12-2013

Improving Time Structure Patterns of Orthogonal Markov Chains and its Consequences in Hydraulic Simulations

Juan C. Jaimes-Correa
University of Nebraska-Lincoln, jcjaimes@huskers.unl.edu

IMPROVING TIME STRUCTURE PATTERNS OF ORTHOGONAL MARKOV CHAINS AND ITS CONSEQUENCES IN HYDRAULIC SIMULATIONS

by

Juan Carlos Jaimes Correa

A THESIS

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Master of Science

Major: Natural Resource Sciences

Under the Supervision of Professor Guillermo A. Baigorria

Lincoln, Nebraska

December, 2013

IMPROVING TIME STRUCTURE PATTERNS OF ORTHOGONAL MARKOV CHAINS AND ITS CONSEQUENCES IN HYDRAULIC SIMULATIONS

Juan Carlos Jaimes Correa, M.S.

University of Nebraska, 2013

Adviser: Guillermo A. Baigorria

Rainfall[1] occurrences understood as rain events are relevant for agricultural practices because temporal distribution of rainfall highly affects yield production. A few stochastic models satisfactorily generate daily rainfall events while preserving temporal and spatial dependence among multiple sites. I evaluated an extension on the traditional Orthogonal Markov chain (TOMC) model in reproducing the temporal structure of rainfall events at multiple sites in Florida (FL), Nebraska (NE) and California (CA). In addition, a simulation of watershed runoff from rainfall events reproduced by a single- and multi-site weather generator was conducted. Results shows that (*i*) a temporal structure extended Orthogonal Markov chain (EOMC) maintained the spatial correlation between observed and generated rainfall events; (*ii*) EOMC used a smaller number of yearlong simulations for generating the observed frequencies of wet spells than TOMC requires for similar accuracy; (*iii*) using EOMC generated rainfall data in SWMM produced similar median runoff values to those generated using observed data; and (*iv*) EOMC reduces 50% of computing time for generating rainfall data. EOMC can benefit modeling of future climate scenarios by economical reduction of hardware need.

[1] Liquid precipitation

DEDICATION

To Melina, your enthusiasm and support are admirable.

To Laura, your energy and kind are unlimited.

To the rest of my family members, especially to Cesar, Mauricio and his family, because you have been a treasure uncovered.

To friends who are from many countries.

ACKNOWLEDGEMENTS

I want to recognize the great contribution from my supervisory committee members: Dr. Betty Walter-Shea, Dr. Martha Shulski, and especially to my advisor Dr. Guillermo Baigorria. His guidance was invaluable feedback to this project.

GRANT INFORMATION

I am really thankful to the Fulbright Foreign Student Program and the Administrative Department of Science, Technology and Innovation of Colombia (Colciencias). This research was possible by the scholarship program “Becas Caldas: Fulbright Colombia – Colciencias”.

TABLE OF CONTENTS

List of Figures

List of Tables

List of Maps

List of Equations

Chapter 1. Introduction

Because rainfall regimes regulate agriculture and limit water availability, rainfall data are a key variable used for water resources management and agricultural modeling. Daily rainfall amounts are important in areas with scarcity of water where rainfall mainly supplies water demands, and also relevant for water management including flood prevention, and meeting water supply demands for cities, industry and agriculture. The availability of rainfall data is sometimes spatially limited. A solution for this issue is to generate synthetic rainfall data that preserves the observed statistics of actual rainfall.

Synthetic rainfall data have several applications such as design and operation of water systems, collection systems (Mhanna and Bauwens 2012), urban drainage systems, land use management, and impacts studies, flood risk assessment, climate scenarios (Tseng et al. 2012; Hayhoe 2000; Semenov and Barrow 1997), and runoff modeling (Kim et al. 2007; Zhang and Switzer 2007).

Rainfall occurrence can be described in a Markov-chain model assuming that the '*state'* of rainfall (event or no event) on a given day relates to the state of the previous days. The use of stochastic models for generating rainfall data has been widely researched, and generally, they reproduce statistical characteristics of event durations (Wan et al. 2005; Roldan and Woolhiser 1982; Richardson 1981). Nevertheless, some rainfall models may underestimate the variance of rainfall, especially in the tropics and lowland tropics (Jones and Thornton 2000). Lowry and Guthrie (1968) recognized that

the first order Markov chain can estimate probabilities of events, but a higher order may be required in areas with diverse climatology. Furthermore, the generation of daily data across multiple sites has been challenging, and few methods have been used satisfactorily to generate daily rainfall occurrence while preserving spatial dependence among different sites (Hughes and Guttorp 1994; Wilks 1998; Khalili et al. 2007; Baigorria and Jones 2010).

For these reasons, some questions are important to consider for generating synthetic rainfall data while preserving spatial correlation: Can an increase in temporal estimation of Markov chain result in a statistical improvement in generating synthetic rainfall data across multiple sites? What is the impact of using a point vs. multi-site weather generator in simulating runoff in a watershed? Do historical and generated rainfall data differ significantly from runoff modeling?

1.1 Objectives

a) To validate the extension of Orthogonal Markov chain algorithms as used in the GiST weather generator to estimate the temporal structure of rainfall occurrences.

b) To analyze the watershed runoff by using rainfall occurrences reproduced by a single- and multi-site weather generators.

1.2 Background

Rainfall data vary widely temporally and spatially (Srikanthan and McMahon 2001) so that a very dense network of precipitation measurements is needed to adequately describe rainfall patterns and occurrences. Such networks are not common. Thus, the generation of rainfall occurrence (Srikanthan et al. 2005; Ng and Panu 2010; Brissette et al. 2007) and amounts (Mhanna and Bauwens 2012; Charles et al. 1999) based on statistical relationships has been intensively researched. Rainfall occurrences are described using a Markov chain assuming that the state of rainfall on a given day relates to the state of prior days, or alternatively, sequences of dry and wet series.

Three main types of models for generating daily rainfall are recognized: conventional Markov chain models (Haan et al. 1976; Wan et al. 2005), non-parametric models (Srikanthan et al. 2005; Lambert 2003) and hybrid models (Thyer and Kuczera 2003).

Conventional stochastic rainfall and non-parametric models have a first component for modeling the occurrence of daily rainfall based on a first-order Markov

chain, and a second component for simulating the amount of rainfall founded on wet days by conventional models; however, non-parametric models use a data resampling that is a conditional boostrap for generating rainfall occurrences (Harrold et al. 2003). The hybrid approaches for generating rainfall consider a hidden-Markov model (Thyer and Kuczera 2003) and a semi-nonparametric model (Kim and Valdés 2005). Hidden-Markov models assume the climate has two stages (wet or dry) and has an independent rainfall distribution. Hidden models can simulate the influence of a complex climate process on long-term hydrological time series, assuming a 'regional' climate state (Thyer and Kuczera 2003). A first-order Markov model represents a useful approach for generating rainfall occurrences, while nonparametric and hybrid models have limitations on extrapolation and application to daily and annual rainfall, respectively (Kim et al. 2007).

Weather generators simulate daily weather data. Most of them generate rainfall data for a single site; but, they are limited to reproduce spatial correlations between several locations, when a strong spatial correlation often exists (Wilks 1999; Srikanthan and McMahon 2001). Jones and Thornton (2000) note that rainfall models such as WGEN, SIMMETEO and WeatherMan tend to underestimate the variance of monthly and annual rainfall for many sites in the tropics and subtropics. Increasing the Markov chain can yield a better simulation of rainfall data where the rainfall regime has high variability. For instance, *MarkSim* software uses a third-order Markov chain to generate daily weather data for Latin America and Africa with satisfactory results (Jones and Thornton 2000), and relatively minor discrepancies with statistical differences at the 5% level between generated and observed rainfall data (Jones and Thornton 1999).

Hayhoe (2000) replaced the conventional first-order Markov chain in WGEN with a second-order Markov chain for reproducing weather data. Generated mean values and standard deviations for both wet and dry spells matched with observed data of thirty years (1960-1989) in three Canadian locations. Wet spell periods between 1 and 4 days occur more frequently in study locations than larger wet periods from 6 to 15 days in those locations (Hayhoe 2000). Additionally, there were no significant differences ($p \leq 0.05$) between observed and generated distribution of dry spells.

1.2.1 Markov chain

The Markov chain model simulates a time series of discrete random variables. This model is described by 'state' or the number of values that the variable can take, and 'order' or the number of previous values used to determine the state of transition probabilities (Schoof and Pryor 2008). A Markov chain specifies the state of each day as 'wet' (1) or 'dry' (0), and develops a relation between the state of the current day (*t*) and the states of preceding days (*t-n*). The order is defined by the number of previous days (*n*) considered. The threshold between dry and wet days is set at 0.1 mm. A wet day must surpass the minimum threshold. Otherwise, the rainfall state for a given day is set to zero (0) or dry state.

1.2.1.1 The original transition probabilities

The original Markov model (Markov 1906) was first used by Gabriel and Neuman (1962) for generating rainfall occurrences. The Markov chain is a two-state, first-order model, described by two pairs of transition probabilities (Schoof and Pryor 2008):

$$\begin{bmatrix} P_{00} = \Pr\{X_t = 0 | X_{t-1} = 0\} \\ P_{10} = \Pr\{X_t = 1 | X_{t-1} = 0\} \end{bmatrix}$$

$$\begin{bmatrix} P_{01} = \Pr\{X_t = 0 | X_{t-1} = 1\} \\ P_{11} = \Pr\{X_t = 1 | X_{t-1} = 1\} \end{bmatrix}$$

Equation 1.1 Original transition probabilities

Where X is the state of occurrence (rain $X = 1$; no rain $X = 0$) for a given day (t) or previous day (t-1). 0 and 1 are the two Boolean values of events. P_{10} is the probability that rainfall occurs for a given day if rainfall did not occur on the previous day, while P_{11} is the probability that rainfall occurs for a given day, if rainfall occurred on the previous day. Then, the model is defined by two transition probabilities: $P_{00} + P_{10} = 1$, and $P_{01} + P_{11} = 1$. These probabilities are calculated from rainfall time series records.

1.2.1.2 Order of Markov chain

Varying results in using first-order, second- or higher orders of Markov chain are found in the literature. Results differ depending on climate characteristics, statistical tests and length of record (Srikanthan and McMahon 2001). For example, a study across the USA showed that using a first-order Markov chain was unjustified; instead, an adequate order should be determined (Chin 1977). Results also underscored that the order depends on seasonal and geographic characteristics. For example, a first-order chain was used to simulate months in summer, while second- or higher-order estimated data in winter (Chin 1977). A second-order Markov chain for reproducing daily rainfall data performed well in terms of short-term temporal dependency, wet/dry spell length and goodness-of-fit (Ng and Panu 2010). A first-order Markov chain reproduced statistics (mean and maximum

lengths) of dry and wet spells in Canada; nevertheless, it was unable to reproduce interannual variability for a number of wet day events (Wan et al. 2005).

In summary, while some researches justify the use of a first-order Markov chain, others show that a second or even higher order can improve the results during certain times of the year, such as winter in temperate zones (Hopkins and Robillard 1964; Chin 1977), and rainy seasons in the tropics (Jones and Thornton 2000).

1.2.2 Orthogonal Markov Chain

Models for generating rainfall data in single-sites reproduce main statistics of the historic data; but, they have constrains in reproducing spatial correlations between multiple locations. Daily single rainfall events at a given station correlates with events at other weather stations during rainy seasons (Baigorria et al. 2007). Spatial dependency exists when a process that links different locations limit observations on a given place to what happens in the surrounding area; spatial correlation refers to the strength with which the occurrence of that process in a given location affects others (Khalili et al. 2007). Spatiotemporal correlation can estimate synthetic rainfall events that preserve both temporal and spatial correlation. As a result, geographical rainfall patterns in a study area are identified.

Baigorria and Jones (2010) compared a TOMC in GiST-wg (see Sec. 1.2.3.2 Geospatial and temporal weather generator GiST for a description of the method) and WGEN two-state first-order Markov chain approach for generating daily rainfall amounts and occurrences at multiple sites in Florida. While WGEN preserved no correlation (ρ = -0.082), generated correlations of daily rainfall events among all pairs of seven location

correlated with monthly observed (30 years record) for a thousand yearlong simulation (ρ = 0.932). The correlations of rain and non-rain events in two locations on the same day (joint probabilities for each month) were 0.980 in both cases using TOMC. Nevertheless, WGEN exhibited correlations equal to 0.045 and 0.131 for rainfall and non-rainfall correlations, respectively (Baigorria and Jones 2010).

The TOMC was also used for generating rainfall data using historic data for 25 weather stations in North Carolina. A 1000 yearlong simulations revealed that correlation for rainfall occurrences was 0.966, and the correlations between joint probabilities were 0.930 and 0.951 (Baigorria and Jones 2010).

1.2.2.1 Transition probabilities

In the Orthogonal Markov chain (Baigorria and Jones 2010), the transition probabilities for rainfall events (P_{i_t}) in a given location *i* at time *t* depends on two conditions: (a) the states in two other locations (*j*, *k*) at the same time *t*, and (b) the state in the same location *i* at time (*t*-1). The probability is indicated as:

$$P_{i_t \mid j_t,\ k_t, i_{(t-1)}}$$

Equation 1.2 Transition probabilities in the Orthogonal Markov chain

Where,

i, j, k locations

t time for a specific day

For example, $P_{1|0,1,1} = \Pr\left(X_{i_t} = 1 | X_{j_t} = 0, X_{k_t} = 1, X_{i_{(t-1)}} = 1\right)$ is the probability of rainfall occurring at location *i* on given day *t.* The probability is dependent on rain or no rain events at two other locations, *j* and *k*. In this example, location *j* did not receive rain on that same day *t*, and location *k* received rain on the day *t*, while location *i* did not receive rain the previous day (*t*-1). Recall, 0 and 1 are the two-state Boolean values of a no-rainfall or a rainfall event, respectively.

Because the TOMC has only two states (rain and no-rain), the transition probabilities of having rainfall and the transition probabilities of having no rainfall must be equal to 1:

$$\Pr\left(X_{i_t} = 1 | X_{j_t}, X_{k_t}, X_{i_{(t-1)}}\right) + \Pr\left(X_{i_t} = 0 | X_{j_t}, X_{k_t}, X_{i_{(t-1)}}\right) = 1.$$

The total number of transition probabilities is given by 2^n, where *n* is the number of variables. The transition probabilities for the TOMC in this example are 16 (= 2^4) or eight pairs. This study will have 32 (= 2^5) probabilities or sixteen pairs of transition probabilities.

1.2.3 Weather generators

Weather generators use existing weather data and random number sampling to produce time series of weather data at daily and hourly steps while preserves statistics in original data. Weather forecast predicts future conditions based on current state of the atmosphere and its processes. On the other hand, climate predictions are usually probabilistic predictions to produce the evolution of the future climate at seasonal, interannual or long-term time scales.

Several weather generators have been developed for generating data on rainfall occurrences and amounts as well air temperature and solar radiation conditions based on wet/dry states and statistical distributions of observed data. These generators are recognized as useful tools for the evaluation of future climate scenarios in agricultural impact assessments (Semenov and Barrow 1997).

The three main types of weather generators are: (1) parametric [e.g., "weather generator" WGEN model (Richardson 1981; Richardson and Wright 1984), "extended version of weather generator" WXGEN (Wallis and Griffiths 1995), "weather generator for climate inputs" CLIGEN (Nicks et al. 1995), "generation of weather elements for multiple applications" GEM (Johnson et al. 2000), "spectral generator" SPECGEN (Schoof et al. 2005), SIMMETEO (Geng et al. 1986), WeatherMan (Pickering et al. 1994), and "Geospatial and Temporal Weather Generator" GiST-wg (Baigorria and Jones 2010)]; (2) semi-parametric/empirical [e.g., "Long Ashton Research Station Weather Generator" LARS-WG (Semenov and Barrow 1997)]; (3) non-parametric, such as with the resampling approach [e.g., (Brandsma and Buishand 1998; Buishand and Brandsma 2001), in which precipitation and temperature are modeled at multiple sites while allowing for a linkage with atmospheric circulation (Brandsma and Buishand 1998)].

1.2.3.1 Reproducing rainfall occurrences

Parametric weather generators that produce a rainfall occurrence or event *state* using a random number with a two-state Markov chain include: (1) first-order Markov chain [WGEN (Richardson and Wright 1984), WXGEN (Hayhoe and Stewart 1996), and CLIGEN (Nicks et al. 1995)]; (2) a second-order Markov chain [SPECGEN (Schoof et al. 2005)]; and (3) Orthogonal first-order Markov chain [GiST-wg (Baigorria and Jones

2010)]. Semi-parametric/empirical weather generators [LARS-WG (Semenov and Barrow 1997)] can simulate rainfall based on distributions of the length of continuous sequences of wet and dry days.

The weather data generators WXGEN and CLIGEN were used for generating 30 years of weather data at five locations in Canada, and compared with 30 years of observed data. WXGEN had differences no higher than 1 day in simulating the number of monthly rainy days; but, CLIGEN differed by 2 days from observed data (Hayhoe and Stewart 1996). Moreover, the mean number of rainy days generated for 28-day periods differed on average by less than 2 days in comparison with 20 years of observed data at three locations in USA (Richardson 1981).

A two-state, second-order Markov model used by SPECGEN reproduced the mean annual number of rainy days with differences lower than 4 days between generated and observed data (at least 1950–2003) for 9 locations in the Southeast USA (Schoof et al. 2005). Based on a *t*-test, the differences are not statistically significant ($\alpha = 0.05$) in any of the weather stations tested. The generated monthly mean number of rainy days differs from observed data by less than 1.5 days for all months and all weather stations. The mean number of generated wet days is not significantly different from the observed means ($\alpha = 0.05$).

The monthly mean of the generated dry spell showed similarities within less than 3 days with the monthly observed data at 3 locations in Europe using the stochastic weather generator LARS-WG (Semenov and Barrow 1997). However, generated values

of monthly dry spell were larger than observed data for about 10 days at 2 other locations from August to October.

Methods that satisfactorily generate daily rainfall occurrence while preserving spatial dependence among different sites are: Space– Time Rainfall Occurrence Model STROM (Kim et al. 2007), GiST-wg (Baigorria and Jones 2010), *RainSim* (Burton et al. 2008), and the Wilks approach (Wilks 1998). While *RainSim* model simulates seasonal cycles and orographic effects (Burton et al. 2008), GiST-wg generates spatially and temporally correlated daily rainfall data in multiple sites based on the TOMC (Baigorria and Jones 2010).

Rainfall data generated in multiple sites were better when using the GiST-wg model than those when using WGEN. GiST reproduced daily rainfall record data in single weather stations as well as spatial correlations between pairs of stations, locations and monthly observed correlations, and the number of days without rain (Baigorria and Jones 2010). Because of the superior GiST-wg performance, this model was used in this study.

1.2.3.2 Geospatial and temporal weather generator GiST-wg

GiST-wg uses four statistical parameters for generating rainfall occurrences (Baigorria and Jones 2010). These are (1) Pearson's coefficient correlations (ρ_{ij}) between all pairs of weather stations, (2) weather station ranking based on Euclidean correlations, (3) transition probabilities of rainfall events, considering events in the study locations on the previous day, and in other locations on the current day, and (4) distribution of the number of daily rainfall events in each month.

1.2.3.2.1 Pearson's coefficient correlation

Pearson's coefficient correlations are calculated as:

$$\rho_{ij} = \frac{1}{\eta} \frac{\sum_{t=1}^{\eta} (X_{i_t} - \mu_i)(X_{j_t} - \mu_j)}{\sigma_i \sigma_j}$$

Equation 1.3 Pearson's coefficient correlations between pairwise weather stations

Where,

μ_i, μ_j: mean of daily rainfall values of locations i, j

X_i, X_j: pairwise observations on day t of locations i, j

σ_i, σ_j: standard deviation of daily observations

η: total number of pairwise daily observations

1.2.3.2.2 Euclidean N-correlation distance

Prior to producing rainfall occurrences in a given location, the rainfall events are simultaneously generated for two core locations with the smallest Euclidean N-correlation distance (G_i). G_i is the degree of association or strength of a variable in a given location with another location (Equation 1.4).

$$G_i = \sum_{j=1}^{n-1} \left(1 - |\rho_{i,j}|\right)^2$$

Equation 1.4 Euclidean N-correlation distance between two core locations

By ranking *Gi* values of all locations in ascending order, the first location (the smallest G_i) has the highest correlation with all remaining locations and the last is the most independent one.

1.2.3.2.3 Generating rainfall occurrences using GiST-wg

A minimum density of three locations, and meteorological record for at least 30 years is desirable to generate rainfall events. It allows analysis of historic records and estimation of main statistics that describe monthly and yearly rainfall patterns. Semenov and Barrow (1997) suggested observed daily data of at least 20 years for estimation of statistical parameters, while Hayhoe (2000) used observed data of 30 years (1960-1989) at three Canadian locations for generating rainfall data.

Two steps are followed using GiST-wg for generating daily rainfall events. The first step considers the selection of two core locations based on the best correlation of rainfall events on day *t*, and spatially correlated number of rainy days per month in these locations. By doing so, the observed spatial correlations between both locations can be reproduced in a specific month. A reordering process is accomplished iteratively by sorting the generated patterns of daily rainfall event that reduces RMSE between the generated Markov transition probabilities and observed Markov transition probabilities in both locations at the same time.

The second step involves the generation of daily rainfall events for a third station using two-state Orthogonal Markov transition probabilities. Thereafter, events are generated for one station at a time, based on two previously most-correlated weather stations and the previous state of the station being generated. For the first generated day, the previous state is calculated by using the unconditional climatological probability of rainfall. Next, to assign a rainfall event for each new day, a random number from the uniform distribution ($r_{unif} \sim U[0, 1]$) is compared to the corresponding two-state Orthogonal Markov transition probability of a rainfall event (Baigorria and Jones 2010).

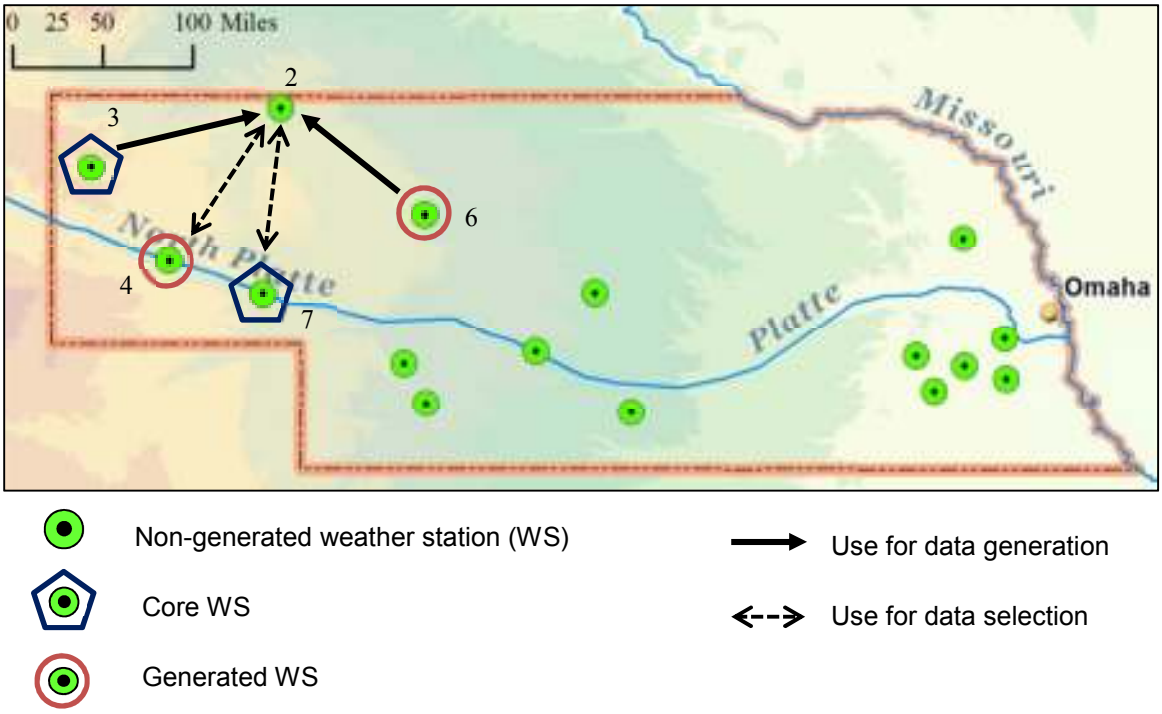


Figure 1.1 Example of the nearest Euclidean three-correlational neighbor assimilation process for weather stations in Nebraska

An ensemble of equally probable realizations is computed for each new location. To select the ensemble member to keep, correlations are calculated between each ensemble member and each of the previously generated locations not used during the generation (Baigorria and Jones 2010). For instance in Figure 1.1, data for location 2 is generated using locations 3 and 6, while it is correlated to observations from locations 4 and 7. The ensemble member that reduces the RMSE is the one that is selected.

1.2.4 Modeling Watershed Runoff

Rainfall records or synthetic data are used in hydraulic modeling watershed as inputs for simulation of the rainfall-runoff process in space-time. Simulation outputs allow the evaluation of future water-related scenarios (flood, evapotranspiration related to

climate change, growing population water demands, pollutant transport) as well as the analysis of possible solutions to water demands in the watershed.

1.2.4.1 Hydrological models using synthetic daily rainfall

Haan et al. (1976) used a first-order Markov chain for reproduction of the daily rainfall data. Similarities between the generated (six simulations of 40 years) and historical data were found. Averages of annual number of wet days differed no more than 2 days in comparison with the historical data at seven different stations in Kentucky, USA.

Results of the simulations and recorded data were used as input in a hydrologic model to generate monthly runoff at the stations. Runoff simulations using the synthetic rainfall data exceeded annual average of stream flows estimated from historical precipitation data by about 1 inch.

Generated long term rainfall data were simulated using a two-stage resampling algorithm (Leander and Buishand 2009), and subsequently used as input in the HBV hydrological model (Lindström et al. 1997) for estimation of synthetic daily catchment discharges. Results showed that distribution of the winter maximum daily discharge for the river Ourthe at Tabreux (Belgium) was not noticeably affected by larger daily precipitation amounts.

1.2.4.2 Storm Water Management Model (SWMM)

Selection of an adequate model for a specific project is always a challenge. Many watershed models have been developed such as DWSM (Dynamic Watershed Simulation Model), HSPF (Hydrological Simulation Program-Fortran), Precipitation-Runoff

Modeling System (PRMS) and Hydrologic Engineering Center-Hydrologic Modeling System (HEC-HMS) [See Borah (2011) for brief descriptions].

SWMM is a storm water modeling program that has applications in hydrology and hydraulics of rainfall-runoff processes in watersheds. It represents a watershed based on the degree of imperviousness, percentage of urban development, and stream networks and sewers. Although such simplicity can be helpful for configuration of a watershed, it might lead to a lack of accuracy on simulating the watershed hydrology.

SWMM model is recommended for watershed management in small scale urban areas (Lee et al. 2010). It simulates runoff responses of watersheds (hydrographs) to rainfall events, which can be compared with observed data (Lee et al. 2010; Wang and Altunkaynak 2012). Applications of the SWMM can include integration with SWAT, for hydrological simulation of watersheds with both rural and urban areas (Kim et al. 2007). SWMM has also been used for simulation of water quality and quantity in small urban catchments (Tsihrintzis and Hamid 1998).

Runoff and flow routing procedures have the most influence on watershed model performance (Borah 2011). SWMM has moderate levels of physical basis and complexity for overland runoff routing, and a strong physical basis and equation complexity for channel or pipe flow routing, when compared with other models (Borah 2011). However, the gain and losses between main stream channel and ground water, many different land-uses and irrigation demands might be limited. Moreover, a large watershed with a broken geography needs long time for setting it up.

SWMM was used in this study for a sensitivity analysis using different sets of rainfall occurrences: the historical records, the EOMC (using 2 previous days) and a point site weather generator (WGEN) as implemented in the decision support system for agrotechnology transfer (DSSAT) (Hoogenboom et al. 2012; Jones et al. 2003).

Chapter 2. Modeling Rainfall Occurrences

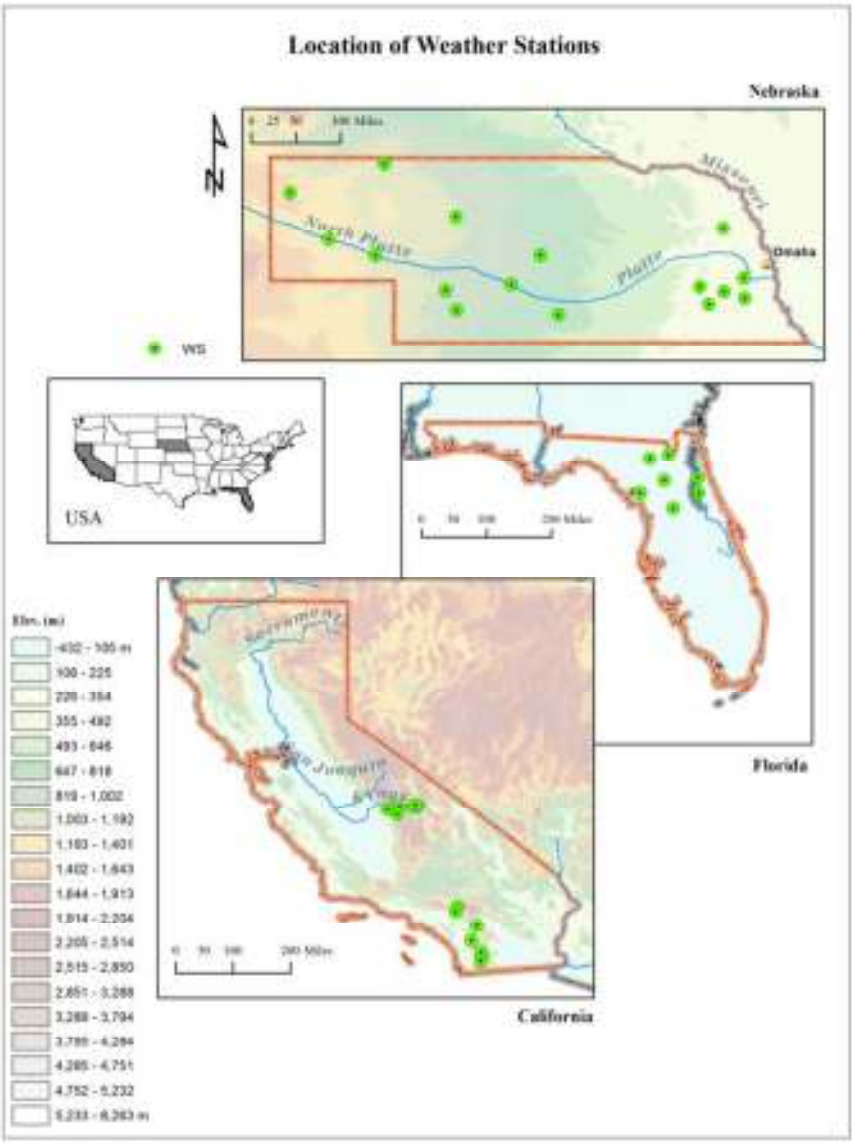
2.1 Methods

2.1.1 Study Sites

Study of three regions of the USA with different rainfall regimes were selected for the evaluation of the EOMC: 1) California (CA) and 2) Nebraska (NE) representative of a mountainous region and a temperate climate respectively, and 3) Florida (FL) characterizing a sub-tropic climate.

A set of weather stations across each state were selected as points of interest. Historical daily rainfall records were obtained from the National Oceanic and Atmospheric Administration (NOAA)/National Climate Data Center and Applied Climate Information System. The Map 2.1 presents the location of weather stations in the study sites. The geographic and meteorological information for each site are summarized in Table 2.1.

Periods of rainfall data were 1971-2000, 1920-2000 and 1911-2010 for weather stations in FL, NE and CA respectively. Weather stations with records less than 30 years in length were excluded. A further identification of missing values revealed some weather stations in NE and CA lacked more than 1,000 daily data points. Seven, six and five weather stations respectively from FL, NE and CA met the selection criteria and were used in generating daily rainfall occurrences.



Map 2.1 Location of the weather stations in the study areas

A set of maps of annual rainfall averages in the contiguous USA states are shown in Appendix C. CA weather stations are located in areas with two different rainfall regimes. One zone characterized by dry conditions with annual rainfall amounts below 125-250 mm/year and a second area that is wet with 1,000-1,250 mm/year of rain. NE has a clear transition of rainfall regimes from drier conditions in the western to wet in the southeastern. Values vary between 250 mm/year and 875 mm/year. Finally, FL in the

southeastern USA has a high annual rainfall. It receives rainfall amounts around 1,250-1,500 mm/year.

Mean values of rainfall for weather stations in FL, CA and NE are presented in Table 2.1. Stations in FL have annual averages of rainfall amounts and rainy days between 1,250-1,516 mm/year (49-59 in/year) and 107-120 days/years respectively. CA has a large difference in annual averages of rainfall amounts and rainy days. These values range from 127 to 1,157 mm/year (5-45 in/year) and rainy days vary from 23 to 70 days/year. Thus, there is a high spatial variation of rainfall regimes between weather stations in CA while weather stations in FL have a rainfall pattern spatially correlated. Locations in NE show annual averages of rainfall amounts about 392-728 mm/year (15-28 in/year). Annual rainy days in NE have less variation between weather stations with values around 62-85 days/year. In general, NE and CA have rainfall regimes with lower annual averages of rainfall amounts and rainy days than locations in FL.

Locations in FL have annual minimum air temperatures that are greater than 12°C. All weather stations in NE and most of those in CA have minimum air temperatures below 0°C. NE has the lowest minimum temperatures (-3 to -12 °C). There are no significant differences in maximum air temperatures between study areas.

Table 2.1 Meteorological information of weather stations in study locations

STATION	ELEV. (MASL)	LAT. N	LONG. W	TMAX (°C)	TMIN (°C)	RTOT (mm)	RNUM (days)
NEBRASKA							
AGATE 3 E	1423	42.424	-103.735	26.3	-12.4	392.1	62.7
ASHLAND 2	326	41.041	-96.378	27.7	-5.8	727.0	78.2
BRIDGEPORT	1117	41.668	-103.104	27.8	-8.6	415.5	68.0
BROKEN BOW 2 W	762	41.408	-99.675	27.6	-8.6	566.5	76.8
CRETE	437	40.619	-96.947	27.2	-5.0	717.1	81.0
GORDON 6 N	1128	42.895	-102.204	26.7	-10.2	456.4	75.6
GOTHENBURG	788	40.939	-100.151	28.3	-6.8	550.9	74.7
HAYES CENTER	928	40.523	-101.035	28.7	-5.6	542.2	71.8
HOLDREGE	707	40.452	-99.380	28.0	-5.0	622.8	77.1
LINCOLN UNIVERSITY POWER PLANT	354	40.823	-96.703	27.5	-3.9	753.4	82.9
MULLEN	981	42.043	-101.046	27.6	-8.5	535.6	65.8
OSHKOSH	1033	41.401	-102.347	27.8	-8.3	438.1	63.0
SEWARD	440	40.900	-97.091	27.1	-5.2	693.8	77.2
WALLACE 2 W	945	40.843	-101.209	28.1	-8.0	475.6	67.6
WEST POINT	399	41.845	-96.714	26.5	-6.7	728.4	85.6
CALIFORNIA							
CUYAMACA	1414	32.990	-116.587	26.1	-1.5	904.0	51.9
DESCANSO RANGER STATION	1067	32.850	-116.617	26.1	-5.9	585.1	47.9
GRANT GROVE	2007	36.733	-118.967	21.2	-5.0	1074.6	63.1
IDYLLWILD FIRE DEPARTMENT	1640	33.757	-116.707	26.7	-3.3	624.3	44.4
INDEPENDENCE	1206	36.800	-118.200	30.5	0.7	127.3	23.0
JULIAN CDF	1285	33.076	-116.593	27.2	1.4	626.5	41.7
LAKE ARROWHEAD	1587	34.247	-117.188	24.1	-1.3	1005.1	41.5
LODGEPOLE	2056	36.600	-118.733	19.7	-8.3	1157.4	70.2
PALOMAR MOUNTAIN OBSVTRY	1692	33.378	-116.840	25.6	0.3	692.4	41.0
SAN BERNARDINO F S 226	348	34.134	-117.254	34.7	4.3	412.7	42.2

FLORIDA							
CRESCENT CITY, PUTNAM	20	29.417	-81.517	27.1	15.9	1315.1	110.2
OCALA, MARION	23	29.083	-82.083	27.7	15.0	1336.4	115.7
LAKE CITY 2 E, COLUMBIA	59	30.183	-82.600	26.5	14.0	1329.6	112.9
GLEN ST MARY NURSERIES, BAKER	39	30.267	-82.183	26.6	12.8	1383.7	103.3
FEDERAL POINT, PUTNAM	2	29.750	-81.533	26.9	15.5	1313.5	120.5
GAINESVILLE MUNI ARPT, ALACHUA	38	29.700	-82.283	26.4	14.1	1250.0	116.0
USHER TOWER, LEVY	10	29.417	-82.817	27.0	13.8	1516.7	107.9

MASL: meters above sea level
TMAX: annual average of maximum temperature
TMIN: annual average of minimum temperature
RTOT: annual average of total liquid rainfall
RNUM: annual average of total number of rainy days

2.1.1.1 Rainfall regimes

The monthly mean rainfall and number of rainy days in the weather stations in FL, NE and CA are shown in Figure 2.1, Figure 2.2 and Figure 2.3 respectively.

2.1.1.1.1 Florida

The rainfall amount varies between 50 and 180 mm/month (2 and 7.1 in./month) for the study locations in FL. A rainy season between June and September is characterized by around 180 mm/month (7.1 in./month), which corresponds to the convective season that generates higher amounts of rainfall than the rest of the year (Ali et al. 2000; Baigorria et al. 2007). These months are marked by cyclone activity that form tropical storms, and become hurricanes. For instance, seven hurricanes and four major hurricanes occurred in 2011 (Avila and Stewart 2013). Tropical cyclones lead to extreme rainfall, exceeding 20 mm/h between 1977-2001 (Black and Hallett 2012), and in turn peak flooding events (Smith et al. 2010).

Between December and March (advection season), rainfall amounts decrease about 50-100 mm/month (2-4 in./month). Rainy days are 6 days/month in advection season (December-March) and 16 days/month in convective season (June-September). April and November have lower rainy days with 5 days/month.

There is a large difference in rainfall regimes between May and June. May has 6-7 rainy days monthly and rainfall amounts of 70-80 mm/month (2.8-3.1 in./month) and June has 12-13 days/month with 180 mm/month (7.1 in./month). September is a transition period between rainy season and dry season.

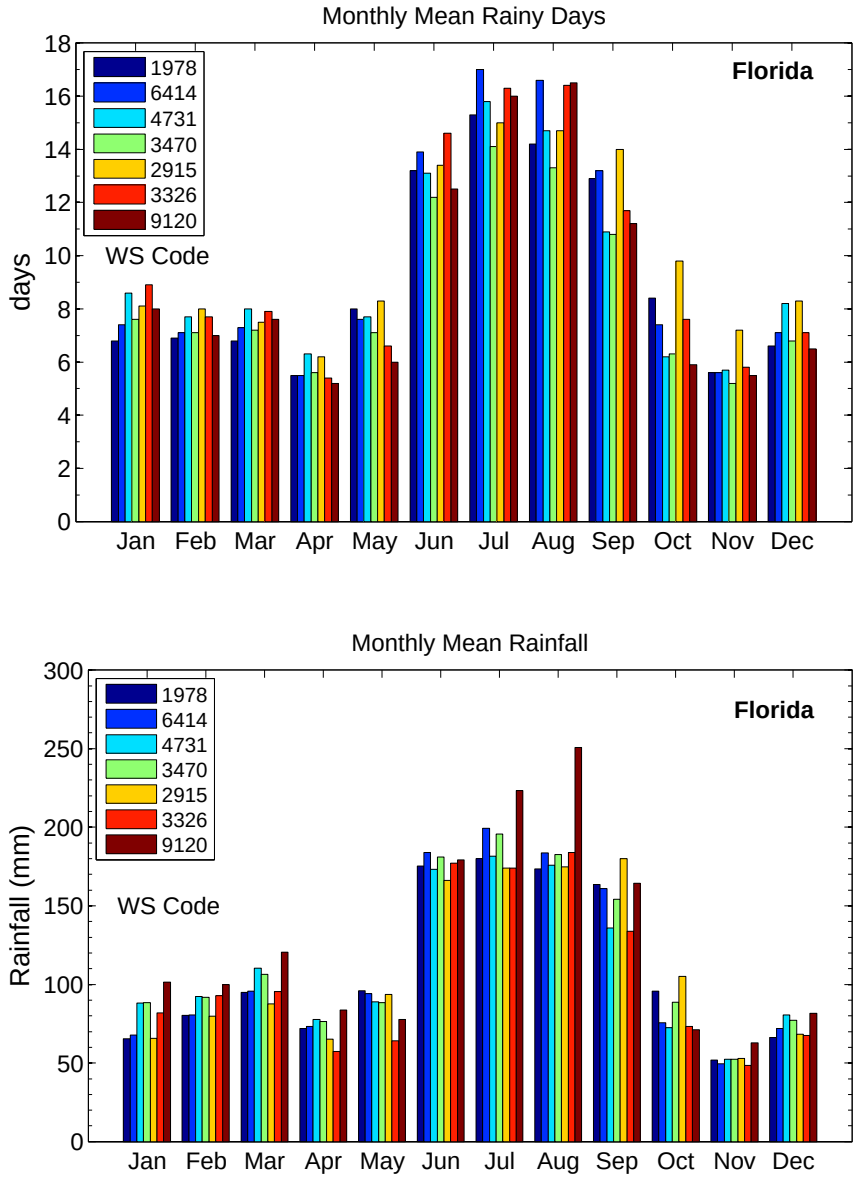


Figure 2.1 Monthly mean of rainfall amounts and rainy days in weather stations of Florida

2.1.1.1.2 Nebraska

Rainfall amounts for the study locations in NE are less than 120 mm/month (4.7 in./month) (Figure 2.2). Rainfall amounts increase from February to June, varying from 10-20 mm/month (0.4-0.8 in./month) to 110 mm/month (4.3 in./month) and decrease between July and November (Sharma and Irmak 2012).

A dry season occurs between November and February with only 2-4 rainy days a month with rainfall amounts below 20 mm/month (0.8 in/month). A rainy season exists during the months of May and June with around 9 days/month and rainfall amounts of 100-120 mm/month (3.9-4.7 in/month).

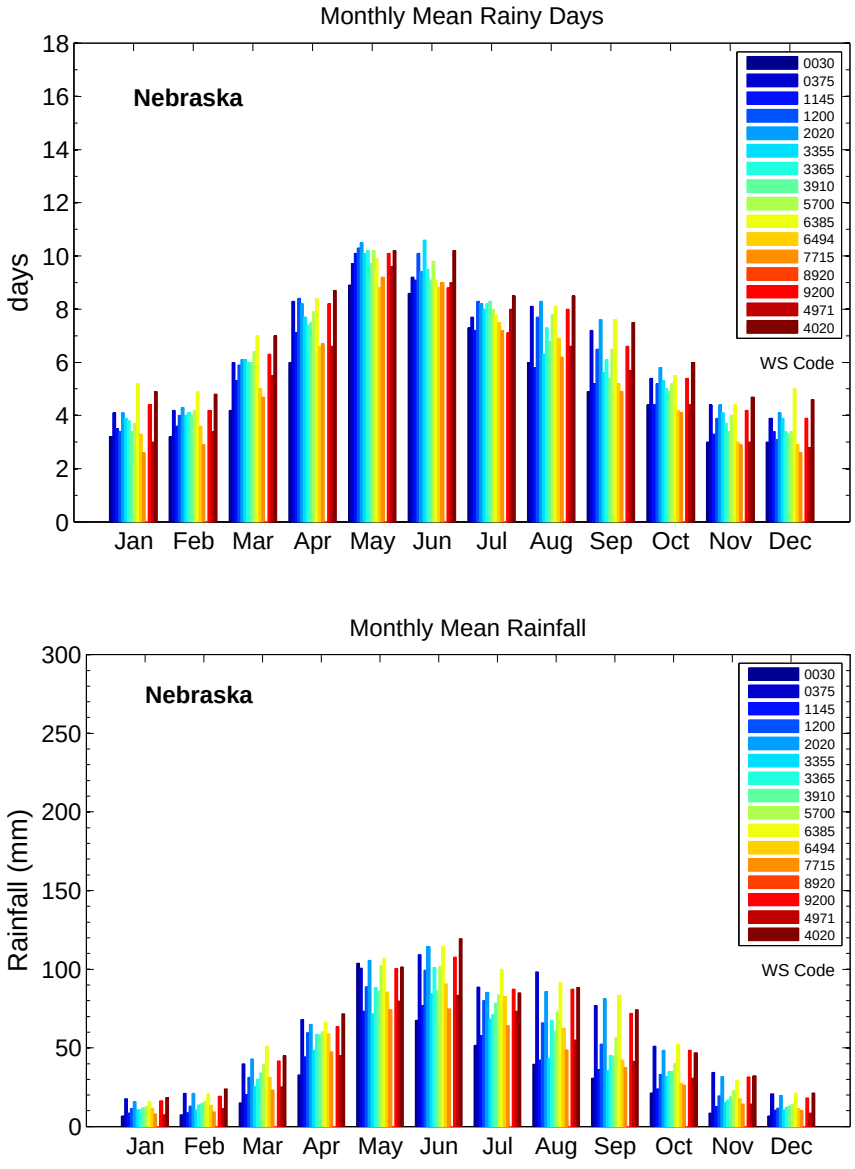


Figure 2.2 Monthly mean of rainfall amounts and rainy days in weather stations of Nebraska

2.1.1.1.3 California

In contrast to NE and FL, the study locations in of CA show a rainy season in winter (January-February). However, there large differences of rainfall amounts between locations, from 30 to 230 mm/month (1.2 to 9.1 in./month).

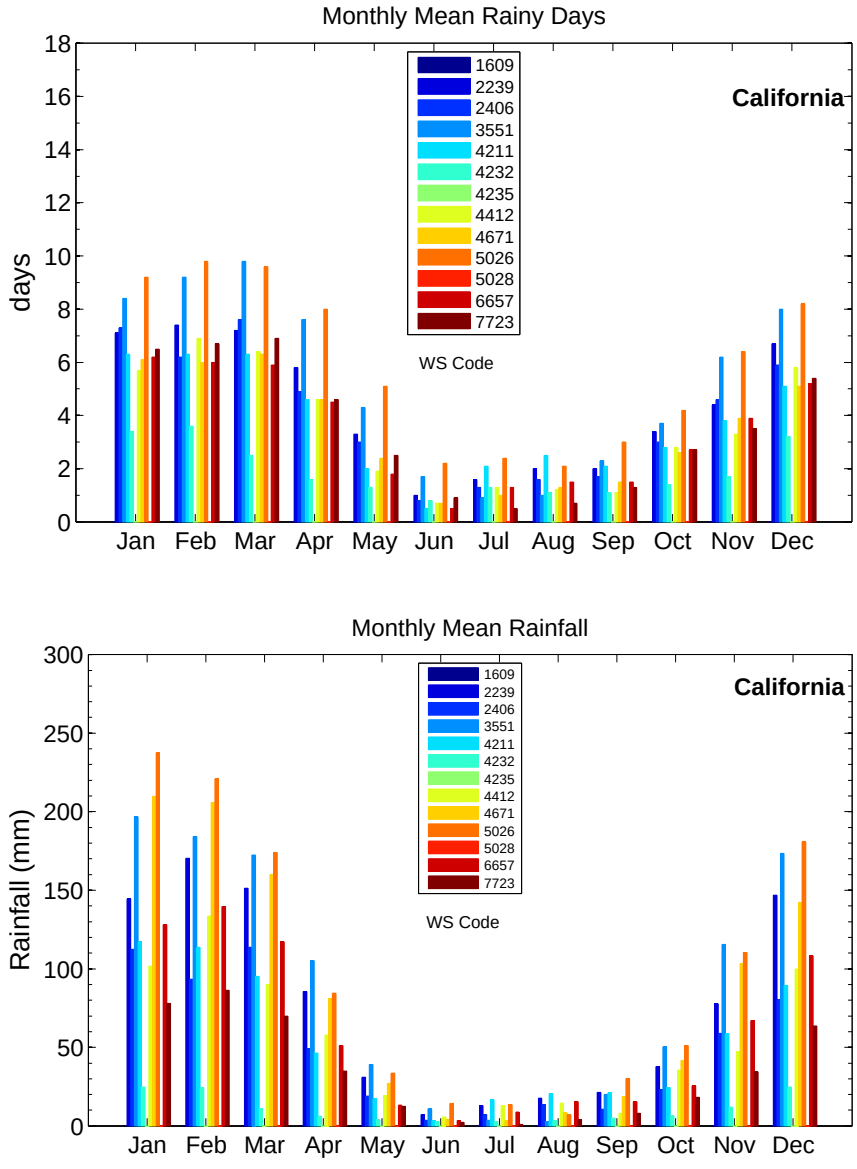


Figure 2.3 Monthly mean of rainfall amounts and rainy days in weather stations of California

Summer (June-August) is a dry season with rainfall amounts less than 20 mm/month (0.8 in./month). Rainy days vary from 1-2 days/month in the dry season (June) to 6-10 days/month in the rainy season (January-March) (Arguez et al. 2012).

2.1.1.1.4 Summary

Study locations for this project show different rainfall regimes. FL and NE have similarities in terms of when rainy and dry seasons occur; but, rainfall amounts are much larger in FL than in NE and CA. Northeast FL is rainy throughout the year, and weather stations in NE are commonly dry, especially in winter. Rainfall monthly amounts in NE and FL have their lowest values in winter (January), when CA has its highest rainfall. Nevertheless, July is a rainy month in NE and FL; but, it is a dry month in CA. The rainfall patterns are clearly shown in Figure C. 2 and Figure C. 3.

A spatial correlation between weather stations in FL and NE was observed. Baigorria et al. (2007) noted that daily rainfall events correlated between stations during the rainy season in FL. However, weather stations in CA evidenced a large spatial variability on monthly rainfall regimes, particularly in the rainy season (December – March). These differences allowed testing of EOMC for generating rainfall data in different rainfall regimes. A very dense network of precipitation measurements is needed to adequately describe rainfall patterns and occurrences in CA. Differences between weather station rainfall amounts and rainy days are 50-100 mm/month (2-4 in/month) and 3-4 days/month respectively (Figure 2.3).

2.1.2 Transition equations of extended Orthogonal Markov chain

This study proposes an extension of the Orthogonal Markov chain (EOMC). The temporal structure of rainfall occurrences used two previous days in contrast to the traditional approach of using one previous day. Transition probabilities (*P*) for rainfall events in a given location *i* at time *t* are expressed as follows:

$$P_{i_t \mid j_t,\ k_t, i_{(t-1)},\ i_{(t-2)}}$$

Equation 2.1 Transition equations of extended Orthogonal Markov chain

Where,

i, j, k locations

t time for a specific day

P was estimated considering events in two other weather stations (*j, k*) on the same day (*t*) and an event state in the location being generated (*i*) on previous two days (*t* -1 and *t* -2).

Transition equations formulated for the EOMC are illustrated in Equation 2.2 and Equation 2.3. These series of equations refer to probabilities for rain and non-rain events respectively.

$$P_{1|0,0,0,0} = Pr\left(X_{i_t} = 1 | X_{j_t} = 0, X_{k_t} = 0, X_{i_{(t-1)}} = 0, X_{i_{(t-2)}} = 0\right)$$
$$P_{1|0,0,0,1} = Pr\left(X_{i_t} = 1 | X_{j_t} = 0, X_{k_t} = 0, X_{i_{(t-1)}} = 0, X_{i_{(t-2)}} = 1\right)$$
$$P_{1|0,0,1,0} = Pr\left(X_{i_t} = 1 | X_{j_t} = 0, X_{k_t} = 0, X_{i_{(t-1)}} = 1, X_{i_{(t-2)}} = 0\right)$$
$$P_{1|0,1,0,0} = Pr\left(X_{i_t} = 1 | X_{j_t} = 0, X_{k_t} = 1, X_{i_{(t-1)}} = 0, X_{i_{(t-2)}} = 0\right)$$
$$P_{1|1,0,0,0} = Pr\left(X_{i_t} = 1 | X_{j_t} = 1, X_{k_t} = 0, X_{i_{(t-1)}} = 0, X_{i_{(t-2)}} = 0\right)$$

$$P_{1|0,0,1,1} = Pr\left(X_{i_t} = 1 | X_{j_t} = 0, X_{k_t} = 0, X_{i_{(t-1)}} = 1, X_{i_{(t-2)}} = 1\right)$$

$$P_{1|0,1,0,1} = Pr\left(X_{i_t} = 1 | X_{j_t} = 0, X_{k_t} = 1, X_{i_{(t-1)}} = 0, X_{i_{(t-2)}} = 1\right)$$

$$P_{1|1,0,0,1} = Pr\left(X_{i_t} = 1 | X_{j_t} = 1, X_{k_t} = 0, X_{i_{(t-1)}} = 0, X_{i_{(t-2)}} = 1\right)$$

$$P_{1|1,0,1,0} = Pr\left(X_{i_t} = 1 | X_{j_t} = 1, X_{k_t} = 0, X_{i_{(t-1)}} = 1, X_{i_{(t-2)}} = 0\right)$$

$$P_{1|1,1,0,0} = Pr\left(X_{i_t} = 1 | X_{j_t} = 1, X_{k_t} = 1, X_{i_{(t-1)}} = 0, X_{i_{(t-2)}} = 0\right)$$

$$P_{1|0,1,1,0} = Pr\left(X_{i_t} = 1 | X_{j_t} = 0, X_{k_t} = 1, X_{i_{(t-1)}} = 1, X_{i_{(t-2)}} = 0\right)$$

$$P_{1|0,1,1,1} = Pr\left(X_{i_t} = 1 | X_{j_t} = 0, X_{k_t} = 1, X_{i_{(t-1)}} = 1, X_{i_{(t-2)}} = 1\right)$$

$$P_{1|1,1,0,1} = Pr\left(X_{i_t} = 1 | X_{j_t} = 1, X_{k_t} = 1, X_{i_{(t-1)}} = 0, X_{i_{(t-2)}} = 1\right)$$

$$P_{1|1,1,1,0} = Pr\left(X_{i_t} = 1 | X_{j_t} = 1, X_{k_t} = 1, X_{i_{(t-1)}} = 1, X_{i_{(t-2)}} = 0\right)$$

$$P_{1|1,0,1,1} = Pr\left(X_{i_t} = 1 | X_{j_t} = 1, X_{k_t} = 0, X_{i_{(t-1)}} = 1, X_{i_{(t-2)}} = 1\right)$$

$$P_{1|1,1,1,1} = Pr\left(X_{i_t} = 1 | X_{j_t} = 1, X_{k_t} = 1, X_{i_{(t-1)}} = 1, X_{i_{(t-2)}} = 1\right)$$

Equation 2.2 Transition equations of extended Orthogonal Markov chain for rain events

$$P_{0|0,0,0,0} = Pr\left(X_{i_t} = 0 | X_{j_t} = 0, X_{k_t} = 0, X_{i_{(t-1)}} = 0, X_{i_{(t-2)}} = 0\right)$$

$$P_{0|0,0,0,1} = Pr\left(X_{i_t} = 0 | X_{j_t} = 0, X_{k_t} = 0, X_{i_{(t-1)}} = 0, X_{i_{(t-2)}} = 1\right)$$

$$P_{0|0,0,1,0} = Pr\left(X_{i_t} = 0 | X_{j_t} = 0, X_{k_t} = 0, X_{i_{(t-1)}} = 1, X_{i_{(t-2)}} = 0\right)$$

$$P_{0|0,1,0,0} = Pr\left(X_{i_t} = 0 | X_{j_t} = 0, X_{k_t} = 1, X_{i_{(t-1)}} = 0, X_{i_{(t-2)}} = 0\right)$$

$$P_{0|1,0,0,0} = Pr\left(X_{i_t} = 0 | X_{j_t} = 1, X_{k_t} = 0, X_{i_{(t-1)}} = 0, X_{i_{(t-2)}} = 0\right)$$

$$P_{0|0,0,1,1} = Pr\left(X_{i_t} = 0 | X_{j_t} = 0, X_{k_t} = 0, X_{i_{(t-1)}} = 1, X_{i_{(t-2)}} = 1\right)$$

$$P_{0|0,1,0,1} = Pr\left(X_{i_t} = 0 | X_{j_t} = 0, X_{k_t} = 1, X_{i_{(t-1)}} = 0, X_{i_{(t-2)}} = 1\right)$$

$$P_{0|1,0,0,1} = Pr\left(X_{i_t} = 0 | X_{j_t} = 1, X_{k_t} = 0, X_{i_{(t-1)}} = 0, X_{i_{(t-2)}} = 1\right)$$

$$P_{0|1,0,1,0} = Pr\left(X_{i_t} = 0 | X_{j_t} = 1, X_{k_t} = 0, X_{i_{(t-1)}} = 1, X_{i_{(t-2)}} = 0\right)$$

$$P_{0|1,1,0,0} = Pr\left(X_{i_t} = 0 | X_{j_t} = 1, X_{k_t} = 1, X_{i_{(t-1)}} = 0, X_{i_{(t-2)}} = 0\right)$$

$$P_{0|0,1,1,0} = Pr\left(X_{i_t} = 0 | X_{j_t} = 0, X_{k_t} = 1, X_{i_{(t-1)}} = 1, X_{i_{(t-2)}} = 0\right)$$

$$P_{0|0,1,1,1} = Pr\left(X_{i_t} = 0 | X_{j_t} = 0, X_{k_t} = 1, X_{i_{(t-1)}} = 1, X_{i_{(t-2)}} = 1\right)$$

$$P_{0|1,1,0,1} = Pr\left(X_{i_t} = 0 | X_{j_t} = 1, X_{k_t} = 1, X_{i_{(t-1)}} = 0, X_{i_{(t-2)}} = 1\right)$$

$$P_{0|1,1,1,0} = Pr\left(X_{i_t} = 0 | X_{j_t} = 1, X_{k_t} = 1, X_{i_{(t-1)}} = 1, X_{i_{(t-2)}} = 0\right)$$

$$P_{0|1,0,1,1} = Pr\left(X_{i_t} = 0 | X_{j_t} = 1, X_{k_t} = 0, X_{i_{(t-1)}} = 1, X_{i_{(t-2)}} = 1\right)$$

$$P_{0|1,1,1,1} = Pr\left(X_{i_t} = 0 | X_{j_t} = 1, X_{k_t} = 1, X_{i_{(t-1)}} = 1, X_{i_{(t-2)}} = 1\right)$$

Equation 2.3 Transition equations of extended Orthogonal Markov chain for non-rain events

Thus,

$$Pr\left(X_{i_t} = 1 | X_{j_t}, X_{k_t}, X_{i_{(t-1)}}, X_{i_{(t-2)}}\right) + Pr\left(X_{i_t} = 0 | X_{j_t}, X_{k_t}, X_{i_{(t-1)}}, X_{i_{(t-2)}}\right) = 1$$

2.1.3 Validation

Synthetic rainfall data was generated using the GiST-wg with the TOMC and with the EOMC version. Results of the synthetic rainfall from both approaches were compared

to observed rainfall data. Statistical analysis included the mean monthly number of rainy days and their standard deviation, correlation of daily rainfall events, and joint probabilities for station pairs with rainfall occurrences. In Addition, frequencies of wet and dry spells, and the RMSE for number of days with and without rain were estimated.

2.2 Results and discussion

2.2.1 Florida

2.2.1.1 Pearson's coefficient for correlations of rainfall events and joint probabilities of weather stations pairs with rainfall events

Figure 2.4 presents the comparisons of monthly correlation between daily observed and synthetic events among all pairs of weather stations in FL.

The extended Orthogonal Markov Chain (EOMC) modifies temporal structure of the traditional Orthogonal Markov Chain (TOMC). EOMC reproduces the monthly correlation of observed daily rainfall events in station pairs with a Pearson's coefficient (ρ) of 0.926. This matches the coefficient from using the TOMC (ρ = 0.927). EOMC preserves the spatial correlation of events between weather stations.

Figure 2.5 shows the comparisons for joint probabilities of rainfall events in station pairs by using observed and generated rainfall data. EOMC reproduced the observed joint probabilities with a Pearson's correlation (ρ) equal to 0.972. The correlation (ρ) was 0.973 when TOMC was used.

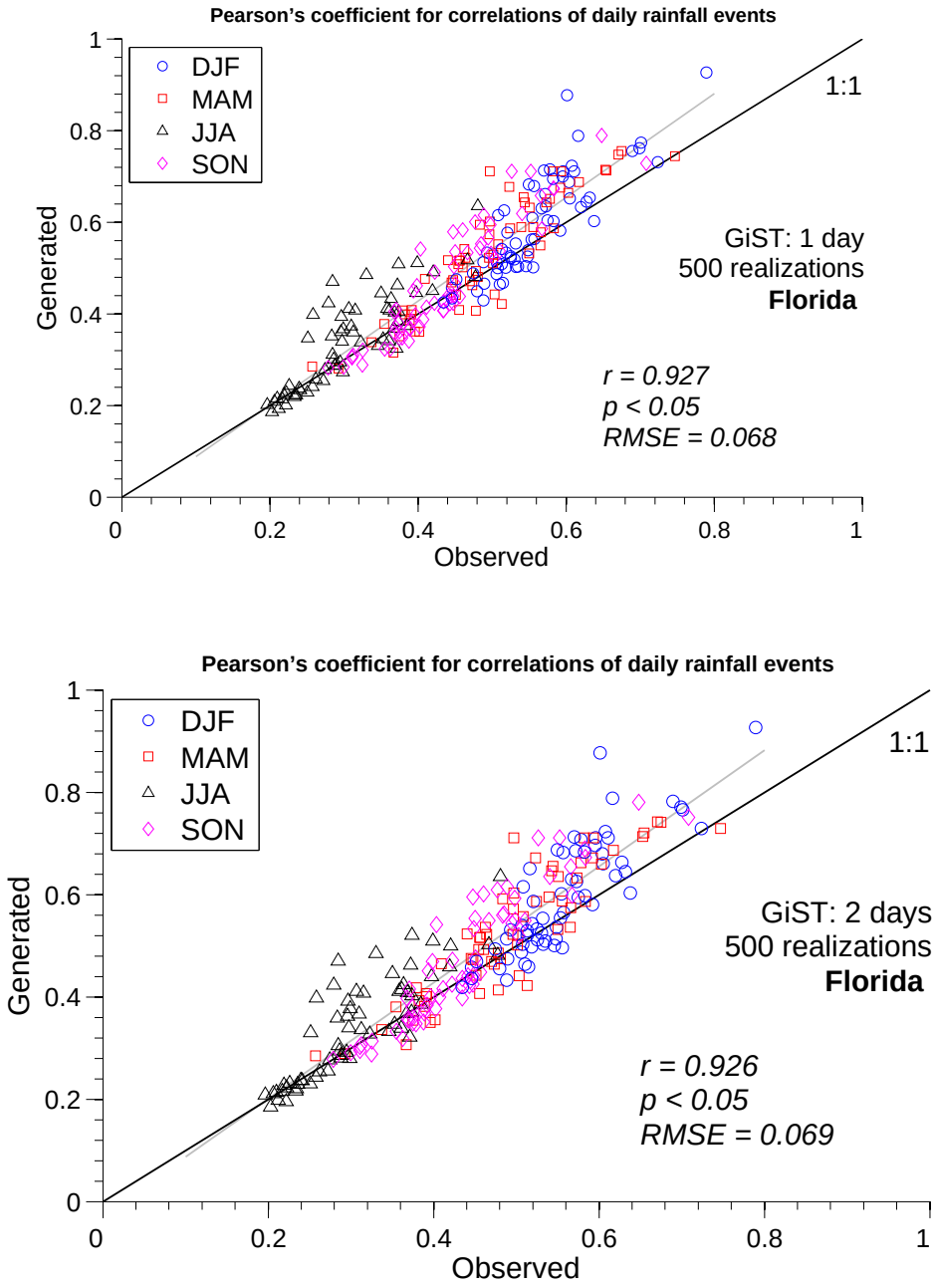


Figure 2.4 Observed and generated monthly correlations of daily rainfall events among pairs of weather stations of Florida, for each month, using traditional and extended Orthogonal Markov chain (500 yearlong simulations)

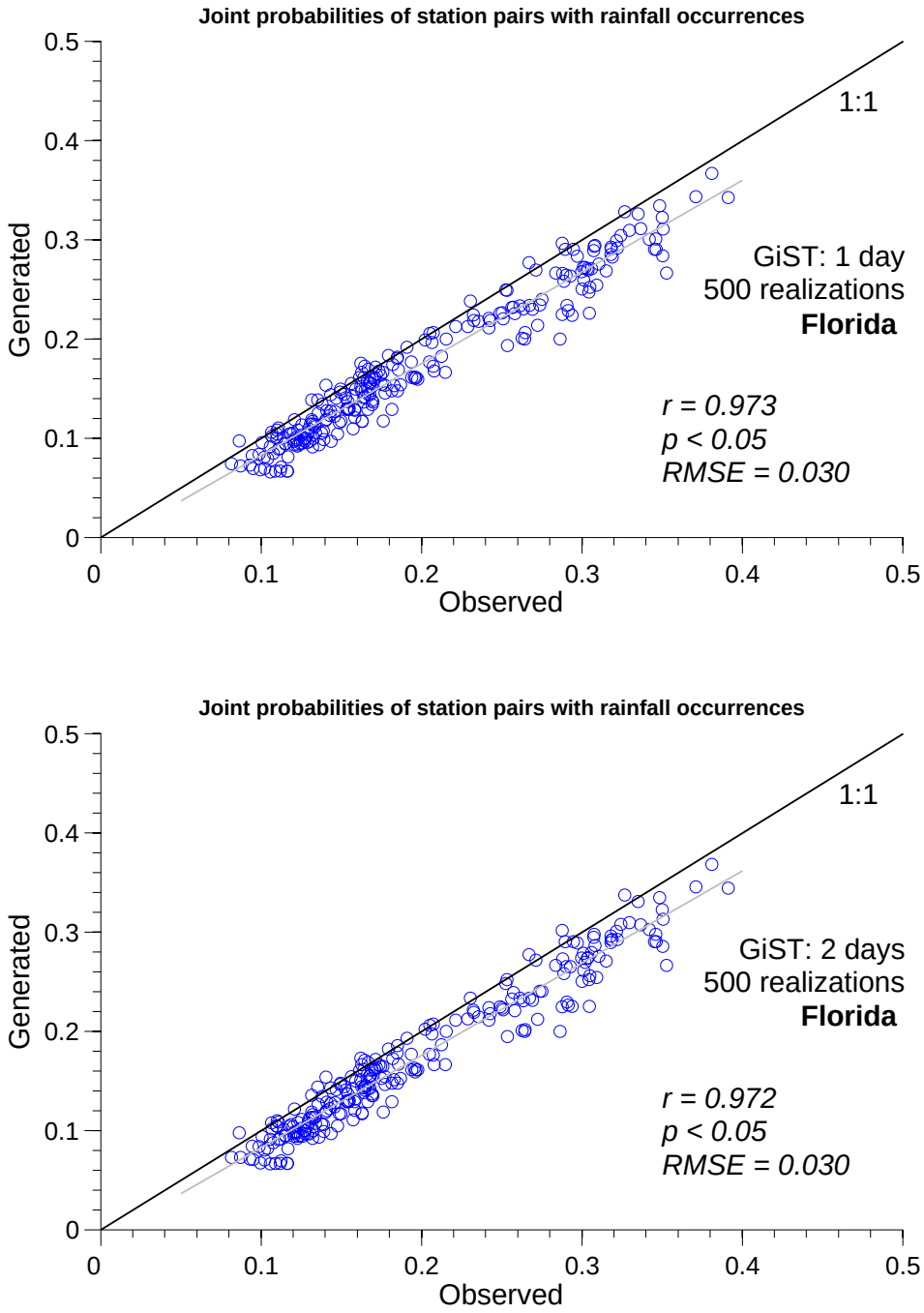


Figure 2.5 Observed and generated monthly joint probabilities among station pairs in Florida with daily rainfall events using traditional and extended Orthogonal Markov chain (500 yearlong simulations)

2.2.1.2 Wet spell length for 50, 100, 500 and 1000 yearlong-simulations and RMSE in estimation of wet spell length

Figure 2.6 shows whisker plots (5th, 25th, 50th, 75th and 95th percentiles) for the observed and generated frequency distribution of wet spells (2 to 20 days) for study locations in FL. Results for January and July include the generation for 50, 100, 500 and 1000 yearlong simulations using EOMC and TOMC.

TOMC did not differ from frequency distribution of wet spells for the month of January (advection season) generated with the EOMC. In fact, the RMSE's were not significantly different between results. On the other hand, EOMC evidences RMSE's below 0.0262 for all yearlong simulations for the month of July (convective rainfall). TOMC had RMSEs of 0.0273. Both EOMC and TOMC generated frequency distributions of dry spells wider than observed dry spells for the month of January (dry season).

As expected, large yearlong simulations had low RMSE in frequency distributions of the wet spells. Values of RMSE from the EOMC simulations for 500 and 1000 yearlong rainfall events were very close, 0.0249 and 0.0248, respectively. EOMC reproduces the observed frequencies of wet spells using fewer yearlong runs than the TOMC requires. This indicates an improvement by EOMC through extension on temporal structure compared to TOMC where the probability of rainfall occurrences at a given weather station is estimated based only on one previous day. EOMC did not need to perform a large number of simulations for accurate results (500 compared to 1000), and reduces by half the computational time.

Comparisons of RMSE in estimations of generated wet spells by EOMC and TOMC in weather stations of FL are illustrated in Figure 2.7. Differences in RMSE of EOMC and TOMC are larger for the month of July (0.04) than for the month of January (below 0.06) where a low RMSE indicates a larger wet spells.

EOMC showed a better performance than TOMC. RMSE of EOMC was smaller than TOMC for generating wet spells, especially for rainfall events with 2 to 6 days length. There are no significant differences for wet spells that lasted more than 4 days for the month of January and 10 days for the month of July.

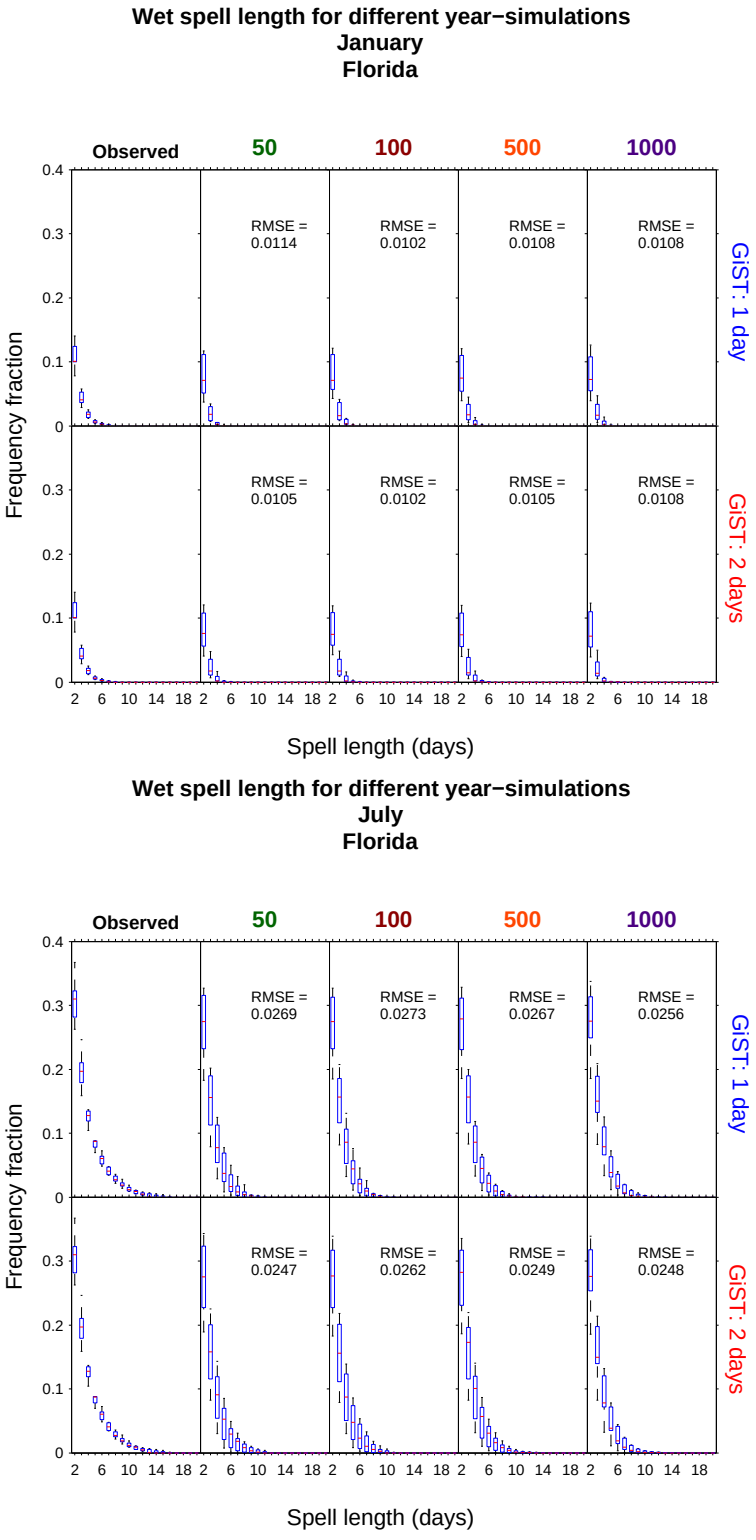


Figure 2.6 Whisker plots (5th, 25th, 50th, 75th and 95th percentiles) for frequencies of observed and generated wet spells in weather stations of Florida, in January and July, using traditional and extended Orthogonal Markov chain (50, 100, 500 and 1000 yearlong simulations)

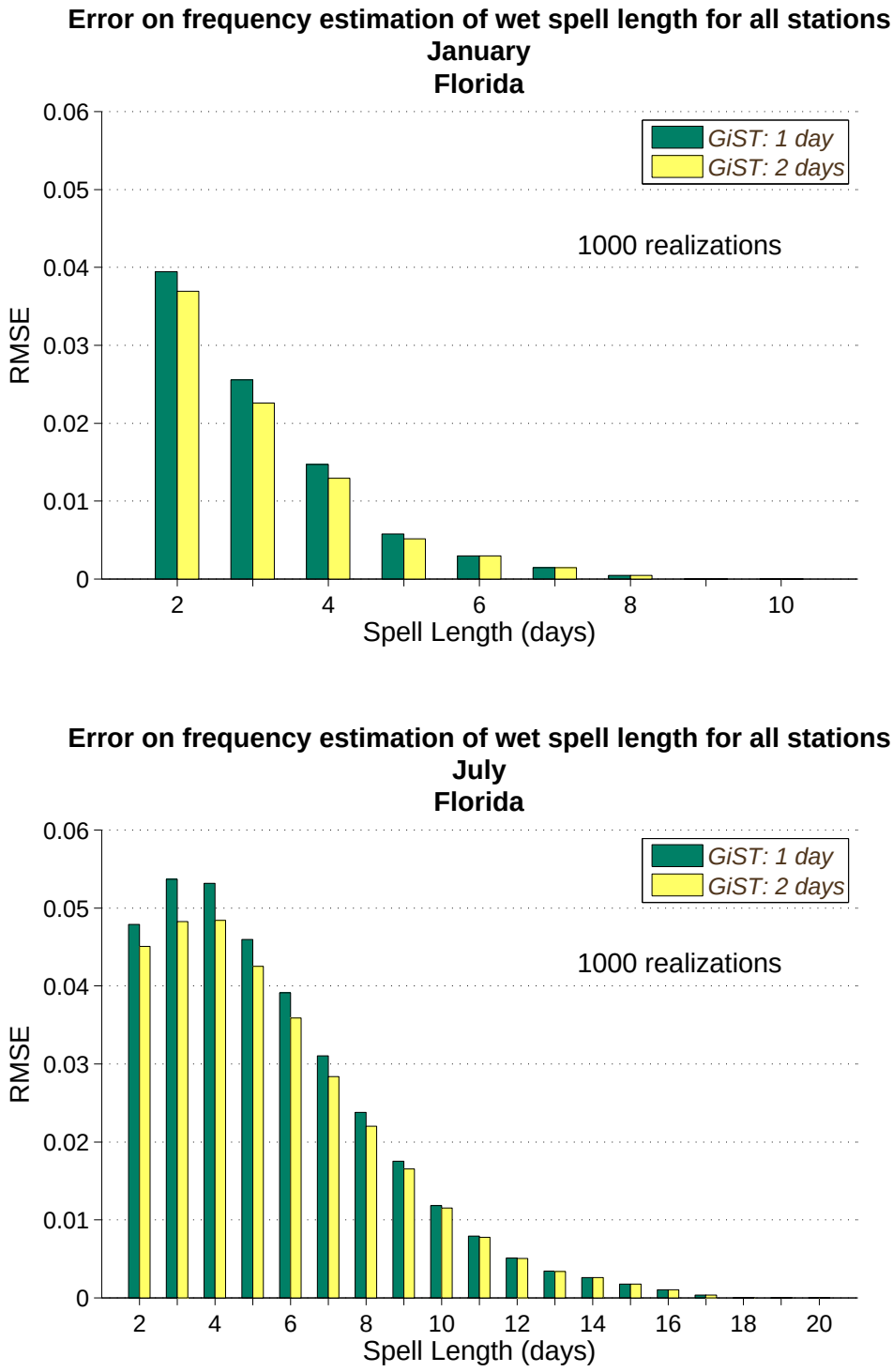


Figure 2.7 Error on estimation for frequencies of generated wet spells in weather stations of Florida, in January and July, using traditional and extended Orthogonal Markov chain (1000 yearlong simulations)

2.2.2 Nebraska

2.2.2.1 Pearson's coefficient for correlations of rainfall events and joint probabilities of weather stations pairs with rainfall events

Monthly correlations of daily observed and synthetic events at station pairs in NE are compared in Figure 2.8. EOMC generated monthly correlations of daily rain events with a Pearson's coefficient (ρ) equal to 0.853 ($p < 0.05$) from the observed correlations. TOMC yielded a Pearson's coefficient (ρ) of 0.859 ($p < 0.05$).

Some monthly correlations of rainfall daily events generated by EOMC in station pairs were close to 1, which is virtually unlikely. These outliers were removed and the Pearson's coefficient recalculated between observed and generated correlations of rainfall events. Joint monthly probabilities of rainfall events in station pairs using observed and generated daily rainfall data are compared in Figure 2.9. For 500 yearlong simulations, EOMC reproduced the observed joint probabilities by TOMC. Pearson's coefficient and RMSE of EOMC and TOMC were very similar, with ρ values of 0.94 and RMSE of 2.8-2.9%, respectively.

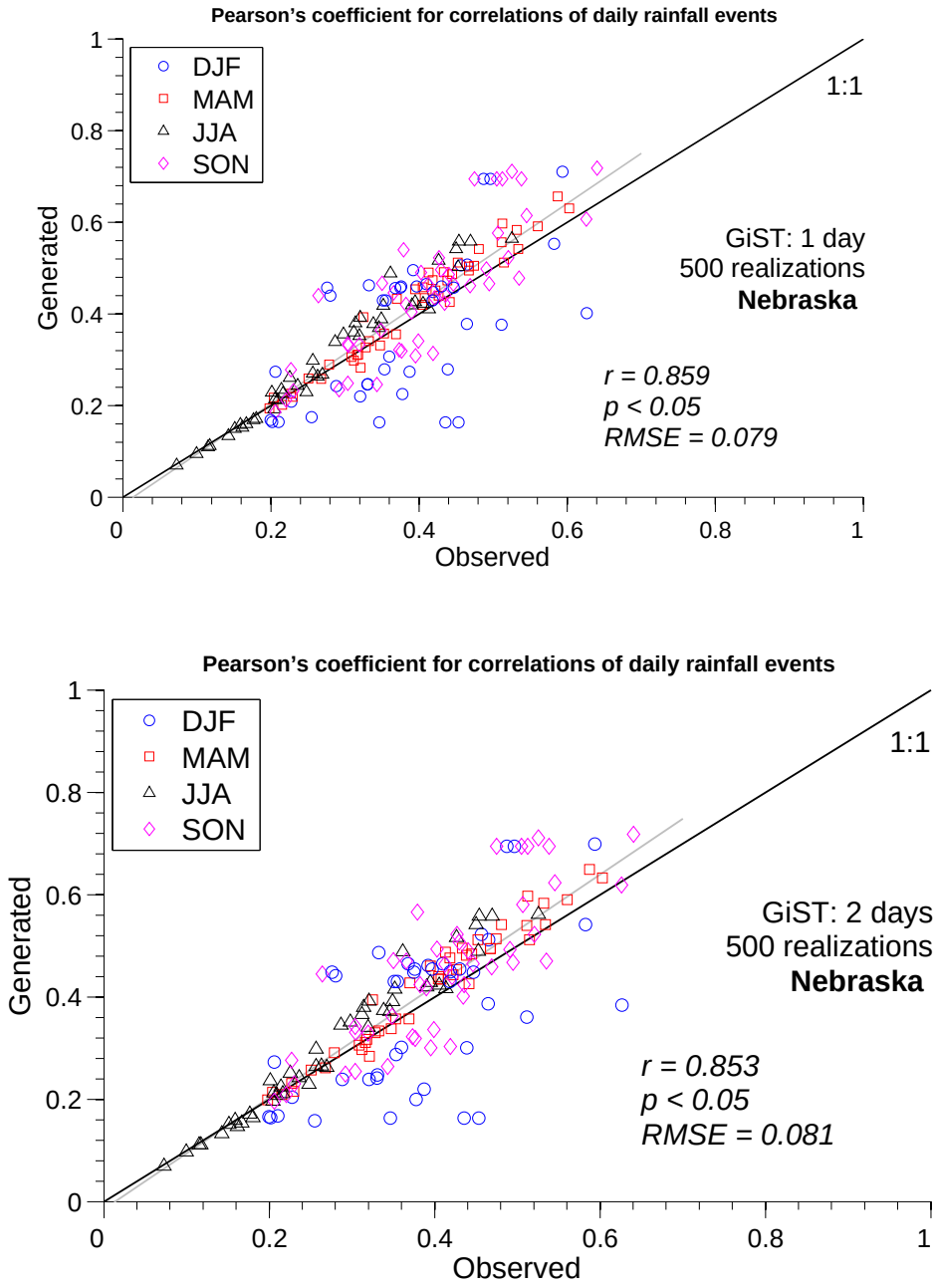


Figure 2.8 Observed and generated monthly correlations of daily rainfall events among pairs of weather stations of Nebraska, for each month, using traditional and extended Orthogonal Markov chain (500 yearlong simulations)

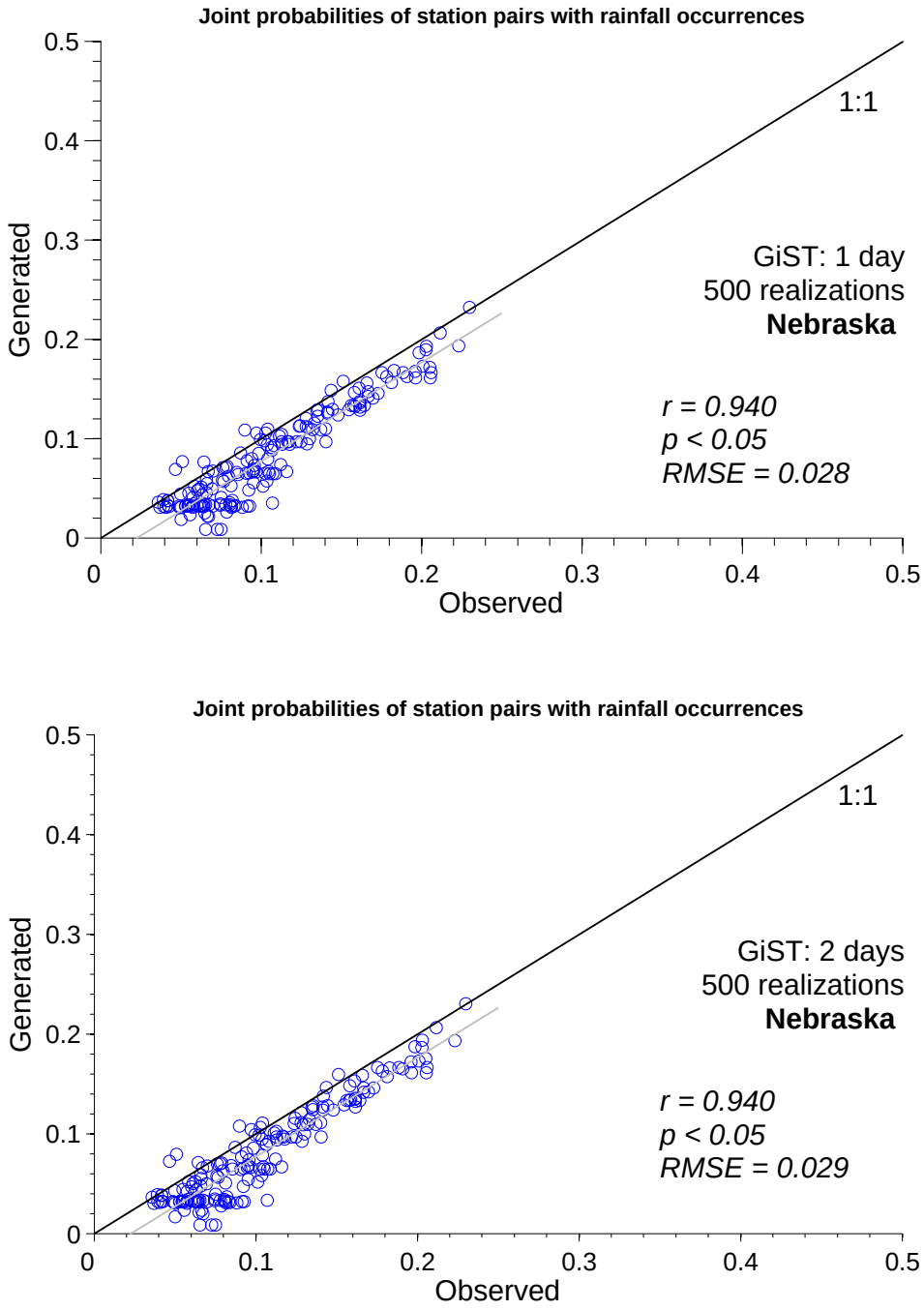


Figure 2.9 Observed and generated monthly joint probabilities among station pairs in Nebraska with daily rainfall events using traditional and extended Orthogonal Markov chain (500 yearlong simulations)

2.2.2.2 Wet and dry spell length for 500 yearlong simulations

The boxplots (5th, 25th, 50th, 75th and 95th percentiles) in Figure 2.10 compares the frequencies of observed and generated wet and dry spells for study locations in NE for 500 yearlong simulations using EOMC and TOMC.

Dry spells occur more frequently than wet spells in NE. Dry spells of 2 days in length have a 0.8 chance of occurring during the dry season (January) and 0.6 in the rainy season (July). On the other hand, wet spells of 2 days have a 0.1 chance of occurring in the rainy season (July).

Comparisons between wet and dry spell generated by EOMC and TOMC indicated no significant differences on RMSE values. For instance, RMSE of TOMC for estimating wet spells was 0.0115 for the month of July and 0.0111 using the EOMC. However, RMSE for dry spells was 0.0296 for the EOMC, 0.0326 when using the TOMC for the July.

The frequency of observed wet spells for the months of January and July were underestimated for some weather stations when using either TOMC or EOMC. The frequencies of observed dry spells were overestimated for the month of January (dry season). This might be due to few weather stations with low correlation of rainy occurrences among them. The number of rainy days was lower than 4 for the dry season (January), however, differences of 2 rainy days between weather stations of NE were observed (Figure 2.2).

Due to a small number of weather stations, the parameterization in the GiST-wg algorithm for generating rainfall data was limited. The transition of rainfall regimes

across large areas need a larger number of weather stations to be represented. On the other hand, the median value between weather stations followed the observed trend of dry spells. Furthermore, the frequencies of generated dry spells match closely with the observed spells for the month of July (rainy season).

NE study locations are characterized by dry conditions during the year. Wet spells normally last 2 days and not longer than 4 days in the rainy season (July). Wet spells of 2 days are less frequent (<0.1) than dry spells of 2 days (0.6-0.8) in NE.

Figure 2.11 shows the boxplots (25^{th}, 50^{th} and 75^{th} percentiles) comparing frequencies of wet and dry spells from observed and generated rainfall data in weather stations of NE. Dry spells are common through the year. Wet spells of 2-3 days occur during the rainy months such as April, May and June with frequencies below 0.2. These patterns were reproduced by EOMC and TOMC. Some differences regarding the frequency of spell length between weather stations of NE occurred during the dry season (December-January). This pattern was also reproduced by EOMC and TOMC. Study locations in NE have a stronger spatial correlation during the rainy season than the dry season.

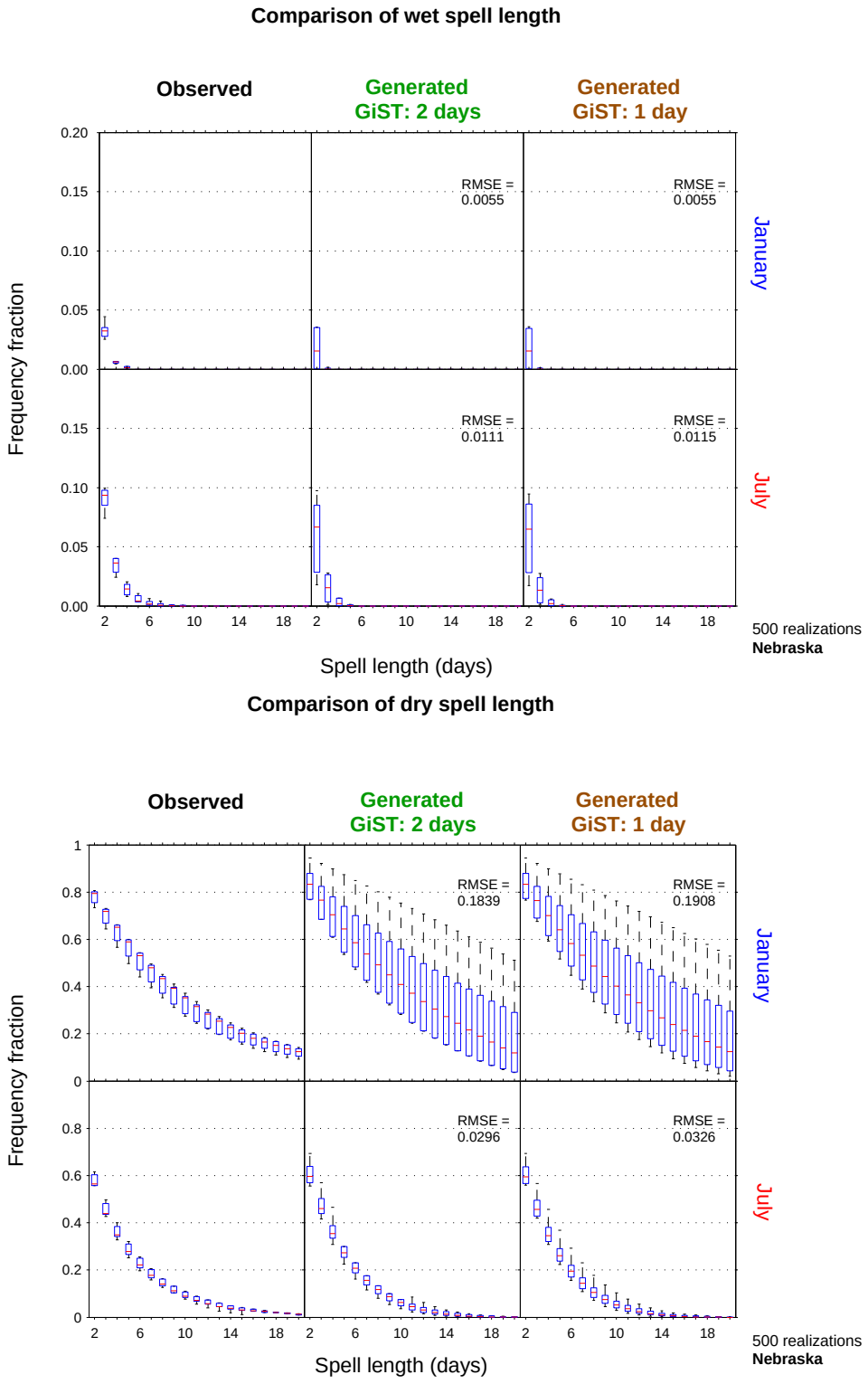


Figure 2.10 Whisker plots (5th, 25th, 50th, 75th and 95th percentiles) for frequencies of observed and generated wet spells in weather stations of Nebraska, in January and July, using traditional and extended Orthogonal Markov chain (500 yearlong simulations)

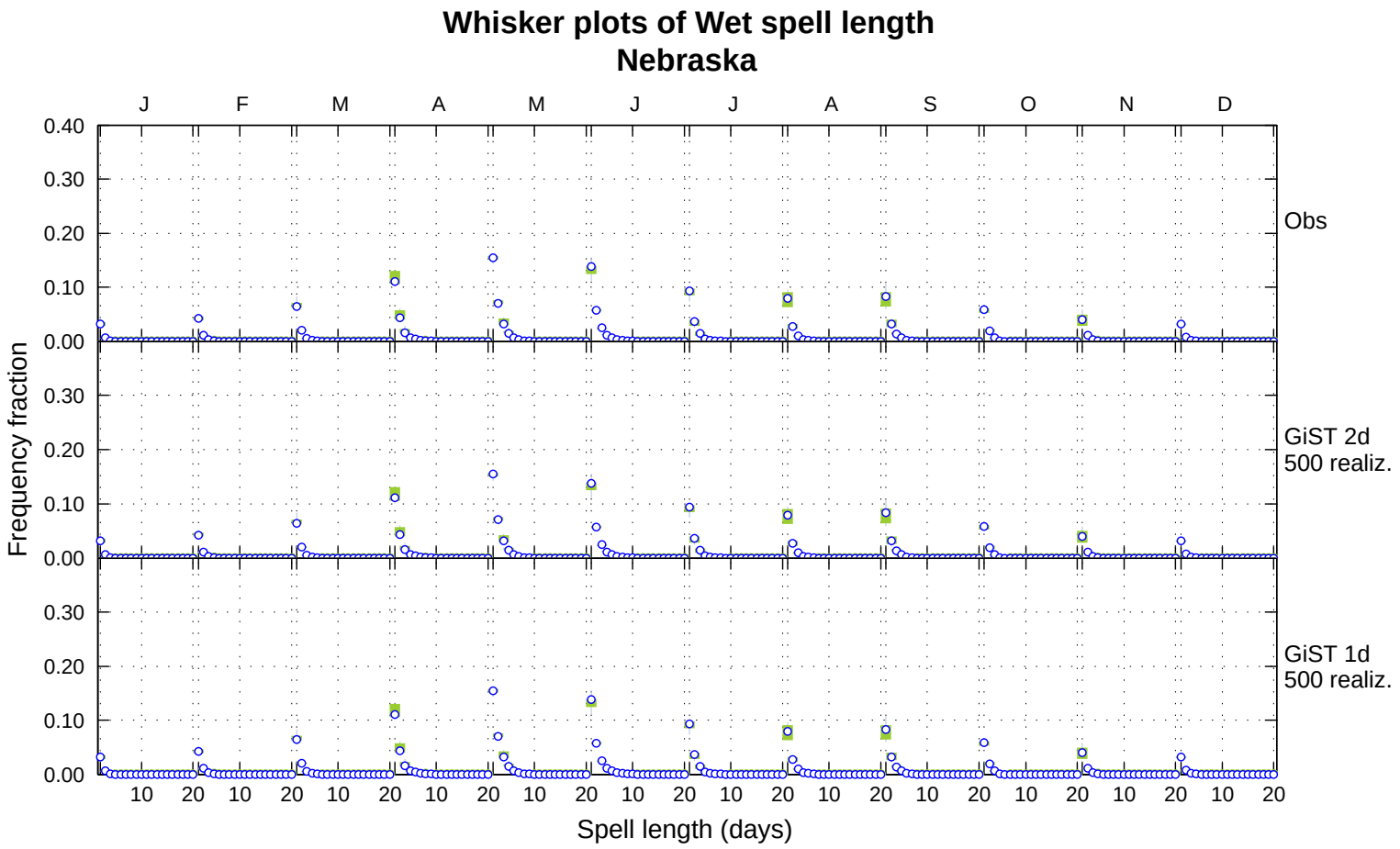


Figure 2.11 Whisker plots (50th (blue), 25th and 75th (green) percentiles) for frequencies of observed and generated wet and dry spells in weather stations of Nebraska, for all months, using traditional and extended Orthogonal Markov chain (500 yearlong simulations)

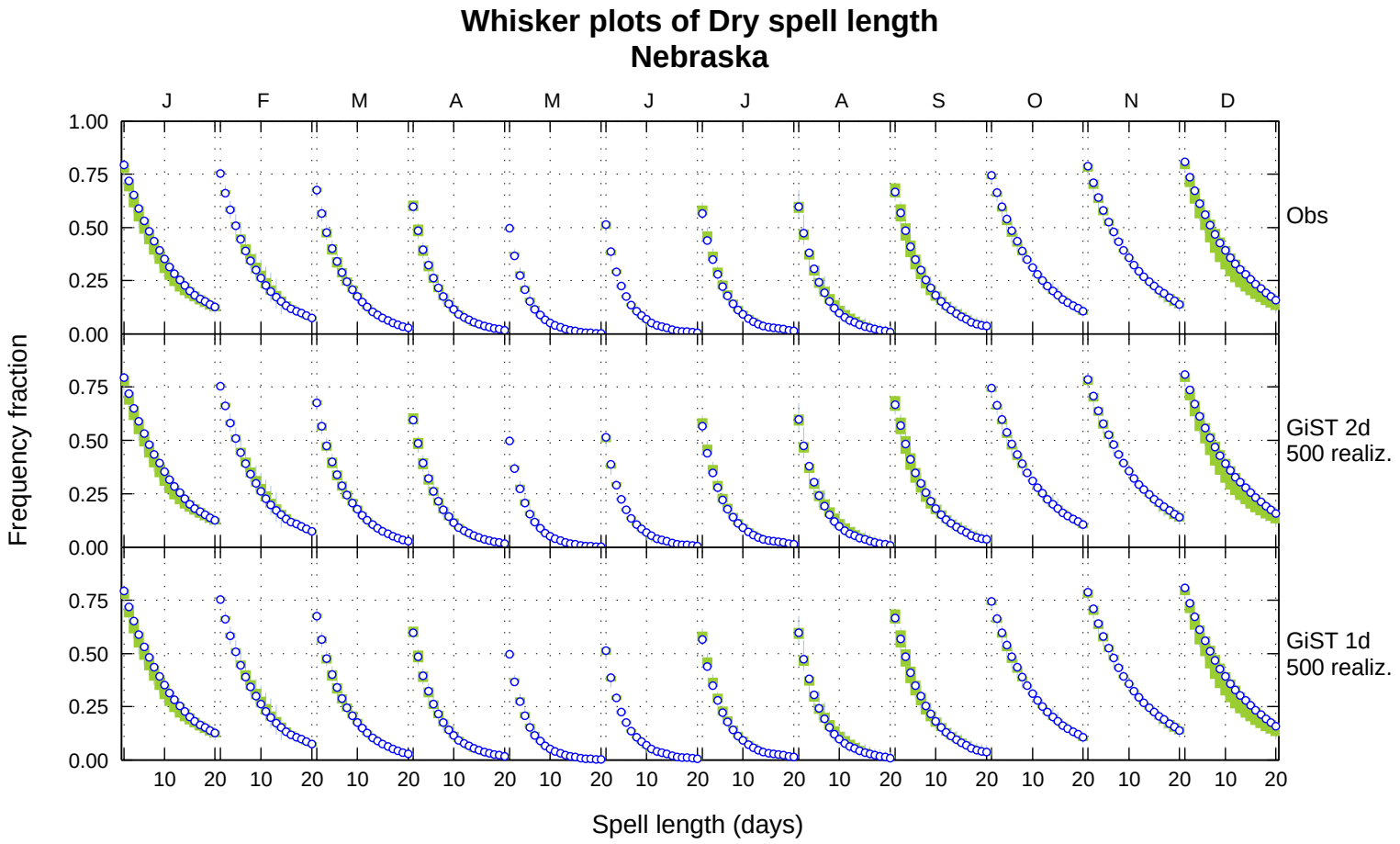


Figure 2.11 Whisker plots (50th (blue), 25th and 75th (green) percentiles) for frequencies of observed and generated wet and dry spells in weather stations of Nebraska, for all months, using traditional and extended Orthogonal Markov chain (500 yearlong simulations) - continued

2.2.2.3 RMSE in estimation of wet and dry spells

Error on generated frequencies of wet and dry spells for weather stations of NE are shown in Figure 2.12. Total of 500 yearlong simulations were performed using EOMC and TOMC. Frequency of wet spells RMSE values are below 0.012. Dry spells have the smallest RMSE (< 0.05) for the months of May and June. In general, there are no significant differences between EOMC and TOMC for estimations of wet and dry spells through the year. EOMC and TOMC generated daily rainfall events that reproduced wet spells with a low monthly RMSE (<0.012) in NE. Dry spells are also well reproduced, especially for the rainy months (May-August).

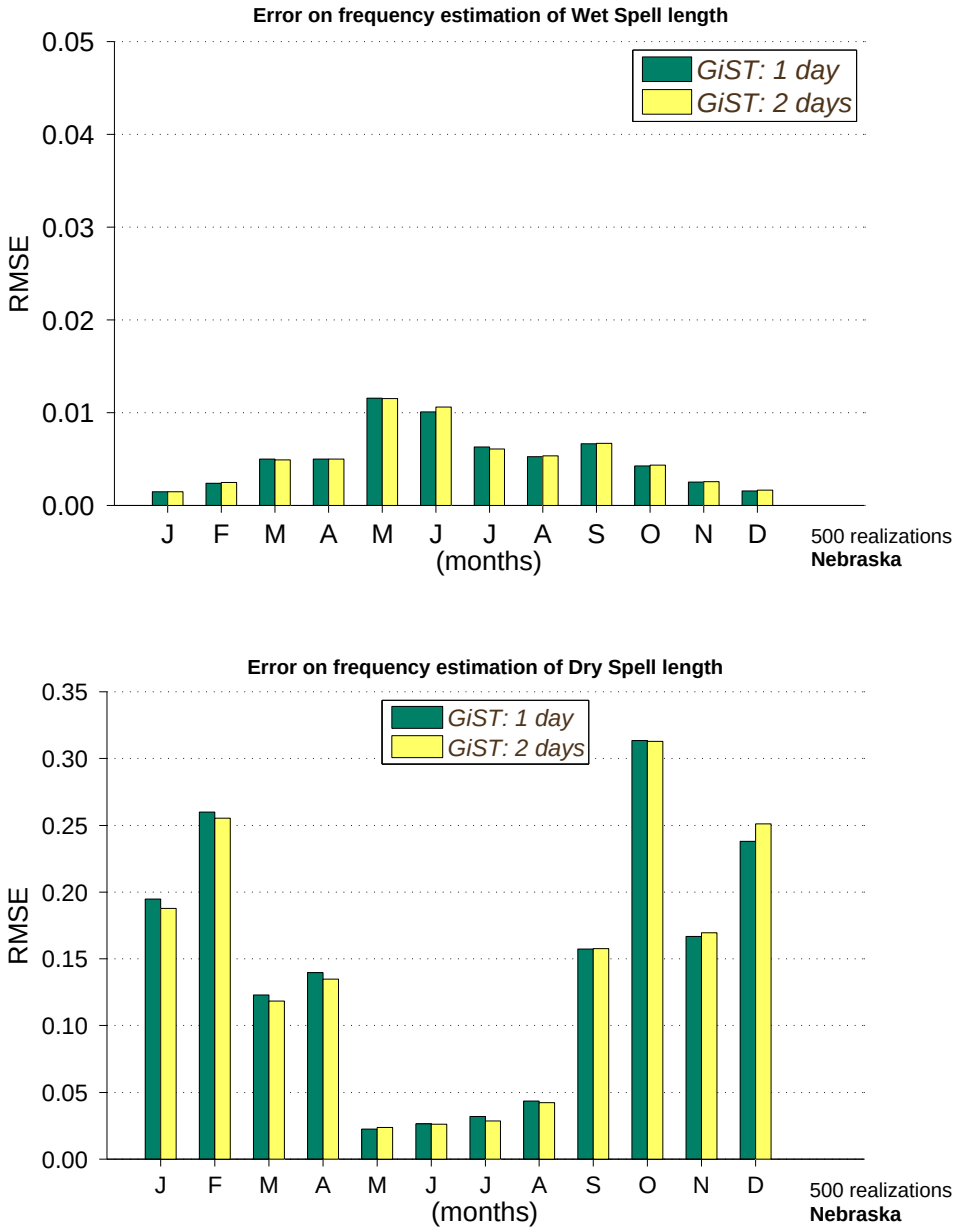


Figure 2.12 Error on estimation for frequencies of generated wet spells in weather stations of Nebraska, in January and July, using traditional and extended Orthogonal Markov chain (500 yearlong simulations)

2.2.3 California

2.2.3.1 Pearson's coefficient for correlations of rainfall events and joint probabilities of weather stations pairs with rainfall events

Monthly correlation of daily observed and synthetic rain events in station pairs of CA are illustrated in Figure 2.13. Pearson's coefficient between observed and generated correlations showed a strong correlation of 0.876 ($p < 0.05$) and RMSE of 0.16), which is similar with results from TOMC. EOMC preserved the spatial correlation of daily rainfall events in station pairs of CA.

Some synthetic monthly correlations of rainfall events between station pairs were not able to be estimated or were equal to one, which is a very unlikely correlation. The correlations were removed and the Pearson's coefficient between observed and generated correlations of daily events was recalculated.

Figure 2.14 illustrates the monthly probabilities that daily rainfall events are in station pairs for CA for 500 yearlong simulations. There are no significant differences for joint probabilities between TOMC and EOMC for study locations in CA. Both EOMC and TOMC reproduced the observed joint probabilities with a RMSE of 0.033.

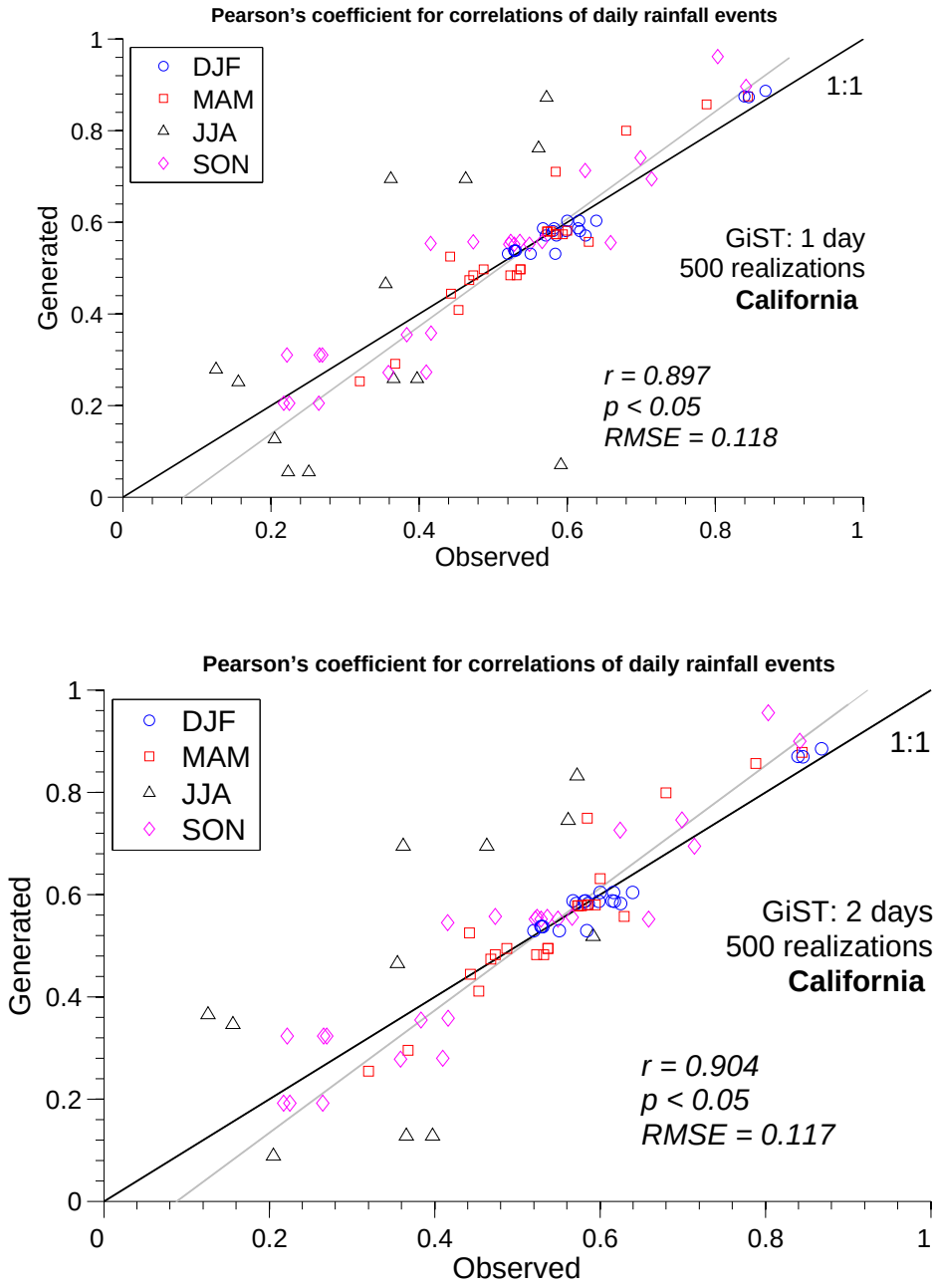


Figure 2.13 Observed and generated monthly correlations of daily rainfall events among pairs of weather stations of California, for each month, using traditional and extended Orthogonal Markov chain (500 yearlong simulations)

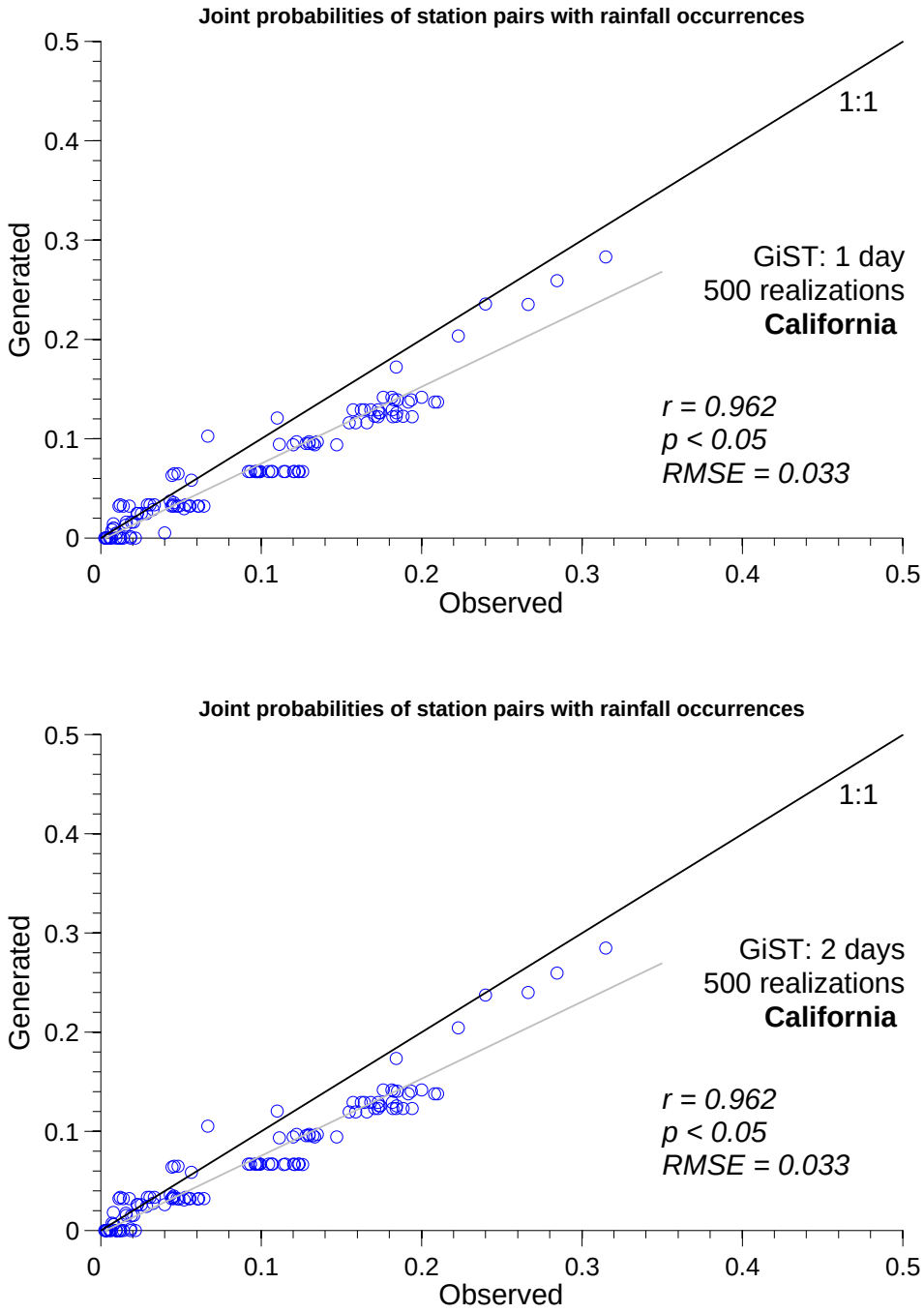


Figure 2.14 Observed and generated monthly joint probabilities among station pairs in California with daily rainfall events using traditional and extended Orthogonal Markov chain (500 yearlong simulations)

2.2.3.2 Wet and dry spell length for 500 yearlong simulations

Wet and dry spells from observed and generated events from weather stations in CA are compared in Figure 2.15 (January and July) and Figure 2.16 (all months). 500 yearlong simulations were performed using EOMC and TOMC.

EOMC reproduced the wet spells of 2 days for the dry season in CA; but, wet spell frequencies of 3-5 days were underestimated by EOMC and TOMC. That reproduced wet spells of 5-6 days for the month of January (rainy season) with frequencies wider than TOMC generates. RMSE was 0.0186 by EOMC, and 0.0193 by TOMC. Synthetic rainfall events using either EOMC or TOMC did not represent wet spells longer than 6 days.

Dry spells for weather stations in CA were reproduced by both EOMC and TOMC. Dry spell frequencies were longer than the observed dry spells for the month of January (rainy season). However, dry spell frequencies were shorter for the month of July (dry season), which indicates an overestimation for weather stations with lower observed frequencies.

The RMSE for EOMC simulations was larger than that for the TOMC simulations for the month of July (0.1648 vs. 0.1553). RMSE of wet and dry spells were smaller for EOMC simulations than for TOMC simulations for the month of January (0.0186 vs. 0.0193 and 0.0688 vs. 0.0729). Although EOMC reproduced observed wet and dry spells in CA, their frequencies were overestimated for the rainy and dry seasons (Figure 2.15). A similar result occurred with the NE dataset. This might occur due to few weather stations with very poor correlations of rainfall events between station pairs for generating

synthetic rainfall data. Rainy days were lower than 4 in dry season (January) and differences until 2 rainy days between weather stations of NE were observed

Although EOMC reproduced the observed wet and dry spells for CA, their frequencies for rainy and dry seasons were overestimated (Figure 2.15). As explained for NE, this might happen due to a few weather stations were used for generating rainfall data, and the poor correlations of rainfall events between station pairs in CA. Differences in weather stations were from 3 to 9 rainy days in rainy season (January). Moreover, rainy events in dry season (June-August) are few with less than 2 days/month (Figure 2.3). Generation of rainfall data was limited by lack of data which were statistically correlated.

Dry spells through the year varied between weather stations of CA (Figure 2.16) due to the distinct rainfall regimes in the state (See 2.1.1).

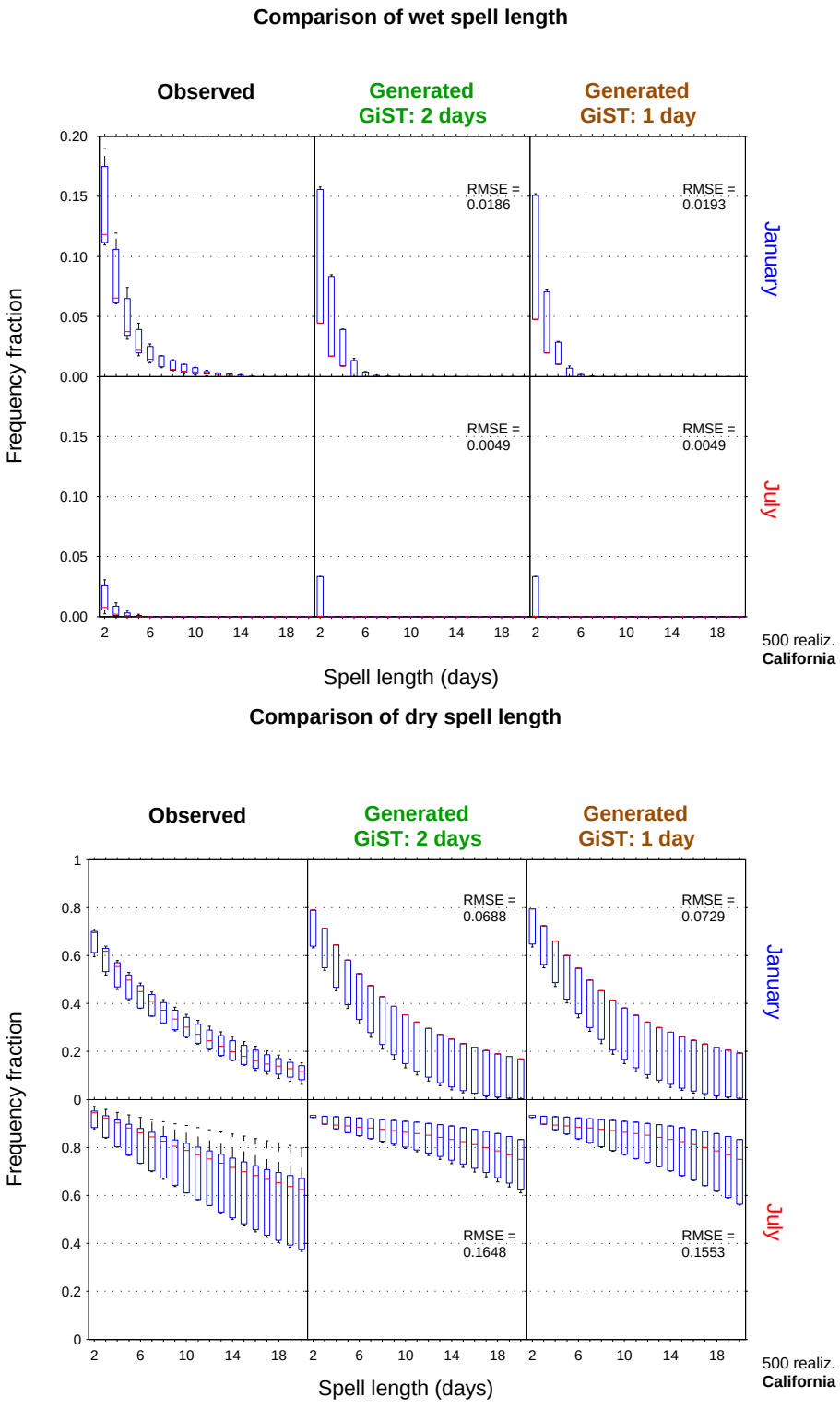


Figure 2.15 Whisker plots (5th, 25th, 50th, 75th and 95th percentiles) for frequencies of observed and generated wet spells in weather stations of California, in January and July, using traditional and extended Orthogonal Markov chain (500 yearlong simulations)

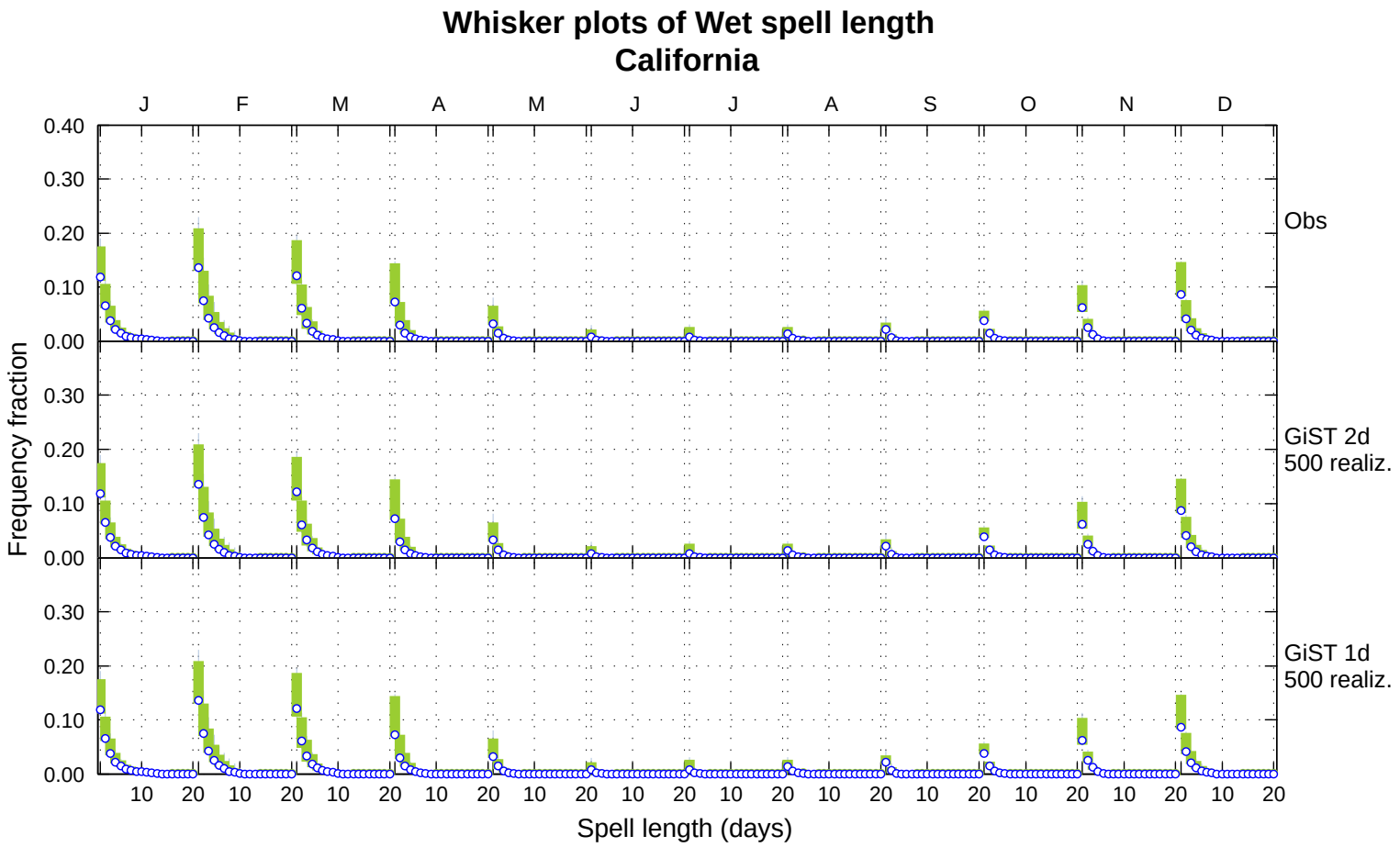


Figure 2.16 Whisker plots (50th (blue), 25th and 75th (green) percentiles) for frequencies of observed and generated wet and dry spells in weather stations of California, for all months, using traditional and extended Orthogonal Markov chain (500 yearlong simulations)

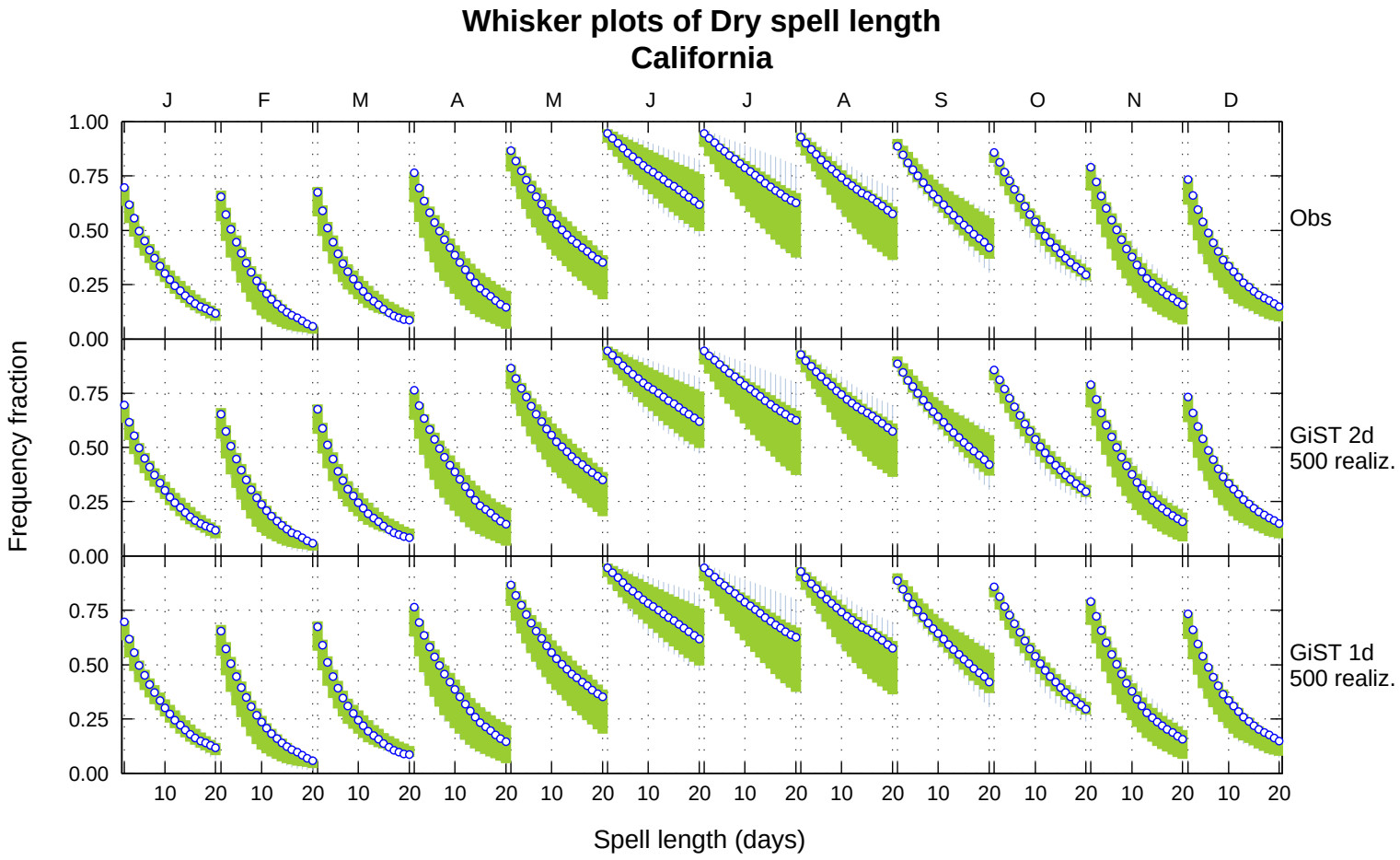


Figure 2.16 Whisker plots (50th (blue), 25th and 75th (green) percentiles) for frequencies of observed and generated wet and dry spells in weather stations of California, for all months, using traditional and extended Orthogonal Markov chain (500 yearlong simulations) - continued

2.2.3.3 RMSE in estimation of wet and dry spells

RMSEs for monthly wet and dry spell frequencies of for CA weather stations are illustrated in Figure 2.17. Wet spell frequency RMSEs were below 0.02 for either EOMC or TOMC. The lowest RMSEs (about 0.07) for generated dry spells were for the rainy months (December-March).

EOMC exhibits an improvement over TOMC for estimating wet spells. Wet spell frequency RMSEs for the rainy months (November-April) were lower when EOMC was used as compared to when TOMC was used. Moreover, frequencies of dry spells generated by EOMC displayed a RMSE higher than TOMC's RMSE in most months; exceptions were for the months of September, October and January when the variation between dry spell frequencies in CA were low (Figure 2.16).

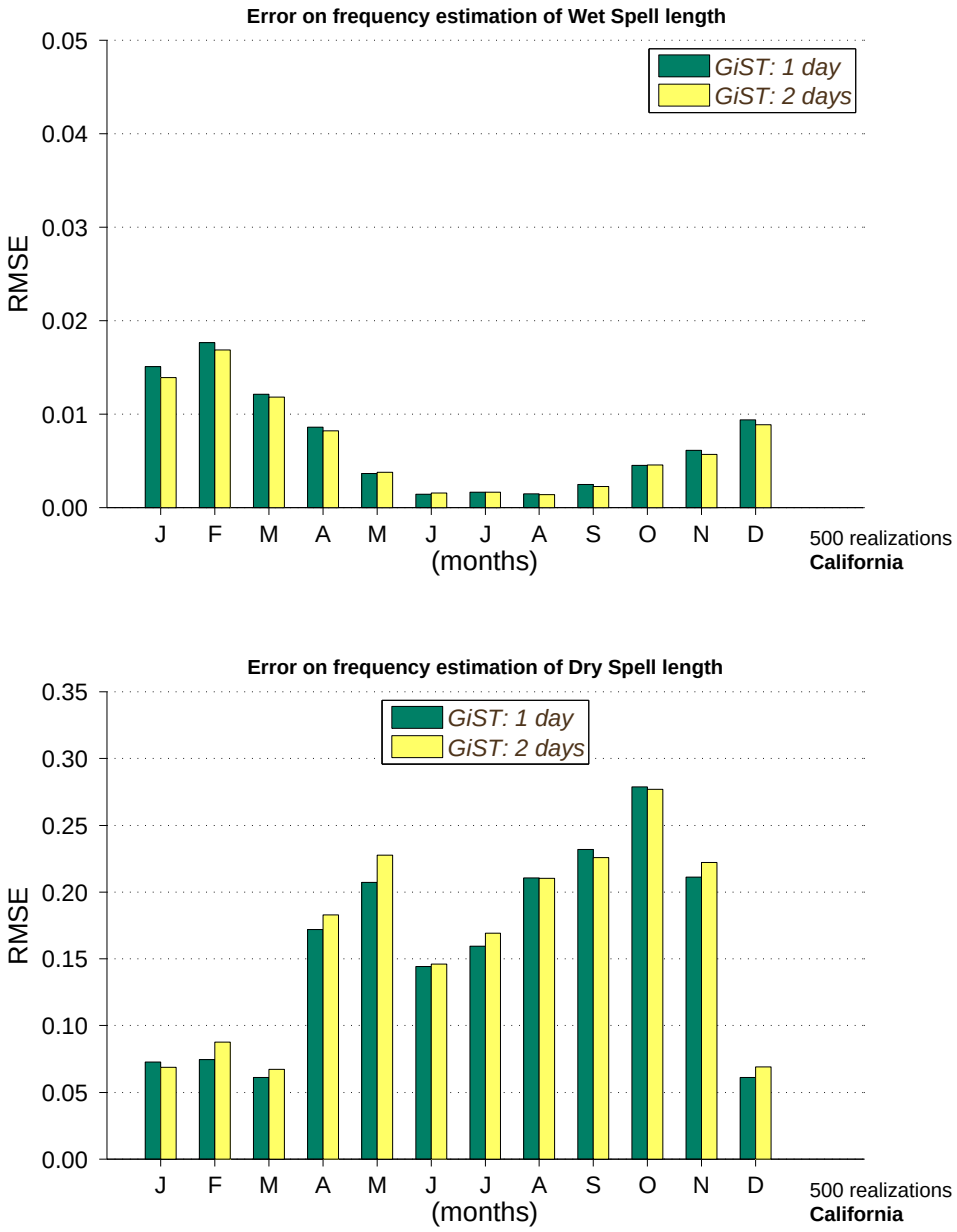


Figure 2.17 Error on estimation for frequencies of generated wet spells in weather stations of California, in January and July, using traditional and extended Orthogonal Markov chain (500 yearlong simulations)

2.2.4 Final remarks

Table 2.2 and Table 2.3 summarizes the monthly correlation coefficients between observed and generated daily rainfall events using TOMC and EOMC with station pairs from the three-state study site.

EOMC preserves the monthly correlations of observed rainfall events among all pairs in weather stations of FL, NE and CA. Values of Pearson's coefficients were 0.853-0.926 for a 500 yearlong simulation (Table 2.2). For 1000 yearlong simulations with TOMC, Baigorria and Jones (2010) found similar correlations equal to 0.932 and 0.966 in weather stations of FL and North Carolina, respectively. However, WGEN did not preserve correlations in FL (ρ = -0.082) (Baigorria and Jones 2010).

Joint monthly probabilities of daily rain events in station pairs from FL, NE and CA are also kept by the EOMC. Pearson's coefficients between observed and generated joint probabilities for 500 yearlong runs ranged between 0.940 and0.972 using EOMC, which does not differ from TOMC results (Table 2.3). For 1000 yearlong simulations with TOMC, Baigorria and Jones (2010) estimated correlations of 0.980 and 0.966 in weather stations of FL and North Carolina; but, WGEN exhibited lower correlation of 0.045 in FL.

Table 2.2 Pearson's coefficient and error between observed and generated monthly correlation of daily rainfall events in station pairs at study sites using traditional and extended Orthogonal Markov chain

Study site	Pearson's coefficient (ρ)		p-value		RMSE	
	TOMC	EOMC	TOMC	EOMC	TOMC	EOMC
FL	0.927	0.926	< 0.05	< 0.05	0.068	0.069
NE	0.859	0.853	< 0.05	< 0.05	0.079	0.081
CA	0.897	0.904	< 0.05	< 0.05	0.118	0.117

Table 2.3 Pearson's coefficient and error between observed and generated monthly joint probabilities among weather stations pairs in study sites with rainfall events using traditional and extended Orthogonal Markov chain

Study site	Pearson's coefficient (ρ)		p-value		RMSE	
	TOMC	EOMC	TOMC	EOMC	TOMC	EOMC
FL	0.973	0.972	< 0.05	< 0.05	0.030	0.030
NE	0.940	0.940	< 0.05	< 0.05	0.028	0.029
CA	0.962	0.962	< 0.05	< 0.05	0.033	0.033

Monthly correlations of daily rainfall events generated using EOMC among station pairs from FL, NE and CA (0.069, 0.081 and 0.117) were lower than values reported by Baigorria and Jones (2010) when TOMC was implemented for weather stations in FL (up to 0.15).

Figure 2.18 shows RMSEs of generated monthly wet spell frequencies in FL weather stations using TOMC (1 day) and EOMC (2 days) with different ensemble members. Wet spell in FL were reproduced with lower RMSE by EOMC for any number of yearlong simulations, in comparison to results using TOMC, especially in rainy months (July).

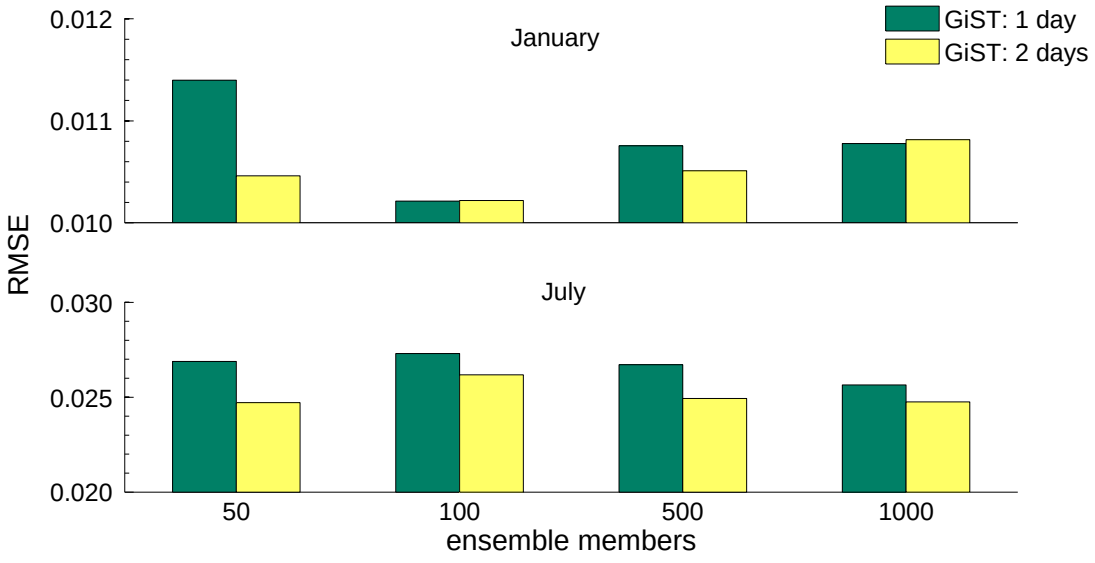


Figure 2.18 RMSE of frequency fraction for generated wet spell length in weather stations of Florida, in January and July, using traditional and extended Orthogonal Markov chain (50, 100, 500 and 1000 yearlong simulations)

Chapter 3. Hydrological Simulating Watershed

3.1 Methods

The runoff response to rainfall events is relevant for watershed hydrology and critical for water management. A sensitivity analysis between different data input (rainfall) was performed for a virtual watershed using the hydrological-hydraulic Storm Water Management Model (SWMM).

The sensitivity of runoff simulations in SWMM to rainfall event inputs produced by different, rain event sources was tested using three different datasets: (*i*) 30 years of observed data, (*ii*) 500 yearlong generated data from the EOMC (2 previous days) using the 'Geospatial and Temporal Weather Generator' (GiST-wg), and (*iii*) 500 yearlong generated data using a point-site 'Weather Generator' (WGEN).

3.1.1 Storm Water Management Model

In SWMM, each sub-watershed receives rainfall through a specific rain gage. This element is associated to a rainfall time series. Then, each sub-watershed relates to a weather station used for generating rainfall (see Chapter 2 for details). Runoff from a sub-watershed drains to junctions nodes interconnected by links. These emulate a virtual watershed configuration with a common surface water channel such as a stream.

Sub-catchments are nonlinear reservoirs that receive inputs from precipitation and/or any upstream sub-catchment. Outflows include infiltration, evaporation and

surface runoff. Ponding, surface wetting and interception of a sub-catchment provide a capacity of the reservoir. When this maximum depression storage is exceeded by depth of water inputs, then surface runoff occurs, as described given by Manning's equation [See Venutelli (2005)]. This empirical formula describes open channel flow based on velocity, hydraulic radius, channel slope, and Manning's coefficient.

Infiltration occurs in pervious zones of unsaturated areas. It is modeled by Horton's equation [See Verma (1982)], Green-Ampt or curve number methods. Conservation of mass and momentum equations control flow routing in conduit links. The equations can be solve by steady flow routing (the simplest), kinematic (normal flow) or dynamic (the most theoretically accurate) wave routings. Evapotranspiration (ET) occurs from sub-catchment surfaces and values can be estimated from daily temperatures using Hargreaves' method and the sites latitude. Other methods to estimate the ET are a single constant value, monthly averages or time-series daily values.

For a more detail explanation about conceptual model refer to the SWMM user's manual (Rossman 2010).

3.1.2 Model Setup

To model the runoff response and analyze spatial variation of rainfall events, the watershed was configured using seven (7) sub-catchments with different percentages of imperviousness. Properties such as extension, slope, elevation and land use were set to different values as shown in Table 3.1, Table 3.2 and Table 3.3. Files created for the simulations are presented in Table A. 1.

Table 3.1 Sub-catchment properties

Sub-catchment	Raingage	Outlet	Area (ha)	Imperv. (%)	Width* (m)	Slope (%)
S1	RG1	J1	353	25	100	2.6
S2	RG2	J1	181	55	60	1.4
S3	RG3	J2	294	38	95	2.8
S4	RG4	J2	399	47	105	1.1
S5	RG5	J3	248	80	85	1.5
S6	RG6	J3	249	72	90	2.7
S7	RG7	J4	416	36	110	2.5

* Characteristic width of the overland flow path

Table 3.2 Conduit properties

Conduits	Inlet	Outlet	Length (m)
C1	J1	J2	350
C2	J2	J3	400
C3	J3	J4	320
C4	J4	Out1	200

Table 3.3 Junction properties

Junction	Invert Elev. (m)
J1	1496
J2	1493
J3	1490
J4	1488
Out1	1486

The model parameters such as pervious and impervious depression storage, Manning's roughness coefficients, and Green-Ampt infiltration were set to their default values in the model. The ET was calculated using Hargreaves' method from daily maximum and minimum temperatures. Because this study did not include any modification for generating air temperatures, ET was modeled by SWMM based on maximum and minimum observed temperatures.

3.1.3 Study area

The configuration of the virtual watershed was set manually in SWMM as shown in Figure 3.1. Conduits from junction J1 to Out1 represented the beginning and end of a surface stream. Junctions J1, J2, J3 and J4 are the outputs of the sub-catchments. Each sub-catchment receives rainfall from a different rain gage over its entire area and drains to an output.

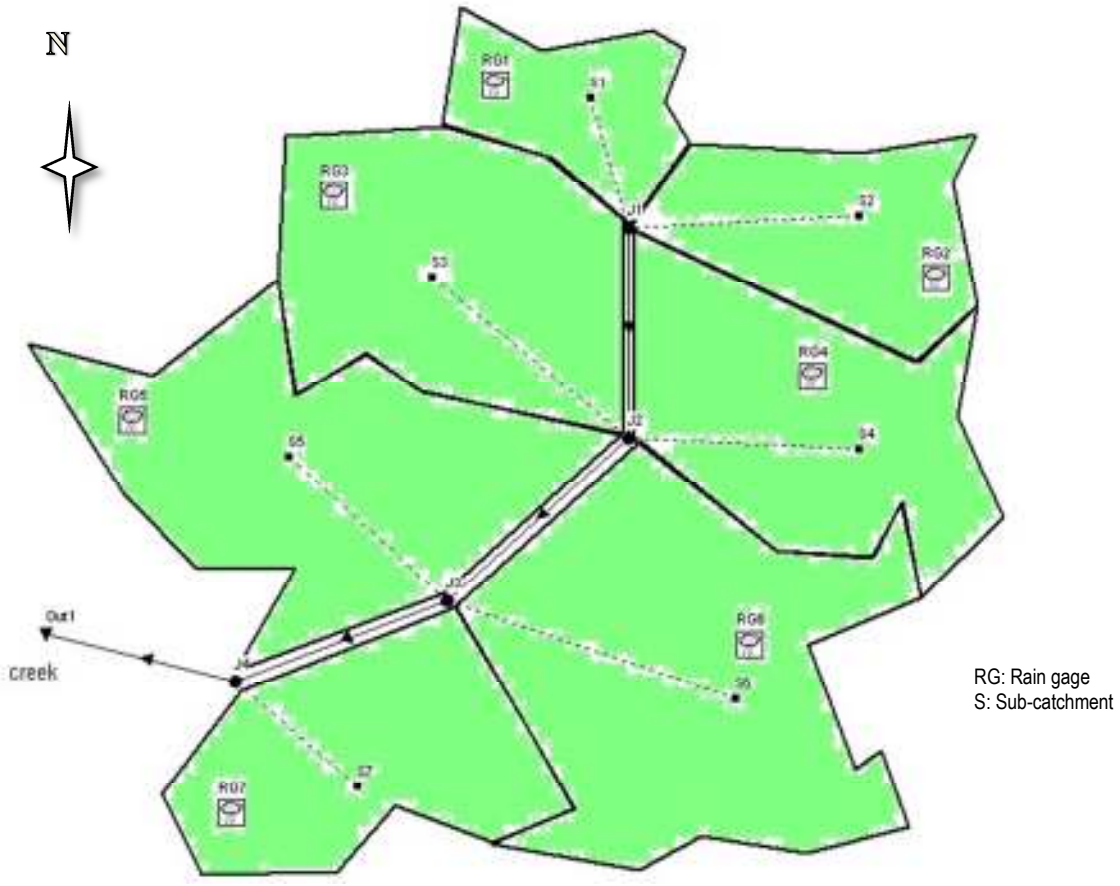


Figure 3.1 Delineation of the virtual watershed

3.2 Results and discussion

3.2.1 Observed rainfall

3.2.1.1 Monthly distribution of ET and runoff, and seasonal accumulated runoff

Monthly values of ET and runoff simulations in SWMM using observed daily rainfall data (30 years) of weather stations in FL are shown in Figure 3.2. Rainfall amounts in the entire virtual watershed had median values around 150 mm/month in the rainy season (June-August). Rainfall was between 50-100 mm/month in dry season (October-May), and amounts showed a variation ±25 mm/month from median values.

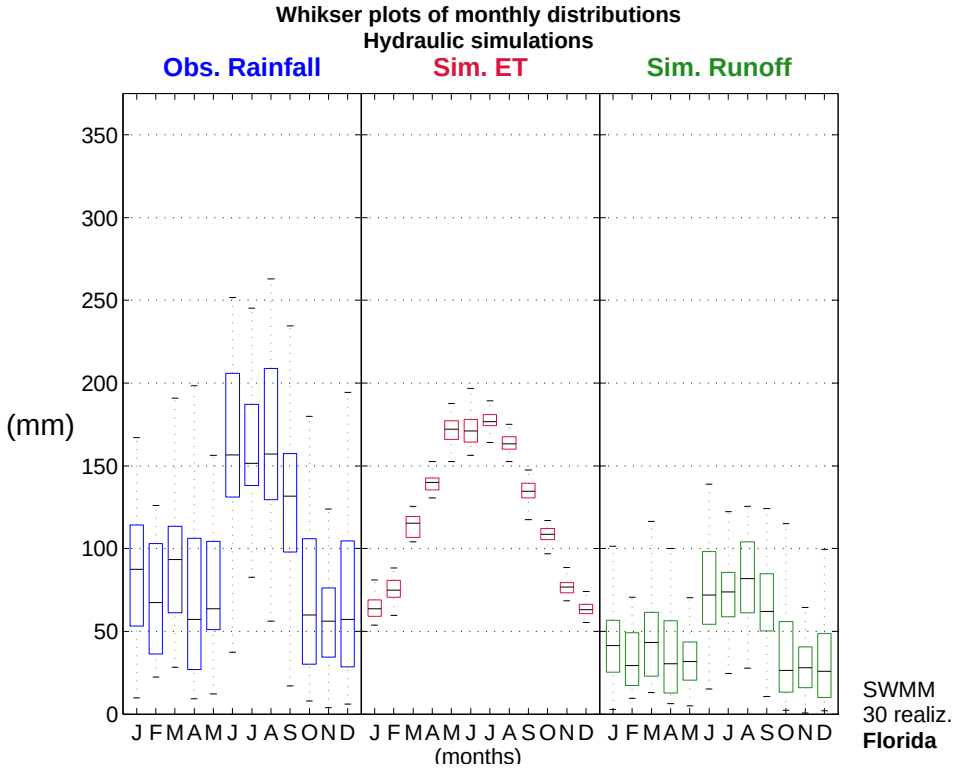


Figure 3.2 Whisker plots (5th, 25th, 50th, 75th and 95th percentiles) for monthly distributions of evapotranspiration (ET) and runoff simulations in a virtual watershed using SWMM with observed daily rainfall data (30 years) of weather stations in Florida

Runoff responses were directly influenced by rainfall amounts and controlled by ET. When rainfall amounts increased in the rainy season (June-August), runoff discharges were the highest (50-100 mm/month). Runoff was lower (25-50 mm/month) when rainfall decreased in dry season (October-May).

ET simulations increased during the colder months (January-February) to warmer ones (May-August), and decreased from September to December. ET was about 175 mm/month in the rainy season, which correspond to summer months in FL (June-August) and it decreased below 80 mm/month in winter (November-February).

Accumulated daily runoff by season in the virtual watershed is shown in Figure 3.3. The steeper trend of accumulation refers to consecutive rainy days, especially in the rainy season (July-September).

Strong changes of accumulated runoff occurred for the months of May and November. These months represented transition periods between dry and rainy seasons and are characterized by occasional dry spells. Thus, accumulated runoff was reduced for this time period.

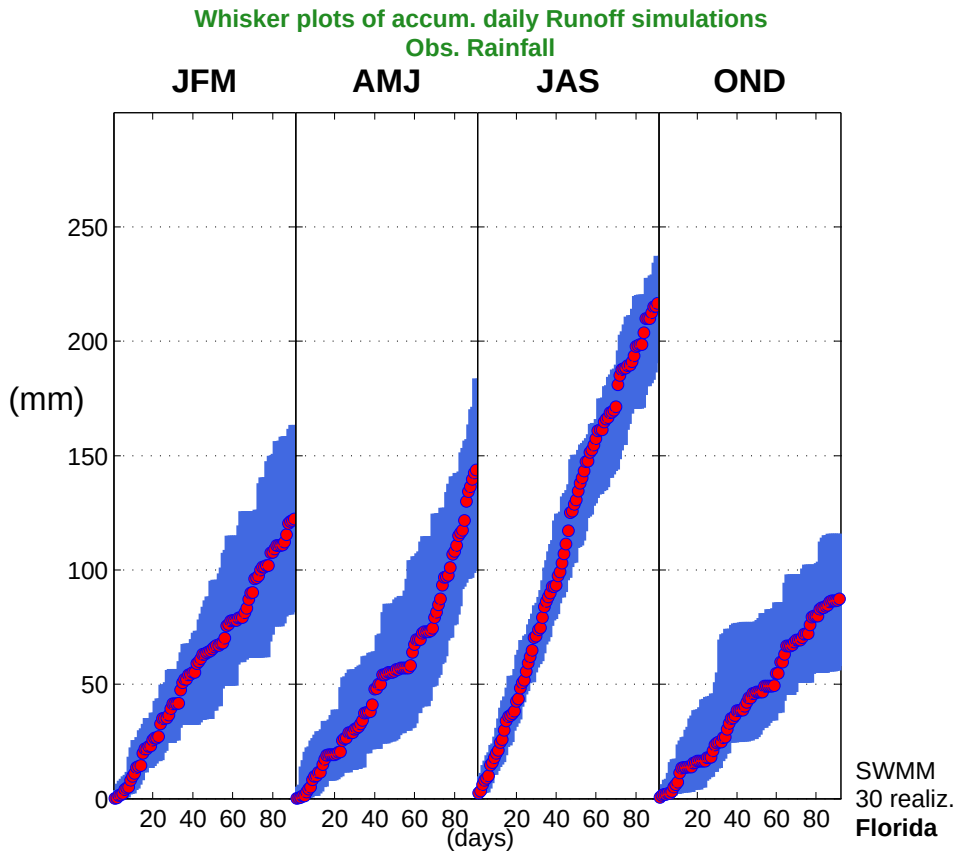


Figure 3.3 Seasonal whisker plots (25th, 50th and 75th percentiles) of accumulated daily runoff simulations in a virtual watershed using SWMM and observed data (30 years) of weather stations in Florida

3.2.2 Generated rainfall using extended Orthogonal Markov chain

3.2.2.1 Monthly distribution of ET and Runoff, and seasonal accumulated runoff

Monthly distributions of rainfall generated by EOMC and simulations of ET and runoff using SWMM are illustrated in Figure 3.4.

July and August had the largest rainfall amounts with 170 mm/month. In the other months, rainfall commonly ranged between 50-90 mm/month. Median rainfall amounts tended to be 10-20 mm/month larger than that of observed data. EOMC produced rainfall distributions were narrower than those of observed rainfall (Figure 3.2 and Figure 3.4).

The interannual variability of the generated rainfall was more homogeneous for large yearlong simulations (500 realizations) than that of the observed data (30 years). This phenomenon, known as 'overdispersion', has been identified in stochastic models (Qian et al. 2002; Baigorria and Jones 2010; Wan et al. 2005; Wilks 1998; Schoof et al. 2005; Hayhoe 2000).

As expected, there were no differences in ET simulations because daily temperatures remained constant. ET was high for the summer months and the rainy season (June-August) and low for the cold months and dry season (November-February). ET increased between February to June, and decrease from August to November, which is to be expected given the temperature trends during these time periods.

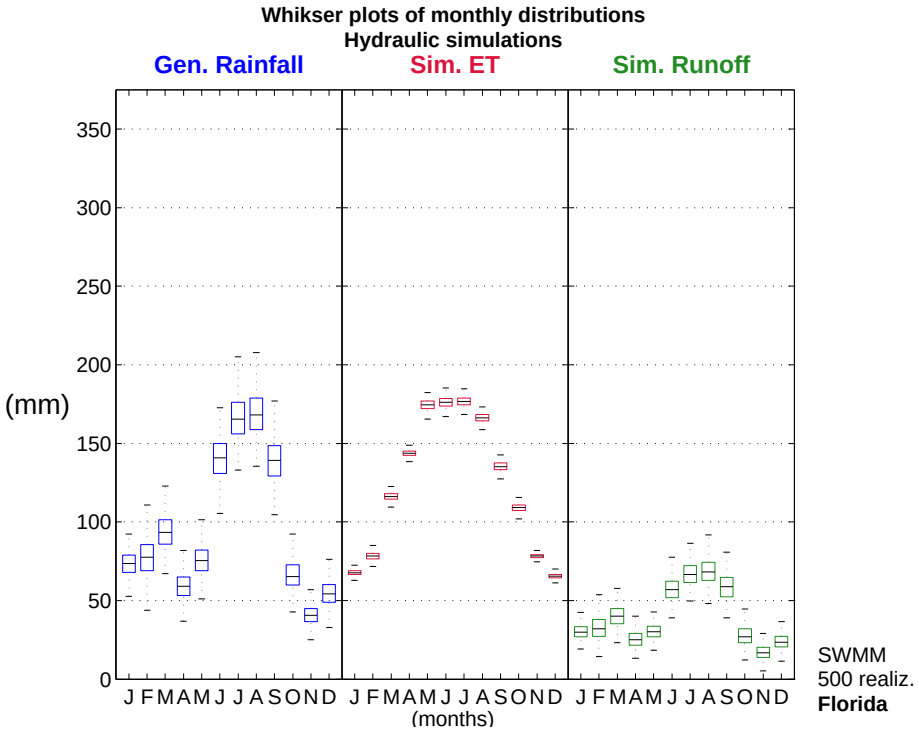


Figure 3.4 Whisker plots (5th, 25th, 50th, 75th and 95th percentiles) for monthly distributions of evapotranspiration (ET) and runoff simulations in a virtual watershed using SWMM with generated daily rainfall data (from extended Orthogonal Markov chain for 500 yearlong simulations) of weather stations in Florida

Median rainfall amounts for the rainy season (June-September) were overestimated by 20 mm/month than the observed data (Figure 3.4 and Figure 3.8). However, this translates into a median simulated runoff 10 mm/month higher than median simulated runoff when observed rainfall data was used (Figure 3.8). Underestimation of the variability of observed monthly rainfall replicates the overdispersion phenomenon in runoff results. This results in higher runoff variability when observed rainfall data were used as input for SWMM (Figure 3.2 and Figure 3.8).

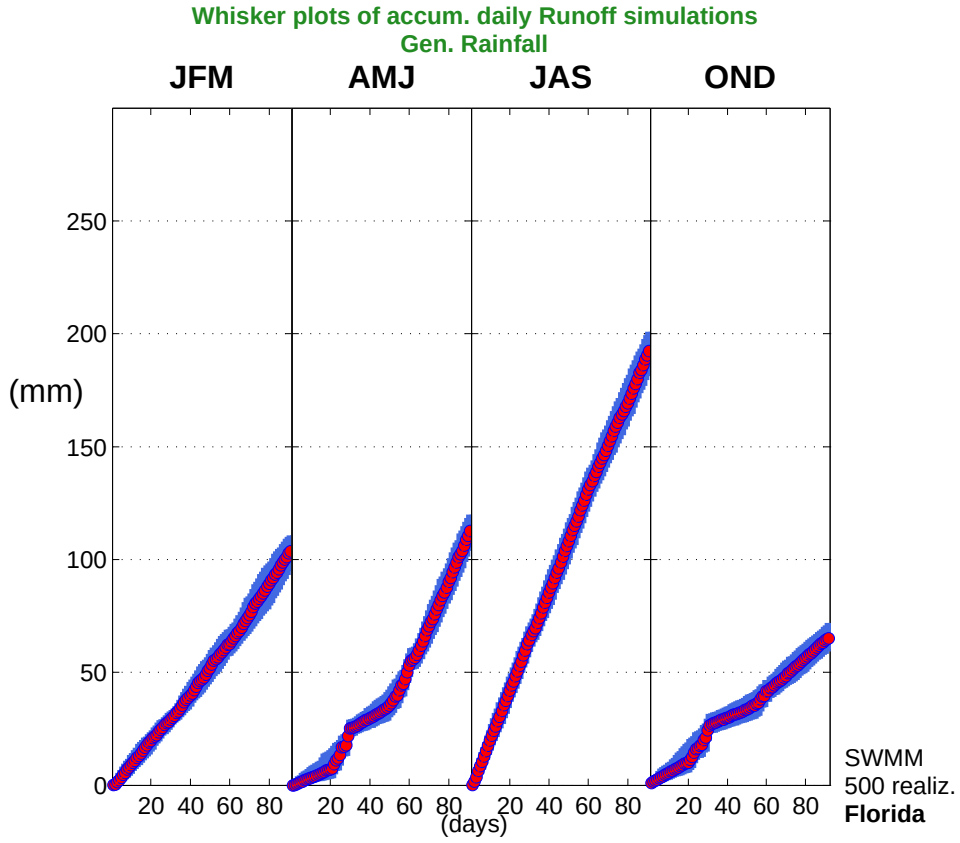


Figure 3.5 Seasonal whisker plots (25th, 50th and 75th percentiles) of accumulated daily runoff simulations in a virtual watershed using SWMM and generated data (from extended Orthogonal Markov chain for 500 yearlong-simulations) of weather stations in Florida

Accumulation of daily runoff from 500 yearlong simulations in SWMM is presented in Figure 3.5. Rainfall data were previously generated by EOMC. In general,

the median accumulated observed runoff was underestimated by about 20 mm at the end of the each season.

3.2.3 Generated rainfall using WeatherMan (WGEN)

3.2.3.1 Monthly distribution of ET and Runoff and seasonal accumulated runoff

Monthly distributions of ET and runoff simulations by SWMM using a rainfall data from WGEN is illustrated in Figure 3.6.

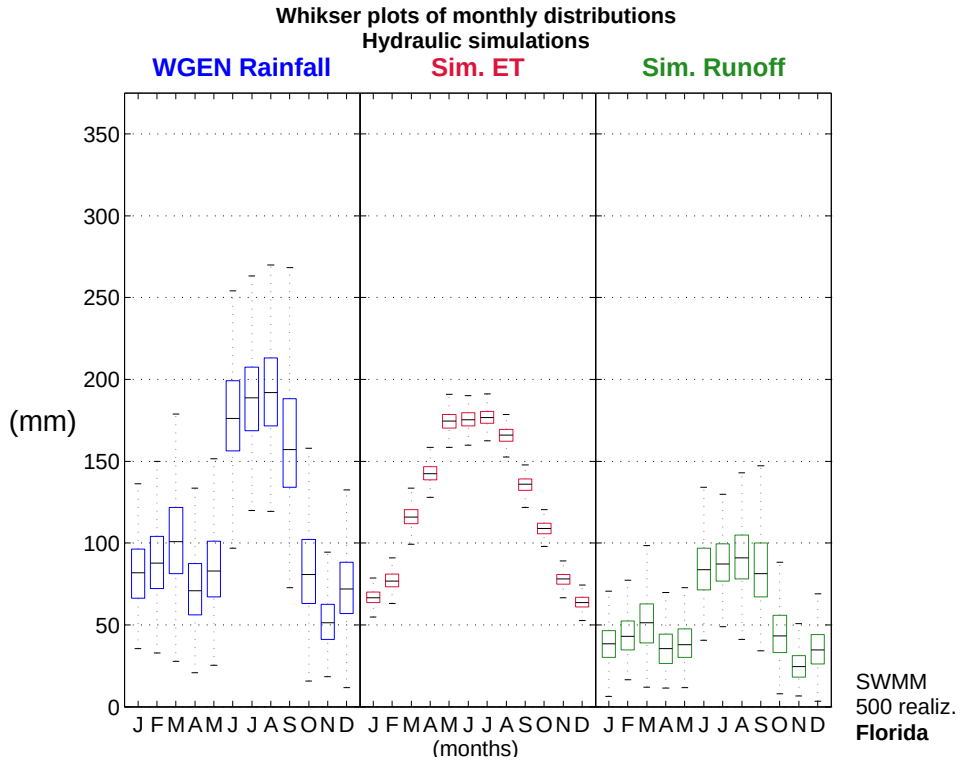


Figure 3.6 Whisker plots (5th, 25th, 50th, 75th and 95th percentiles) for monthly distributions of evapotranspiration (ET) and runoff simulations in a virtual watershed using SWMM with generated daily rainfall data (from WGEN for 500 yearlong-simulations) of weather stations in Florida

Rainfall data from WGEN preserved the pattern of observed rainfall. A rainy season occurred in June-September with median values between 160-190 mm/month. Rainfall amounts were lower around 50-100 mm/month for the remainder of the year.

These values were larger by about 30 mm/month than the simulations using observed rainfall data, and 15-20 mm/month larger than results from EOMC.

SWMM runoff simulations using WGEN rainfall data yielded median runoff values higher than those when EOMC rainfall data was implemented and higher than those when observed data were used (Figure 3.8). WGEN also underestimated the monthly variability of observed rainfall and its runoff. The median runoff was overestimated by 30 mm/month for the rainy months (June-September) and 15 mm/month for the dry season (Figure 3.6 and Figure 3.8).

The accumulated simulated daily runoff for the four seasons using WGEN was about 30-50 mm higher than those simulated using EOMC rainfall data (Figure 3.7). Simulations using WGEN generated rainfall did not represent the observed dry spells. This is in contrast to simulations using rainfall data from EOMC in which dry spells were identified for the months of March and November (See Figure 3.5).

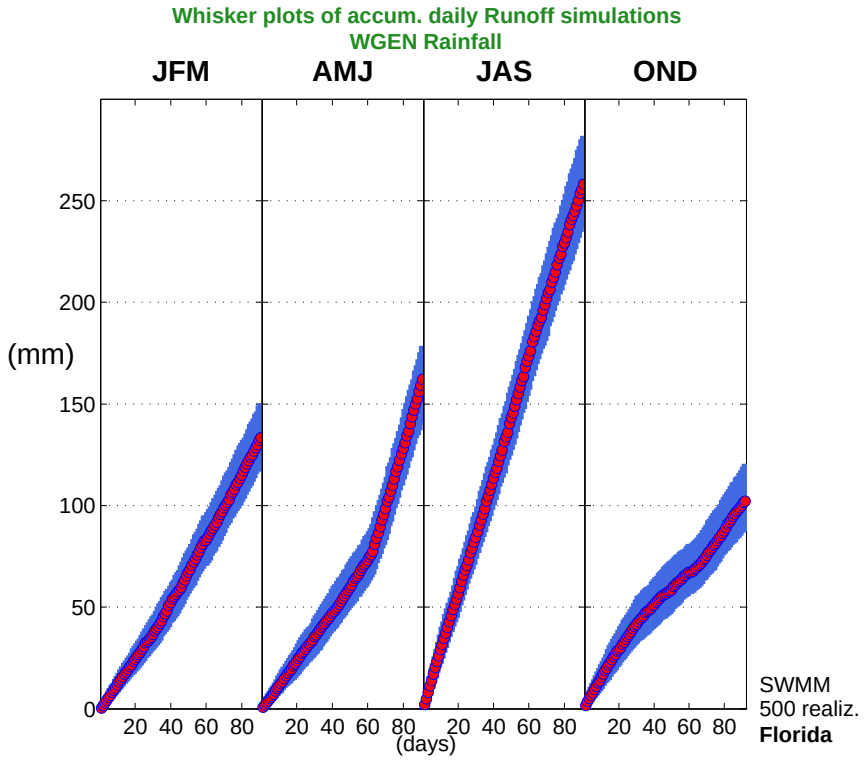


Figure 3.7 Seasonal whisker plots (25th, 50th and 75th percentiles) of accumulated daily runoff simulations in a virtual watershed using SWMM and generated data (from WGEN for 500 yearlong-simulations) of weather stations in Florida

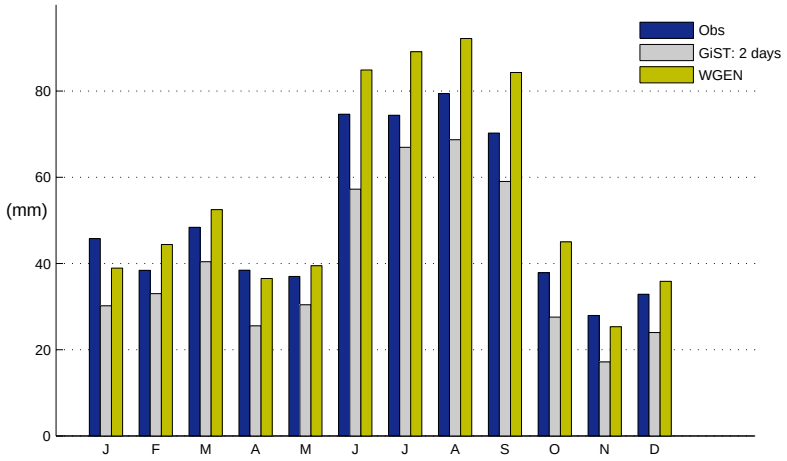


Figure 3.8 Monthly mean runoff simulated in a virtual watershed using SWMM with observed (30 years), and generated daily rainfall data (500 yearlong-simulations) by EOMC and WGEN from weather stations in Florida

Chapter 4. Conclusions

EOMC uses a small number of yearlong simulations (500) than TOMC requires (1000) to represent observed rainfall events. Thus, EOMC reduces by half the computing time. It will allow the evaluation on forecasting and modeling future climate scenarios and also the economical reduction of hardware need for such modeling. EOMC represents an improvement of the GiST-wg for generating rainfall data while preserving spatial correlation between events at stations.

Generated dry and wet spells using EOMC were similar to observed data from weather stations in FL. However, dry spell lengths in NE and CA were overestimated by EOMC around 10-20% in January (Figure 2.10 and Figure 2.15), and wet spell lengths in CA were underestimated 10-20% in July (Figure 2.15). This can be caused because few weather stations with low correlation of rainy occurrences, especially during the dry season (January in NE, and July in CA), among stations were used. For instance, CA had monthly differences of 3-9 rainy days between weather stations during the rainy season (January) (Figure 2.3). Then, the number of weather stations to use by EOMC for generating rainfall daily data in transitional rainfall regimes has to be bigger when large rainfall spatial variability exists.

EOMC preserves the monthly correlation of observed daily rainfall events in station pairs of FL, NE and CA with a Pearson's coefficient (ρ) of 0.926, 0.853 and 0.904 respectively. Moreover, joint monthly probabilities for 500 yearlong simulations of

observed daily rain events in station pairs of FL, NE and CA are also maintained by EOMC. EOMC reproduces the observed joint probabilities in FL, NE and CA with a Pearson's correlation (ρ) equal to 0.973, 0.940 and 0.962 correspondingly.

The hydrological simulation of a virtual watershed with generated rainfall occurrences using EOMC in the GiST-wg showed that simulations in SWMM, using EOMC for generating rainfall data, produces similar median runoff values to those generated using observed rainfall data (±20 mm/month). However, using synthetic rainfall data from WGEN overestimates the median runoff values by 30 mm/month in the rainy season (June-September) and 15 mm/month in the dry season.

References

Ali, A., W. Abtew, S. Van Horn, and N. Khanal, 2000: Temporal and spatial characterization of rainfall over Central and South Florida. *J. Am. Water Resour. Assoc.*, **36**, 833–848.

Arguez, A., I. Durre, S. Applequist, R. S. Vose, M. F. Squires, X. Yin, R. R. Heim Jr., and T. W. Owen, 2012: NOAA's 1981-2010 US Climate Normals: An Overview. *Bull. Am. Metereological Soc.*, **93**, 1687–1697, doi:10.1175/BAMS-D-11-00197.1.

Avila, L. A., and S. R. Stewart, 2013: Atlantic Hurricane Season of 2011*. *Mon. Weather Rev.*, **141**, 2577–2596, doi:10.1175/MWR-D-12-00230.1.

Baigorria, G. A., and J. W. Jones, 2010: GiST: A Stochastic Model for Generating Spatially and Temporally Correlated Daily Rainfall Data. *J. Clim.*, **23**, 5990–6008, doi:10.1175/2010JCLI3537.1.

Baigorria, G. A., J. W. Jones, and J. J. O'Brien, 2007: Understanding rainfall spatial variability in southeast USA. *Int. J. Climatol.*, **760**, 749–760, doi:10.1002/joc.

Black, R. A., and J. Hallett, 2012: Rain Rate and Water Content in Hurricanes Compared with Summer Rain in Miami, Florida. *J. Appl. Meteorol. Climatol.*, **51**, 2218–2235, doi:10.1175/JAMC-D-11-0144.1.

Borah, D. K., 2011: Hydrologic procedures of storm event watershed models: a comprehensive review and comparison. *Hydrol. Process.*, **25**, 3472–3489, doi:10.1002/hyp.8075.

Brandsma, T., and T. Buishand, 1998: Simulation of extreme precipitation in the Rhine basin by nearest-neighbour resampling. *Hydrol. Earth Syst. Sci.*, **2**, 195–208.

Brissette, F. P., M. Khalili, and R. Leconte, 2007: Efficient stochastic generation of multi-site synthetic precipitation data. *J. Hydrol.*, **345**, 121–133, doi:10.1016/j.jhydrol.2007.06.035.

Buishand, T., and T. Brandsma, 2001: Multisite simulation of daily precipitation and temperature in the Rhine basin by nearest-neighbor resampling. *Water Resour. Res.*, **37**, 2761–2776.

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. *Environ. Model. Softw.*, **23**, 1356–1369, doi:10.1016/j.envsoft.2008.04.003.

Charles, S., B. Bates, and J. Hughes, 1999: A spatiotemporal model for downscaling precipitation occurrence and amounts. *J. Geophys. Res.*, **104**, 31657–31669.

Chin, E., 1977: Modeling daily precipitation occurrence process with Markov chain. *Water Resour. Res*, **13**, 949–956.

Gabriel, K. R., and J. Neumann, 1962: A Markov Chain Model for Daily Rainfall Occurrence at Tel-Aviv. *Q. J. R. Meteorol. Soc.*, **88**, 90–95.

Geng, S., F. P. de Vries, and I. Supit, 1986: A simple method for generating daily rainfall data. *Agric. For. Meteorol.*, **36**, 363–376.

Haan, C. T., D. M. Allen, and J. O. Street, 1976: A Markov Chain Model of daily rainfall. *Water Resour. Res.*, **12**, 443–449, doi:10.1029/WR012i003p00443.

Harrold, T. I., A. Sharma, and S. J. Sheather, 2003: A nonparametric model for stochastic generation of daily rainfall occurrence. *Water Resour. Res.*, **39**, doi:10.1029/2003WR002182.

Hayhoe, H., 2000: Improvements of stochastic weather data generators for diverse climates. *Clim. Res.*, **14**, 75–87.

Hayhoe, H., and D. Stewart, 1996: Evaluation of CLIGEN and WXGEN weather data generators under Canadian conditions. *Can. Water Resour. J.*, **21**, 53–67.

Hoogenboom, G. and Coauthors, 2012: Decision Support System for Agrotechnology Transfer (DSSAT) [CD-ROM].

Hopkins, J., and P. Robillard, 1964: Some Statistics of Daily Rainfall Occurrence for the Canadian Prairie Provinces. *J. Appl. Meteorol.*, **3**, 600–602.

Hughes, J., and P. Guttorp, 1994: A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. *Water Resour. Res.*, **30**, 1535–1546.

Johnson, G. L., C. Daly, G. H. Taylor, and C. L. Hanson, 2000: Spatial variability and interpolation of stochastic weather simulation model parameters. *J. Appl. Meteorol.*, **39**, 778–796.

Jones, J. W. and Coauthors, 2003: The DSSAT cropping system model. *Eur. J. Agron.*, **18**, 235–265, doi:http://dx.doi.org/10.1016/S1161-0301(02)00107-7.

Jones, P., and P. Thornton, 1999: Fitting a third-order Markov rainfall model to interpolated climate surfaces. *Agric. For. Meteorol.*, **97**, 213–231.

Jones, P., and P. Thornton, 2000: MarkSim: Software to generate daily weather data for Latin America and Africa. *Agron. J.*, **453**, 445–453.

Khalili, M., R. Leconte, and F. Brissette, 2007: Stochastic Multisite Generation of Daily Precipitation Data Using Spatial Autocorrelation. *J. Hydrometeorol.*, **8**, 396–412, doi:10.1175/JHM588.1.

Kim, T., and J. Valdés, 2005: Synthetic generation of hydrologic time series based on nonparametric random generation. *J. Hydrol. Eng.*, **10**, 395–404.

Kim, T., H. Ahn, G. Chung, and C. Yoo, 2007: Stochastic multi-site generation of daily rainfall occurrence in south Florida. *Stoch. Environ. Res. Risk Assess.*, **22**, 705–717, doi:10.1007/s00477-007-0180-8.

Lambert, M., 2003: A non-parametric hidden Markov model for climate state identification. *Hydrol. Earth Syst. Sci.*, **7**, 652–667.

Leander, R., and T. A. Buishand, 2009: A daily weather generator based on a two-stage resampling algorithm. *J. Hydrol.*, **374**, 185–195, doi:10.1016/j.jhydrol.2009.06.010.

Lee, S.-B., C.-G. Yoon, K. W. Jung, and H. S. Hwang, 2010: Comparative evaluation of runoff and water quality using HSPF and SWMM. *Water Sci. Technol.*, **62**, 1401–1409, doi:10.2166/wst.2010.302.

Lindström, G., B. Johansson, M. Persson, M. Gardelin, and S. Bergström, 1997: Development and test of the distributed HBV-96 hydrological model. *J. Hydrol.*, **201**, 272–288.

Lowry, W., and D. Guthrie, 1968: Markov Chains of Order Greater than ONE*. *Mon. Weather Rev.*, **96**, 798–801.

Markov, A. A., 1906: Rasprostranenie zakona bol'shih chisel na velichiny, zavisyaschie drug ot druga (in Russian). *Izv. Fiz. Obs. pri Kazan. Univ.*, **15**, 135–156.

Mhanna, M., and W. Bauwens, 2012: A stochastic space-time model for the generation of daily rainfall in the Gaza Strip. *Int. J. Climatol.*, **32**, 1098–1112, doi:10.1002/joc.2305.

Ng, W. W., and U. S. Panu, 2010: Comparisons of traditional and novel stochastic models for the generation of daily precipitation occurrences. *J. Hydrol.*, **380**, 222–236, doi:10.1016/j.jhydrol.2009.11.002.

Nicks, A. D., L. J. Lane, and G. A. Gander, 1995: CLIGEN Weather generator. *USDA–Water Erosion Prediction Project hillslope profile and watershed model documentation*, D.C. Flanagan and M.A. Nearing, Eds., USDA–ARS National Soil Erosion Research Laboratory, 2.1–2.22.

Pickering, N., J. Hansen, and J. Jones, 1994: WeatherMan: a utility for managing and generating daily weather data. *Agron. J.*, **86**, 332–337.

Qian, B., J. Corte-Real, and H. Xu, 2002: Multisite stochastic weather models for impact studies. *Int. J. Climatol.*, **22**, 1377–1397, doi:10.1002/joc.808.

Richardson, C. W., 1981: Stochastic Simulation of Daily Precipitation, Temperature, and Solar Radiation. *Water Resour. Res*, **17**, 182–190.

Richardson, C. W., and D. A. Wright, 1984: *WGEN: A model for generating daily weather variables*. U.S. Department of Agriculture, Agricultural Research Service, ARS-8,.

Roldan, J., and D. Woolhiser, 1982: Stochastic daily precipitation models: 1. A comparison of occurrence processes. *Water Resour. Res.*, **18**, 1451–1459.

Rossman, L. A., 2010: *Storm Water Management Model User's Manual, Version 5.0*. National Risk Management Reserach Laboratory, Office of Research and Development, US Environmental Protection Agency, Cincinnati, OH,.

Schoof, J. T., and S. C. Pryor, 2008: On the Proper Order of Markov Chain Model for Daily Precipitation Occurrence in the Contiguous United States. *J. Appl. Meteorol. Climatol.*, **47**, 2477–2486, doi:10.1175/2008JAMC1840.1.

Schoof, J. T., A. Arguez, J. Brolley, and J. J. O'Brien, 2005: A new weather generator based on spectral properties of surface air temperatures. *Agric. For. Meteorol.*, **135**, 241–251, doi:10.1016/j.agrformet.2005.12.004.

Semenov, M., and E. Barrow, 1997: Use of a stochastic weather generator in the development of climate change scenarios. *Clim. Change*, **35**, 397–414, doi:10.1023/A:1005342632279.

Sharma, V., and S. Irmak, 2012: Interpolated precipitation, reference evapotranspiration, actual crop evapotranspiration, and net irrigation requirements in Nebraska: Part I. Precipitation and reference. *Trans. ASABE*, **55**, 907–921.

Smith, J. A., G. Villarini, and M. L. Baeck, 2010: Mixture Distributions and the Hydroclimatology of Extreme Rainfall and Flooding in the Eastern United States. *J. Hydrometeorol.*, **12**, 294–309, doi:10.1175/2010JHM1242.1.

Srikanthan, R., and T. a. McMahon, 2001: Stochastic generation of annual, monthly and daily climate data: A review. *Hydrol. Earth Syst. Sci.*, **5**, 653–670, doi:10.5194/hess-5-653-2001.

Srikanthan, R., T. I. Harrold, A. Sharma, and T. a. McMahon, 2005: Comparison of two approaches for generation of daily rainfall data. *Stoch. Environ. Res. Risk Assess.*, **19**, 215–226, doi:10.1007/s00477-004-0226-0.

Thyer, M., and G. Kuczera, 2003: A hidden Markov model for modelling long-term persistence in multi-site rainfall time series 1. Model calibration using a Bayesian approach. *J. Hydrol.*, **275**, 12–26, doi:10.1016/S0022-1694(02)00412-2.

Tseng, H.-W., T.-C. Yang, C.-M. Kuo, and P.-S. Yu, 2012: Application of Multi-site Weather Generators for Investigating Wet and Dry Spell Lengths under Climate Change: A Case Study in Southern Taiwan. *Water Resour. Manag.*, **26**, 4311–4326, doi:10.1007/s11269-012-0146-6.

Tsihrintzis, V., and R. Hamid, 1998: Runoff quality prediction from small urban catchments using SWMM. *Hydrol. Process.*, **12**, 311–329.

Venutelli, M., 2005: A constitutive explanation of Manning's formula. *Meccanica*, **40**, 281–289, doi:10.1007/s11012-005-6529-5.

Verma, S. C., 1982: Modified Horton Infiltration Equation. *J. Hydrol.*, **58**, 383–388, doi:10.1016/0022-1694(82)90047-6.

Wallis, T. W. R., and J. F. Griffiths, 1995: An assessment of the weather generator (WXGEN) used in the erosion/productivity impact calculator (EPIC). *Agric. For. Meteorol.*, **73**, 115–133, doi:10.1016/0168-1923(94)02172-G.

Wan, H., X. Zhang, and E. Barrow, 2005: Stochastic modelling of daily precipitation for Canada. *Atmosphere-Ocean*, **43**, 23–32.

Wang, K., and A. Altunkaynak, 2012: Comparative Case Study of Rainfall-Runoff Modeling between SWMM and Fuzzy Logic Approach. *J. Hydrol. Eng.*, **17**, 283–291, doi:10.1061/(ASCE)HE.1943-5584.0000419.

Wilks, D. S., 1998: Multisite generalization of a daily stochastic precipitation generation model. *J. Hydrol.*, **210**, 178–191, doi:10.1016/S0022-1694(98)00186-3.

Wilks, D. S., 1999: Simultaneous stochastic simulation of daily precipitation, temperature and solar radiation at multiple sites in complex terrain. *Agric. For. Meteorol.*, **96**, 85–101, doi:10.1016/S0168-1923(99)00037-4.

Zhang, Z., and P. Switzer, 2007: Stochastic space-time regional rainfall modeling adapted to historical rain gauge data. *Water Resour. Res.*, **43**, W03441, doi:10.1029/2005WR004654.

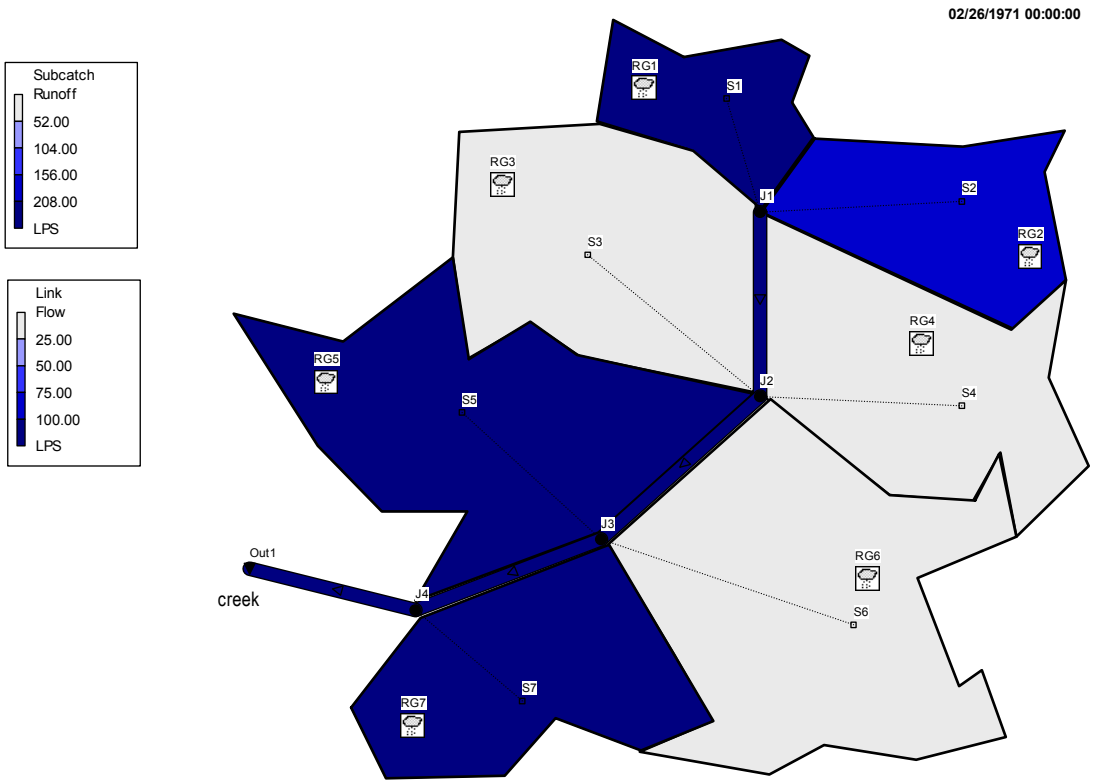
APPENDIX A. ELEMENT PROPERTIES IN VIRTUAL WATERSHED

Table A. 1 List of files created for modeling the virtual watershed in EPA-SWMM

Source of rainfall	.inp file	.ini file
Historic data	virtual_watershed.inp	virtual_watershed.ini
EOMC*	2d_500y_Virtual_Watershed.inp	2d_500y_Virtual_Watershed.ini
WGEN	WGEN_500y_Virtual_Watershed.inp	WGEN_500y_Virtual_Watershed.ini

* Extended orthogonal Markov chain

APPENDIX B. RUNOFF SIMULATIONS IN EPA-SWMM



LPS: Liters per Second

Figure B. 1 Visualization of runoff modeling in EPA-SWMM

APPENDIX C. AVERAGE RAINFALL (1981-2010) IN CONTIGUOUS STATES OF USA

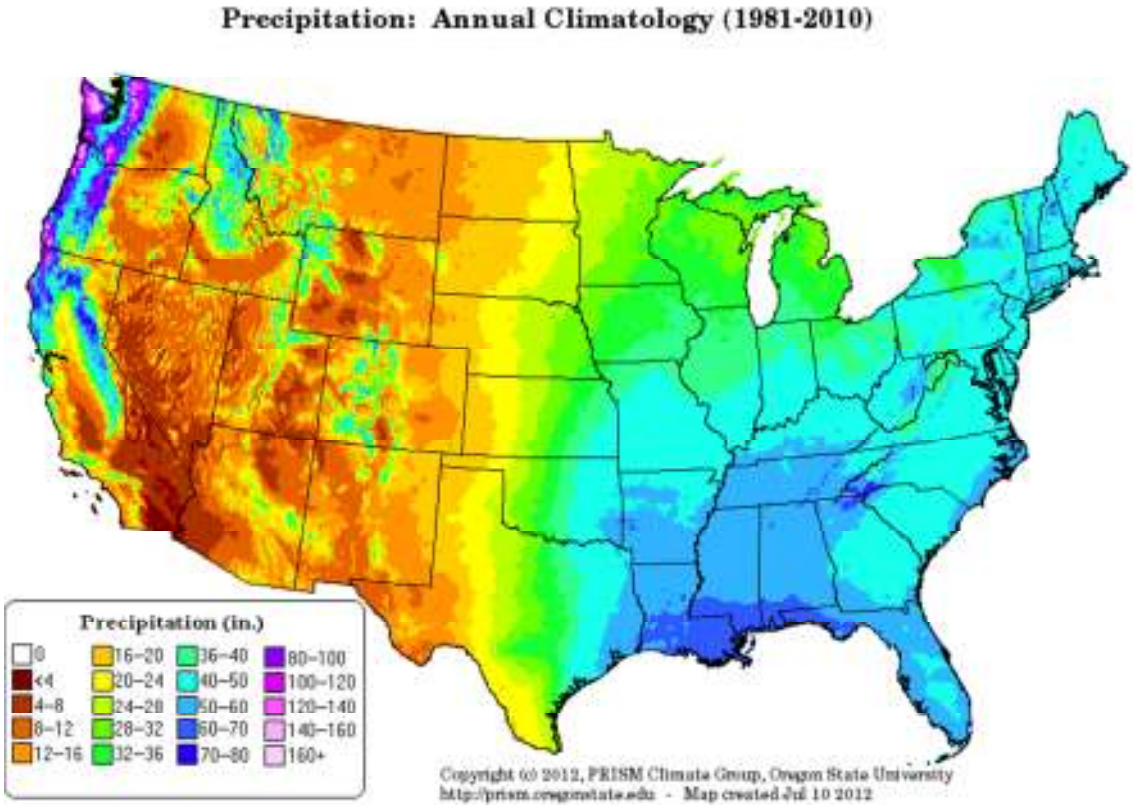


Figure C. 1 Annual average precipitation (1981-2010) in contiguous states of USA*

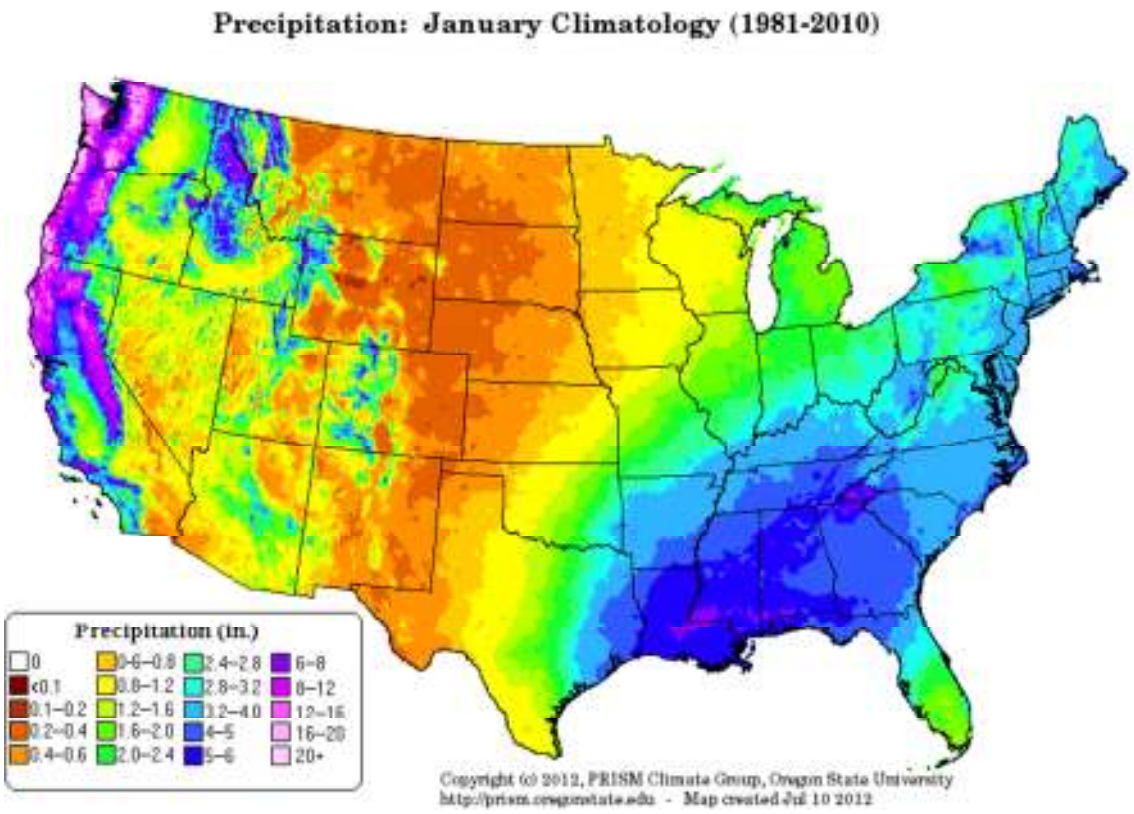


Figure C. 2 January average precipitation (1981-2010) in contiguous states of USA[*]

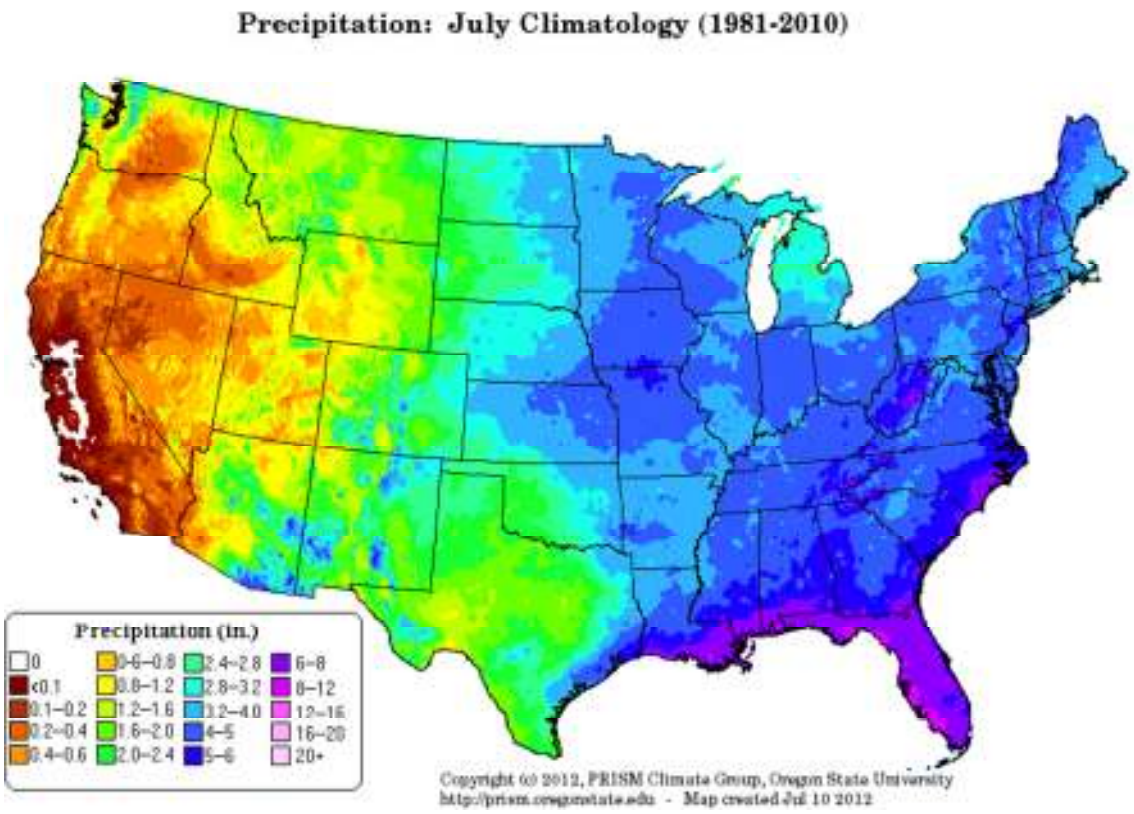


Figure C. 3 July average precipitation (1981-2010) in contiguous states of USA[*]

[*] Figure C. 1, Figure C. 2 and Figure C. 3: "Copyright © 2012, PRISM Climate Group, Oregon State University, http://prism.oregonstate.edu - Map created July 10 2012".