, and 89% for the curve number of 50, 60, 70, 80, and 90, respectively. The same values for the traditional approach increased along with the increase of the curve number, which are 0%, 3%,

14%, 24%, and 24%, respectively. It is noteworthy that the success proportion of peak flow reproduction increased from the counterpart of monthly maximum rainfall depth reproduction when THM was used. The opposite was observed when the traditional approach was applied. This discrepancy can be caused by the following reasons: the antecedent soil moisture conditions before the large rainfall events of the synthetic rainfall time series that is generated by the traditional approach do not reflect the ones of the observed rainfall time series; the internal rainfall structure that is right before and after the peak rainfall depth are not close to what is observed in reality. The solution for the first cause was discussed by Cowpertwait [1991], as mentioned above, which derived an equation for the probability of an arbitrary interval of any chosen length being dry for Neyman-Scott rectangular pulse model. Considering that the MBLRP model that is used in this study do not consider this variable, applying the principle suggested by Cowpertwait [1991] will enhance the performance of the traditional approach in reproduction of the distribution of the peak flows. The solution for the second cause of the traditional approach will require further developments on the fundamental model structure, which will involve additional mathematical deliberation to the current approach of Poisson cluster rainfall modeling that is already complicated. Conversely, THM successfully reproduced the distribution of peak flows simply by providing more information about rainfall time series while keeping the fundamental structure of the current Poisson cluster models.

Table 4-3 Proportion that THM and the traditional approach succeeded / failed in reproducing the distribution of the monthly peak flows.

Curve Number	50	60	70	80	90
Proportion both approaches succeeded	0	0.0192	0.1048	0.2019	0.2019
Proportion only THM succeeded	0.9808	0.9519	0.8269	0.7212	0.6928
Proportion only traditional approach succeeded	0	0.0096	0.0385	0.0385	0.0385
Proportion both approach failed	0.0192	0.0192	0.0288	0.0385	0.0673

4.6.3 Reproduction of Runoff Depths

Table 4-4 shows the proportion, among 104 gage-months, that THM and the traditional approach succeeded / failed in reproducing the distribution of the runoff volumes derived from the observed rainfall time series. Overall, similar result was observed as the one corresponding to the reproduction of peak flow values. THM showed the success proportion of 98%, 97%, 92%, 92%, and 87% for the curve number of 50, 60, 70, 80, and 90, respectively. The same values for the traditional approach increased along with the increase of the curve number, which are 0%, 3%, 14%, 25%, and 26%, respectively.

Table 4-4 Proportion that THM and the traditional approach succeeded / failed in reproducing the distribution of the monthly runoff volumes.

Curve Number	50	60	70	80	90
Proportion both approaches succeeded	0	0.0192	0.1058	0.2019	0.2115
Proportion only THM succeeded	0.9808	0.9519	0.8173	0.7212	0.6538
Proportion only traditional approach succeeded	0	0.0096	0.0385	0.0481	0.0481
Proportion both approach failed	0.0192	0.0192	0.0385	0.0288	0.0865

4.6.4 Relationship with Mean Rainfall Depth

Figure 4-9 shows how the performance of the model is related to mean rainfall depth. The x-axis of the plot shows if THM and the traditional approach succeeded in reproducing 1-hour duration monthly maximum rainfall depth at a specific gage-month. The y-axis of the plot shows the mean hourly rainfall of the corresponding gage-month.

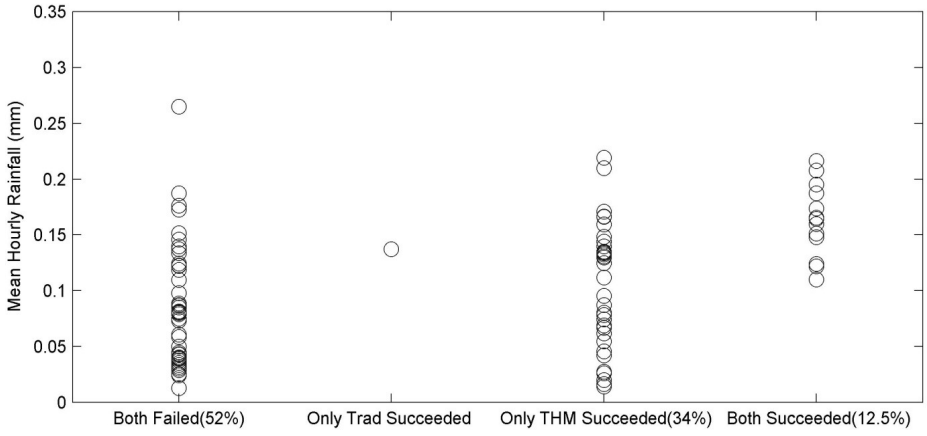


Figure 4-9 Relationship between the success proportion of reproducing the distribution of the 1 hour duration monthly maximum rainfall depths and mean rainfall depth.

It can be noted that the traditional approach does not succeed in reproducing the distribution of the monthly maximum rainfall depth when mean rainfall is less than 0.1mm/hr. Also, the concentration of the circles in the lower portion of the first column of the plot indicates that both approaches of rainfall generation are likely to fail when mean rainfall is small. The circles that are evenly distributed in the third column of the plot indicate that the performance of the THM is not highly influenced by the mean rainfall depth.

Figure 4-10 shows how the success proportion of peak flows (upper) and runoff depths (lower) relate to mean rainfall depth. The evenly distributed circles in the third column of both plots indicate that THM succeeded in reproducing peak flows and runoff depths regardless of mean rainfall depths. In the mean time, the success of the traditional approach in reproducing peak flows and runoff depths was limited by the narrow range of the mean rainfall depth distribution.

4.7 Conclusion and Discussion

The result of the hydrologic modeling study based on 150 rain gages across the United States suggests that the variability of watershed response characteristics calculated based on the traditional framework of Poisson cluster rainfall models are systematically underestimated. The measure of underestimation was 40% and 27% for monthly runoff depth and monthly peak flow, respectively. This result indicates that the runoffs and peak flows of a given month derived from the synthetic rainfall time series can miss the upper and lower extremes that are observed in reality, which typically leads to water resources related problems in practice.

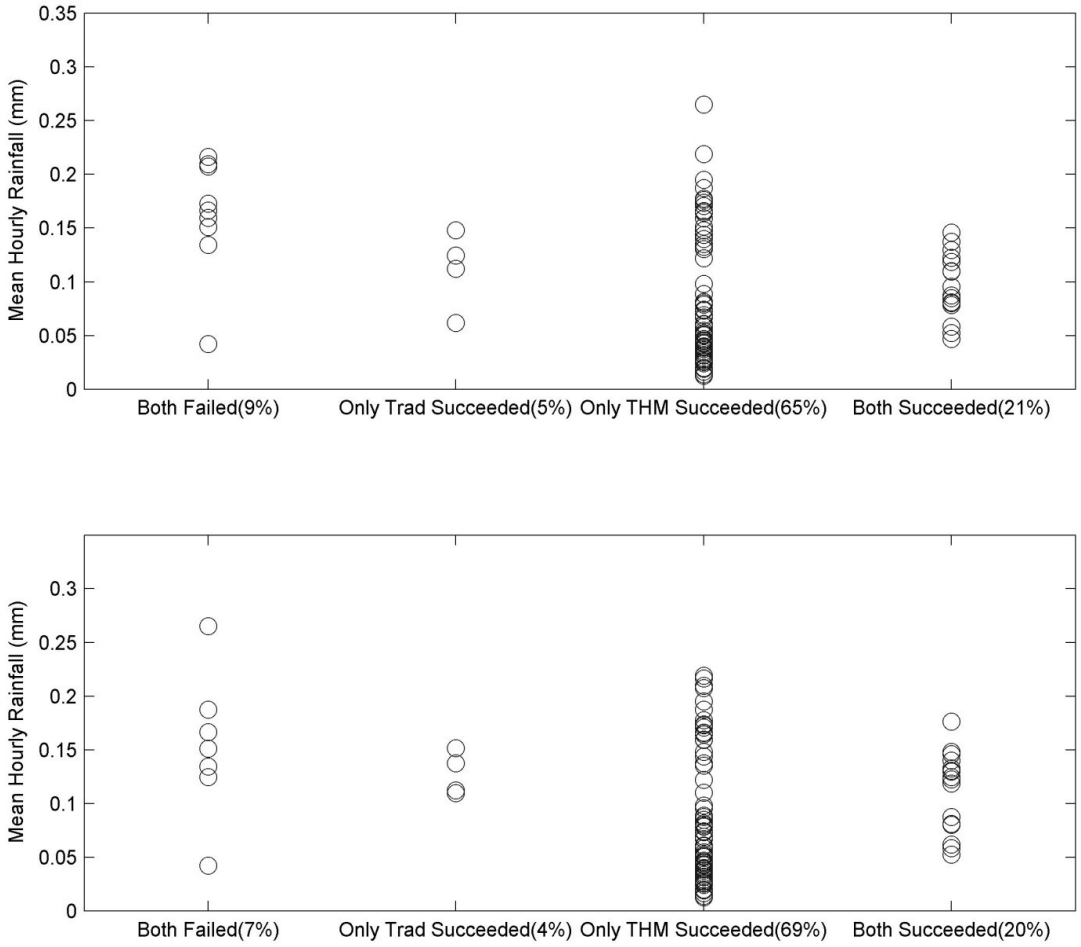


Figure 4-10 Relationship between the success proportion of reproducing the distribution of the peak flows and mean rainfall depth (top) and the relationship between the success proportion of reproducing the distribution of the runoff depths and mean rainfall depth (bottom)

To resolve the identified problems, a novel approach based on Modified Bartlett-Lewis Rectangular Pulse model (MBLRPM) [*Rodriguez-Iturbe et al.*, 1988] that can account for the year-to-year variability of the rainfall statistics is suggested. The suggested approach incorporates this variability by simulating the short-term rainfall statistics based on the correlation between the observed rainfall statistics. Then, the model generates the rainfall time series using MBLRPM based on the short-term rainfall

statistics simulated in the first part of the model. The suggested approach was termed "The Hybrid Model – THM" because it combines the process of generating rainfall statistics and generating rainfall time series.

The result of validation study based on 104 gages across the United States suggests that THM outperforms the traditional framework of Poisson cluster rainfall modeling. In reproduction of the monthly maximum rainfall depths with the duration of 1, 3, 6, 12, and 24 hours at 104 gages, THM succeeded 47%, 72%, 84%, 86%, and 76%, while the traditional approach succeeded 13%, 22%, 25%, 27%, and 29%, respectively. In reproduction of monthly peak flows at the virtual watershed with curve number of 50, 60, 70, 80, and 90, THM succeeded 98%, 97%, 93%, 92%, and 89% while the traditional approach succeeded 0%, 3%, 14%, 24%, and 24%, respectively. In reproduction of monthly runoff depths at the virtual watershed with curve number of 50, 60, 70, 80, and 90, THM succeeded 98%, 97%, 92%, 92%, and 87% while the traditional approach succeeded 0%, 3%, 14%, 25%, and 26%, respectively.

Of course, additional endeavors are crucial to clarify that THM is a superior to the traditional approach in rainfall modeling. Firstly, it should be noted that this study adopted rather simple methodology [*SCS*, 1972] to derive watershed responses from rainfall time series. Considering that this methodology is the extremely simplified version of the complex hydrologic phenomena occurring in nature, a more detailed analysis based on more realistic hydrologic models will signify the result of the present study. However, it should be noted that THM outperformed the traditional approach in reproducing the monthly maximum rainfall depths, which are not filtered through

hydrologic models. Secondly, THM requires significantly more model parameters and computing powers than the traditional approach does; THM requires 40 parameters to simulate rainfall statistics, which is significantly greater than what is required by the traditional approach; In THM, parameters of MBLRPM should be separately estimated for every single month on which rainfall time series is synthesized. Also, the additional selection process of THM (selection of a monthly rainfall time series out of 20 pre-generated ones) demands more time and computing power compared to the traditional approach. Lastly, in the light of the fact that the fundamental purpose of most stochastic rainfall models is to generate the rainfall time series at ungaged locations, regionalization of the model is crucial. Regionalization of 40 parameters compared to 5 to 6 of the traditional approach will require significantly more efforts.

5. CONCLUSION

In the light of the fact that one of the fundamental purposes of a stochastic rainfall model is to synthesize the rainfall at ungaged locations and that no regionalization study has been performed in the United States on Poisson cluster rainfall models with acceptable and validated accuracy. The parameter maps of the MBLRPM that is suggested by this study will tremendously improve the applicability of the model in the United States. Thus, this study will benefit many hydrologists and water resources engineers aiming to assess the risks related to hydrologic phenomena occurring in the contiguous United States. Also, the maps of the MBLRP compartments will provide a useful tool for hydrologists and meteorologists to intuitively understand the regional and seasonal rainfall characteristics.

Most of the former studies on Poisson cluster rainfall models assessed the performance of the model focusing on how well the rainfall statistics are reproduced. The major contribution of this study is that it identified another benchmark of performance measure focusing on watershed response characteristics and that it quantitatively estimated which statistics exerts more influence in the generation of watershed response characteristics such as mean monthly runoff and mean monthly peak flow. The outcome of the study can be applied in the calibration of the Poisson cluster models in weighing rainfall statistics differently according to the purpose of hydrologic modeling study. However, it is important to stress that the specific weigh values used in this study should not be considered as recommended values for other applications. The value of the weights depends on the hydrologic process being modeled. Also, it should

be kept in mind that the results were derived using rather simple hydrologic models (SCS curve number method, and SCS unit hydrograph method). The detailed analysis using more realistic hydrologic models will solidify the outcome of this study.

The traditional approach of Poisson cluster rainfall modeling is based on the statistics that is calculated over the entire recording period. This study suggests that the year-to-year variability residing in those statistics is significant and identifies the problems of the current approaches in which this year-to-year variability is neglected. The newly suggested approach (THM) incorporates this variability by simulating the short-term rainfall statistics based on the correlation between the observed rainfall statistics. Then, the model generates the rainfall time series using the modified Bartlett-Lewis rectangular pulse model (MBLRPM) based on the simulated short-term rainfall statistics. The result of validation study based on 104 gages across the United States suggests significantly improved performance of THM over the traditional approach in reproducing monthly maximum rainfall depth, monthly maximum flow peak, and monthly runoff depth. However, it should be noted that this study adopted rather simple methodology [*SCS*, 1972] to derive watershed responses from rainfall time series, so a more detailed analysis based on more realistic hydrologic models might be necessary. Also, it should be noted that THM requires significantly more model parameters and computing powers than the traditional approach does. Lastly, whether the 40 model parameters can be regionalized or not will determine the further applicability of the model in ungaged basins.

REFERENCES

Arnaud, P., and J. Lavabre (1999), Using a stochastic model for generating hourly hyetographs to study extreme rainfalls, *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, *44*(3), 433-446.

Arnaud, P., and J. Lavabre (2002), Coupled rainfall model and discharge model for flood frequency estimation, *Water Resources Research*, *38*(6), 1075, doi:10.1029/2001WR000474

Aronica, G. T., and A. Candela (2007), Derivation of flood frequency curves in poorly gauged Mediterranean catchments using a simple stochastic hydrological rainfall-runoff model, *Journal of Hydrology*, *347*(1-2), 132-142.

Bae, D. H., I. W. Jung, and H. Chang (2008), Potential changes in Korean water resources estimated by high-resolution climate simulation, *Climate Research*, *35*(3), 213-226.

Benda, L., and T. Dunne (1997), Stochastic forcing of sediment supply to channel networks from landsliding and debris flow, *Water Resources Research*, *33*(12), 2849-2863.

Bo, Z., S. Islam, and E. A. B. Eltahir (1994), Aggregation-disaggregation properties of a stochastic rainfall model, *Water Resources Research*, *30*(12), 3423-3435.

Botter, G., T. Settin, M. Marani, and A. Rinaldo (2006), A stochastic model of nitrate transport and cycling at basin scale, *Water Resources Research*, *42*(4), W04415.

Bouvier, C., L. Cisneros, R. Dominguez, J. P. Laborde, and T. Lebel (2003), Generating rainfall fields using principal components (PC) decomposition of the covariance matrix: a case study in Mexico City, *Journal of Hydrology*, *278*(1-4), 107-120.

Brath, A., A. Montanari, and G. Moretti (2006), Assessing the effect on flood frequency of land use change via hydrological simulation (with uncertainty), *Journal of Hydrology*, *324*(1-4), 141-153.

Burton, A., C. G. Kilsby, H. J. Fowler, P. S. P. Cowpertwait, and P. E. O'Connell (2008), RainSim: A spatial-temporal stochastic rainfall modelling system, *Environmental Modelling & Software*, *23*(12), 1356-1369.

Calenda, G., and F. Napolitano (1999), Parameter estimation of Neyman-Scott processes for temporal point rainfall simulation, *Journal of Hydrology*, *225*(1-2), 45-66.

Chandler, R. E. (1997), A spectral method for estimating parameters in rainfall models, *Bernoulli*, *3*(3), 301-322.

Chatterjee, S., A. S. Hadi, and Hoboken (2006), *Regression Analysis by Example (4th ed.)*, Wiley-Interscience, Hoboken, NJ.

Cho, H., F. Olivera, S. D. Guikema, and D. Kim, Enhanced speciation in particle swarm optimization for multi-modal problems, *European Journal of Operational Research* (In Review).

Choi, J., S. A. Socolofsky, and F. Olivera (2008), Hourly disaggregation of daily rainfall in Texas using measured hourly precipitation at other locations, *Journal of Hydrologic Engineering*, *13*(6), 476-487.

Cowpertwait, P., V. Isham, and C. Onof (2007), Point process models of rainfall: developments for fine-scale structure, *Proceedings of the Royal Society A-Mathematical Physical and Engineering Sciences*, *463*(2086), 2569-2587.

Cowpertwait, P. S. P. (1991), Further developments of the Neyman-Scott clustered point process for modeling rainfall, *Water Resources Research*, *27*(7), 1431-1438.

Cowpertwait, P. S. P. (1994), A generalized point process model for rainfall, *Proceedings of the Royal Society of London Series A-Mathematical Physical and Engineering Sciences*, *447*(1929), 23-37.

Cowpertwait, P. S. P. (1995), A generalized spatial-temporal model of rainfall based on a clustered point process, *Proceedings of the Royal Society of London Series A-Mathematical and Physical Sciences*, *450*(1938), 163-175.

Cowpertwait, P. S. P., P. E. O'Connell, A. V. Metclafe, and J. A. Mawdsley (1996a), Stochastic point process modelling of rainfall, I, Single-site fitting and validation, *Journal of Hydrology*, *175*(1-4), 17-46.

Cowpertwait, P. S. P., P. E. O'Connell, A. V. Metclafe, and J. A. Mawdsley (1996b), Stochastic point process modeling of rainfall, II, Regionalization and disaggregation, *Journal of Hydrology*, *175*(1-4), 47-65.

Cowpertwait, P. S. P., and P. E. O'Connell (1997), A regionalised Neyman-Scott model of rainfall with convective and stratiform cells, *Hydrology and Earth System Sciences*, *1*, 71-80.

Cowpertwait, P. S. P. (1998), A Poisson-cluster model of rainfall: high-order moments and extreme values, *Royal Society of London Series A-Mathematical Physical and Engineering Sciences*, *454*(1971), 885-898.

Dibike, Y. B., and P. Coulibaly (2005), Hydrologic impact of climate change in the Saguenay watershed: comparison of downscaling methods and hydrologic models, *Journal of Hydrology*, *307*(1-4), 145-163.

Econopouly, T. W., D. R. Davis, and D. A. Woolhiser (1989), Parameter transferability for a daily rainfall disaggregation model, *Journal of Hydrology*, *118*(1-4), 209-228.

Favre, A. C., A. Musy, and S. Morgenthaler (2004), Unbiased parameter estimation of the Neyman-Scott model for rainfall simulation with related confidence interval, *Journal of Hydrology*, *286*(1-4), 168-178.

Fernandez-Galvez, J., E. Barahona, A. Iriarte, and M. D. Mingorance (2007), A simple methodology for the evaluation of groundwater pollution risks, *Science of the Total Environment*, *378*(1-2), 67-70.

Fowler, H. J., C. G. Kilsby, and P. E. O'Connell (2000), A stochastic rainfall model for the assessment of regional water resource systems under changed climatic conditions, *Hydrology and Earth System Sciences*, *4*(2), 263-282.

Fowler, H. J., C. G. Kilsby, and P. E. O'Connell (2003), Modeling the impacts of climatic change and variability on the reliability, resilience, and vulnerability of a water resource system, *Water Resources Research*, *39*(8), 1222, doi:10.1029/2002WR001778.

Gabellani, S., G. Boni, L. Ferraris, J. von Hardenberg, and A. Provenzale (2007), Propagation of uncertainty from rainfall to runoff: a case study with a stochastic rainfall generator, *Advances in Water Resources*, *30*(10), 2061-2071.

Glasbey, C. A., G. Cooper, and M. B. McGehan (1995), Disaggregation of daily rainfall by conditional simulation from a point-process model, *Journal of Hydrology*, 165(1-4), 1-9.

Green, W. H., and G. A. Ampt (1911), Studies on soil physics Part I - The flow of air and water through soils, *Journal of Agricultural Science*, *4*, 1-24.

Gyasi-Agyei, Y., and G. Willgoose (1997), A hybrid model for point process modelling, *Water Resources Research*, *33*(7), 1699-1706.

Gyasi-Agyei, Y. (1999), Identification of regional parameters of a stochastic model for rainfall disaggregation, *Journal of Hydrology*, *223*(3-4), 148-163.

Gyasi-Agyei, Y., and S. M. P. B. Mahbub (2007), A stochastic model for daily rainfall disaggregation into fine time scale for a large region, *Journal of Hydrology*, *347*(3-4), 358-370.

Haberlandt, U., A. D. E. von Eschenbach, and I. Buchwald (2008), A space-time hybrid hourly rainfall model for derived flood frequency analysis, *Hydrology and Earth System Sciences*, *12*(6), 1353-1367.

Harrold, T. I., A. Sharma, and S. J. Sheather (2003), A nonparametric model for stochastic generation of daily rainfall occurrence, *Water Resources Research*, *39*(10), 1300, doi:10.1029/2003WR002182, 2003

Hawk, K., and P. S. Eagleson (1992), Climatology of station storm rainfall in the continental United States: Parameters of the Bartlett Lewis and Poisson rectangular pulses model, Rep. 336, Ralph M. Parsons Lab, Mass. Inst. of Technol. Cambridge, MA, 1992

Hershenhorn, J., and D. A. Woolhiser (1987), Disaggregation of daily rainfall, *Journal of Hydrology*, *95*(3-4), 299-322.

Hoerl, A. E., and R. W. Kennard (1970), Ridge regession - Applications to nonorthogonal problems, *Technometrics*, *12*(1), 69-82.

Hreiche, A., W. Najem, and C. Bocquillon (2007), Hydrological impact simulations of climate change on Lebanese coastal rivers, *Hydrological Sciences*, *52*(6), 1119-1133.

Iida, T. (2004), Theoretical research on the relationship between return period of rainfall and shallow landslides, *Hydrological Processes*, *18*(4), 739-756.

Isham, S., D. Entekhabi, and R. L. Bras (1990), Parameter estimation and sensitivity analysis for the modified Bartlett-Lewis rectangular pulses model of rainfall, *Journal of Geophysical Research*, *95*(D3), 2093-2100.

James, W. P., J. Warinner, and M. Reedy (1992), Application of the Green-Ampt infiltration equation to watershed modeling, *Water Resources Bulletin*, *28*(3), 623-635.

Journel, A. G., and C. J. Huijbregts (1978), *Mining Geostatistics*, Academic Press, London.

Kakou, A. (1998), A point process model for rianfall with dependent duration and intensity, Research Report n = 196, Dept. of Statistical Science, University College, London.

Kang, B., and J. A. Ramirez (2007), Response of streamflow to weather variability under climate change in the Colorado Rockies, *Journal of Hydrologic Engineering*, *12*(1), 63-72.

Kavvas, M. L., and J. W. Delleur (1975), The stochastic and chronologic structure of rainfall sequences - application to Indiana, Tech. Rep. 57, Water Resources Research Center, Purdue Univ., West Lafayette, IN.

Khaliq, M., and C. Cunnane (1996), Modelling point rainfall occrrences with the modified Bartlett-Lewis rectangular pulses model, *Journal of Hydrology*, *180*, 109-138.

Kilsby, C. G., P. D. Jones, A. Burton, A. C. Ford, H. J. Fowler, C. Harpham, P. James, A. Smith, and R. L. Wilby (2007), A daily weather generator for use in climate change studies, *Environmental Modelling & Software*, *22*(12), 1705-1719.

Koutsoyiannis, D., and C. Onof (2001), Rainfall disaggregation using adjusting procedures on a Poisson cluster model, *Journal of Hydrology*, *246*(1-4), 109-122.

Laio, F., S. Tamea, L. Ridolfi, P. D'Odorico, and I. Rodriguez-Iturbe (2009), Ecohydrology of groundwater-dependent ecosystems: 1. Stochastic water table dynamics, *Water Resources Research*, *45*, W05419.

Leander, R., A. Buishand, P. Aalders, and M. De Wit (2005), Estimation of extreme floods of the River Meuse using a stochastic weather generator and a rainfall-runoff model, *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, *50*(6), 1089-1103.

Lettenmaier, D. P., and D. P. Sheer (1991), Climatic sensitivity of California water-resources, *Journal of Water Resources Planning and Management-Asce*, *117*(1), 108-125.

Marani, M., G. Grossi, F. Napolitano, and M. Wallace (1997), Forcing, intermittency, and land surface hydrologic partitioning, *Water Resources Research*, *33*(1), 167-175.

Marani, M. (2003), On the correlation structure of continuous and discrete point rainfall, *Water Resources Research*, *39*(5), 1128, doi: 10.1029/2002WR001456, 2003

Marani, M., and S. Zanetti (2007), Downscaling rainfall temporal variability, *Water Resources Research*, *43*, W09415, doi: 10.1029/2006WR005505

McCuen, R. H., S. L. Wong, and W. J. Rawls (1984), Estimating urban time of concentration, *Journal of Hydraulic Engineering*, *110*(7), 887-904.

McGrath, G. S., C. Hinz, and M. Sivapalan (2008), Modelling the impact of within-storm variability of rainfall on the loading of solutes to preferential flow pathways, *European Journal of Soil Science*, *2009*(59), 24-33.

McMillan, H. K., and J. Brasington (2008), End-to-end flood risk assessment: a coupled model cascade with uncertainty estimation, *Water Resources Research*, *44*, W03419, doi: 10.1029/2007WR005995

Mehrotra, R., and A. Sharma (2007), A semi-parametric model for stochastic generation of multi-site daily rainfall exhibiting low-frequency variability, *Journal of Hydrology*, *335*(1-2), 180-193.

Michel, C., V. Andreassian, and C. Perrin (2005), Soil Conservation Service Curve Number method: How to mend a wrong soil moisture accounting procedure? *Water Resources Research*, *41*, W02011, doi: 10.1029/2004WR003191

Mohan, S., and P. K. Sahoo (2008a), Stochastic simulation of droughts. Part 2: regional droughts, *Hydrological Processes*, *22*(6), 863-872.

Mohan, S., and P. K. Sahoo (2008b), Stochastic simulation of droughts. Part 1: point droughts, *Hydrological Processes*, *22*(6), 854-862.

Muneepeerakul, C. P., F. Miralles-Wilhelm, S. Tamea, A. Rinaldo, and I. Rodriguez-Iturbe (2008), Coupled hydrologic and vegetation dynamics in wetland ecosystems, *Water Resources Research*, *44*(7), 15.

NCDC (2008), NCDC: National Climate Data Directory, http://www.ncdc.noaa.gov/oa/climate/climatedata.html#hourly, *2008*(1/16).

Nearing, M. A., B. Y. Liu, L. M. Risse, and X. Zhang (1996), Curve numbers and Green-Ampt effective hydraulic conductivities, *Water Resources Bulletin*, *32*(1), 125-136.

Nolan, B. T., I. G. Dubus, N. Surdyk, H. J. Fowler, A. Burton, J. M. Hollis, S. Reichenberger, and N. J. Jarvis (2008), Identification of key climatic factors regulating the transport of pesticides in leaching and to tile drains, *Pest Management Science*, *64*(9), 933-944.

Nunes, J. P., J. Seixas, and N. R. Pacheco (2008), Vulnerability of water resources, vegetation productivity and soil erosion to climate change in Mediterranean watersheds, *Hydrological Processes*, *22*(16), 3115-3134.

Olsson, J., and P. Burlando (2002), Reproduction of temporal scaling by a rectangular pulses rainfall model, *Hydrological Processes*, *16*(3), 611-630.

Onof, C., and H. S. Wheater (1994a), Improved fitting of the Bartlett-Lewis rectangular pulse model for hourly rainfall, *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, *39*(6), 663-680.

Onof, C., and H. S. Wheater (1994b), Improvements to the modeling of British rainfall using a modified random parameter Bartlett-Lewis rectangular pulse model, *Journal of Hydrology*, *157*(1-4), 177-195.

Onof, C., D. Faulkner, and H. S. Wheater (1996a), Design rainfall modelling in the Thames catchment, *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, *41*(5), 715-733.

Onof, C., P. Northrop, H. S. Wheater, and V. Isham (1996b), Spatiotemporal storm structure and scaling property analysis for modeling, *Journal of Geophysical Research-Atmospheres*, *101*(D21), 26415-26425.

Onof, C., R. E. Chandler, A. Kakou, P. Northrop, H. S. Wheater, and V. Isham (2000), Rainfall modelling using Poisson-cluster processes: a review of developments, *Stochastic Environmental Research and Risk Assement*, *14*(6), 384-411.

Ozturk, A. (1981), On the study of a probability-distribution for precipitation totals, *Journal of Applied Meteorology*, *20*(12), 1499-1505.

Pearson, K. (1901), On lines and planes of closest fit to systems of points in space, *Philosophical Magazine*, *2*(7-12), 559-572.

Ponce, V. M., and R. H. Hawkins (1996), Runoff curve number: Has it reached maturity? *Journal of Hydrologic Engineering*, *1*(1), 11-19.

Porporato, A., E. Daly, and I. Rodriguez-Iturbe (2004), Soil water balance and ecosystem response to climate change, *American Naturalist*, *164*(5), 625-632.

Richards, L. A. (1931), Capillary conduction of liquids through porous mediums, *Physics-A Journal of General and Applied Physics*, *1*(1), 318-333.

Rodriguez-Iturbe, I., D. R. Cox, and V. Isham (1987), Some models for rainfall based on stochastic point processes, *Proceedings of Royal Society of London Series A-Mathematical Physical and Engineering Sciences*, *410*(1839), 269-288.

Rodriguez-Iturbe, I., D. R. Cox, and V. Isham (1988), A point process model for rainfall: further developments, *Proceedings of Royal Society of London Series A-Mathematical Physical and Engineering Sciences*, *417*(1853), 283-298.

SCS (1972), Soil Conservation Service (SCS), *National Engineering Handbook - Section 4* - Hydrology, Dept. of Agriculture, Washington DC.

Semenov, M. A., and E. M. Barrow (1997), Use of a stochastic weather generator in the development of climate change scenarios, *Climatic Change*, *35*(4), 397-414.

Shamir, E., D. M. Meko, N. E. Graham, and K. P. Georgakakos (2007), Hydrologic model framework for water resources planning in the Santa Cruz River, southern Arizona, *Journal of the American Water Resources Association*, *43*(5), 1155-1170.

Socolofsky, S., E. Adams, and D. Entekhabi (2001), Disaggregation of daily rainfall for continuous watershed modeling, *Journal of Hydrologic Engineering*, *6*(4), 300-309.

van Werkhoven, K., T. Wagener, P. Reed, and Y. Tang (2008), Rainfall characteristics define the value of streamflow observations for distributed watershed model identification, *Geophysical Research Letters*, *35*(11), L11403.

Vanmullem, J. A. (1991), Runoff and peak discharges using green-ampt infiltration-model, *Journal of Hydraulic Engineering*, *117*(3), 354-370.

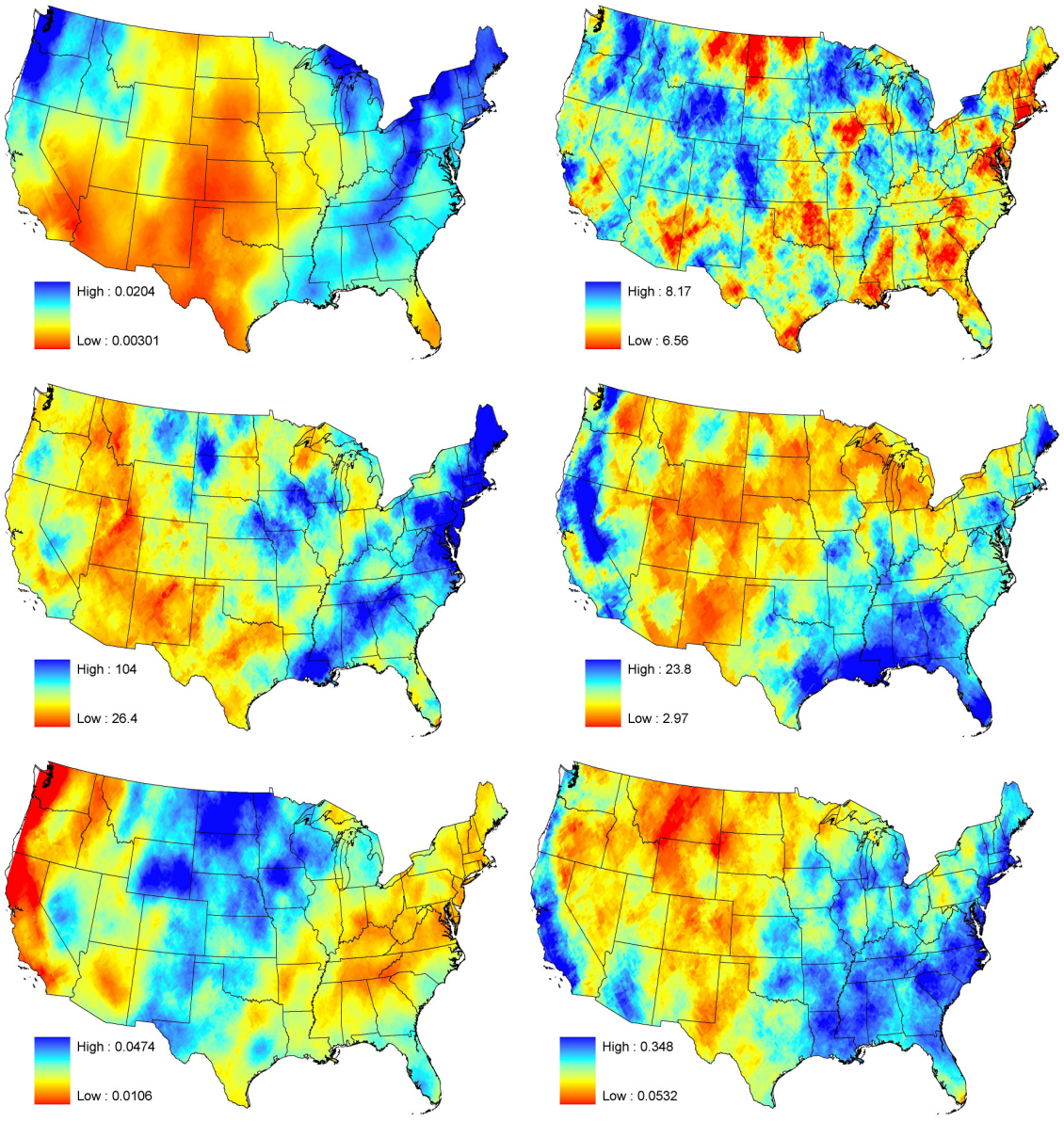
Verhoest, N., P. A. Troch, and F. P. De Troch (1997), On the applicability of Bartlett-Lewis rectangular pulses models in the modeling of design storms at a point, *Journal of Hydrology*, *202*(1-4), 108-120.

Wheater, H. S., R. E. Chandler, C. J. Onof, V. S. Isham, E. Bellone, C. Yang, D. Lekkas, G. Lourmas, and M. L. Segond (2005), Spatial-temporal rainfall modelling for flood risk estimation, *Stochastic Environmental Research and Risk Assessment*, *19*(6), 403-416.

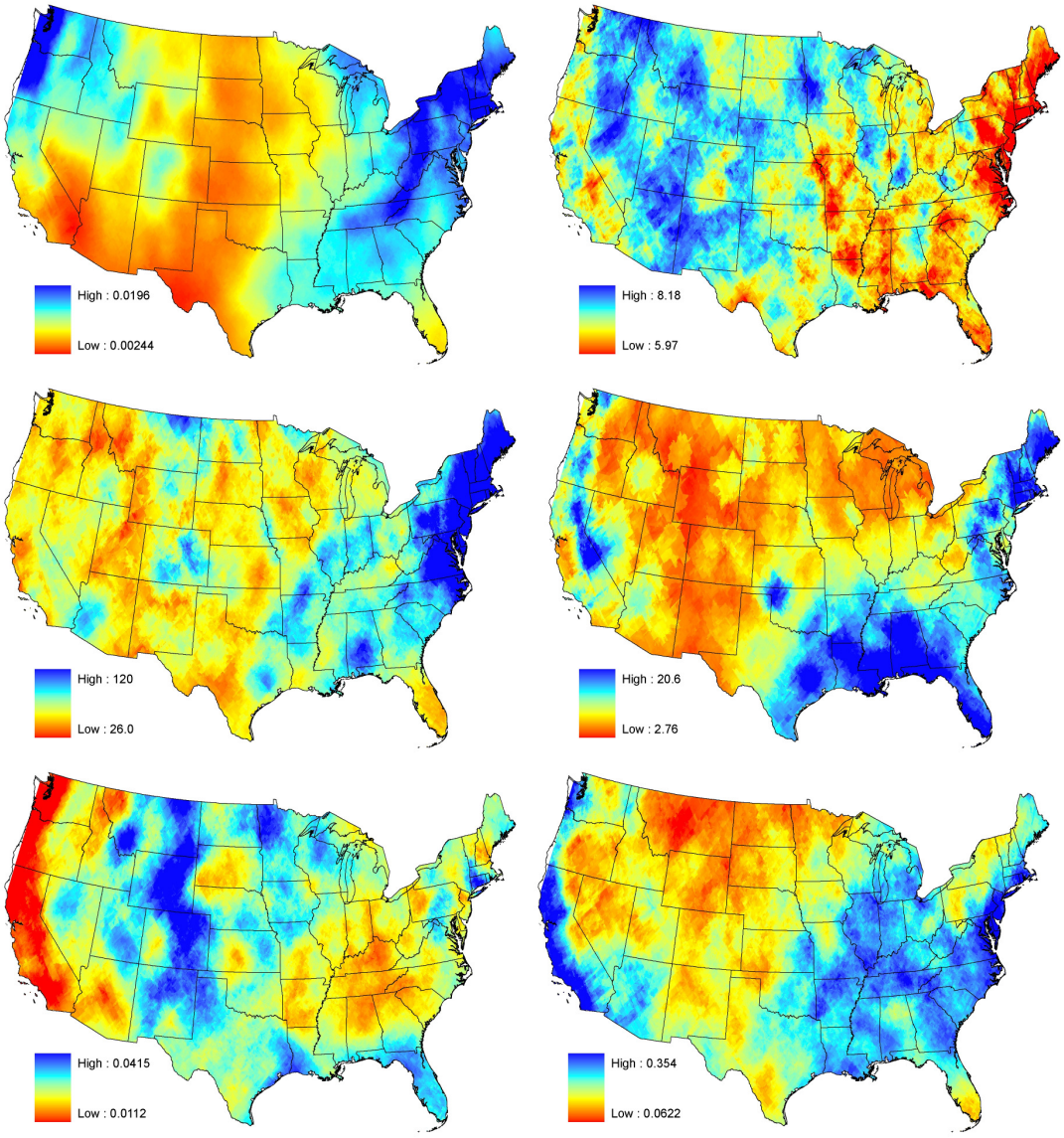
Wilby, R. L. (1994), Stochastic weather type simulation for regional climate-change impact assessment, *Water Resources Research*, *30*(12), 3395-3403.

Yoo, C., D. Kim, T. W. Kim, and K. N. Hwang (2008), Quantification of drought using a rectangular pulses Poisson process model, *Journal of Hydrology*, *355*(1-4), 34-48.

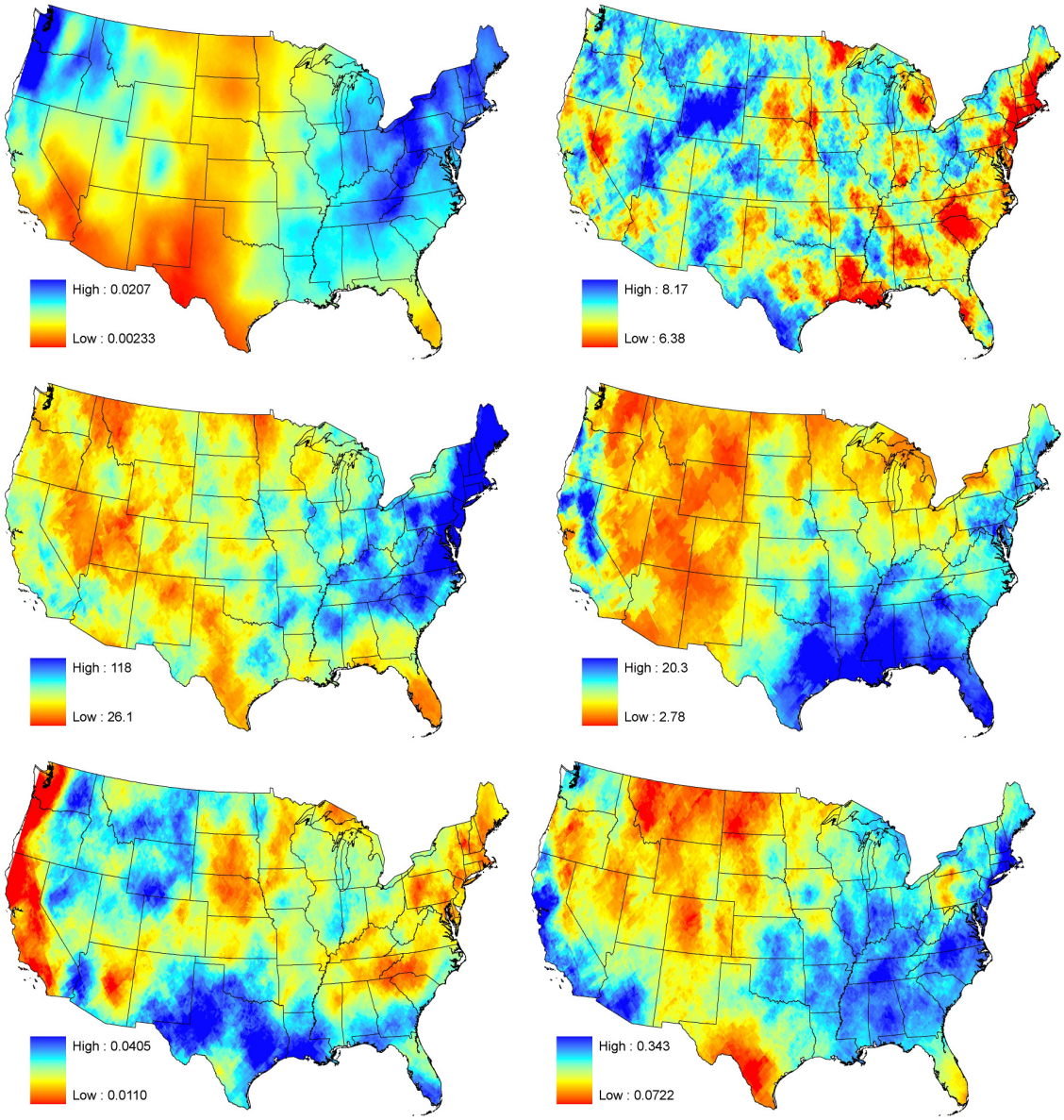
APPENDIX A. PARAMETERS OF MBLRPM FOR 12 CALENDAR MONTHS



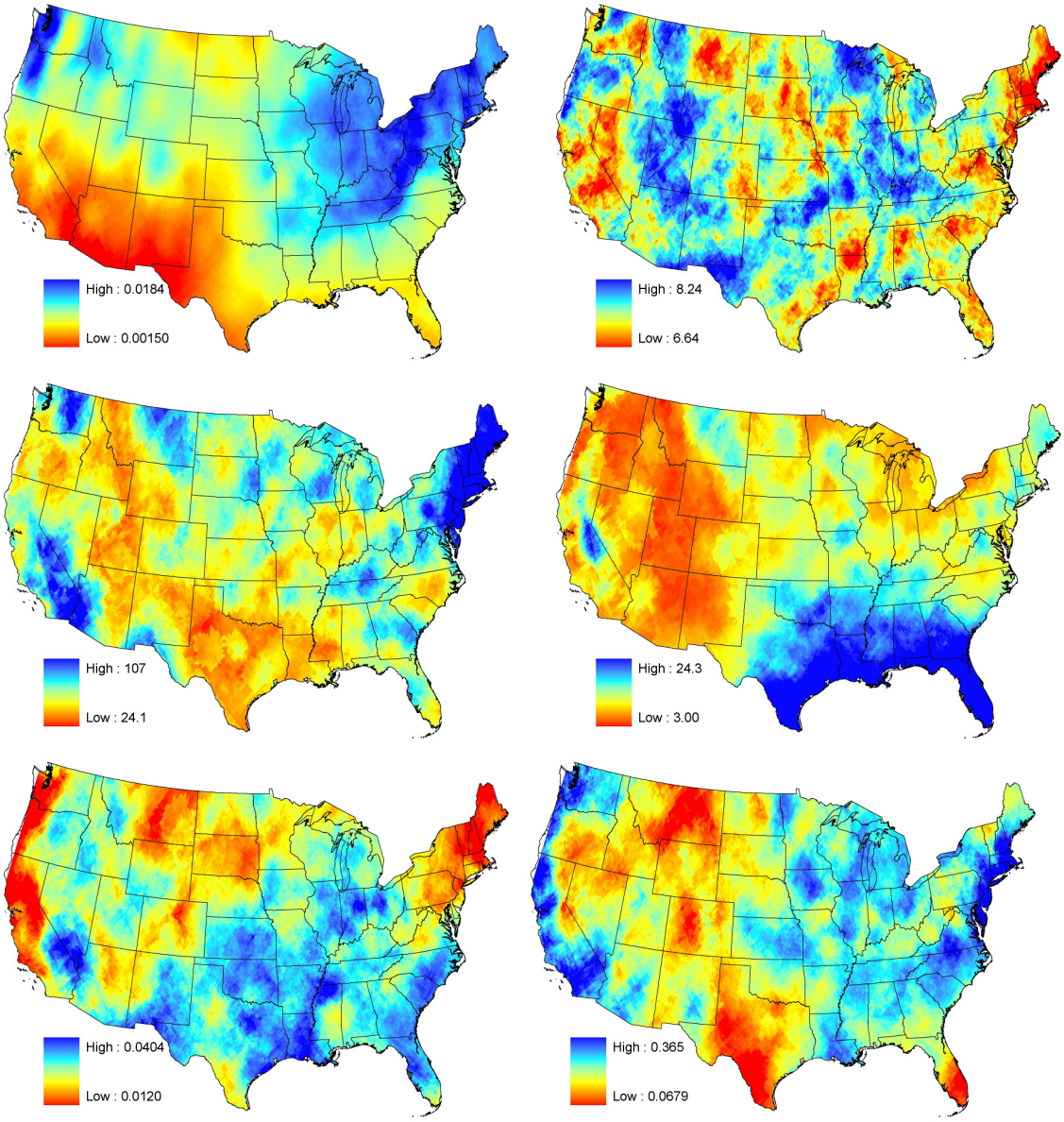
MBLRP model parameters for the month of January. From the top to bottom and left to right, λ (1/hr), ν (hr), α, μ (mm/hr), ϕ and κ.



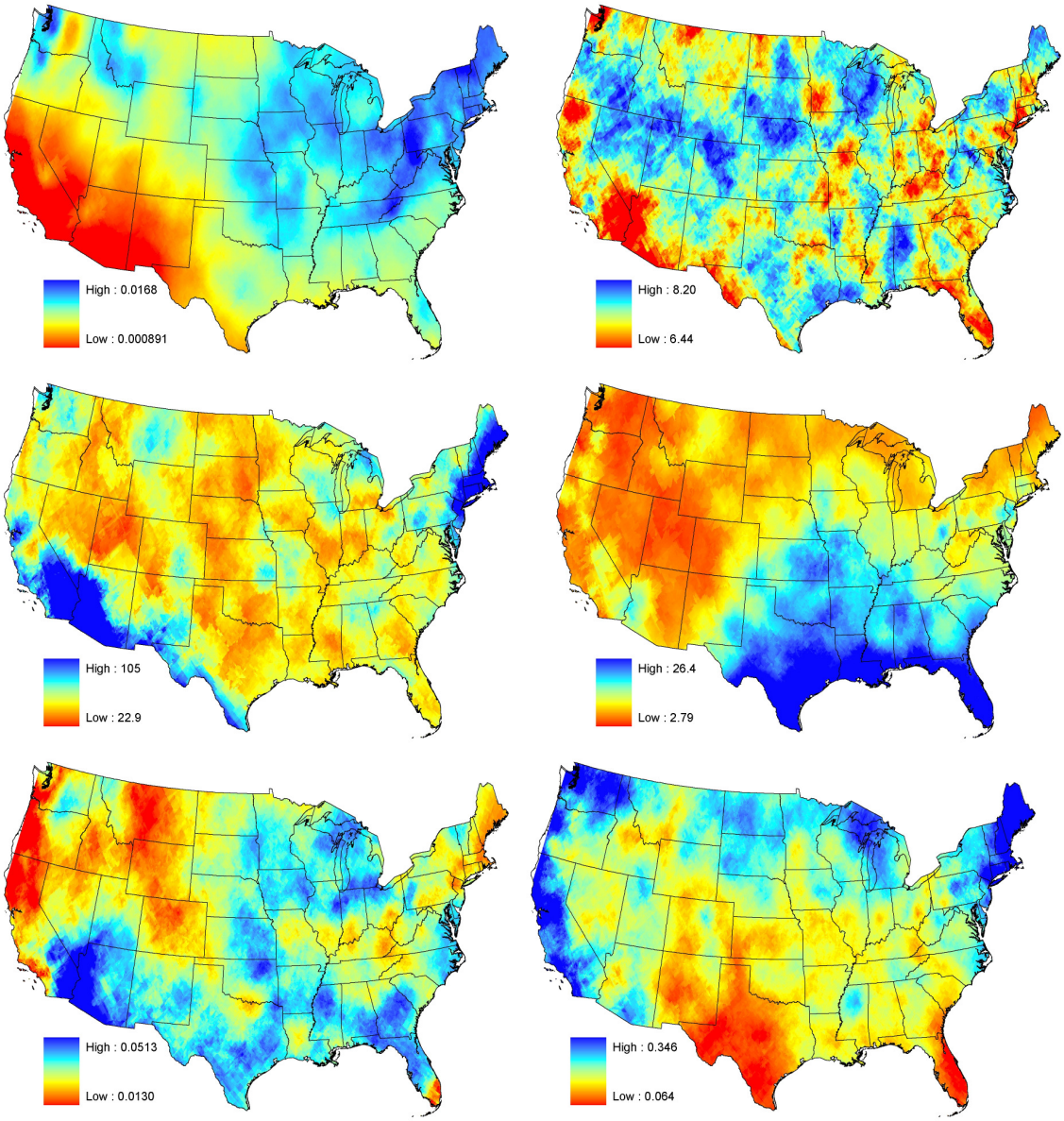
MBLRP model parameters for the month of February. From the top to bottom and left to right, λ (1/hr), ν (hr), α, μ (mm/hr), ϕ and κ.



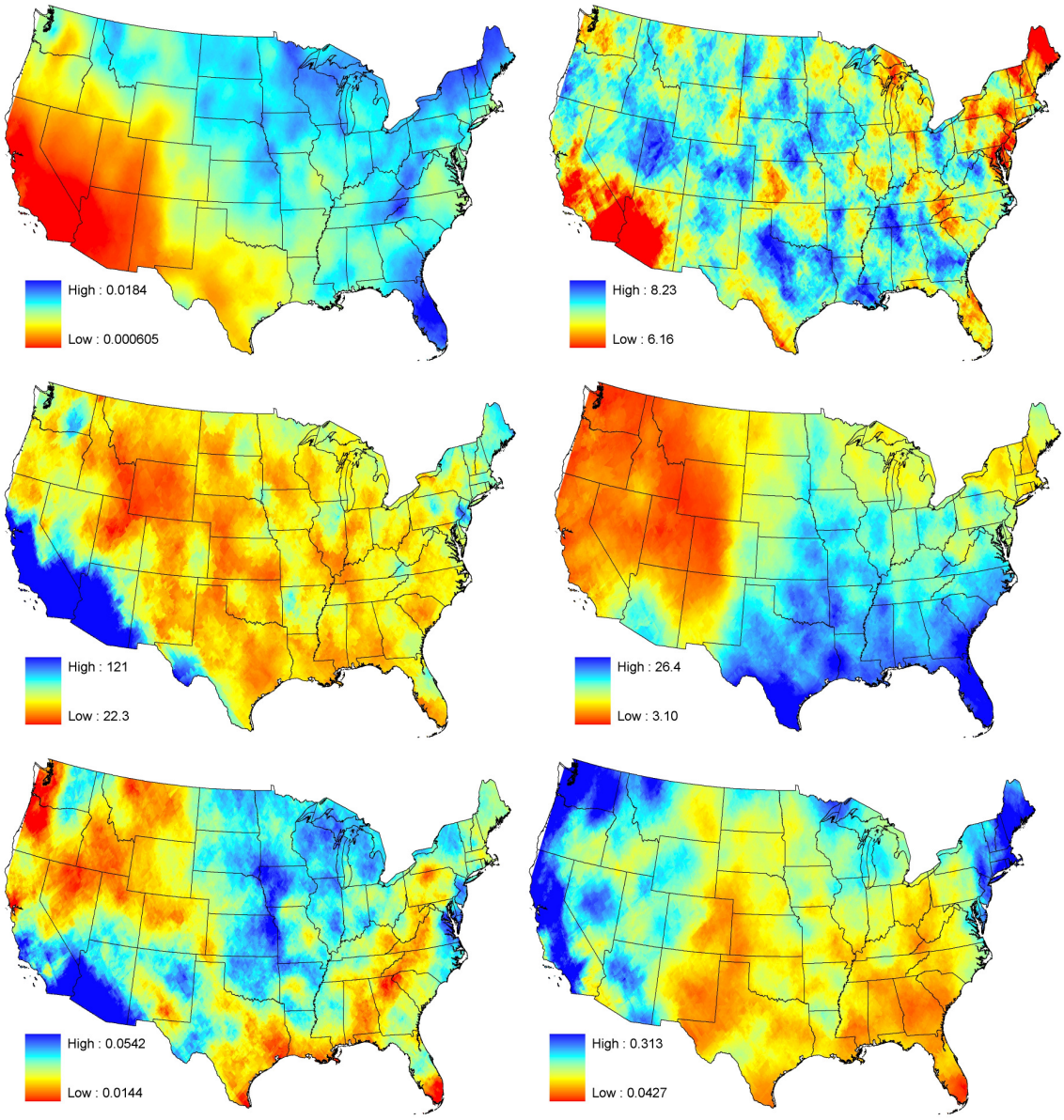
MBLRP model parameters for the month of March. From the top to bottom and left to right, λ (1/hr), ν (hr), α, μ (mm/hr), ϕ and κ.



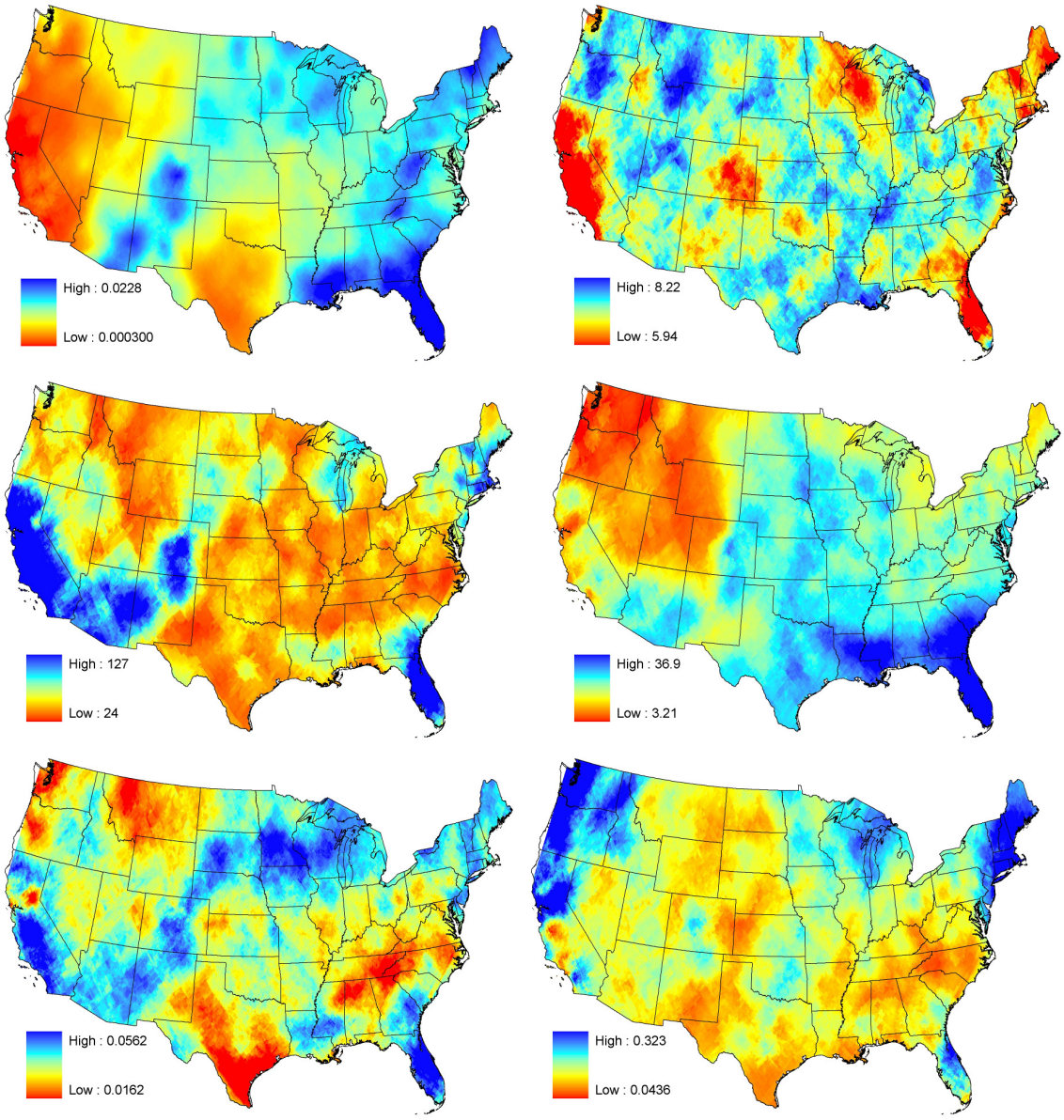
MBLRP model parameters for the month of April. From the top to bottom and left to right, λ (1/hr), ν (hr), α, μ (mm/hr), ϕ and κ.



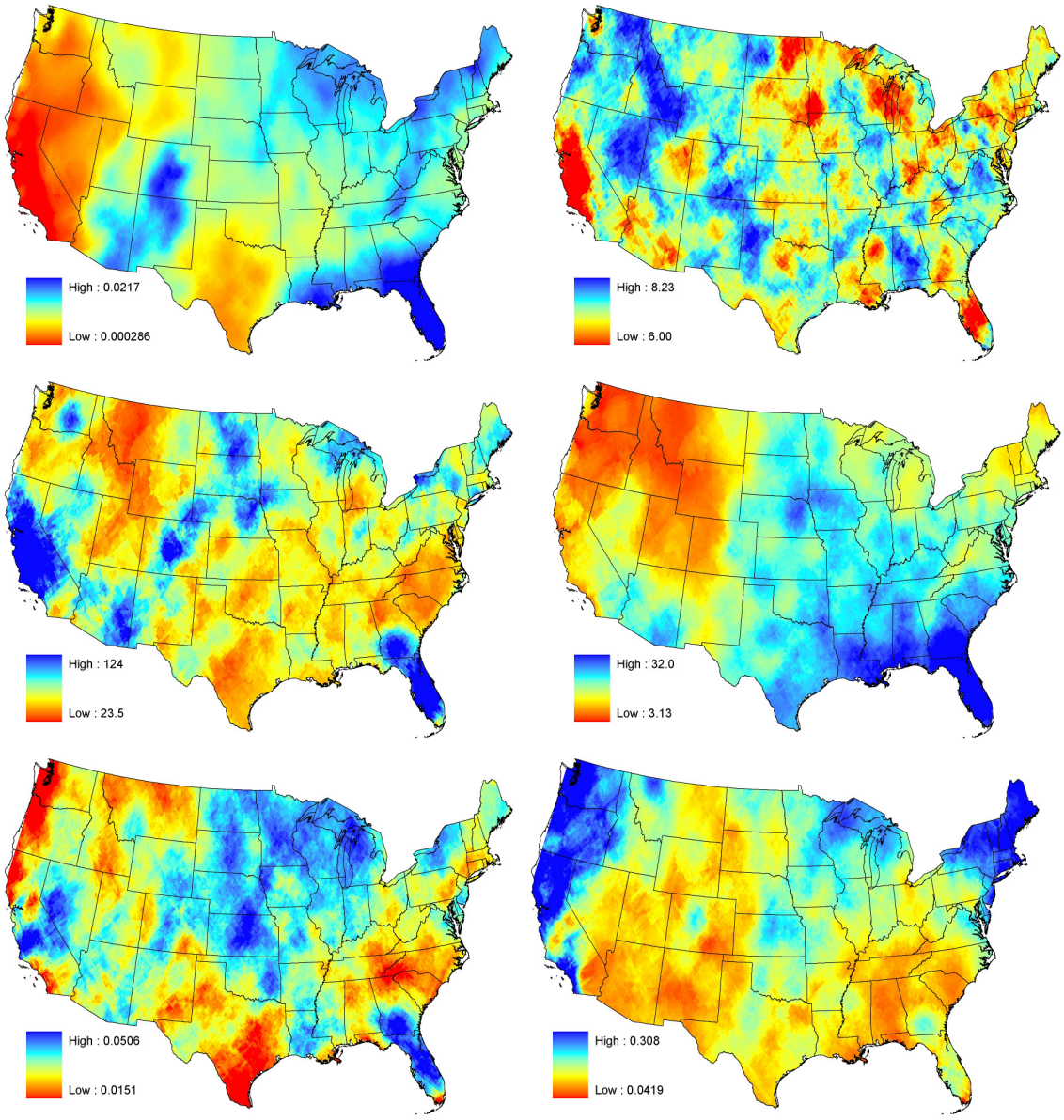
MBLRP model parameters for the month of May. From the top to bottom and left to right, λ (1/hr), ν (hr), α, μ (mm/hr), ϕ and κ.



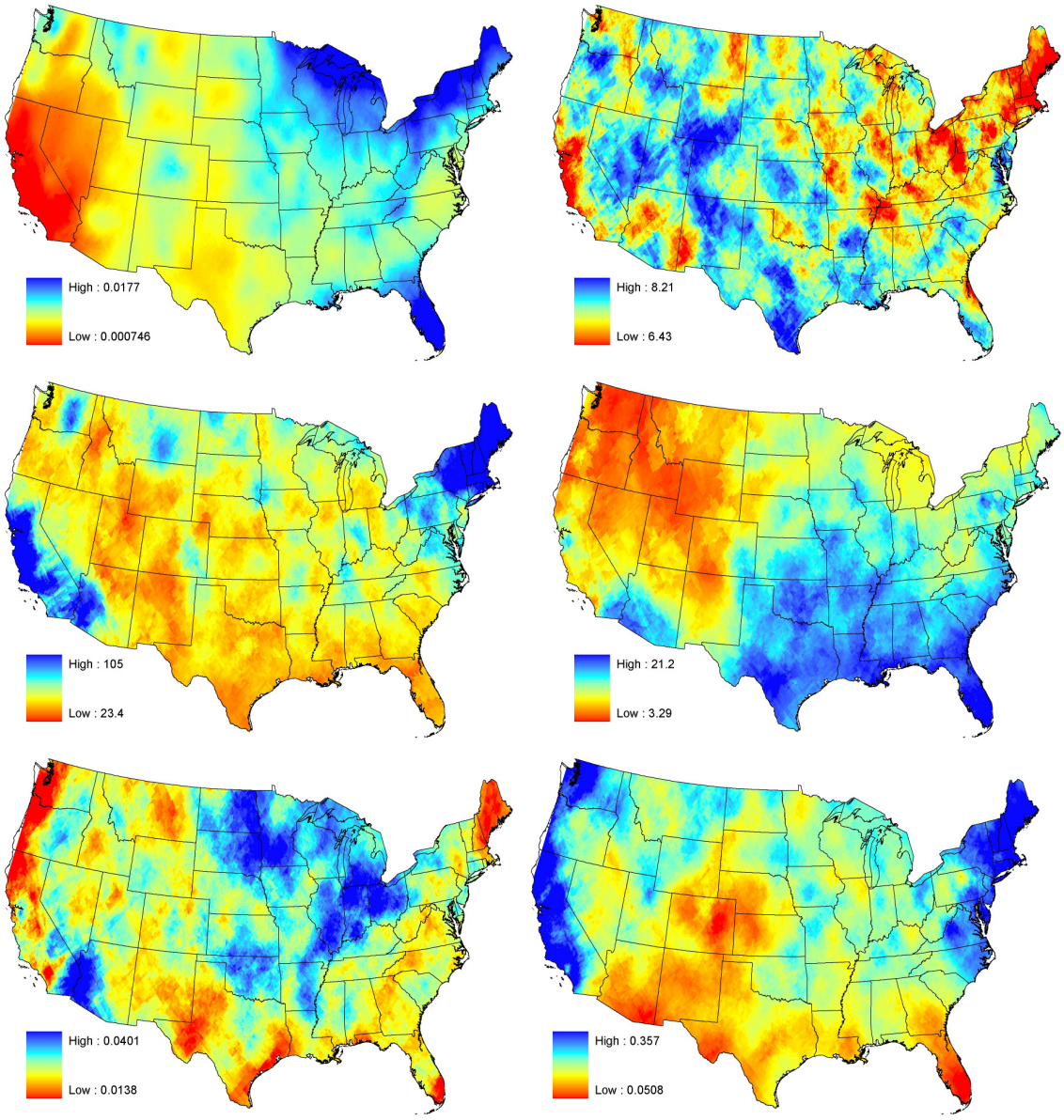
MBLRP model parameters for the month of June. From the top to bottom and left to right, λ (1/hr), ν (hr), α, μ (mm/hr), ϕ and κ.



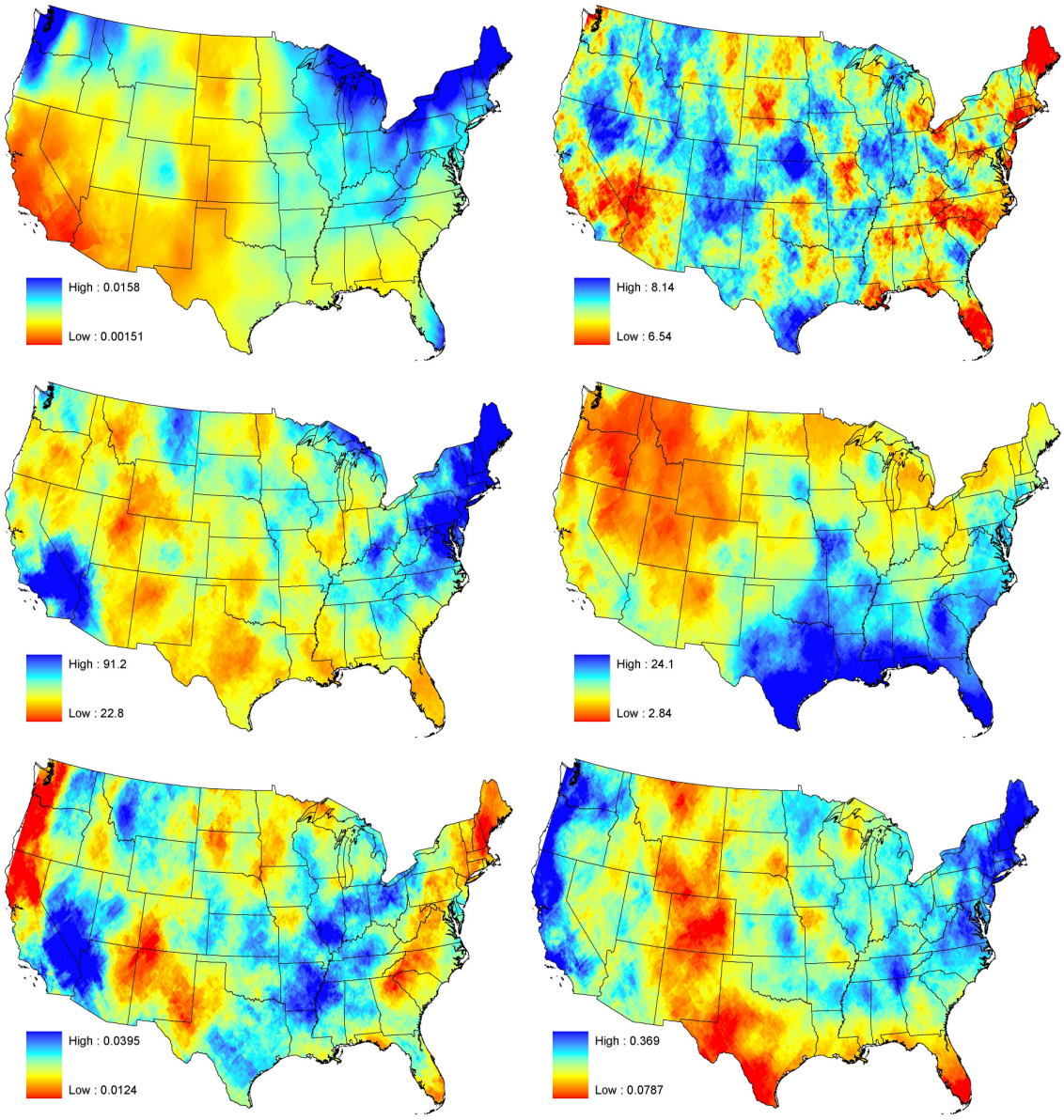
MBLRP model parameters for the month of July. From the top to bottom and left to right, λ (1/hr), ν (hr), α, μ (mm/hr), ϕ and κ.



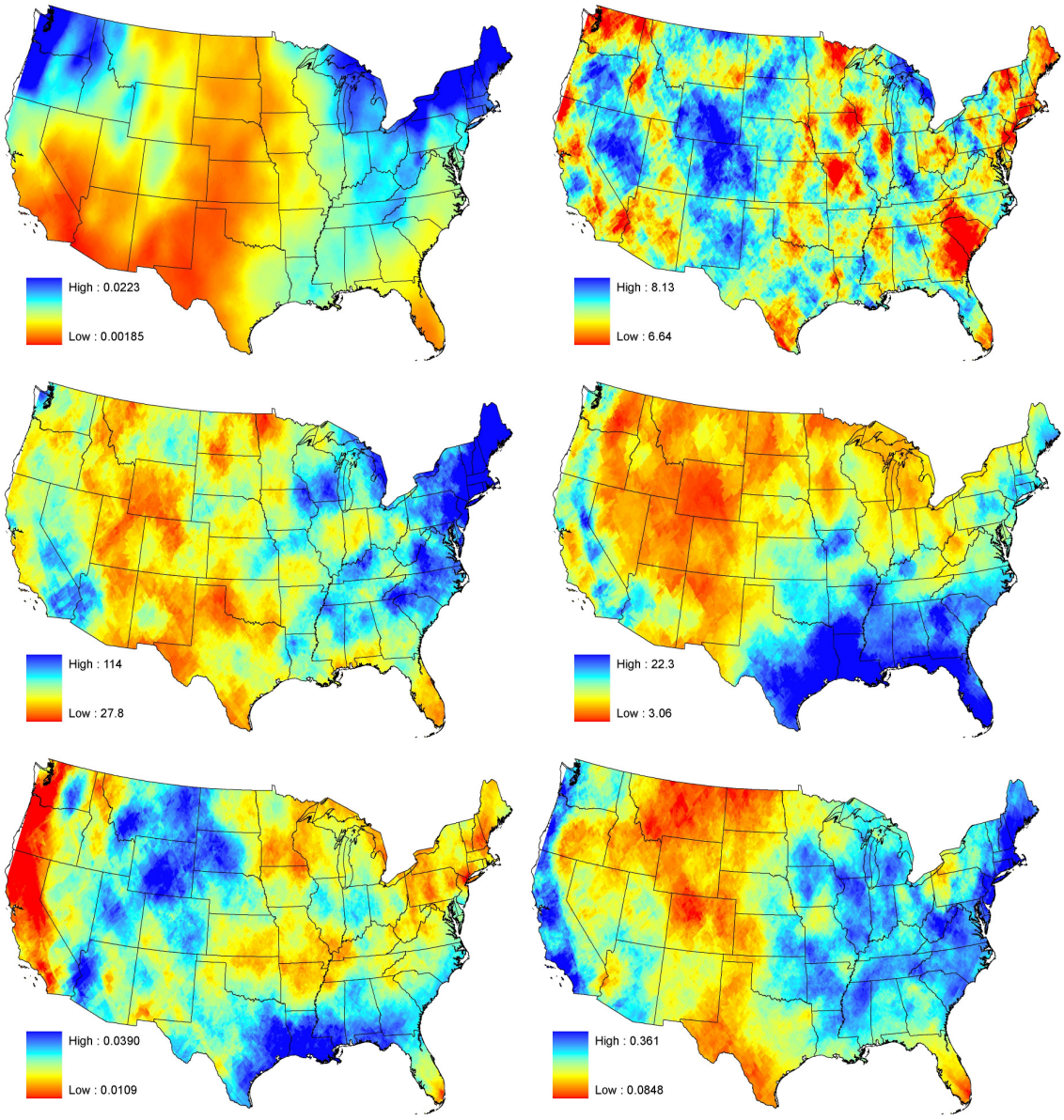
MBLRP model parameters for the month of August. From the top to bottom and left to right, λ (1/hr), ν (hr), α, μ (mm/hr), ϕ and κ.



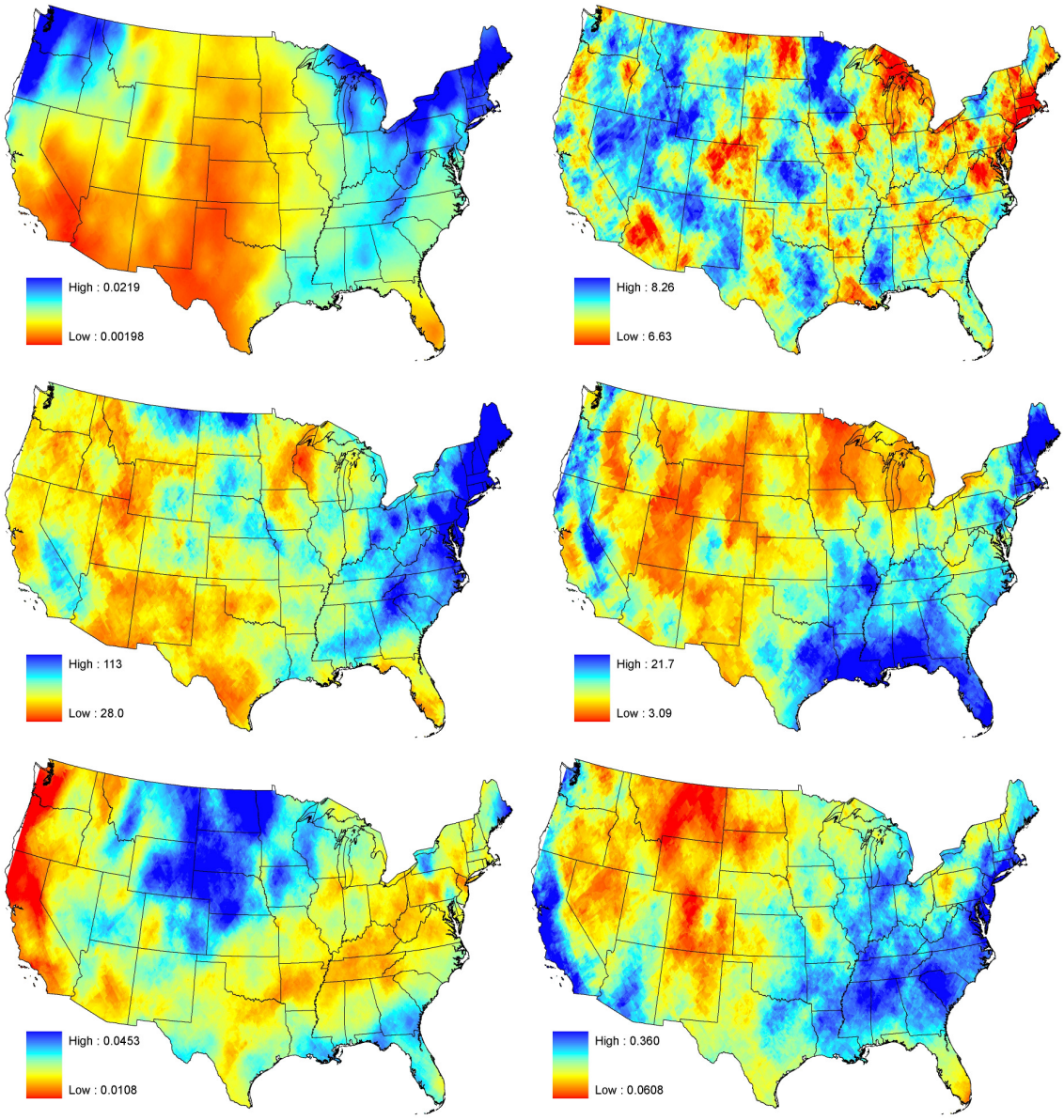
MBLRP model parameters for the month of September. From the top to bottom and left to right, λ (1/hr), ν (hr), α, μ (mm/hr), ϕ and κ.



MBLRP model parameters for the month of October. From the top to bottom and left to right, λ (1/hr), ν (hr), α, μ (mm/hr), ϕ and κ.

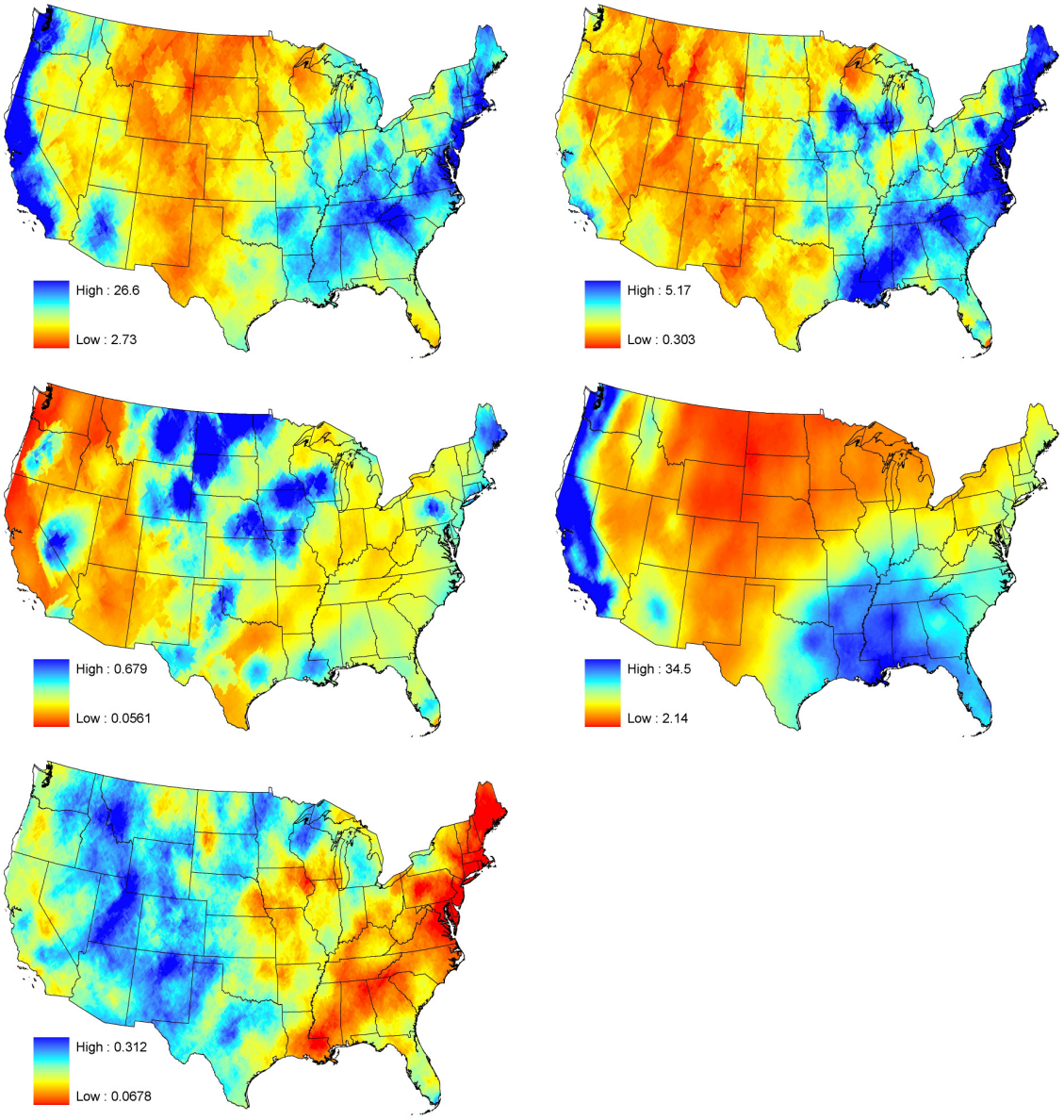


MBLRP model parameters for the month of November. From the top to bottom and left to right, λ (1/hr), ν (hr), α, μ (mm/hr), ϕ and κ.

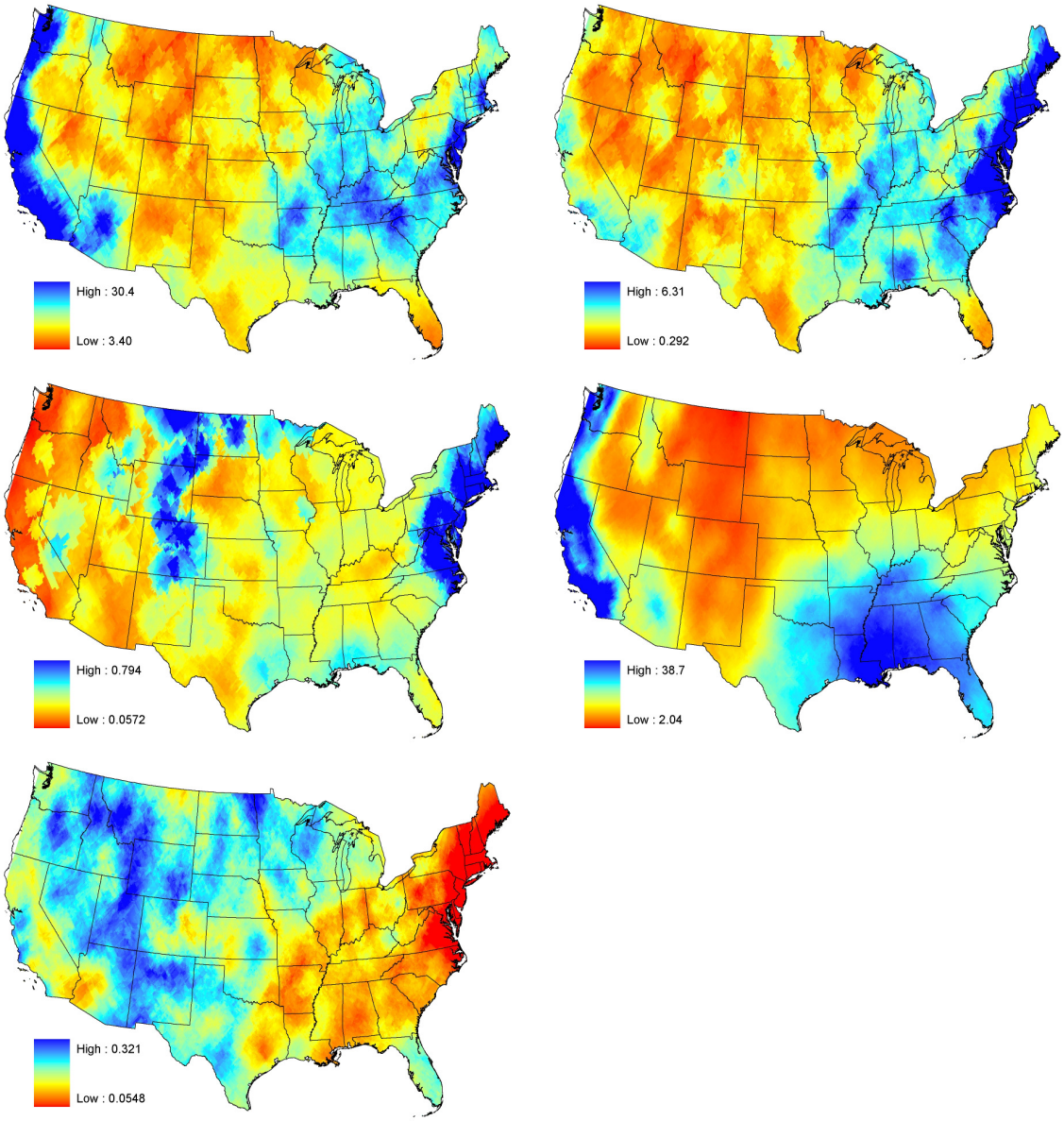


MBLRP model parameters for the month of December. From the top to bottom and left to right, λ (1/hr), ν (hr), α, μ (mm/hr), ϕ and κ.

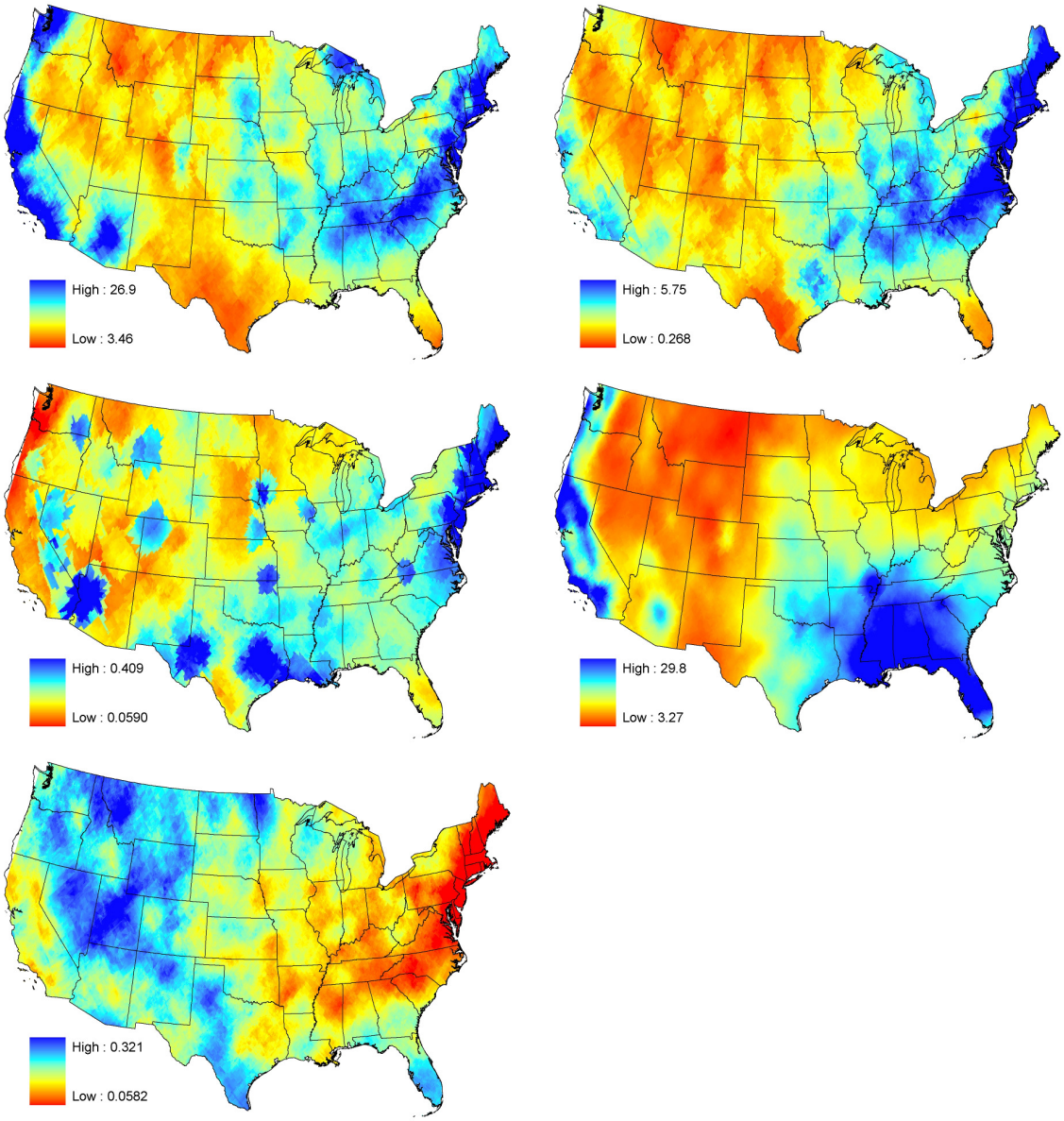
APPENDIX B. MBLRPM COMPARTMENTS OF 12 CALENDAR MONTHS



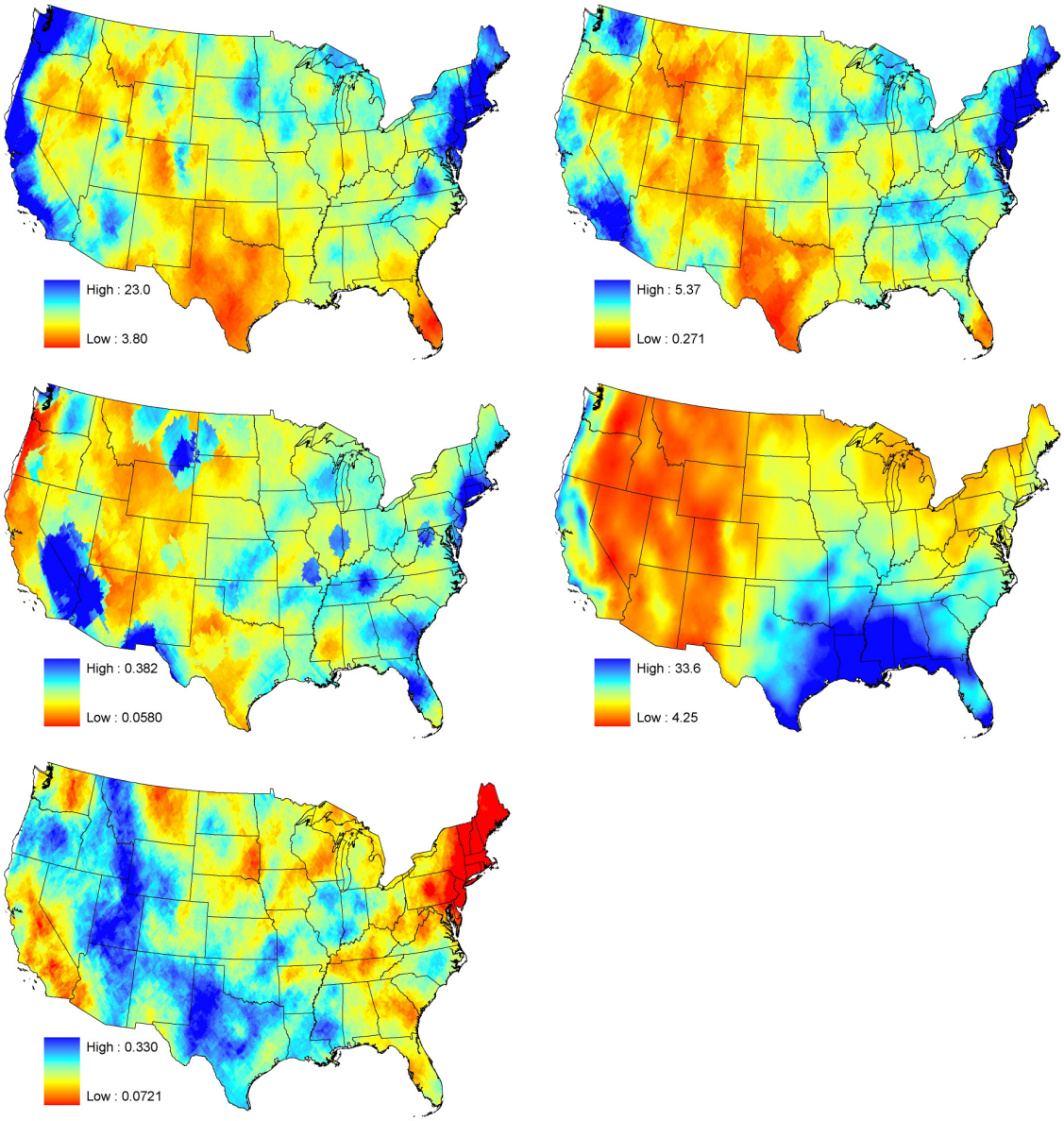
Storm and rain cell characteristics for the month of January according to the MBLRP model parameters. From top to bottom and left to right, average number of rain cells per storm, average rain cell arrival rate (1/hr), average storm duration (hr), average rainfall depth per storm (mm), and average rain cell duration (hr).



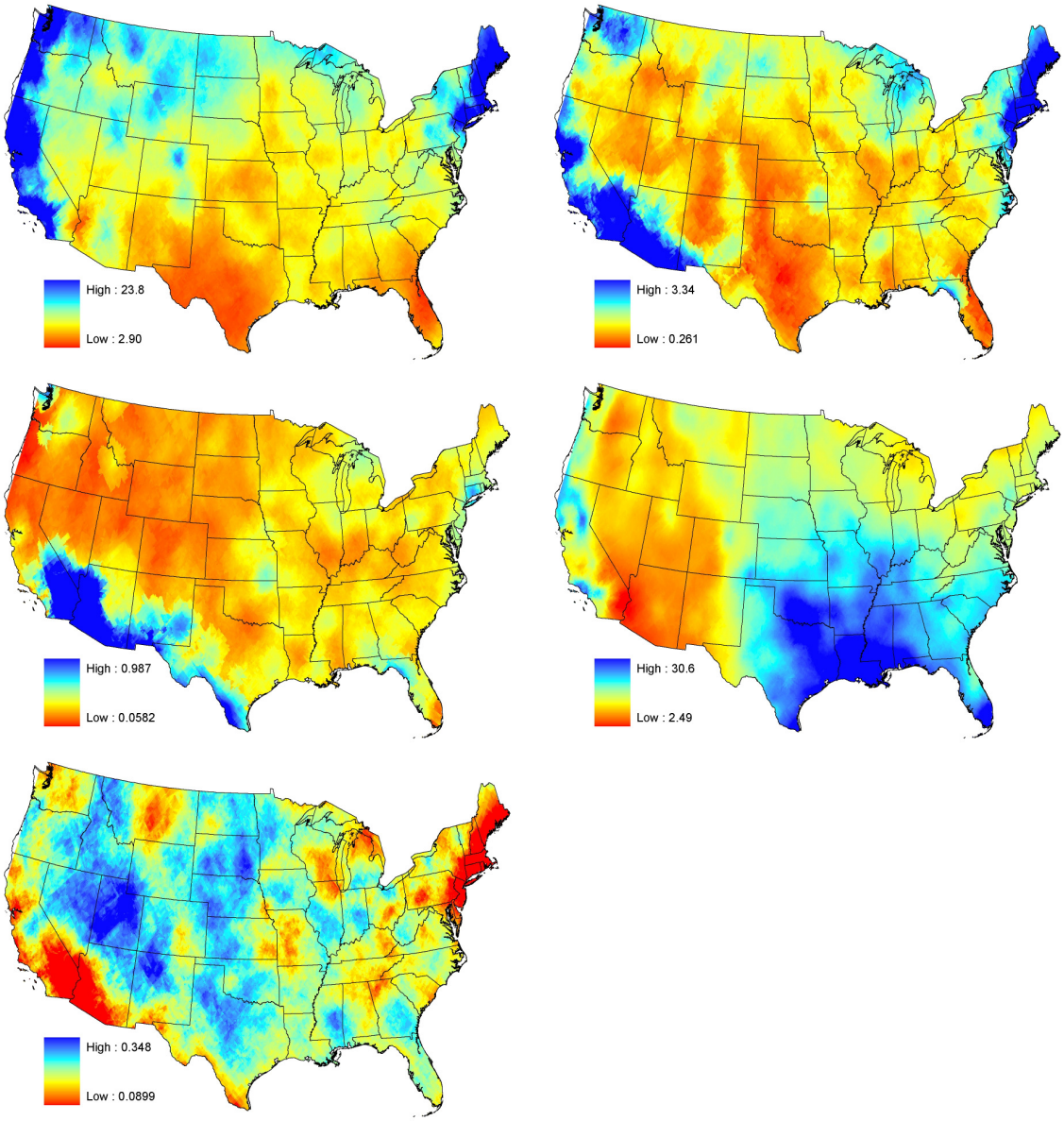
Storm and rain cell characteristics for the month of February according to the MBLRP model parameters. From top to bottom and left to right, average number of rain cells per storm, average rain cell arrival rate (1/hr), average storm duration (hr), average rainfall depth per storm (mm), and average rain cell duration (hr).



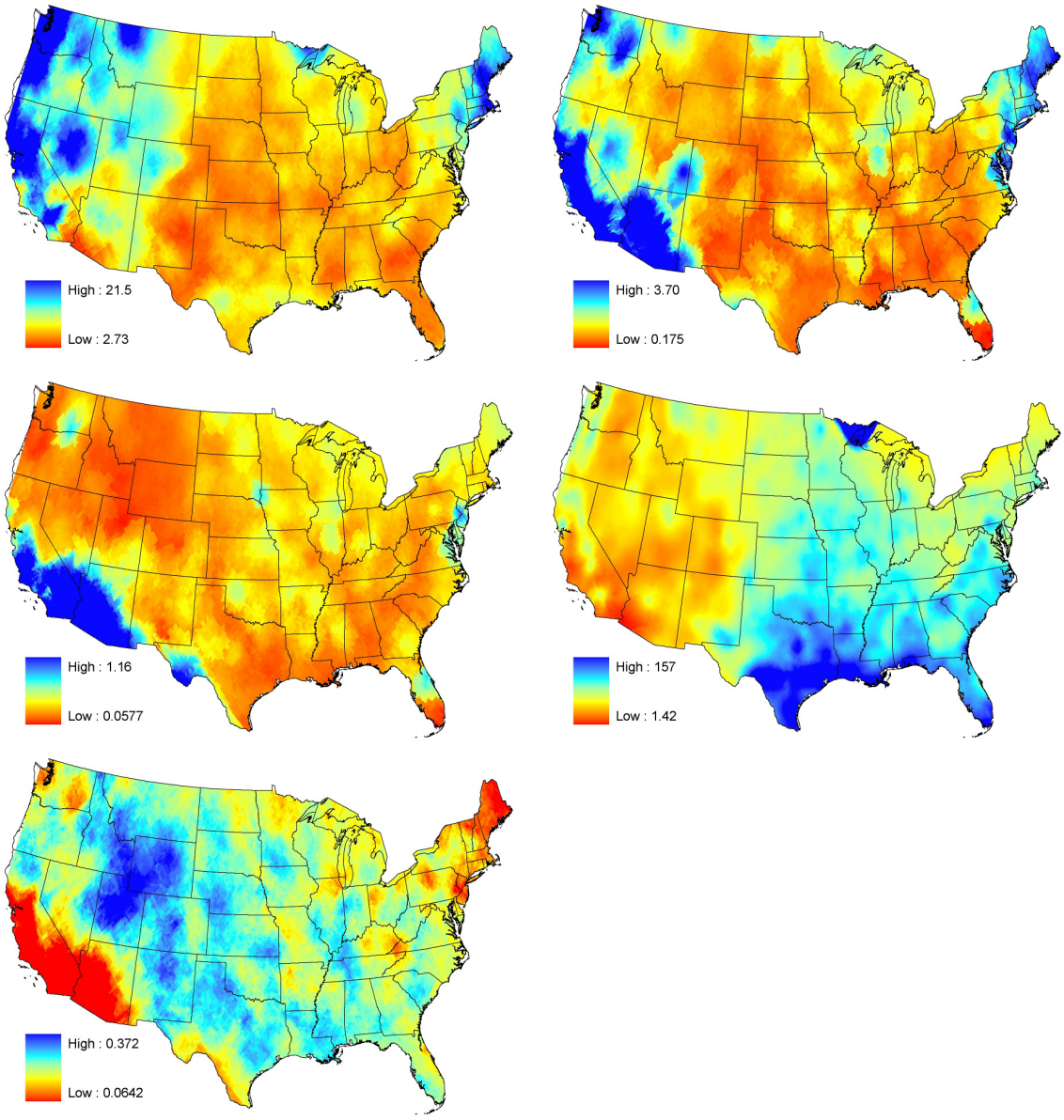
Storm and rain cell characteristics for the month of March according to the MBLRP model parameters. From top to bottom and left to right, average number of rain cells per storm, average rain cell arrival rate (1/hr), average storm duration (hr), average rainfall depth per storm (mm), and average rain cell duration (hr).



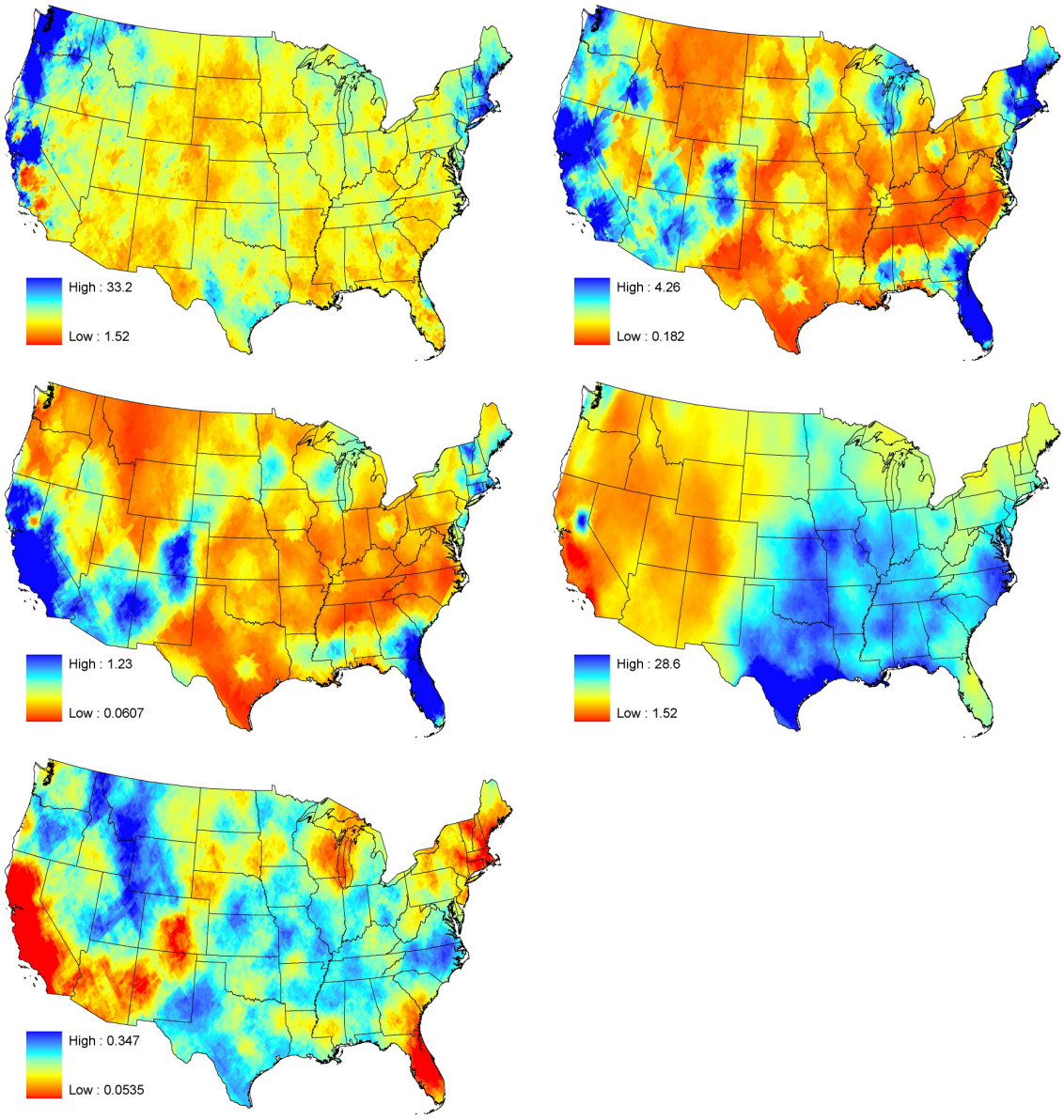
Storm and rain cell characteristics for the month of April according to the MBLRP model parameters. From top to bottom and left to right, average number of rain cells per storm, average rain cell arrival rate (1/hr), average storm duration (hr), average rainfall depth per storm (mm), and average rain cell duration (hr).



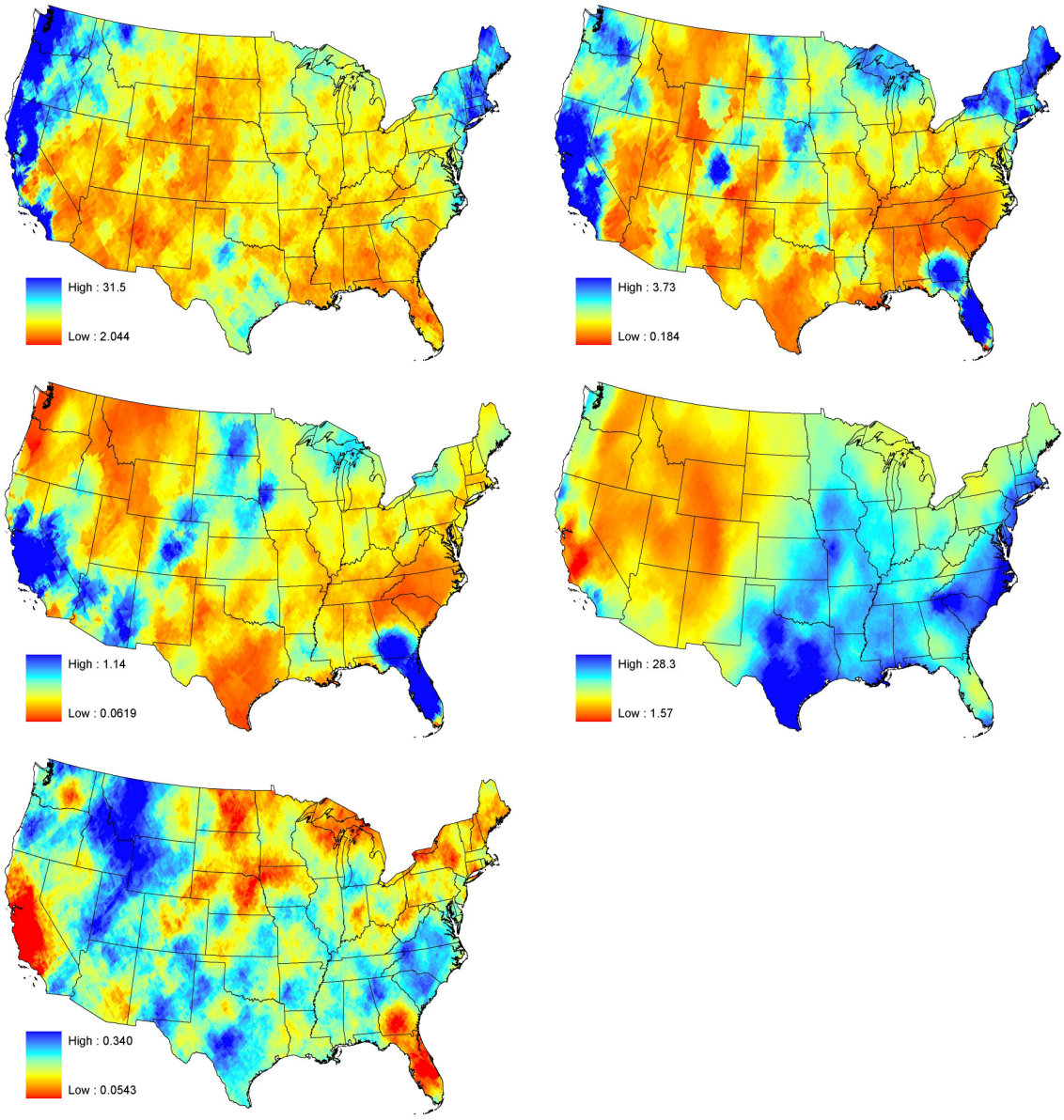
Storm and rain cell characteristics for the month of May according to the MBLRP model parameters. From top to bottom and left to right, average number of rain cells per storm, average rain cell arrival rate (1/hr), average storm duration (hr), average rainfall depth per storm (mm), and average rain cell duration (hr).



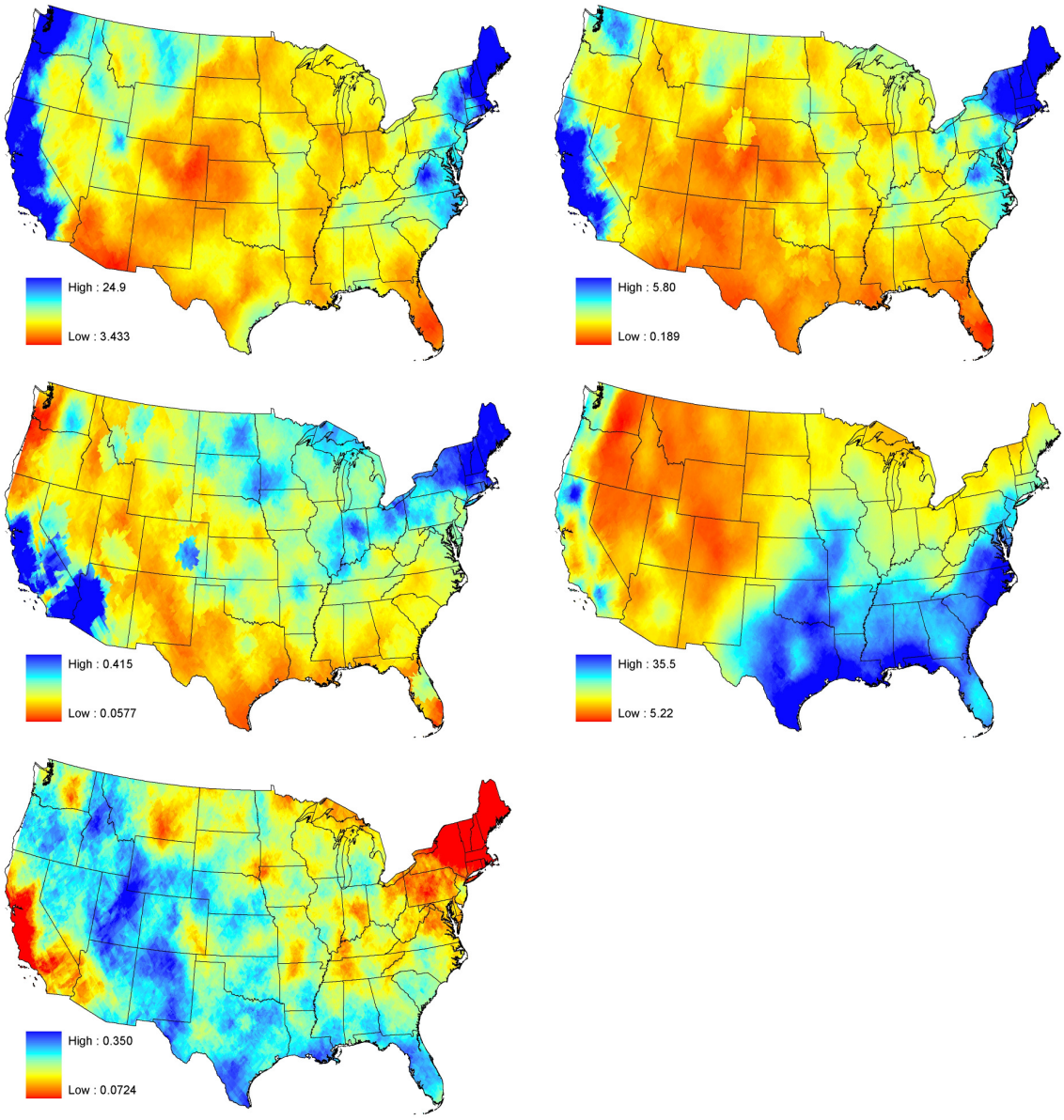
Storm and rain cell characteristics for the month of June according to the MBLRP model parameters. From top to bottom and left to right, average number of rain cells per storm, average rain cell arrival rate (1/hr), average storm duration (hr), average rainfall depth per storm (mm), and average rain cell duration (hr).



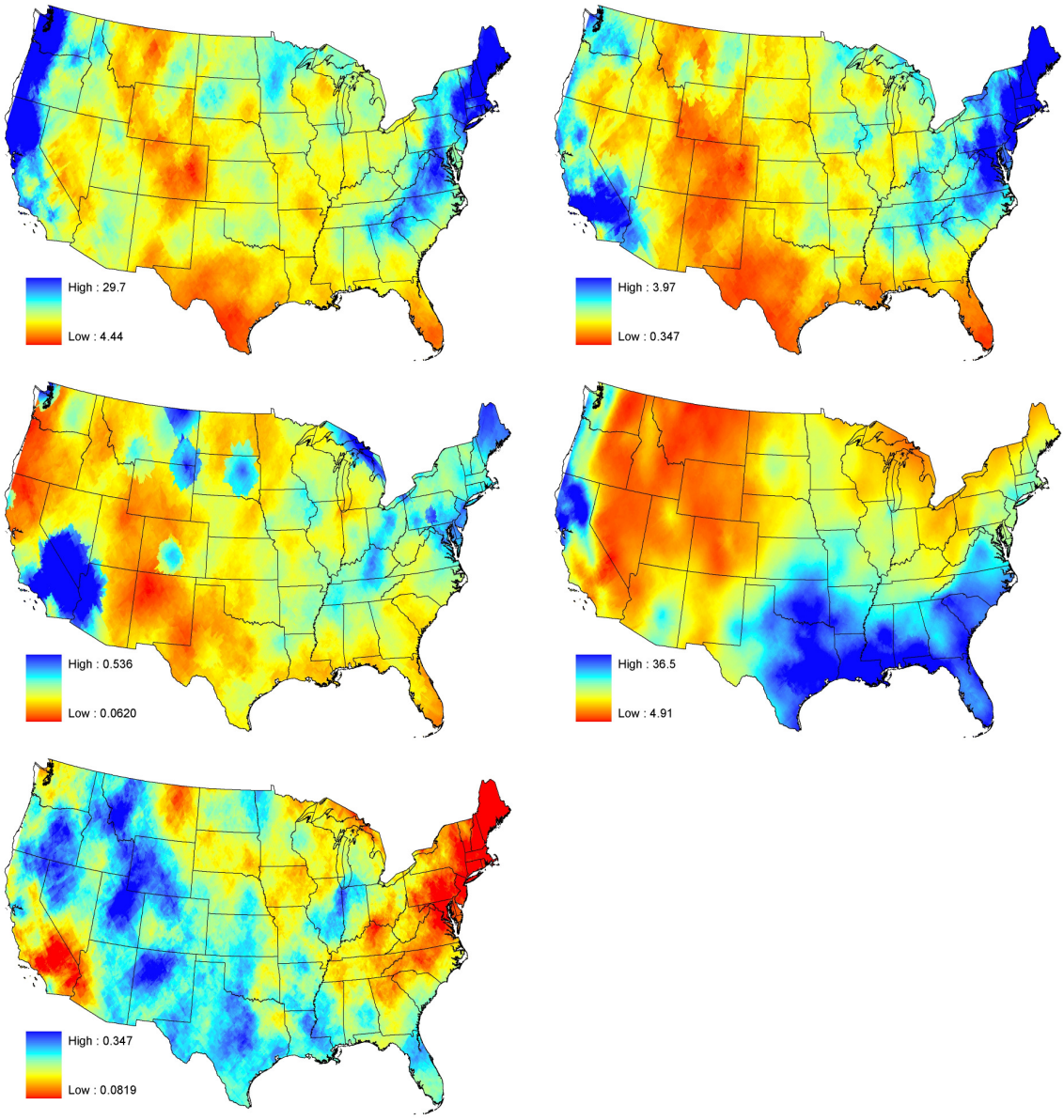
Storm and rain cell characteristics for the month of July according to the MBLRP model parameters. From top to bottom and left to right, average number of rain cells per storm, average rain cell arrival rate (1/hr), average storm duration (hr), average rainfall depth per storm (mm), and average rain cell duration (hr).



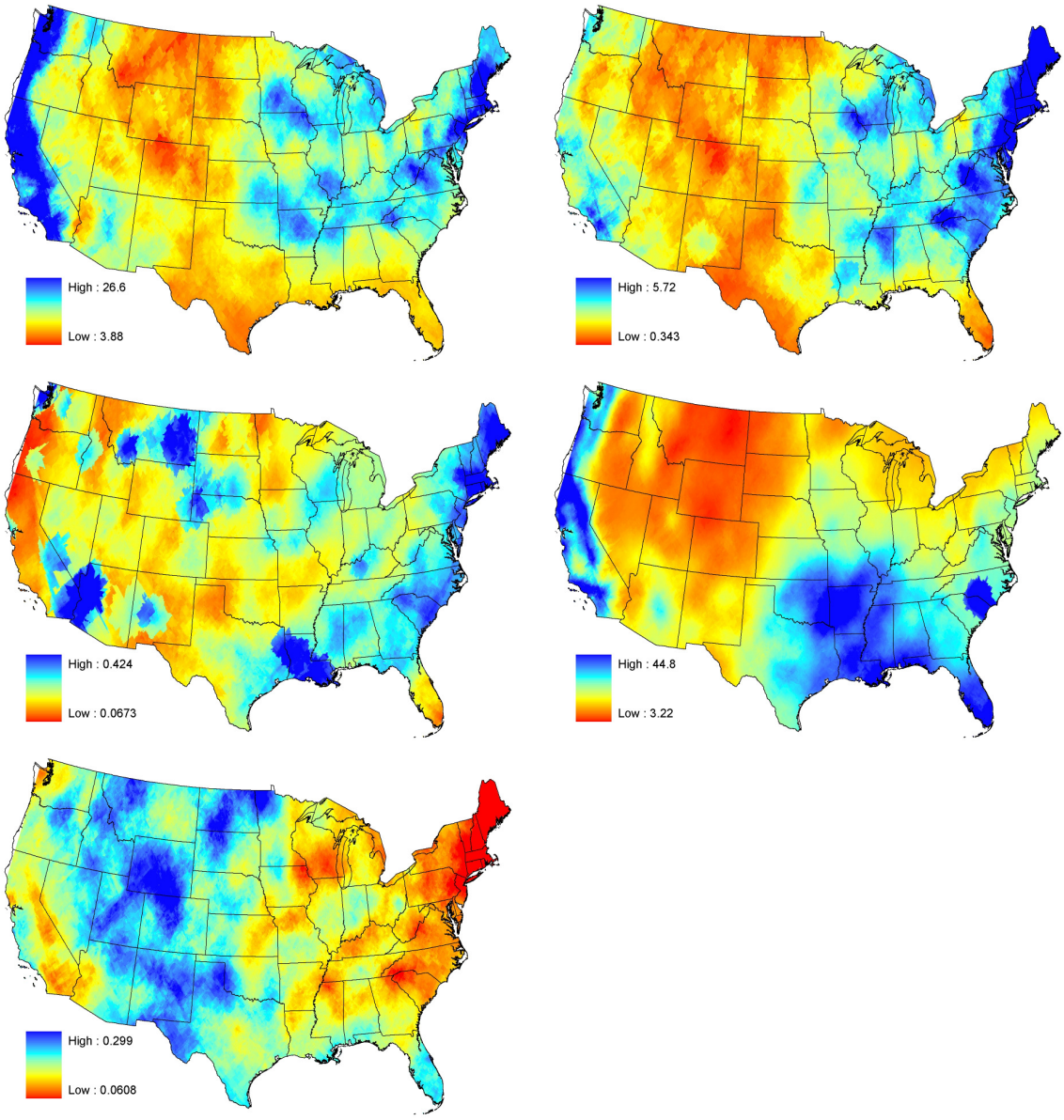
Storm and rain cell characteristics for the month of August according to the MBLRP model parameters. From top to bottom and left to right, average number of rain cells per storm, average rain cell arrival rate (1/hr), average storm duration (hr), average rainfall depth per storm (mm), and average rain cell duration (hr).



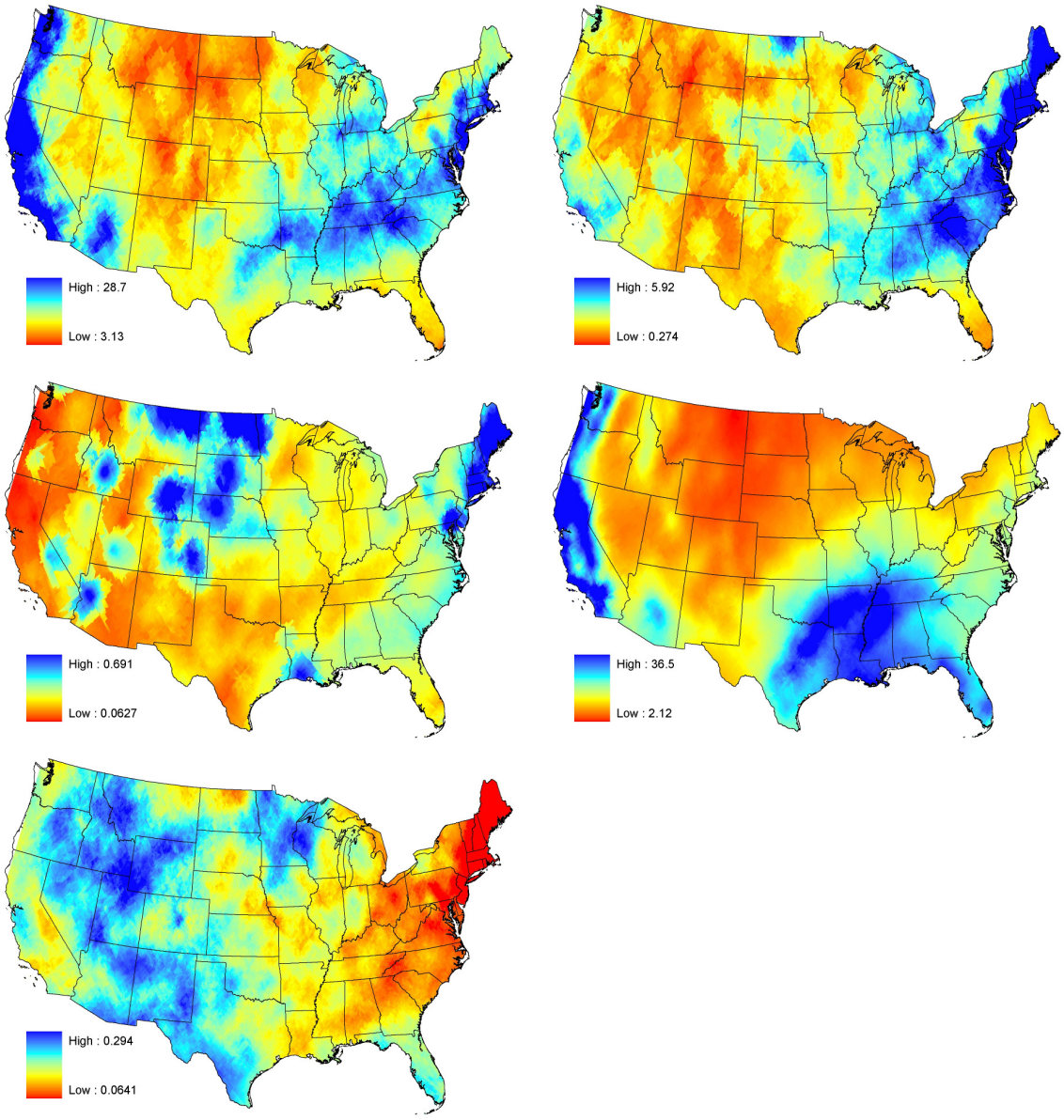
Storm and rain cell characteristics for the month of September according to the MBLRP model parameters. From top to bottom and left to right, average number of rain cells per storm, average rain cell arrival rate (1/hr), average storm duration (hr), average rainfall depth per storm (mm), and average rain cell duration (hr).



Storm and rain cell characteristics for the month of October according to the MBLRP model parameters. From top to bottom and left to right, average number of rain cells per storm, average rain cell arrival rate (1/hr), average storm duration (hr), average rainfall depth per storm (mm), and average rain cell duration (hr).



Storm and rain cell characteristics for the month of November according to the MBLRP model parameters. From top to bottom and left to right, average number of rain cells per storm, average rain cell arrival rate (1/hr), average storm duration (hr), average rainfall depth per storm (mm), and average rain cell duration (hr).



Storm and rain cell characteristics for the month of December according to the MBLRP model parameters. From top to bottom and left to right, average number of rain cells per storm, average rain cell arrival rate (1/hr), average storm duration (hr), average rainfall depth per storm (mm), and average rain cell duration (hr)

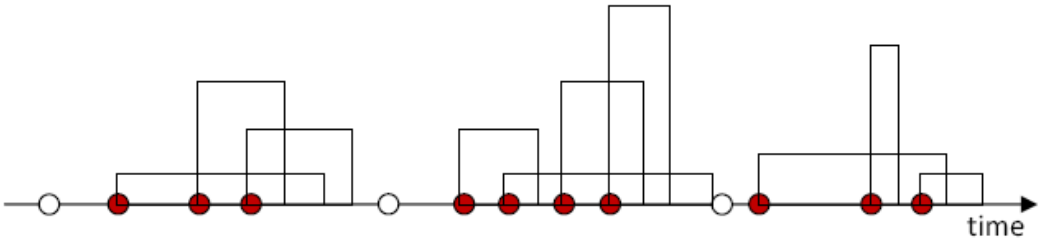
APPENDIX C. MATLAB AND R FUNCTION USAGE

Introduction

In this dissertation, the computer software tools MATLAB and R were utilized to generate stochastic rainfall time series using the Modified Bartlett-Lewis Rectangular Pulse model (MBLRPM). This section explains how to use the codes of Matlab and R designed to generate stochastic rainfall time series based on MBLRPM.

MBLRPM needs 6 parameters to generate rainfall time series. Detailed explanation about these parameters and the way to estimate them are as follow:

In the MBLRP rainfall simulation model, rainfall time series are represented as sequences of storms comprised of rain cells. In the following figure, white circles corresponds to the arrival of storms in time, and the red cells following the white circles represents the arrival of rain cells within each storm.



In the model, X1 is a random variable that represents the number of storms that arrive in a given period, which is governed by a Poisson process with parameter λ [1/T]; X2 is a random variable that represents the duration of storm activity (i.e., the time window after the beginning of the storm within which rain cells can arrive), which varies according to

an exponential distribution with parameter γ [1/T]; X3 is a random variable that represents the number of rain cells that arrive within the duration of storm activity, which is governed by a Poisson process with parameter β [1/T]; X4 is a random variable that represents the duration of the rain cells, which varies according to an exponential distribution with parameter η [1/T] that, in turn, is represented by a gamma distribution with parameters ν and α; and X5 is a random variable that represents the rain cell intensity, which varies according to an exponential distribution with parameter $1/\mu$ [T/L]. From the physical viewpoint, λ is the expected number of storms that arrive in a given period, γ is the inverse of the expected duration of storm activity, β is the expected number of rain cells that arrive within the duration of storm activity, η is the inverse of the expected duration of the rain cells, and μ is the average rain cell intensity. Therefore, the model has six parameters: λ, γ, β, ν, α and μ; however, it is customary to use the ratios $\phi = \gamma/\eta$ and $\kappa = \beta/\eta$ as parameters instead of γ and β.

The estimation of the parameters is done by matching statistics of the simulated and observed rainfall time series. Some commonly used statistics are the mean, the variance, the lag-s covariance and the probability of zero rainfall at various time scales (Khaliq and Cunnane, 1996). According to Rodriguez-Iturbe et al. (1987, 1988), the statistics of the synthetically generated rainfall time series at an accumulation interval of T are:

$$E\left[Y_t^{(t)}\right] = \lambda \mu \mu_c \frac{\nu}{\alpha\text{-}1} T \qquad (1)$$

$$Var\left[Y_t^{(T)}\right] = \frac{2\nu^{2-\alpha}T}{\alpha\text{-}2}\left(k_1 - \frac{k_2}{\phi}\right) - \frac{2\nu^{3-\alpha}}{(\alpha\text{-}2)(\alpha\text{-}3)}\left(k_1 - \frac{k_2}{\phi^2}\right) + \frac{2}{(\alpha\text{-}2)(\alpha\text{-}3)}\left[k_1(T+\nu)^{3-\alpha} - \frac{k^2}{\phi^2}(\phi T+\nu)^{3-\alpha}\right] \qquad (2)$$

$$\mathrm{Cov}\left[\mathrm{Y_t}^{(\mathrm{T})},\mathrm{Y_{t+s}}^{(\mathrm{T})}\right]=\frac{k_1}{(a-2)(a-3)}\left\{[T(s-1)+\nu]^{3-a}+[T(s+1)+\nu]^{3-a}-2(Ts+\nu)^{3-a}\right\}$$
$$+\frac{k_2}{\phi^2(a-2)(a-3)}\left\{2(\phi Ts+\nu)^{3-a}-[\phi T(s-1)+\nu]^{3-a}-[\phi T(s+1)+\nu]^{3-a}\right\} \quad (3)$$

$$\mathrm{P(zero\ rainfall)} = \exp\left\{-\lambda\mathrm{T}-\frac{\lambda\nu}{\phi(\alpha\text{-}1)}\left[1+\phi(\kappa+\phi)-\frac{1}{4}\phi(\kappa+\phi)(\kappa+4\phi)+\frac{\phi(\kappa+\phi)(4\kappa^2+27\kappa\phi+72\phi^2)}{72}\right]\right.$$
$$+\frac{\lambda\nu}{(\alpha\text{-}1)(\kappa+4\phi)}\left(1-\kappa-\phi+\frac{3}{2}\kappa\phi+\phi^2+\frac{\kappa^2}{2}\right)$$
$$\left.+\frac{\lambda\nu}{(\alpha\text{-}1)(\kappa+\phi)}\left[\frac{\nu}{\nu+(\kappa+\phi)T}\right]^{\alpha-1}\frac{\kappa}{\phi}\left(1-\kappa-\phi+\frac{3}{2}\kappa\phi+\phi^2+\frac{\kappa^2}{2}\right)\right\} \quad (4)$$

where

$$\mathrm{k_1}=\left(2\lambda\mu_c\mu^2+\frac{\lambda\mu_c\kappa\phi\mu^2}{\phi^2\text{-}1}\right)\left(\frac{\nu^\alpha}{\alpha-1}\right) \quad (5)$$

$$\mathrm{k_2}=\left(\frac{\lambda\mu_c\kappa\phi\mu^2}{\phi^2\text{-}1}\right)\left(\frac{\nu^\alpha}{\alpha-1}\right) \quad (6)$$

$$\mu_c=1+\frac{\kappa}{\phi} \quad (7)$$

s is the lag time in number of accumulation intervals, and $\mathrm{Y_t}^{(\mathrm{T})}$ is the rainfall time series at an accumulation interval T.

Parameter Estimation

Because it is not possible to analytically solve equations (1) through (4), they are obtained by minimizing the following objective function, which represents the disagreement between the statistics of the observed and simulated rainfall time series:
Because it is not possible to analytically solve equations (1) through (4) (Onof et al. 2000), they are typically obtained by minimizing the following objective function, which represents the disagreement between the statistics of the observed and simulated rainfall time series:

$$\sum_{k=1}^{n} w_k \left[1 - \frac{F_k(\vec{\theta})}{f_k} \right]^2 \qquad (8)$$

where $\vec{\theta}$ is the parameter set (λ, ν, α, μ, ϕ, κ), n is the number of statistics being matched, $F_k(\vec{\theta})$ is the k^{th} statistic of the simulated rainfall time series, fk is the kth statistics of the observed rainfall time series and w_k is a weight factor given to the kth statistic.

The surface of the objective function in parameter space is not smooth, so heuristic approach of optimization should be performed. This study uses the Isolated Speciation-Based Particle Swarm Optimization (ISPSO, Cho et al. 2008). ISPSO is implemented in R source code. This code optimizes any given objective function. The objective function for ISPSO (in the case of this study, Equation 8) should also be defined in R source code. Also, there should be another source code of R that (1) integrates the source code

of ISPSO and of objective function and that (2) takes care of input and output. Therefore, 3 following R source codes should be prepared to estimate the parameters of ISPSO:

1. The source code for ISPSO
   - A complete optimization package
   - Does not have to be modified.
   - \\BLRP_SourceCodes\Parameter Estimation\pso.r
2. The source code for the objective function
   - \\BLRP_SourceCodes\Parameter Estimation\blrp_many_moments_weighted_qp.r
3. The source code that integrates the source code 1, source code 2, and handles file input / output
   - \\BLRP_SourceCodes\Parameter Estimation\doit_2009-06-02_weighted_qp.r

These files can be opened and edited with typical text editors such as UltraEdit. Before reading the following paragraphs explaining details of codes, readers are recommended to be acquainted with the basic grammars of computer language R by visiting the following internet website (http://www.r-project.org/), which also provides a completely free copy of the software tool R.

Firstly, the source code for the objective function (\\BLRP_SourceCodes\Parameter Estimation\blrp_many_moments_weighted_qp.r) is investigated. The purpose of this function is to calculate Equation (8).

```
blrp_many_moments_weighted_qp <- function(x, moments){

l <- x[1]
v <- x[2]
a <- x[3]
E <- x[4]
o <- x[5]
k <- x[6]

mc <- (1+k/o)
k1 <- (( 2 * l * mc * E^2 + (l * mc * k * o * E^2)/(o^2-1)) * (v^a/(a-1)))
k2 <- ((l * mc * k * E^2) / (o^2-1) * v^a/(a-1))

# Hourly Mean
Mean1 <- l * E * mc * v / (a - 1)

# Variance of the rainfall at T-hourly time scale (T = 1, 3, 6, 12, 24)
T <- 1
Var1 <- (2 * v ^ (2-a) * T)/(a-2) * ( k1 - k2 / o) - (2*v^(3-a))/((a-2)*(a-3))*( k1 - k2 / o^2) + 2 / ((a-2)*(a-3)) *
T <- 3
Var3 <- (2 * v ^ (2-a) * T)/(a-2) * ( k1 - k2 / o) - (2*v^(3-a))/((a-2)*(a-3))*( k1 - k2 / o^2) + 2 / ((a-2)*(a-3)) *
T <- 12
Var12 <- (2 * v ^ (2-a) * T)/(a-2) * ( k1 - k2 / o) - (2*v^(3-a))/((a-2)*(a-3))*( k1 - k2 / o^2) + 2 / ((a-2)*(a-3)) *
T <- 24
Var24 <- (2 * v ^ (2-a) * T)/(a-2) * ( k1 - k2 / o) - (2*v^(3-a))/((a-2)*(a-3))*( k1 - k2 / o^2) + 2 / ((a-2)*(a-3)) *

# Lag T hour correlation coefficient of T-hourly time scale (T = 1, 3, 6, 12, 24)
s <- 1
T <- 1
Lag1 <- ((k1/((a-2)*(a-3)))*(((T*(s-1)+v)^(3-a))+((T*(s+1)+v)^(3-a))-(2*(T*s+v)^(3-a))) + (k2/(o^2*(a-2)*(a-3)))*((2*
T <- 3
Lag3 <- ((k1/((a-2)*(a-3)))*(((T*(s-1)+v)^(3-a))+((T*(s+1)+v)^(3-a))-(2*(T*s+v)^(3-a))) + (k2/(o^2*(a-2)*(a-3)))*((2*
T <- 12
Lag12 <- ((k1/((a-2)*(a-3)))*(((T*(s-1)+v)^(3-a))+((T*(s+1)+v)^(3-a))-(2*(T*s+v)^(3-a))) + (k2/(o^2*(a-2)*(a-3)))*((2*
T <- 24
Lag24 <- ((k1/((a-2)*(a-3)))*(((T*(s-1)+v)^(3-a))+((T*(s+1)+v)^(3-a))-(2*(T*s+v)^(3-a))) + (k2/(o^2*(a-2)*(a-3)))*((2*

#F5 <- exp( - l * 1 - ( l * v / (o * (a-1))) * (( 1 + o * (k + o) - 0.25 * o * (k+o) * (k+4*o) + (o * (k+o)*(4*k^2 + 2
#F6 <- exp( - l * 2 - ( l * v / (o * (a-1))) * (( 1 + o * (k + o) - 0.25 * o * (k+o) * (k+4*o) + (o * (k+o)*(4*k^2 + 2
T <- 1
Prob1 <- exp( - l * T - ( l * v / (o * (a-1))) * (( 1 + o * (k + o) - 0.25 * o * (k+o) * (k+4*o) + (o * (k+o)*(4*k^2 +
T <- 3
Prob3 <- exp( - l * T - ( l * v / (o * (a-1))) * (( 1 + o * (k + o) - 0.25 * o * (k+o) * (k+4*o) + (o * (k+o)*(4*k^2 +
T <- 12
Prob12 <- exp( - l * T - ( l * v / (o * (a-1))) * (( 1 + o * (k + o) - 0.25 * o * (k+o) * (k+4*o) + (o * (k+o)*(4*k^2
T <- 24
Prob24 <- exp( - l * T - ( l * v / (o * (a-1))) * (( 1 + o * (k + o) - 0.25 * o * (k+o) * (k+4*o) + (o * (k+o)*(4*k^2

1.19 * (Mean1/moments[1] - 1)^2 +
1.09 * (Var1/moments[2] - 1)^2 +
1.09 * (Var3/moments[3] - 1)^2 +
1.09 * (Var12/moments[4] - 1)^2 +
1.09 * (Var24/moments[5] - 1)^2 +
0.246 * (Lag1/moments[6] - 1)^2 +
0.246 * (Lag3/moments[7] - 1)^2 +
0.246 * (Lag12/moments[8] - 1)^2 +
0.246 * (Lag24/moments[9] - 1)^2 +
0.490 * (Prob1/moments[10] - 1)^2 +
0.490 * (Prob3/moments[11] - 1)^2 +
0.490 * (Prob12/moments[12] - 1)^2 +
0.490 * (Prob24/moments[13] - 1)^2
}
```

Line 1 defines the name of the function and input. In the case of this study, the input of the function is parameter values (x) and the observed rainfall statistics (moments). “x” is a vector with 6 arguments (number of parameters), and “moments” is a vector with 13

arguments, each of which corresponds to rainfall statistics (mean at 1-hour accumulation level, variance at 1, 3, 12, and 24 hour accumulation level, lag-1 autocorrelation at 1, 3, 12, and 24 hour accumulation level, and probability of zero rainfall at 1, 3, 12, and 24 hour accumulation level)

Line 3 to Line 8 assigns each element of "x" to the variable with the name l, v, a, e, o, and k, which corresponds to λ, ν, α, μ, ϕ, and κ.

Line 10 to Line 12 corresponds to Equation (5) to Equation (7)

Line 15 corresponds to Equation (1)

Line 18 to Line 25 corresponds to Equation (2)

Line 28 to Line 37 corresponds to Equation (3). *Here, it is noteworthy that Equation (3) represents the "covariance", which is the product of "correlation coefficient" and variance. Line 28 to Line 37 represents the correlation coefficients, so at the end of the line 31, 33, 35, and 37, there are terms that divide covariance by variance.*

Line 41 to Line 48 corresponds to Equation (4).

Line 52 to Line 62 corresponds to Equation (8). The first element of the line 52 through 62 (e.g. 1.19, 1.09, etc.) corresponds to w_k of Equation (8)

Secondly, the source code that integrates the source code 1, source code 2, and handles file input / output is investigated. Open the file "\\BLRP_SourceCodes\Parameter Estimation\doit_2009-06-02_weighted_qp.r" Only the lines that need to be changed will be explained. Changing the remaining lines can cause the malfunction of the optimization procedures.

```
#!/home/fyns95/usr/local/bin/Rscript --restore --save
source("pso.R")
source("blrp_many_moments_weighted_qp.R")

s <- list()
scale_x <- function(x){
  # real ranges
    xmin <- c(0.00001, 2, 3.0001, 0.5, 0.0, 0.01)
    xmax <- c(0.05, 6, 300, 150, 0.4, 0.999)

  if(is.matrix(x))
    sx <- t((xmax-xmin)/(s$xmax-s$xmin)*(t(x)-s$xmin)+xmin)
  else
    sx <- (xmax-xmin)/(s$xmax-s$xmin)*(x-s$xmin)+xmin
}

s$D <- 6
s$xmin <- rep(-1, s$D)
s$xmax <- rep(1, s$D)

################################################################################
s$S <- 10 + floor(2*sqrt(s$D))
s$vmax <- (s$xmax-s$xmin)*0.5
s$vmax0 <- diagonal(s)*0.01
s$maxiter <- 500

s$rspecies <- diagonal(s)*0.1
s$rprey <- diagonal(s)*0.002
s$rnest <- diagonal(s)*0.05
s$.distance_to_solution <- s$rnest

s$age <- 10
s$xeps <- 0.00001
s$feps <- 0.0001
```

```
#s$.plot_method <- "species,movement"
#s$.plot_method <- "density"
s$.plot_method <- ""
s$.plot_x <- c(1, 2)

################################################################################
s$mycontrib <- ISPSO
s$pso <- "spso"
s$vupdate <- "constriction"
s$c1 <- 2.05
s$c2 <- 2.05
s$w <- 2/abs(2-s$c1-s$c2-sqrt((s$c1+s$c2)^2-4*(s$c1+s$c2)))

################################################################################
stats <- read.csv("stats_192stations.csv")
sol <- c()

old_sol <- c()

for(gage in 1:nrow(stats)){
  for(month in 1:12){
    s$moments <- as.numeric(stats[gage, seq(3+month, ncol(stats), 12)])
    print(c(s$moments))
    s$f <- function(x) blrp_many_moments_weighted_qp(scale_x(x), s$moments)

    if(any(s$moments==0)){
      sol0 <- c(gage=gage, long=stats[gage, "Longitude"], lat=stats[gage, "Latitude"], month=month,
        rep(-9999, 9))
    }else{
      ret <- pso(s)#, initx=old_sol)
      bestidx <- which(ret$pop[,"f"]==min(ret$pop[,"f"], na.rm=TRUE))[1]
      old_sol <- rbind(ret$pop[bestidx,1:s$D], ret$nest[,1:s$D])
      ret$nest[,1:s$D] <- scale_x(ret$nest[,1:s$D])
      ret$pop[,1:s$D] <- scale_x(ret$pop[,1:s$D])

      sol0 <- c(gage=gage, long=stats[gage, "Longitude"], lat=stats[gage, "Latitude"], month=month,
        ret$pop[bestidx,1:(s$D+1)])
      if(!is.null(ret$nest)){
        for(i in 1:nrow(ret$nest))
          sol0 <- rbind(sol0, c(gage=gage, long=stats[gage, "Longitude"], lat=stats[gage, "Latitude
            ret$nest[i,1:(s$D+1)]), deparse.level=0)
      }
    }

  sink("weightfactor_verification.txt", append=TRUE)
  if(any(s$moments==0) || is.null(ret$nest)){
   cat(sol0, "\n")
  }else{
   for(i in 1:nrow(sol0))
    cat(sol0[i,], "\n")
  }
  sink()

  sol <- rbind(sol, sol0, deparse.level=0)
  }
}[EOF]
```

Line 2 imports the source code of ISPSO (the optimization tool developed by Cho et al. (2008))

Line 3 imports the source code of the objective function.

Line 8 and Line 6 defines the lower and upper bound of the parameter space corresponding to λ, ν, α, μ, ϕ, and κ. For example, 0.00001 and 0.05 are the lower and upper bound of the search space of the parameter λ. 2 and 6 are the lower and upper bound of the search space of the parameter ν.

Line 50 reads in the comma-delimited text file containing the values of the observed statistics that are used in Equation (8). Open the file "\\BLRP_SourceCodes\Parameter Estimation\stats_192stations.csv" using Microsoft Excel.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1	Latitude	Longitude	Age	Mean1_1	Mean1_2	Mean1_3	Mean1_4	Mean1_5	Mean1_6	Mean1_7	Mean1_8	Mean1_9	Mean1_10	Mean1_11	Mean1_12	Var
2	33.9378	-117.885	56	0.1137	0.1174	0.0851	0.0387	0.0087	0.0033	0.001	0.0021	0.0092	0.0107	0.0528	0.0799	
3	47.0139	-116.266	51	0.2113	0.1855	0.158	0.1237	0.1263	0.1067	0.0517	0.0524	0.0804	0.1085	0.1976	0.1718	
4	40.4356	-85.2892	55	0.095	0.0978	0.1359	0.1479	0.1614	0.1954	0.1837	0.1674	0.1435	0.1219	0.143	0.1219	
5	48.5969	-102.928	55	0.0239	0.0208	0.0242	0.0544	0.0882	0.14	0.11	0.0846	0.0748	0.041	0.0205	0.0203	
6	32.6169	-106.741	58	0.0242	0.023	0.013	0.0116	0.0152	0.032	0.0866	0.0935	0.063	0.0384	0.0203	0.0357	
7	37.69	-90.78	51	0.0895	0.1134	0.1486	0.1669	0.1549	0.1622	0.1456	0.1509	0.1493	0.1099	0.1685	0.1149	
8	33.8167	-85.7664	54	0.1993	0.2107	0.2309	0.1986	0.1815	0.1875	0.1848	0.1303	0.1658	0.1216	0.1556	0.1793	
9	45.4825	-120.724	55	0.0769	0.0623	0.0525	0.0414	0.0409	0.0312	0.0131	0.0158	0.0233	0.0427	0.0752	0.0825	
10	36.8486	-121.422	55	0.1112	0.0885	0.0879	0.0361	0.0132	0.0028	0.0012	0.0008	0.0103	0.025	0.06	0.0756	
11	45.6983	-110.441	61	0.0196	0.0169	0.0275	0.0539	0.1038	0.105	0.0597	0.0573	0.0608	0.0493	0.025	0.0173	
12	40.1164	-95.7567	55	0.0333	0.0447	0.0865	0.1078	0.1882	0.2111	0.2017	0.1746	0.1667	0.1076	0.0666	0.0389	
13	40.6675	-89.6839	57	0.0734	0.0756	0.1168	0.1755	0.1765	0.1722	0.186	0.148	0.159	0.1146	0.1151	0.0927	
14	33.2942	-85.7789	57	0.2048	0.2316	0.2585	0.2097	0.1858	0.1983	0.2148	0.1699	0.1441	0.1316	0.1631	0.1907	
15	44.5333	-70.5333	40	0.1275	0.1181	0.1392	0.14	0.1412	0.1604	0.1517	0.1556	0.129	0.1562	0.1746	0.1596	
16	34.0358	-93.4217	55	0.1261	0.1505	0.1755	0.2129	0.2215	0.1689	0.1274	0.1045	0.1364	0.1464	0.1703	0.1485	
17	45.9336	-90.4506	57	0.0411	0.0351	0.0602	0.0929	0.1234	0.1546	0.1505	0.1472	0.1429	0.0856	0.0726	0.0463	
18	33.7164	-94.3814	56	0.1266	0.161	0.1938	0.209	0.2135	0.2026	0.1706	0.1317	0.1638	0.1475	0.1839	0.1642	
19	47.2614	-123.715	55	0.7219	0.6674	0.539	0.3542	0.2322	0.1564	0.104	0.1326	0.2183	0.4803	0.7617	0.7421	
20	36.1278	-79.4069	55	0.1273	0.1265	0.1493	0.1355	0.1346	0.1506	0.1586	0.1622	0.1421	0.1137	0.1101	0.1115	
21	42.8278	-90.7889	57	0.0412	0.0466	0.0889	0.1398	0.1678	0.1921	0.1674	0.1735	0.1629	0.0909	0.0861	0.0534	
22	42.1225	-78.2064	51	0.108	0.1136	0.1327	0.1491	0.142	0.2026	0.1724	0.1695	0.1825	0.1423	0.155	0.1386	
23	34.1864	-90.5572	52	0.1819	0.2218	0.214	0.2147	0.2295	0.1732	0.1676	0.1025	0.1244	0.1013	0.1921	0.1946	
24	38.9703	-89.0922	56	0.0859	0.0969	0.1328	0.1613	0.1662	0.1499	0.1514	0.0929	0.1103	0.1062	0.1133	0.0965	
25	42.9217	-93.2828	57	0.0324	0.038	0.0831	0.1371	0.1692	0.1931	0.1946	0.1612	0.1564	0.0951	0.0726	0.035	
26	38.0053	-91.3706	55	0.0716	0.0973	0.1287	0.1456	0.1734	0.1424	0.1258	0.107	0.1235	0.1089	0.112	0.0877	
27	36.4081	-80.9964	63	0.1615	0.1994	0.2079	0.1843	0.2155	0.2146	0.2623	0.248	0.264	0.1879	0.1846	0.1737	

Each column of the first line of the file has the text entry indicating the information that each column has. For example, the column named "Latitude" contains the latitude information of the gage on which MBLRPM is calibrated. The column named "Mean1_1" contains the mean rainfall (in millimeter) of the calendar month January at hourly accumulation level. The first 1 in Mean1_1 represents the accumulation level,

and the second 1 in Mean 1_1 represents the calendar month. In similar manner Var12_5 represents the variance of rainfall (in mm^2) of the month of May at the accumulation of 12 hours. The first 3 columns do not have to be accurate. It is not used in the optimization process but used in file input and output process. *The sequence of the column entry should correspond with the Line 57 of the file "\\BLRP_SourceCodes\Parameter Estimation\doit_2009-06-02_weighted_qp.r"*

Line 55 means that the variable "gage" runs from 1 to the number of rows of the file "\\BLRP_SourceCodes\Parameter Estimation\stats_192stations.csv"

Line 56 means that the variable "month" runs from 1 to 12.

Line 57 means that the "moment" property of the variable "s" are assigned with the entries in "\\BLRP_SourceCodes\Parameter Estimation\stats_192stations.csv" in the following manner: (1) "gageth row, (2) 3+monthth columns with the step of 12. In other words, if the variable "gage" is 2 and the variable "month" is 3, the variable "s$moment" will be assigned with the following elements of the file "\\BLRP_SourceCodes\Parameter Estimation\stats_192stations.csv": (row, column) = (2, 6), (2, 9), (2, 12), (2, 15), (2, 18), (2, 21), (2, 24), (2, 27), …, (2, number of columns - 9). *Line 57 should be manipulated with caution because it can provide wrong input to optimization algorithm, which will result in wrong parameter estimates.*

Line 80 defines the name of the output file. In this case, the name of the output file is "weightfactor_verification.txt"

Generation of Rainfall Time Series

Once the parameters of the MBLRPM are estimated, synthetic rainfall time series can be generated using a MATLAB code. The code is developed by the author of this dissertation. The code is located at "\\BLRP_SourceCodes\Synthetic Rainfall Generation\BLRP_Simulator.m." This file is also a complete package, thus does not need any further modification. The function can only be used in MATLAB Window. The input of the function is as follows:

```
A = BLRP_Simulator(par, simulation_length, plot_index);
```

,where the variable "par" is 1x6 matrix containing the parameters of the MBLRPM in the following sequence: λ, ν, α, μ, ϕ, and κ; the variable "simulation_length" is the length of the desired synthetic rainfall time series *in the unit of hour*; If the variable "plot_index" is 1, hyetograph of the synthetic rainfall time series in drawn as an output of the function. Otherwise, set the variable "plot_index" as zero. The following example shows the usage of the MATLAB function to generate synthetic rainfall time series based on MBLRPM. Generate 50 years of synthetic rainfall time series with a given parameter set (λ, ν, α, μ, ϕ, κ) = (0.02534, 2.1452, 14.81, 5.21, 0.01549, 0.13198).

```
synthetic_hyetograph = zeros(50, 8760); % set up initial variable

par = [0.02534 2.1452 14.81 5.21 0.01549 0.13198]; % define parameter values

for i = 1:50 % loop for year (repeat 50 times)
    synthetic_hyetograph(i,:) = BLRP_Simulator(par, 8760, 0); % synthesize yearly hyetograph
end

% Plot the hyetograph year by year
for i = 1:50 % loop for year (repeat 50 times)
    figure(i) % open the new figure
    bar(synthetic_hyetograph(i,:)); % draw the hyetograph
end
```

VITA

Dong Kyun Kim received his Bachelor's degree in civil engineering from Hanyang University, Korea in 2001. Then, he received Master of Science degree in environmental fluid mechanics and hydrology from Stanford University in 2003. His research interest includes space-time stochastic rainfall modeling, distributed hydrologic modeling using GIS, and development of GUI-based custom software tools that are related to hydrology and water resources engineering.

He is currently the research engineer at Dewberry Inc., 8401 Arlington Blvd., Fairfax, VA, 22031, USA. His email is dekaykim@gmail.com.

Education: B.S., Civil Engineering, Hanyang University, 2001

M.S., Environmental Fluid Mechanics and Hydrology, Stanford University, 2003

Ph.D., Water Resources Engineering, Texas A&M University