Stochastic precipitation generation based on a multivariate autoregression model

Oleg V. Makhnin *

New Mexico Institute of Mining and Technology

Devon L. McAllister

New Mexico Institute of Mining and Technology

* *Corresponding author address:* Oleg V. Makhnin, Dept. of Mathematics, New Mexico Institute of Mining and Technology, 801 Leroy St., Socorro, NM 87801

E-mail: olegm@nmt.edu

ABSTRACT

The problem of stochastic precipitation generation has long been of interest. A good generator should produce time series with statistical properties to match those of the real precipitation. Here, we present a multivariate autoregression model designed to capture the covariance and lag-1 cross-covariance structure of the precipitation measurements. We use truncated and power-transformed normal distribution to simultaneously model both occurrences and amounts of daily precipitation. We illustrate the methodology using daily rain gauge data sets for three areas in Continental US.

Keywords: Gibbs sampler, Bayesian inference, precipitation generation, multivariate autoregression, cross-covariance, spatial-temporal process

1. Introduction

Precipitation is an extremely complex spatial-temporal process. Synthetic precipitation generators, developed in order to simulate this process, have been an active topic of research for more than three decades. They have found applications as inputs for agriculture and hydrology models (Semenov and Porter, 1995), and are important for climate change studies (Brissette et al, 2007). An ideal precipitation generator is supposed to possess statistical properties that are very close to those of the real precipitation process, and it allows to inexpensively generate very long precipitation sequences.

Simpler generators work for a single site at which simulation is desired, and more sophisticated versions are multisite generators. The latter are designed to not only match the precipitation statistics for each observation site separately, but also to capture and reproduce dependencies between sites. Finally, a high-density precipitation generator is desirable, that would interpolate between the observation sites and provide the user with simulated precipitation data over the entire study area.

Most of the recently introduced generators employ a separate occurrence process (which is usually modeled by a Markov chain), and, for the sites where precipitation is deemed to occur, a continuously distributed intensity process, for which various distributions have been used (Gamma, Exponential and their mixtures, empirical and so on). Wilks (1998) introduces correlations between sites through a delicate system of correlated Gamma random variables. Yang et al (2005) consider a smaller study region within which all the stations are assumed to be equally dependent on each other. Apipattanavis et al (2007) introduce region-wide dry and wet weather states, and use a non-parametric multisite generator con-

ditioned on these states. Mehrotra and Sharma (2007) propose a non-parametric multisite precipitation amounts generator, coupled with a separate precipitation occurrence generator based on Markov Chains, and dependent on the long-memory "wetness" indicator. Furrer and Katz (2007) use a generalized linear model (GLM) instead of a Markov Chain for the occurrence process, followed by another GLM for the intensity process.

Bárdossy and Plate (1992) followed by Stehlík and Bárdossy (2002), and Sansó and Guenni (2000) use a different approach based on truncated and power transformed (TPT) normal distribution. For this type of model, the negative values of the distribution correspond to the periods of no precipitation. We propose a version of this approach using multivariate autoregression (MAR), designed to reproduce correlations and lag-1 cross-correlations between sites.

The use of multivariate normal (MVN) distribution in conjunction with TPT makes is convenient to model multivariate processes, on one hand, and to unify the occurrence and intensity processes, on the other. This helps borrowing information between the two processes, thus potentially improving the model performance.

However, some questions have been raised (e.g. Yang et al(2005)) about possibly different natures of the occurrence and intensity processes. Here, we investigate feasibility of TPT modeling. We also discuss the difficulties in creating a synthetic precipitation product over a dense grid of locations.

To fit the TPT-based models, we use Gibbs sampler. It is a Bayesian inference tool suited well to the problem of filling in missing data (such as the unknown negative values we have to deal with when no precipitation occurred). The Gibbs sampler is discussed in Section 3.c.

The data sets used for illustration are for three areas in Continental US: a set of 22 gauge stations in Southwestern Colorado (CO), 27 stations in Minnesota (MN) and 25 stations in teh state of New York (NY), see Fig. 1. These areas exhibit different climatic behaviors, especially in Colorado. The data sets are taken from `http://www.cdc.noaa.gov/USstation/` and contain daily precipitation amounts for the period of 1950-1999 (Eischeid et al, 2000). To reflect seasonal variability of precipitation, we fit separate models for each month, judging the conditions within each month to change little enough to be considered stationary.

In Section 2, we discuss applicability of the univariate TPT model for the joint occurrence/amounts modeling.

In Section 3, we describe two autoregressive models for the daily precipitation amounts. In Section 4, we present the results of the model application to our study areas.

2. Preliminary model fitting and statistical issues of precipitation generation

The multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}, \mathbf{V})$ for a vector $\mathbf{x}$ of dimension d, mean $\boldsymbol{\mu}$ and variance matrix $\mathbf{V}$ has the density

$$\frac{1}{(2\pi)^{d/2}|\mathbf{V}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{V}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

We can use the same distribution for describing both the intensity (the amount of precipitation on wet days) and occurrence (dry or wet days) simultaneously. Consider the *precipitation*

potential W_t that falls below 0 if no precipitation has occurred on that day. For a positive precipitation value R_t, we infer $W_t = R_t^{1/\beta}$, where β is a transformation parameter depending on the skewness of the distribution, with higher values of β corresponding to heavier-tail distributions for R.

The idea goes back to Stidd (1973) and is also discussed in Sansó and Guenni (2000) and Stehlík and Bárdossy (2002). The latter introduce the spatial-temporal distribution of W based on weather patterns identified in the area, as inferred from atmospheric pressure measured over a network of sites. However, their approach demands detailed weather information. Our model is conceptually simpler and is comparable with Yang et al (2005).

The difference from Yang et al (2005) is that they introduce a special process for the precipitation occurrence. This is the classical approach in synthetic precipitation (Todorovic and Woolhiser, 1975). In case of Yang et al, their model possesses some nice characteristics, but it lacks the ability to fine-tune the occurrence process (the stations are assumed to be interchangeable within the area). Also, most approaches that separate occurrence and amounts have to deal with the *spatial intermittence* effect (Bárdossy & Plate (1992), Koutsoyannis (2006)), where smooth transitions are required between wet and dry areas. For example, Wilks (1998) notes that lower rainfall intensity is observed when more neighboring stations are dry. This can be partially accounted for by using ad-hoc methods, but they result in a more complex model.

Otherwise, our modeling concept is somewhat similar to the GLM's employed by Yang et al (2005) and Furrer and Katz (2007). To model the interactions between stations we introduce multivariate autoregression (MAR) terms that link the precipitation potential for the day t

and site i to the potential for the previous day (or days) for all sites.

We would like to investigate the feasibility of using the TPT normal distribution for the simultaneous occurence/intensity modeling. First, we will examine how the univariate TPT represents precipitation amounts.

a. TPT model fit to empirical data

Many authors used different distributions to describe precipitation amounts. Notably, Wilks (1999) uses mixtures of exponentials, Yang et al (2005) and Furrer and Katz (2007) use transformed Gamma distributions etc.

Here, we check the validity of the TPT normal distribution for daily precipitation amounts data, one month at a time.

To assess the fit, we produce the empirical CDF plots for the actual data, overlaid with the fitted TPT-model CDF. (The empirical CDF, or cumulative distribution function, $\hat{F}(x)$ is the proportion of days for which the precipitation $\leq x$.)

The parameters fitted are the normal distribution mean μ, standard deviation σ and the transformation parameter β. Some examples of the plots obtained are shown in Fig.6 and Fig.7. These plots also show the precipitation probability at the same time, which is given by $P(R > 0) = 1 - F(0)$. We can conclude that the TPT fits the data satisfactorily, even though the assumption that the rainfall distribution is constant for different climatic conditions is hardly going to be satisfied. Also, the TPT normal distribution applies to different spatial scales, both the single site (Fig.6) and the total over all sites (Fig.7), albeit with

different parameter values. The estimated parameter values for these distributions are given in Table 1. For example, the values of μ and β are higher for the total precipitation.

In addition, we note that rounding of the precipitation values (in the present data set they are rounded to the closest 0.01 inch) poses some difficulties, and the fitted parameters are noticeably different with and without taking rounding into account. This is especially important for the precipitation values close to 0.

b. Challenges of modeling and statistical tools used for assessment

Some descriptive statistics for both multisite occurrence and intensity were offered by Wilks (1998). In his opinion, a successful synthetic precipitation generator should match these statistics with those of real precipitation. However, this task is extremely difficult. It is not enough to generate single-site (univariate) precipitation values well.

A difficulty in modeling multi-site precipitation can be illustrated by the following "correlation inversion" effect. This was observed when computing lag-1 cross-correlations of precipitation between sites. For NY and MN study areas, we observe that high lag-1 cross-correlations from site i to site j are accompanied by *low* lag-1 cross-correlations from site j to site i. See Fig.3 for the MN study area.

This effect might partially be explained by the general movement of weather systems in the direction from West to East. This systematic movement will be addressed in Section 3.a. On the other hand, this effect was not observed for CO study area (Fig. 4). Also, the statistical relationship between sites is not completely determined by their locations. Measurements

for some pairs of sites exhibit fairly high lag-1 cross-correlations at a distance, while the others do not, and this effect is surprisingly stable over different months. See for example, Fig.5. This idiosyncratic behavior presents a considerable difficulty in modeling, and also in producing high-resolution precipitation plots. This is the reason we focus on lag-1 cross-correlations.

3. Autoregression models using TPT

We may start with a simple single-site autoregression, in the form of a state-space model

$$\begin{cases} Z_t = rZ_{t-1} + \xi_t \\ W_t = \mu + Z_t + \varepsilon_t \end{cases}$$

Here, the first equation is an *evolution equation* involving "states" Z_t and the second one is the *observation equation* (of course, only positive values of W are observed directly), and ε_t represent transient effects (white noise).

One attraction of using a linear model is that other terms (covariates) can be easily added: time trends, elevation, some global climate index like El-Niño Southern Oscillation (ENSO), etc.

To account for dependencies between sites, we will introduce to variants of this model: one is area-based ("Convolution model") and the other is site-based ("Multisite model").

a. Convolution model

In order to convey large-scale spatial information, we might use a convolution model. First, we will make a regular grid over the study area, and consider the area-averaged values of precipitation potential, $\boldsymbol{\theta}_t$, over the grid nodes $i = 1, ..., m$. (See Section 3.b below for the explanation of terms in the equation.)

$$\boldsymbol{\theta}_t = \mathbf{r}_\theta \boldsymbol{\theta}_{t-1} + \boldsymbol{\xi}_t \tag{1}$$

Then, the fitted values of $\boldsymbol{\theta}_t$ will extend to the sites $\mathbf{Z}_t$ via a convolution method described by Higdon (2007),

$$Z_{tj} = \sum_{i=1,...,m} \psi(d_{ij})\theta_{ti} \tag{2}$$

where d_{ij} is the distance between grid location i and site j, the kernel function is the Gaussian $\psi(x) = exp(-x/(2\sigma^2))$, suitably normalized so that $\sum_{i=1,...,m} \psi(d_{ij}) = 1$.

The equations (1) and (2) are coupled with the observation equation

$$\mathbf{W}_t = \mathbf{Z}_t + \mathbf{S} + \boldsymbol{\varepsilon}_t,$$

where $\mathbf{S}$ are the Site effects, representing the average precipitation potential values at particular sites. The residuals $\boldsymbol{\varepsilon}_t$ are assumed to be $\mathcal{N}(0, \tau^2\mathbf{I})$.

As an example, we will consider a two-node grid ($m = 2$). For MN data, this is shown in Fig.2. If there is any persistent movement of weather systems in a direction between nodes, it should be reflected in the $\mathbf{r}_\theta$ coefficients. We indeed observe this behavior over all months except November and December in MN, and all the time in NY, see Section 4. Not observed

over Colorado.

Further refining the grid, we can take into account smaller-scale spatial information. The convolution model also makes it easy to interpolate precipitation values over the entire study area, which should be important for applications. However, due to the idiosyncratic behavior of correlations between sites, we cannot hope to reproduce them very well with this model.

b. Multisite model

This model focuses closely on reproducing covariances and lagged cross-covariances between sites, including correlation inversion effect. Consider multivariate autoregression (MAR)

$$\begin{cases} \mathbf{Z}_t = \mathbf{r}\mathbf{Z}_{t-1} + \boldsymbol{\xi}_t \\ \mathbf{W}_t = \mathbf{Z}_t + \mathbf{S} + \boldsymbol{\varepsilon}_t, \end{cases} \tag{3}$$

where $\mathbf{Z}_t$ is a vector of precipitation potentials at the time t, $\mathbf{r}$ is an $N \times N$ matrix with (possibly complex) eigenvalues λ_i such that $|\lambda_i| < 1$, for all i. This ensures that the process $\mathbf{Z}$ is stationary. Innovations in the model, $\boldsymbol{\xi}_t$, are multivariate normal with mean 0 and variance matrix $\mathbf{V}_\xi$.

Note that the number of parameters in the model describing $\mathbf{r}$ and $\mathbf{V}_\xi$ is $N^2 + N(N+1)/2$, this exactly equals the number of parameters in the zero- and first-order autocovariance matrices of the process $\mathbf{Z}$, $\hat{\Sigma}_0$ and $\hat{\Sigma}_1$. Thus, if $\hat{\Sigma}_0$ and $\hat{\Sigma}_1$ were known, we could fit the parameters in (3) as follows:

$$\mathbf{r}^T = \hat{\Sigma}_0^{-1}\hat{\Sigma}_1, \quad \mathbf{V}_\xi = \hat{\Sigma}_0 - \mathbf{r}\hat{\Sigma}_0\mathbf{r}^T \tag{4}$$

However, in reality the values of $\mathbf{Z}$ are generated according to the Gibbs sampling algorithm and then we get samples from conditional posterior distributions of $\mathbf{r}$ and $\mathbf{V}_\xi$. For example,

$$\mathbf{r} \sim \mathcal{N}\left(\hat{\Sigma}_0^{-1}\hat{\Sigma}_1, \hat{\Sigma}_0^{-1}\right)$$

The multisite model does not directly account for site locations. Spatial information may still be used for between-site interpolation in the way described by Wilks (2008). One may expect some benefit in combining multisite model and convolution model as a way to obtain a high-density precipitation simulator.

c. Fitting the models

The models are fitted using the Bayesian approach, via the Gibbs sampler. See, for example, Gelman et al (2003). This is a powerful approach, just as needed due to a large number of missing variables (negative W values) and a complex hierarchical model structure. One advantage of the Gibbs sampler is that various parameters and missing variables are fitted in independent blocks, based on FCP's (full conditional posteriors). The FCP are defined as conditional density (or, for discrete random variables, their conditional probability mass function) of some model variable, given all the other model variables combined. The variable of interest is drawn from its FCP, based on current samples of all other variables. Then, we move on to another variable etc. The process is repeated until the distributions are determined to have converged, and then is run long enough to obtain an MCMC (Markov Chain Monte-Carlo) sample from the joint posterior distribution of all model variables and parameters. See Appendix for more information.

4. Results

a. Convolution model

The results for Convolution model over MN study area are presented in Table 2. Note the asymmetry in the estimated $\mathbf{r}_\theta$ coefficients. They indicate movement in the West-East direction. For example, the estimates for August are

$$\mathbf{r} = \begin{bmatrix} 0.402 & -0.144 \\ 0.236 & 0.323 \end{bmatrix}, \quad \mathbf{V}_\xi = \begin{bmatrix} 0.542 & 0.087 \\ 0.087 & 0.627 \end{bmatrix}.$$

The covariance matrix of $\boldsymbol{\theta}$, $\hat{\Sigma}_0$, and the lag-1 cross-covariance, $\hat{\Sigma}_0$, implied by these estimates and equations (4) are

$$\hat{\Sigma}_0 = \begin{bmatrix} 0.648 & 0.125 \\ 0.125 & 0.762 \end{bmatrix}, \quad \hat{\Sigma}_1 = \begin{bmatrix} 0.243 & 0.193 \\ -0.059 & 0.276 \end{bmatrix}.$$

Or, in terms of lag-1 cross-correlation,

$$\hat{\Sigma}_1^* = \begin{bmatrix} 0.374 & 0.275 \\ -0.085 & 0.362 \end{bmatrix}.$$

This implies persistence in day-to-day area potentials, and also movement as the value on the Day t at the Western node (θ_1) is positively correlated with both Western and Eastern nodes on Day $t+1$ (0.374 and 0.275, respectively); but the value on the Day t at the Eastern node (θ_2) is only positively correlated with the Eastern node on the next Day.

We could have also used grids with more nodes, but for them, the coefficients are far less interpretable. Also, this approach does not address the idiosyncratic behavior at the sites.

b. Multisite model

We fit the Multisite model for several months in each study area. We assessed the fit primarily using three statistics: the average daily simulated precipitation at the sites, covariances between daily (square-root transformed) precipitation at the pairs of sites, and lag-1 cross-covariances. Note that we do not aim for the exact replication of the site statistics, like means etc. This would mean overfitting our model, which in turn might produce simulated monthly total precipitation with unrealistically small variability (Katz and Parlange, 1998).

Examples of the multisite model simulated means, compared to real results, are shown in Fig.8. To give an idea about variability, we added horizontal intervals showing middle 90% of the sample of 100 50-year simulated means. They are in good agreement with the observed means.

The covariance and lag-1 cross-covariance patterns are reproduced fairly well. However, synthetic covariances tend to have higher values, as seen in Fig.9a. When we consider the covariances for the occurrence process only (that is, 0 if no precipitation, and 1 if there is precipitation), the agreement becomes much better (Fig.9b). This might serve as an indication of a certain discrepancy between the occurrence and intensity processes.

Compared to the Convolution model, the residual standard deviation τ is greatly reduced, for example from 0.456 to 0.202 for July, MN, and from 0.238 to 0.067 for January, CO! This shows that only a little portion of between-sites spatial-temporal dependence is explained

by the systematic movement described in previous Section.

Wet spells (durations of consecutive days for which precipitation was observed) were represented adequately by this model. However, the *dry spell* distribution have differed from empirical data. One possible way of dealing with this is to introduce dry spell durations directly into the model. We have implemented this in a form of a region-wide dry spell. Namely, whenever a dry spell starts (that is, all the sites in the area have 0 precipitation), we can model its length by using either a parametric or the empirical distribution of dry spell lengths. After ending the dry spell, we will resume the normal regime given by eq.(3), which will have been re-fit to the wet-spell data fragments. Examples of this correction are given in Fig. 10.

Another option is to introduce two or more weather regimes (for example, dry and wet regimes), each with its own multisite model. This approach is reminiscent of Stehlík and Bárdossy (2002) and Apipattanavis et al (2007). It awaits future testing.

5. Summary

We proposed a new precipitation generator based on truncated and power-transfromed normal distribution, with the spatial-temporal dependence represented by multivariate autoregression.

This generator integrates the occurrence and amounts processes, thus poised to overcome the spatial intermittence problem. It also gives realisic spatial behavior, not only for covariances between sites, but also for lag-1 and cross-covariances. The differences between occurrence

and amounts processes do not appear to be systematic and can be minimized through the introduction of additional dry spell periods.

The generator is conceptually simple. The parameter values should be estimated using computationally expensive Gibbs sampler; however, once the values are fitted, the precipitation sequences can be generated straight from equation (3) at very little cost. This could be an advantage over non-parametric approaches that require to keep the entire historical precipitation record. Also, non-parametric approaches are harder to implement when the precipitation record is short, the station locations are changing, or data are missing.

Another advantage of a linear model is that exogenous variables can be incorporated easily. These might include time trends, as well as global and regional climatic information, elevation etc.

Several extensions of the described algorithm appear promising. One, introduction of two or more weather regimes on which the multivarite model would be conditioned. This could be useful for downscaling General and Regional Circulation Models (Wilby et al, 1999). Two, it could be useful to adapt our algorithm (possibly combining the Convolution and Multisite approaches) for high-resolution simulation.

Also, further testing and application of this algorithm (for example, in reservoir studies) is desirable.

APPENDIX

Details of MCMC model fitting

In this section, we describe details of the model fitting via the Gibbs sampler, full conditional posteriors (FCP's) etc.

The FCP is defined as a conditional distribution of one parameter (or a block of parameters) given all the other parameters and variables in the model. The unknown parameters and variables are given some starting values. Gibbs sampler is then implemented in iterating draws from FCP's of one parameter at a time, then cycling through the rest of parameters, and repeat the procedure. This usually results in a correlated sample from the joint distribution of the parameters of interest, called *Markov Chain Monte-Carlo (MCMC) sample*. After an initial period of convergence, we can use this sample (thinned out, if necessary, to avoid autocorrelation) for the inference about the parameters. We can use, for example, sample means as point estimates, or medians in case of skewed distributions. We can also produce Bayesian credible intervals by taking quantiles of the MCMC samples. The estimates are subject to Monte Carlo error, but we can make this error small by obtaining large enough samples.

For example, the FCP for the transformation parameter β, $f(\beta \,|\, \{R_{it}\}, \{Z_{it}\}, \tau)$ is propor-

tional to

$$\prod_{R_{it}>0} \frac{1}{\beta} R_{it}^{\frac{1}{\beta}-1} \exp\left[-\frac{(R_{it}^{\frac{1}{\beta}} - Z_{it})^2}{2\tau^2}\right]$$

The details of fitting the area potentials $\boldsymbol{\theta}_t$ in the convolution model are discussed in Higdon (2007). Most of the other FCP's we used are well-known in statistical practice and are discussed in Gelman et al (2003).

A program in `R` statistical language (R Development Core Team, 2008) was written to implement the Gibbs sampler.

REFERENCES

[1] Apipattanavis, S., G. Podestá, B. Rajagopalan, and R. W. Katz, 2007. A semiparametric multivariate and multisite weather generator, Water Resour. Res., 43, W11401

[2] Bárdossy, A. and Plate, E.J., 1992. Space-time model of daily rainfall using atmospheric circulation patterns. Water Resour. Res. 285, pp. 1247–1259.

[3] Brissette, F.P., M. Khalili and R. Leconte, 2007. Efficient stochastic generation of multi-site synthetic precipitation data, J. Hydrol., 345, 121–133

[4] Eischeid J. K., P. A. Pasteris, H. F. Diaz, M. S. Plantico, and N. J. Lott, 2000. Creating a serially complete, national daily time series of temperature and precipitation for the western United States. J. Appl. Meteor., 39, 1580-1591.

[5] Furrer, E. and Katz, R.W., 2007. Generalized linear modeling approach to stochastic weather generators. Clim. Res., 34, 129-144

[6] Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B., 2003. Bayesian Data Analysis, Second Edition. New York: Chapman and Hall.

[7] Higdon, D., 2007. A primer on space-time modeling from a Baeysian perspective, in Statistical Methods for Spatio-Temporal Systems, Finkenstadt B., Held L. and Isham, V., eds., Chapman & Hall, London.

[8] Koutsoyiannis D., 2006, An entropic-stochastic representation of rainfall intermittency: The origin of clustering and persistence, Water Resour. Res., 42, W01401.

[9] Lee, H.H.K., Higdon, D., Calder, K. and Holloman, C., 2005. Efficient models for correlated data via convolutions of intrinsic processes. Statistical Modelling 5, 53–74

[10] Mehrotra, R. and Sharma, A., 2007. A semi-parametric model for stochastic generation of multi-site daily rainfall exhibiting low-frequency variability. J. Hydrology 335, 180–193.

[11] R Development Core Team, 2007. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

[12] Sansó, B., and Guenni, L. 2000. A Nonstationary Multisite Model for Rainfall. J. Amer. Stat. Assoc., 95, 1089–1100

[13] Semenov, M.A., Porter, J.R., 1995. Climatic variability and the modelling of crop yields. Agric. Forest Meteorol. 73, 265–283.

[14] Stehlík and Bárdossy, 2002. Multivariate stochastic downscaling model for generating daily precipitation series based on atmospheric circulation, J. Hydrol. 256 (2002), pp. 120–141.

[15] Stidd, C.K., 1973. Estimating the precipitation climate. Water Resour. Res. 9, 1235-1241.

[16] Todorovic, P., Woolhiser, D.A., 1975. A stochastic model of n-day precipitation. J. Appl. Meteor. 14, 17–24.

[17] Wilby R. L., T. M. L. Wigley, D. Conway, P. D. Jones, B. C. Hewitson, J. Main, and D. S. Wilks, 1998. Statistical downscaling of general circulation model output: A comparison of methods. Water Resour. Res., 34, 2995—3008.

[18] Wilks D. S., 1998. Multisite generalization of a daily stochastic precipitation generation model. J. Hydrol., 210, 178–191.

[19] Wilks D. S., 1999. Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. Agric. For. Meteor., 93, 153–169.

[20] Wilks D. S., 2008. High-resolution spatial interpolation of weather generator parameters using local weighted regressions, Agric. For. Meteor., 148, 111–120

[21] Yang, C., Chandler, R.E., Isham, V. and Wheater, H.S., 2005. Spatial-temporal rainfall simulation using Generalized Linear Models. Water Resour. Res., 41, W11415.

List of Figures

1 Map of the study areas with observation sites shown 22

2 MN study sites with 1x2 grid 23

3 Correlation inversion for the MN study area 24

4 No correlation inversion for the CO study area 25

5 Idiosyncratic lag-1 cross-correlations for NY area, months May- October. Gray arrows: lag-1 cross-correlations above 0.45, black arrows: lag-1 cross-correlations above 0.55 26

6 Empirical CDF (thick gray) and TPT normal fit (thin black lines) for sites 1-12, CO study area, August 27

7 Empirical CDF (thick gray) and TPT normal fit (thin black lines) for total precipitation at all sites, MN study area, July-December 28

8 Comparison of simulated and real means, (a) January, MN (b) January, CO (c) July, MN (with dry spell correction) (d) July, CO 29

9 Comparison of simulated and real covariances and lag-1 cross-covariances, January, CO. (a) sqrt-transformed precipitation (b) occurrence only 30

10 Comparison of simulated and real covariances and lag-1 cross-covariances, with dry-spell correction. (a) July, CO (b) July, NY 31

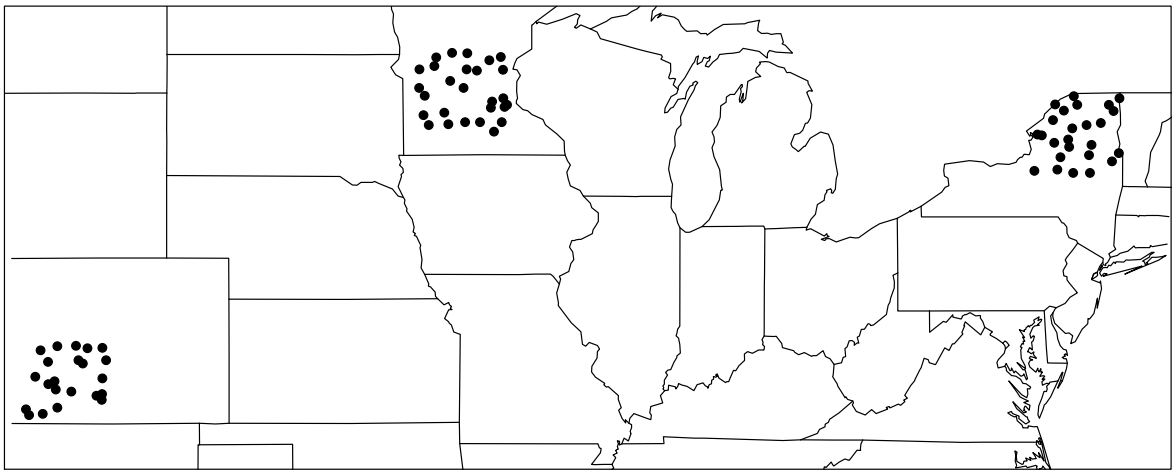

FIG. 1. Map of the study areas with observation sites shown

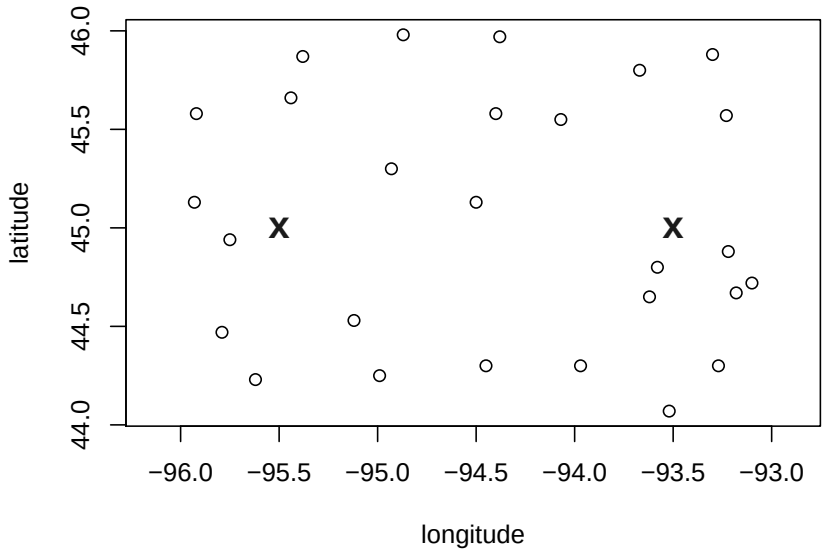

FIG. 2. MN study sites with 1x2 grid

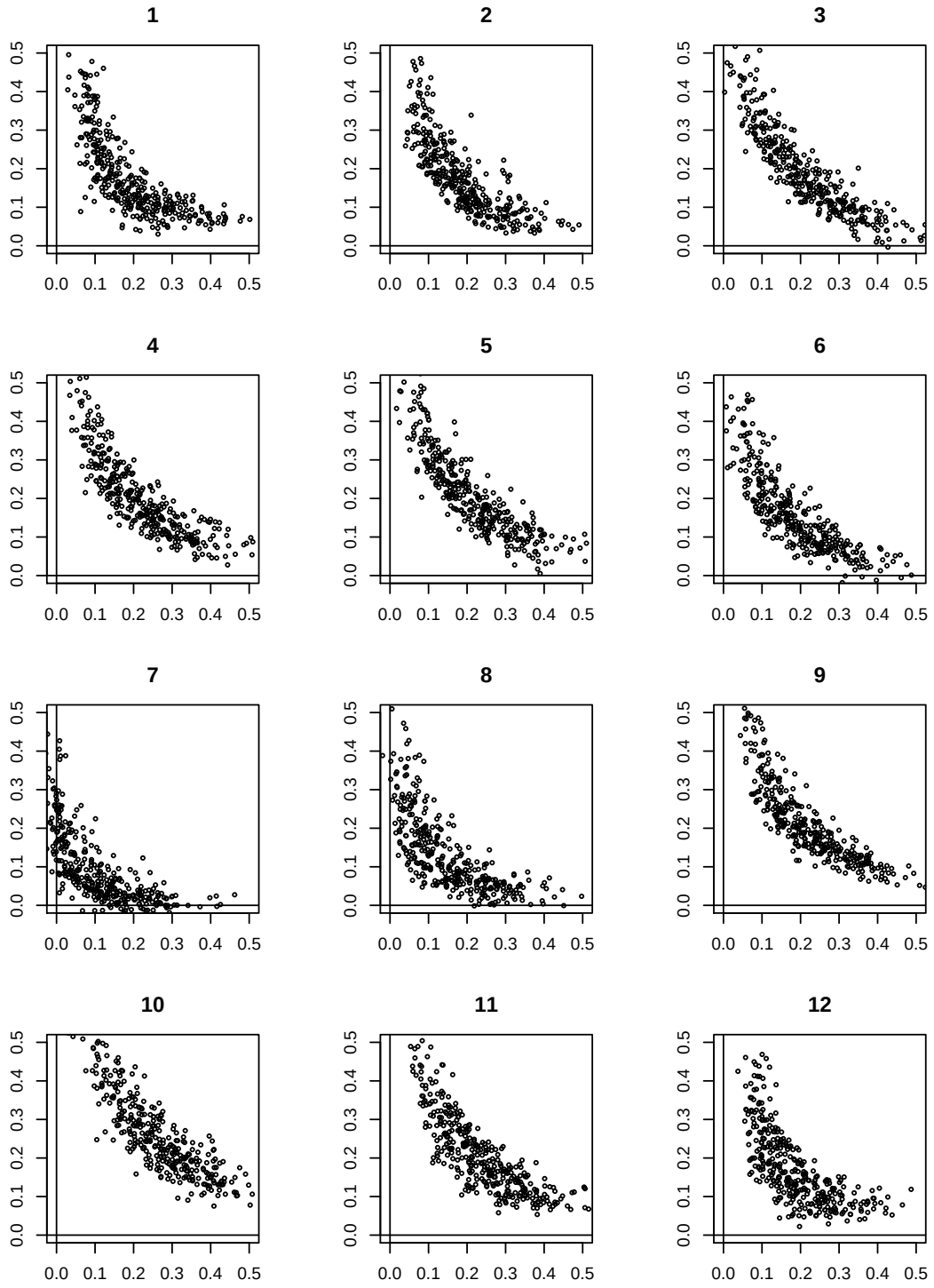

FIG. 3. Correlation inversion for the MN study area

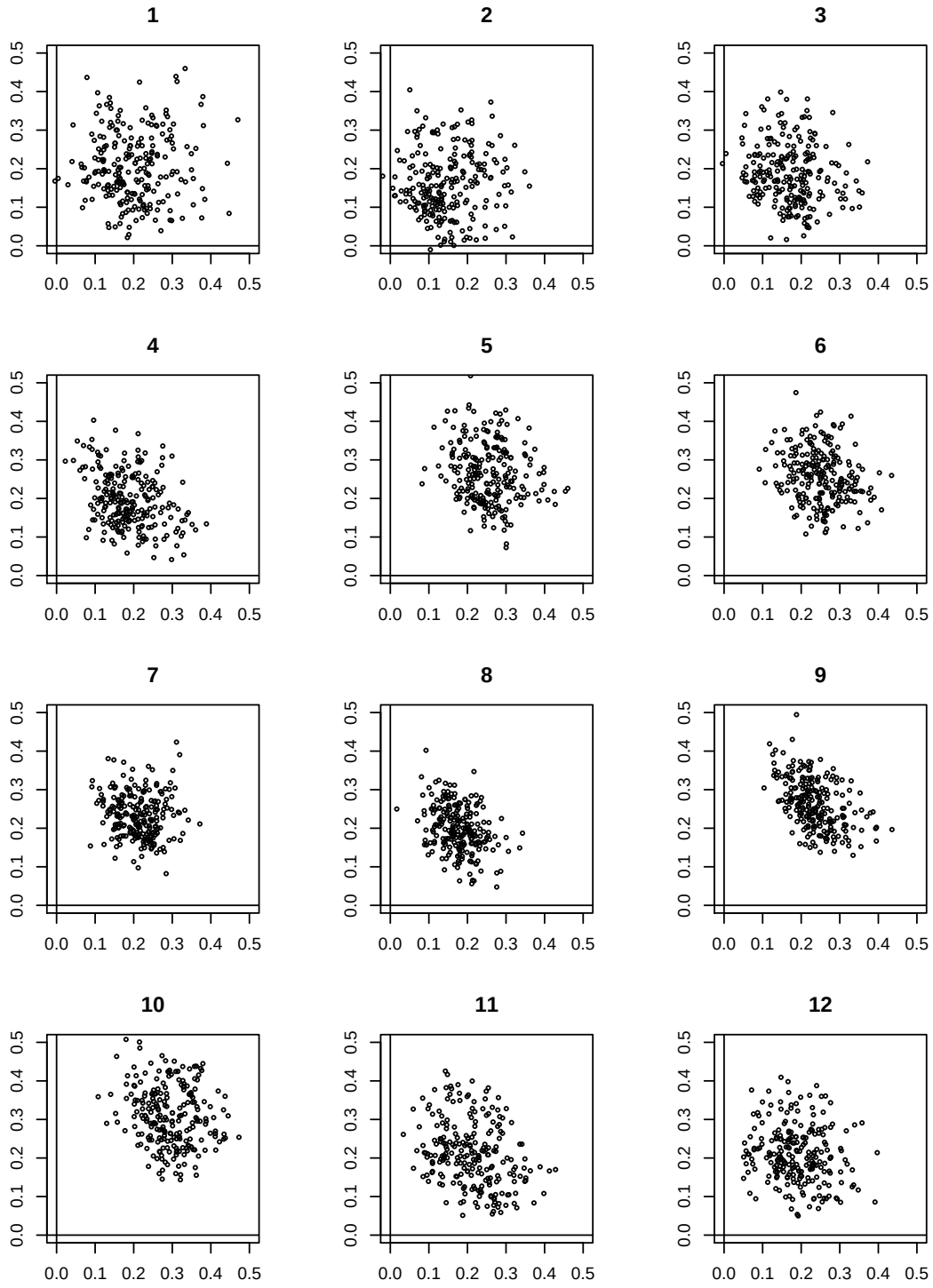

FIG. 4. No correlation inversion for the CO study area

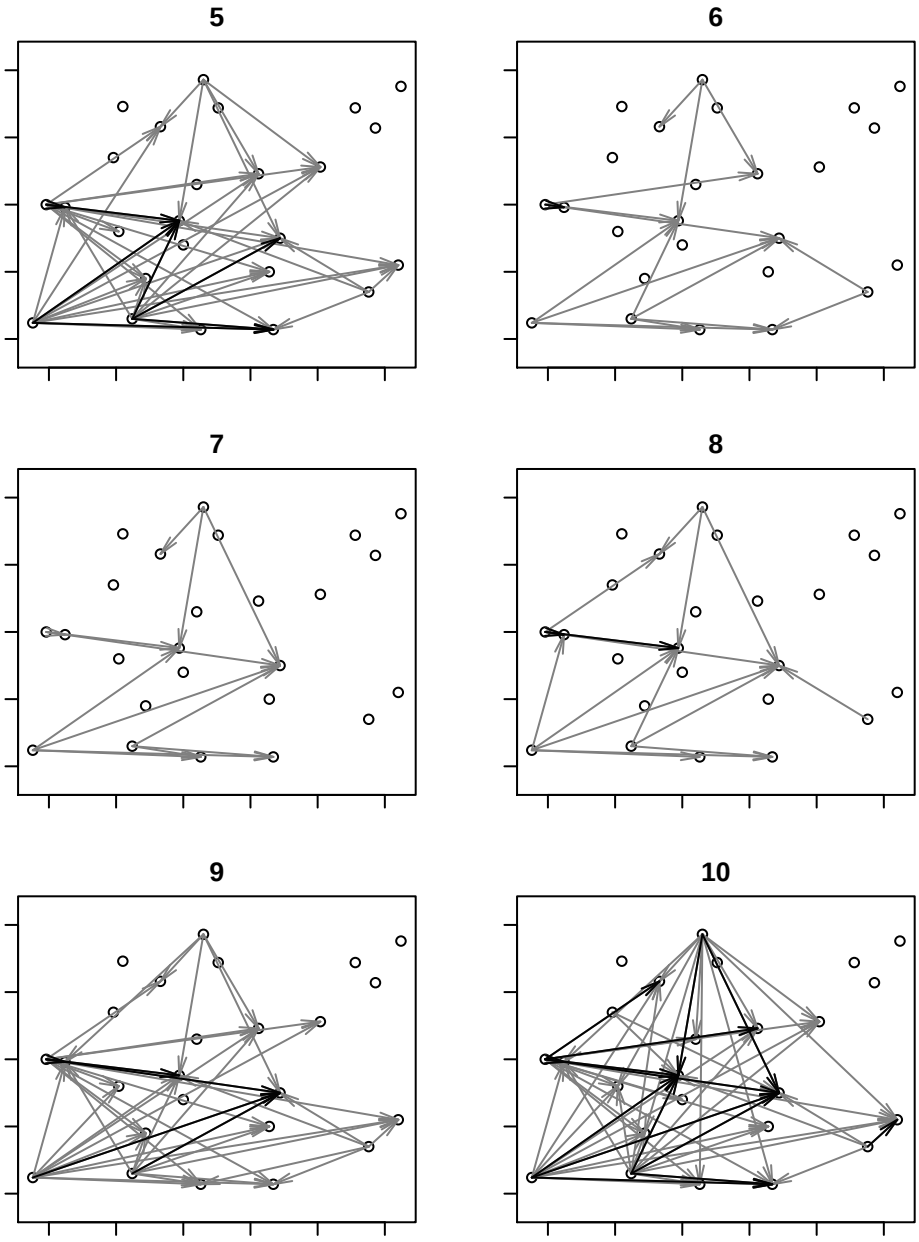

FIG. 5. Idiosyncratic lag-1 cross-correlations for NY area, months May- October. Gray arrows: lag-1 cross-correlations above 0.45, black arrows: lag-1 cross-correlations above 0.55

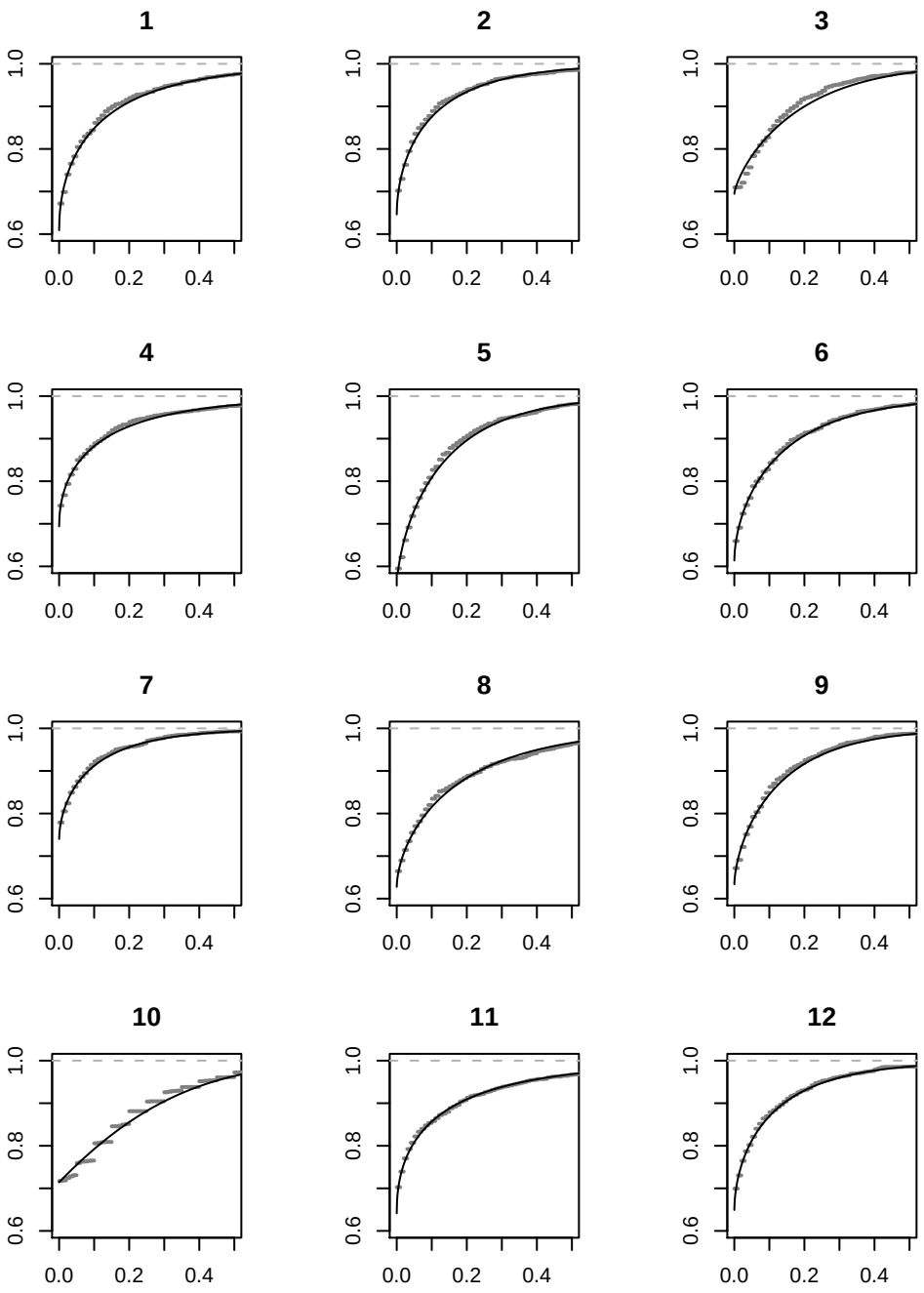

FIG. 6. Empirical CDF (thick gray) and TPT normal fit (thin black lines) for sites 1-12, CO study area, August

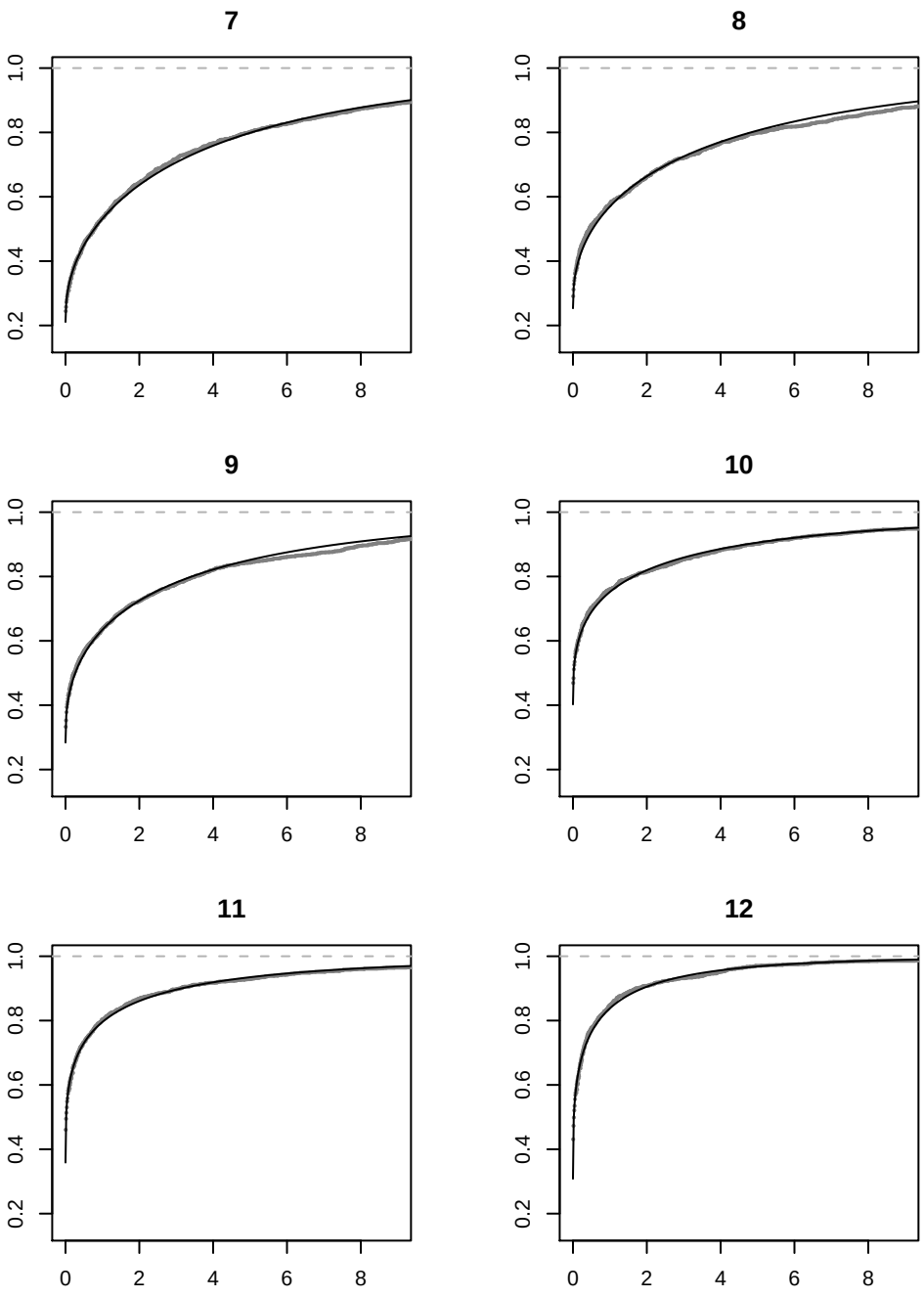

FIG. 7. Empirical CDF (thick gray) and TPT normal fit (thin black lines) for total precipitation at all sites, MN study area, July-December

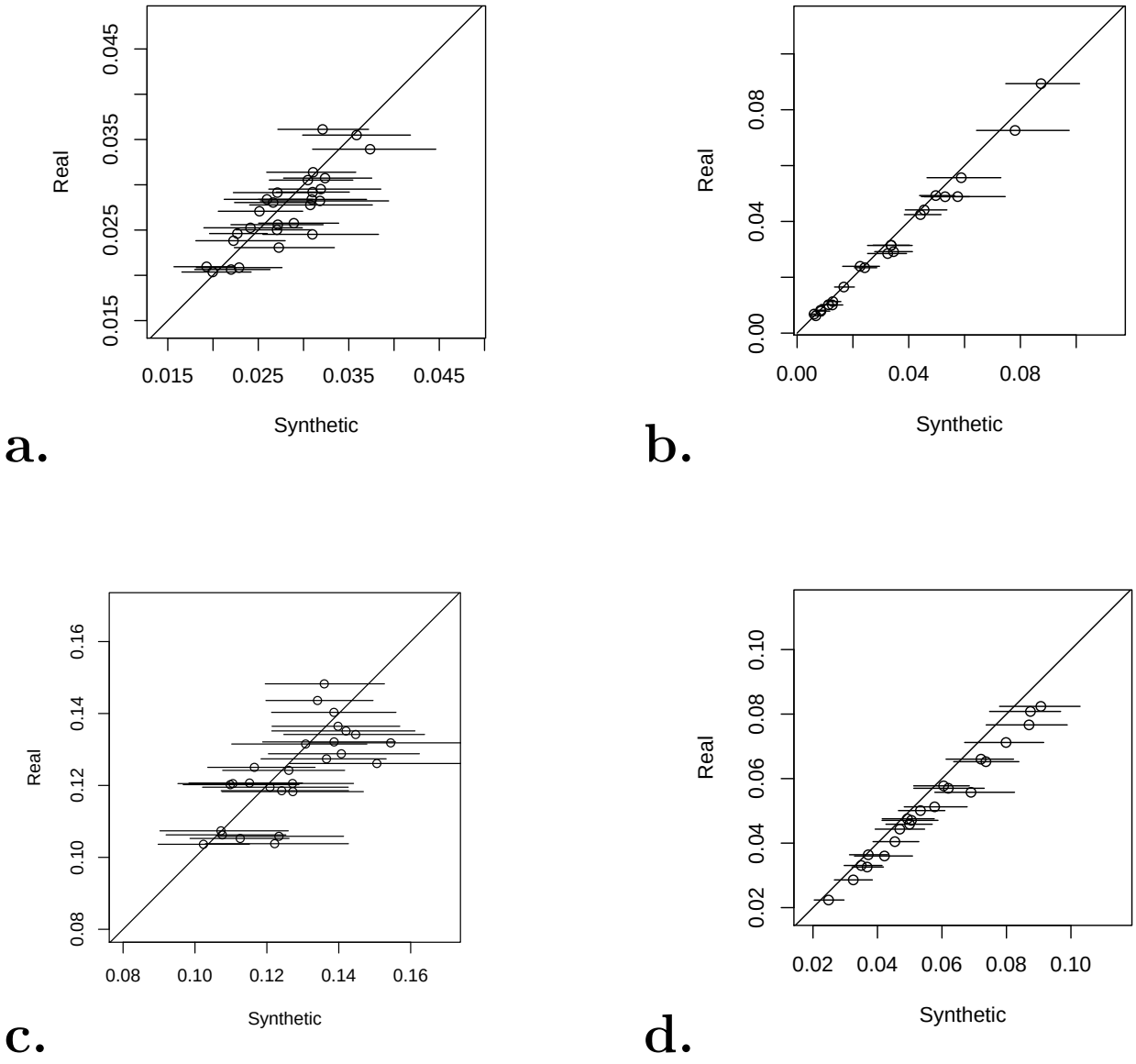

FIG. 8. Comparison of simulated and real means, (a) January, MN (b) January, CO (c) July, MN (with dry spell correction) (d) July, CO

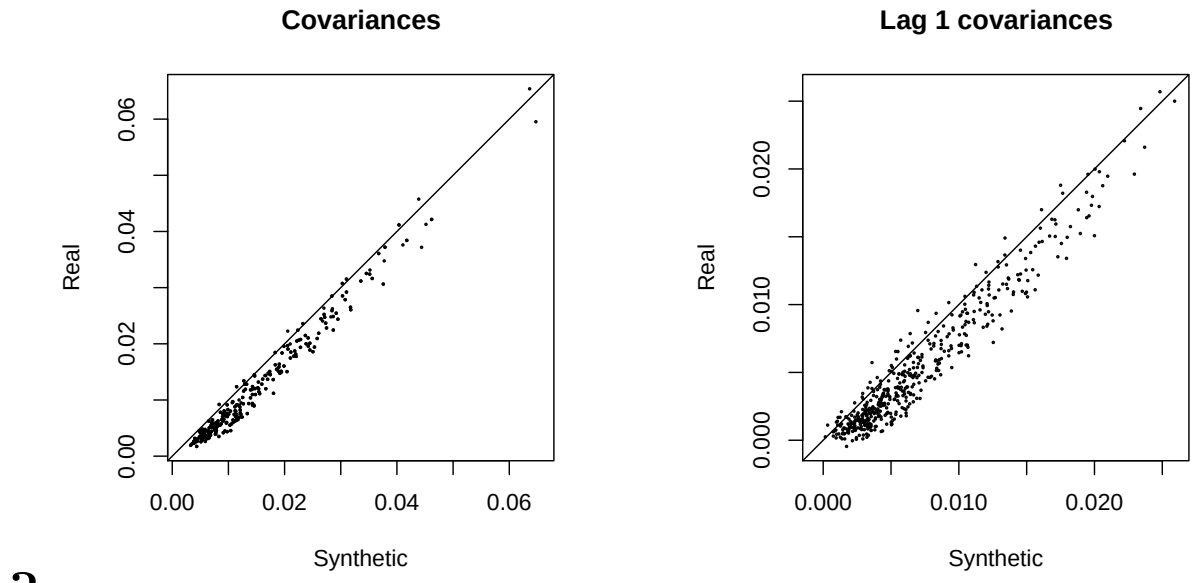

FIG. 9. Comparison of simulated and real covariances and lag-1 cross-covariances, January, CO. (a) sqrt-transformed precipitation (b) occurrence only

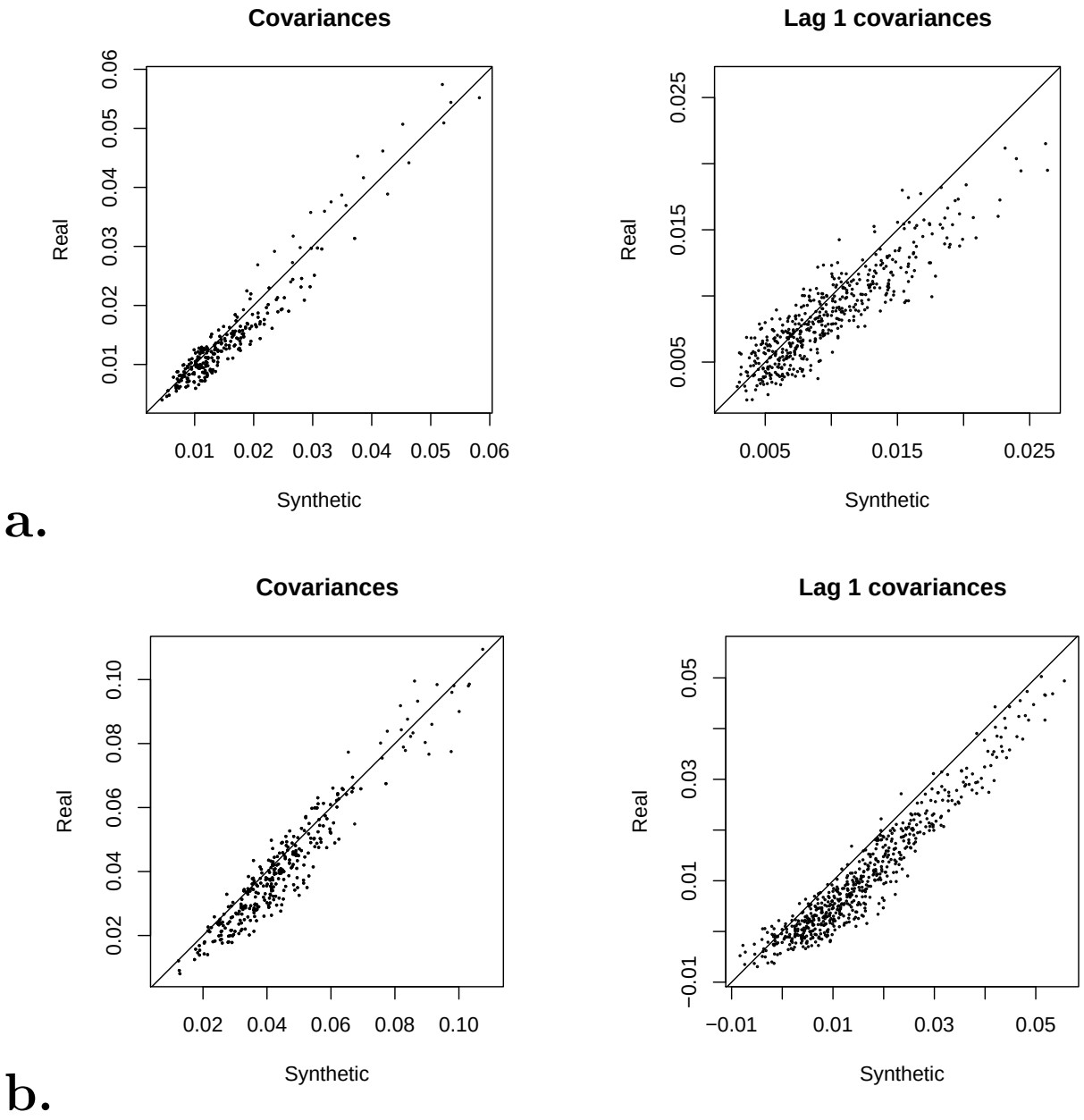

FIG. 10. Comparison of simulated and real covariances and lag-1 cross-covariances, with dry-spell correction. (a) July, CO (b) July, NY

List of Tables

1 Fitted TPT normal distribution values 33

2 The parameter estimates for the Convolution model, MN study area 34

TABLE 1. Fitted TPT normal distribution values

a) Daily precipitation for sites 1-6, August, CO

Site	μ	σ	β	precipitation probability
1	–0.115	0.418	2.008	0.392
2	–0.139	0.37	1.856	0.353
3	–0.199	0.393	1.349	0.306
4	–0.23	0.463	2.013	0.31
5	–0.039	0.337	1.68	0.454
6	–0.11	0.382	1.736	0.386

b) Total daily precipitation for all sites, months 1–12, MN

Month	μ	σ	β	precipitation probability
1	0.336	0.696	3.254	0.685
2	0.267	0.712	3.411	0.646
3	0.405	0.835	3.425	0.686
4	0.595	1.037	2.843	0.717
5	0.791	1.148	2.564	0.755
6	1.027	1.121	2.67	0.82
7	0.91	1.133	2.585	0.789
8	0.79	1.191	2.697	0.746
9	0.623	1.098	2.824	0.715
10	0.261	1.092	3.084	0.595
11	0.303	0.845	3.478	0.64
12	0.339	0.67	3.491	0.693

TABLE 2. The parameter estimates for the Convolution model, MN study area

Month	1	2	3	4	5	6
τ	0.198	0.208	0.252	0.287	0.334	0.403
$\bar{S}$	–0.203	–0.255	–0.353	–0.266	–0.256	–0.191
β	1.844	1.841	1.815	1.812	1.882	2.003
$\mathbf{r}_\theta$	0.377 0.093	0.483 –0.046	0.530 –0.031	0.519 –0.061	0.482 –0.102	0.386 –0.044
	0.191 0.331	0.157 0.373	0.121 0.360	0.207 0.344	0.214 0.392	0.200 0.328
$\mathbf{V}_\xi$	0.113 0.046	0.177 0.022	0.224 0.073	0.343 0.034	0.381 0.000	0.488 0.038
	0.046 0.139	0.022 0.181	0.073 0.279	0.034 0.318	0.000 0.436	0.038 0.425
Month	7	8	9	10	11	12
τ	0.456	0.442	0.351	0.308	0.254	0.207
$\bar{S}$	–0.315	–0.378	–0.315	–0.435	–0.336	–0.245
β	1.986	2.022	2.091	1.95	1.9	1.918
$\mathbf{r}_\theta$	0.348 –0.09	0.402 –0.144	0.513 –0.086	0.621 –0.056	0.459 0.095	0.493 0.055
	0.145 0.273	0.236 0.323	0.179 0.37	0.147 0.426	0.093 0.447	0.022 0.445
$\mathbf{V}_\xi$	0.48 –0.005	0.542 0.087	0.526 0.050	0.460 0.061	0.197 0.071	0.115 0.039
	–0.005 0.711	0.087 0.627	0.050 0.346	0.061 0.326	0.071 0.319	0.039 0.193