A Multi-Site Stochastic Weather Generator for High-Frequency Precipitation Using Censored Skew-Symmetric Distribution

Yuxiao Li and Ying Sun*

Statistics Program, King Abdullah University of Science and Technology

September 22, 2020

Abstract

Stochastic weather generators (SWGs) are digital twins of complex weather processes and widely used in agriculture and urban design. Due to improved measuring instruments, an accurate SWG for high-frequency precipitation is now possible. However, high-frequency precipitation data are more zero-inflated, skewed, and heavy-tailed than common (hourly or daily) precipitation data. Therefore, classical methods that either model precipitation occurrence independently of their intensity or assume that the precipitation follows a censored meta-Gaussian process may not be appropriate. In this work, we propose a novel multi-site precipitation generator that drives both occurrence and intensity by a censored non-Gaussian vector autoregression model with skew-symmetric dynamics. The proposed SWG is advantageous in modeling skewed and heavy-tailed data with direct physical and statistical interpretations. We apply the proposed model to 30-second precipitation based on the data obtained from a dense gauge network in Lausanne, Switzerland. In addition to reproducing the high-frequency precipitation, the model can provide accurate predictions as the long short-term memory (LSTM) network but with uncertainties and more interpretable results.

Keywords: Censored data, digital twins, non-Gaussian processes, skewed distribution, spatio-temporal model

*This research was supported by King Abdullah University of Science and Technology (KAUST), Office of Sponsored Research (OSR) under Award No: OSR-2019-CRG7-3800.

1 Introduction

Tremendous efforts have been made to model, forecast, and reproduce local and global precipitation patterns. Among these efforts, the stochastic weather generator (SWG) makes use of statistical tools to simulate random sequences and reproduce atmospherical variables efficiently (Wilks and Wilby, 1999). Typically, an ideal stochastic precipitation generator (SPG) should be able to reproduce the statistical properties of occurrence, intensity, and dry spell length of precipitation. The idea of reproducing physical processes with a virtual replica applied in other areas, such as “digital twin” (Boschert and Rosen, 2016; Söderberg et al., 2017; El Saddik, 2018). A Digital twin integrates many modern techniques such as internet of things and artificial intelligence. It is also the critical sector of smart city (Falconer and Mitchell, 2012) and industry 4.0 (Gilchrist, 2016).

SPGs have proven to be important in water resources research. As an example of digital twins, SPGs provide valuable information for water resource management for a smart city (Parra et al., 2015). In hydrologic and agricultural science, SPGs can serve as an input in further simulations of erosion, flood and crop growth (Mary et al., 2009). Moreover, as the primary atmospheric variable in SWGs, precipitation is typically used to generate other variables, such as wind and solar irradiance, due to their close association with rainfall occurrence (Richardson, 1981). Techniques in modeling precipitation data can be also beneficial to other fields of study, such as sociology (Heckman, 1976) and economics (McDonald and Moffitt, 1980), where the data properties are often similar to those of precipitation, e.g., they are zero-inflated, nonnegative, right-skewed, heavy-tailed, and correlated in space and time.

Due to their wide applicability and the intriguing challenges, studies of SPGs have drawn attentions since 1960s (Gabriel and Neumann, 1962) and been systematic reviewed in Wilks and

Wilby (1999), Srikanthan and McMahon (2001), and Ailliot et al. (2015). Traditional studies mainly focus on the SPGs with a low temporal resolution, usually on a daily scale, due to the data available. For instance, the most popular chain-dependent model (Katz, 1977; Richardson, 1981) assumes that the occurrence processes can be modeled independently of intensity by Markov chains, and the intensity processes can be estimated conditionally on wet events using Gamma or exponential distributions. However, their assumption of independent occurrence is not appropriate for high-frequency precipitation, since greater quantities of rainfall in the past may lead to a significantly higher probability of occurrence (Koch and Naveau, 2015).

Recently, acoustic rain gauges are able to provide precise and high-frequency data which are undetectable by the classical measurement devices, such as tipping-bucket rain gauge, satellite-based radar, and terrestrial radar. The high-frequency dataset is valuable to many rainfall-related phenomena, such as rapid surface water runoff, flash flooding, and small river catchments (Chan et al., 2016). However, the rainfall model at this scale has been rarely developed and assessed. To our knowledge, only Benoit et al. (2018) has performed an investigation of the minutely and sub-kilometer SPGs with the same device; in their study, the spatial dependence was the major concern. Even though broad literatures exist on the development of hourly and daily SPGs, we cannot assume that their results can extend to our timescales. In this study, our objective is to use this single model to reproduce precipitation on a very fine scale both spatially (within one radar pixel, 10-100 meters) and temporally (less than a minute). The data of our interest contain 30-second rainfall that were collected on eight acoustic rain gauges located within a radius of $1km$ in the University of Lausanne campus in Switzerland (see Figure 1).

To handle the high-frequency data, the common practice is to make use of censored models to drive both the occurrence and the intensity processes. The models are also called truncated models by Allard (2012), and the Tobit models (McDonald and Moffitt, 1980) from econometrics.

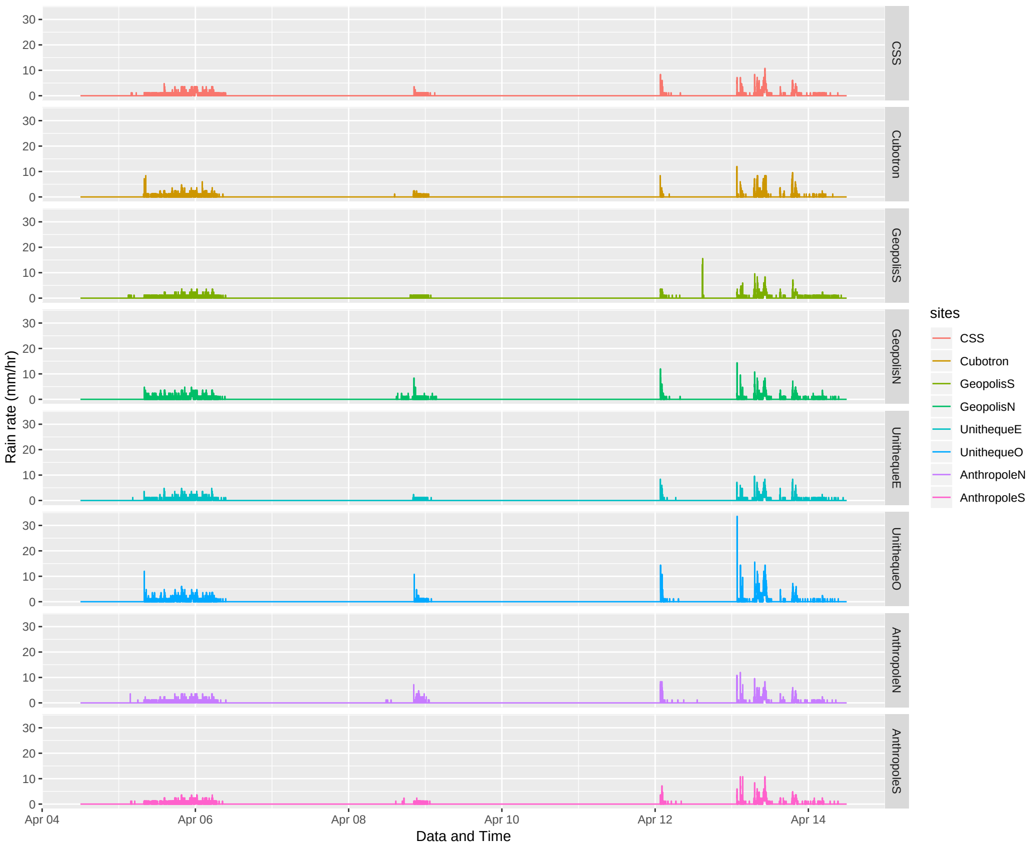

Figure 1: The eight time series of rain rate (unit: mm/hr) collected by Dryptich Pluvimate acoustic rain gauges every 30 seconds from April 4th to April 14th, 2016.

In the censored models, both the occurrence and intensity are driven by uncensored processes, where the dry events are zero values left-censored at a certain threshold, and the wet events are modeled by the process with a positive support.

Meta-Gaussian processes are the most popular way for the spatio-temporal dependence and non-Gaussian features of rainfall data (Ailliot et al., 2009; Kleiber et al., 2012; Baxevani and Lennartsson, 2015). Although the methodologies provide high flexibility with regard to the spatio-temporal dependence, several limitations remain. First, the transformation of Gaussian processes is ad-hoc, ranging from simple power or logarithm transformations (Bell, 1987; Glasbey and Nevison, 1997; Durbán and Glasbey, 2001) to complex Tukey g-and-h or hybrid Gamma transformations (Baxevani and Lennartsson, 2015; Xu and Genton, 2017). The efforts to achieve normality can be tedious. Second, the latent Gaussian processes make the interpretation difficult, especially when different transformations are suggested (Ailliot et al., 2015). Third, Gaussian processes are more suitable for denser spatial data rather than spatially sparse data, such as the multi-site high frequency time series.

Another popular way to quantify the spatio-temporal dependence in SPG is the vector autoregressive (VAR) processes (Hamilton, 1994; Sigrist et al., 2012; Rasmussen, 2013; Koch and Naveau, 2015). The VAR framework has been widely used in a set of high-frequency data such as stock returns and brain signals. For precipitation, the state-of-arts censored VAR model is proposed by Koch and Naveau (2015), for which the interpretation is easy and its likelihood-based inference is straightforward. However, this model only considers heteroscedastic Gaussian and meta-Gaussian error, which may not be adequate for data with skewed or heavy tailed distributions.

Therefore, we propose to use the skew-symmetric families (Azzalini, 1985) in the censored VAR model. The skew-symmetric distributions have been generalized and applied in many fields of study (Azzalini and Capitanio, 1999; Genton, 2004; Azzalini, 2013) in modeling the obviously non-Gaussian features, but rarely used in the SWG. The only application to SWG was investigated by Flecher et al. (2010), where the multivariate skew-normal distribution was adopted to model multiple atmospheric variables.

As an example, the density function of a univariate skew-t random variable is specified as

$$f_{\mathcal{ST}}(x;\xi,\omega,\alpha,\nu) = 2t(u;\nu)T(\alpha u\sqrt{(\nu+1)(\nu+u^2)};\nu+1), u = \frac{x-\xi}{\omega}, \tag{1}$$

where ξ and ω are the location and scale parameters, α is the skewness parameter, ν, is the degree of freedom, $t(\cdot)$ is the standard t density function, and $T(\cdot)$ is the standard t distribution function.

The skew-normal distribution is a special case of the skew-t distribution when $\nu = \infty$. Similar to the normal or student-t distributions, x can take any real value. In contrast, two extra parameters in the skew-t distribution control the skewness and tail behavior. Another attractive feature of the skew-t distribution is the stochastic representation (Azzalini and Regoli, 2012). A skew-t random variable X can be expressed by hidden selective mechanism such that $X = (X_1|X_2 > 0)$, where X_1, X_2 are correlated t random variables with the same degree of freedom.

Since the generation mechanism behind the precipitation can be viewed as a hidden selection process, using a skew-symmetric distribution in modeling precipitation data produces nice physical interpretations, as will be further explained in Section 2.3. By incorporating skew-symmetric distributions and a censored VAR framework, we propose a new stochastic precipitation generator that

1) uses a single spatio-temporal model to simultaneously drive the precipitation occurrence and its intensity;

2) has direct interpretation to the precipitation process;

3) allows for flexible and tractable tail behaviors, which is crucial in modeling high-frequency precipitation data;

4) implies parsimonious parametrization and efficient data generation.

For the application to high-frequency precipitation data, recent studies also propose to use deep learning methods such as long short-term memory (LSTM) or other recurrent neural networks (RNN) (Shi et al., 2015, 2017). Although SPG focuses on reproducing the precipitation in the functional perspective, it is also important to show the results of prediction, especially comparing to these deep learning methods. We choose the multivariate LSTM networks as our competing methods to show the difference between our stochastic generators and the deep learning forecasters.

The rest of our paper is organized as follows. In Section 2, we introduce the new class of SPGs, provide model properties, and describe the inference procedure. In Section 3, we present simulation studies to validate our inference method and compare its performance to other models. In Section 4, we show the performance of the proposed SPGs on the high-frequency rainfall dataset collected at the University of Lausanne campus. In Section 5, we summarize our main results and discuss the potential limitations.

2 Censored skew-symmetric VAR models

2.1 Censored VAR models

Let $Y_t(s)$ be a random variable describing the precipitation rate at a site s and time t, $s = 1, \dots, N, t = 1, 2, 3, \dots, T$. Collect $Y_t(s)$ as an $N \times 1$ random vector $\mathbf{Y}_t$. Then we specify the multi-site precipitation generator based on censored VAR as

$$Y_t(s) = \begin{cases} \boldsymbol{\beta}_s' \mathbf{Y}_{t-1} + \varepsilon_t(s), & \boldsymbol{\beta}_s' \mathbf{Y}_{t-1} + \varepsilon_t(s) > u_t(s), \\ 0, & \boldsymbol{\beta}_s' \mathbf{Y}_{t-1} + \varepsilon_t(s) \leq u_t(s), \end{cases} \tag{2}$$

where $\boldsymbol{\beta}_s$ is an $N \times 1$ vector of autoregression coefficients for site s, $u_t(s)$ is the cutoff vector representing the censoring threshold associated with the rain probability, and $\varepsilon_t(s)$ is the random error.

The censored VAR generator in (2) is originally proposed by Koch and Naveau (2015). The model (2) differs from other SPGs since it avoids applying any ad-hoc transformation to the rainfall intensity when $Y_t(s) > 0$. Therefore, the likelihood of $Y_t(s)$ has an explicit form and parameters in the model have direct interpretation to $Y_t(s)$.

The model (2) also differs from classical VAR model (Sims, 1980; Hamilton, 1994), where $Y_t(s)$ is not censored and the error term $\varepsilon_t(s)$ is assumed to be white noise with zero mean and constant variance. Here, a more general case by spliting $\varepsilon_t(s)$ as $\varepsilon_t(s) = \sigma_t Z_t(s)$, where σ_t is stochastic and represents the conditional standard deviation of positive $Y_t(s)$ given that $\mathbf{Y}_{t-1} = \mathbf{y}_{t-1}$ and $Z_t(s)$ are independent random variables both in space and time.

2.2 Censored VAR models with skew-symmetric errors

The independent random variables $Z_t(s)$ in Koch and Naveau (2015) is assumed to follow a standard normal distribution and the standard deviations are modeled with other atmospheric explanatory variables by a linear regression. However, these explanatory variables are often either unavailable or hard to choose in practice. This issue becomes more problematic when they assume that $Z_t(s)$ is normally distributed because the right-skewed and heavy-tailed features of the rainfall data can be only explained by σ_t. Then the entire dynamics of the SPG will be heavily influenced by the selected explanatory variables.

Therefore, we consider another random error that is flexible with the skewness and tail behavior so that the explanatory variables are not required. Instead of a normal distribution, we assume that $Z_t(s)$ follows a family of skew-symmetric distributions (Azzalini, 2013) with zero mean and unit variance. As we mentioned in (1), $Z_t(s)$ itself already describes the skewed and heavy-tailed features, rather than relying on the other explanatory variables.

For the conditional standard deviations σ_t, we allow for heteroscedasticity with a temporally varying standard deviation. The idea of heteroscedasticity is widely used to predict high-frequency data such as wind power and stock-returns (Engle, 1982; Bollerslev, 1986; Taylor et al., 2009), such as the (generalized) autoregressive conditionally heteroscedastic (ARCH/GARCH) models. Specifically, when $Y_{t-1}(s) = y_{t-1}(s)$ is observed, we assume that $\sigma_t = b_0 + b_1 \bar{y}_{t-1}$, where

$\bar{y}_{t-1} = \sum_{s=1}^{N} y_{t-1}(s)/N$ and $b_0, b_1 \geq 0$ to avoid a negative variance.

To describe the distribution of $Z_t(s)$, two candidates in the skew-symmetric family are the most commonly used: the skew-normal distribution (Azzalini, 1985) and the skew-t distribution (Branco and Dey, 2001; Azzalini and Capitanio, 2003). Since the skew-t distribution with the degree of freedom ν includes the skew-normal distribution as special cases, hereafter we mainly discuss the properties of the skew-t distribution. The results from the skew-normal can be obtained by replacing ν with ∞. Note that some values of ν are inadmissible. For example, the mean of $Z_t(s)$ is infinite and the variance of $Z_t(s)$ is negative if $\nu = 1$, and the variance of $Z_t(s)$ is infinite if $\nu = 2$.

We assume that $Z_t(s) \overset{\text{iid}}{\sim} \mathcal{ST}(\xi, \omega, \alpha, \nu)$ with density function (1). For the skew-t distribution, ξ is not the mean and ω is not the standard deviation. Instead, $\mathbb{E}\{Z_t(s)\} = \xi + \omega b_\nu \delta$ and $\text{var}\{Z_t(s)\} = \omega^2 \left\{\frac{\nu}{\nu-2} - (b_\nu \delta)^2\right\}$, where $b_\nu = \frac{\sqrt{\nu}\Gamma(\frac{1}{2}(\nu-1))}{\sqrt{\pi}\Gamma(\frac{1}{2}\nu)}$ and $\delta = \frac{\alpha}{\sqrt{1+\alpha^2}}$. To obtain zero mean and unit variance, we set $\omega = \frac{1}{\sqrt{\frac{\nu}{\nu-2} - (b_\nu \delta)^2}}$ and $\xi = -\frac{b_\nu \delta}{\sqrt{\frac{\nu}{\nu-2} - (b_\nu \delta)^2}}$. Therefore, the scaled skew-t distribution only depends on the skewness parameter α and the degree of freedom ν. The corresponding density function and distribution are denoted by $f_{\mathcal{SST}}(\cdot) = f_{\mathcal{SST}}(x; \alpha, \nu)$ and $F_{\mathcal{SST}}(\cdot) = F_{\mathcal{SST}}(x; \alpha, \nu)$, respectively.

2.3 Model implications and interpretations

The proposed model is flexible and all the parameters in model (2) have natural interpretations. For example, a higher $\nu \in \mathbb{R}^+$ and $\alpha \in \mathbb{R}$ imply a lighter tail and larger right-skewness, respectively; σ_t introduces the heteroscedasticity; the autoregression matrix $B = (\boldsymbol{\beta}_s)_{s=1}^N = (\beta_{ij})_{N \times N}$ controls the spatio-temporal dependence, and $u_t(s)$ is the threshold that determines the wet or dry probability.

We can derive several important precipitation probabilities conditional on previous observations. First, the conditional dry probability at site s and time t is

$$\begin{aligned} \mathbb{P}\{Y_t(s) = 0 | \mathbf{Y}_{t-1} = \mathbf{y}_{t-1}\} &= \mathbb{P}\{\varepsilon_t(s) \le u_t(s) - \boldsymbol{\beta}_s' \mathbf{y}_{t-1}\} \\ = \mathbb{P}\{Z_t(s) \le \tfrac{u_t(s) - \boldsymbol{\beta}_s' \mathbf{y}_{t-1}}{b_0 + b_1 \bar{y}_{t-1}}\} &= F_{\mathcal{SST}}\left(\tfrac{u_t(s) - \boldsymbol{\beta}_s' \mathbf{y}_{t-1}}{b_0 + b_1 \bar{y}_{t-1}}\right). \end{aligned} \tag{3}$$

Hence, a higher dry probability can be reached by either decreasing b_0, b_1 or α, or by increasing $u_t(s)$ or ν. In particular, if $\mathbf{Y}_{t-1} = \mathbf{0}$, then the consecutive dry probability can be written as $\mathbb{P}(Y_t(s) = 0 | \mathbf{Y}_{t-1} = \mathbf{0}) = F_{\mathcal{SST}}\left(\frac{u_t(s)}{b_0}\right)$.

Since the random variables $Z_t(s)$ are independent, conditional on $\mathbf{Y}_{t-1} = \mathbf{y}_{t-1}$, we can obtain the simultaneous dry probability at multiple sites to be the product of marginal conditional dry probabilities using Equation (3). Therefore, once we plug in the estimated parameters for those probabilities, which are conditional on previous events, we can immediately obtain the dry/wet probability (rainfall occurrence), consecutive dry/wet probability (distribution of the dry/wet spell length), and the simultaneous dry/wet probability for multiple sites (rainfall spatial pattern).

The unconditional dry probability can be calculated using the law of total probability, i.e., $\mathbb{P}\{Y_t(s) = 0\} = \int \cdots \int_{\mathbf{y}} \mathbb{P}\{Y_t(s) = 0 | \mathbf{Y}_{t-1} = \mathbf{y}\} \mathbb{P}\{\mathbf{Y}_{t-1} = \mathbf{y}\} dy_1 \cdots dy_N$, which cannot be solved explicitly but can be obtained recursively.

The SPG in model (2) possesses both flexible statistical properties and nice physical interpretations. We illustrate these by representing model (2) as a state-space model, i.e., a two-layer model where the transition from zero to positive values is driven by the selection mechanism of the skew-t distribution. The equivalent model can be specified as:

$$Y_t(s) = \begin{cases} X_t(s), & X_t(s) > u_t(s), \\ 0, & X_t(s) \le u_t(s), \end{cases} \quad (4) \qquad X_t(s) = g(\mathbf{X}_{t-1}) + \begin{cases} Z_t(s), & W_t(s) > 0, \\ -Z_t(s), & W_t(s) \le 0, \end{cases} \quad (5)$$

where the threshold $u_t(s)$ is deterministic, the latent autoregressive process $X_t(s)$ depends on a function of $\mathbf{X}_{t-1}$ called $g(\mathbf{X}_{t-1})$, and the random errors are controlled by two processes, $Z_t(s)$ and $W_t(s)$. The proof for the equivalence between Equation (2) and Equations (4) and (5) is given in the Appendix.

In meteorology, it is well known that the precipitation comes from condensed atmospheric water vapor, which is formed at certain temperatures and moisture conditions, and then falls as observable rainfall due to gravity. Our equations (4) and (5) describe this physical process.

Equation (4) is the measurement equation, which we call the ground layer model. It describes the amount of condensed atmospheric water vapor $X_t(s)$ that becomes precipitation $Y_t(s)$ at time t and location s. The wet-dry threshold $u_t(s)$ represents the necessary conditions for rainfall, defined as the minimum condensed water vapor required for observable rainfall, i.e., to reach the detection limit of the measuring instrument. Equation (5) is the transition equation, which we call the atmospheric layer model. It describes the formation of condensed water vapor in the atmosphere. The current condensed water vapor $X_t(s)$ is modeled using past observations at all locations $g(\mathbf{X}_{t-1})$, with random fluctuations $Z_t(s)$ that represent the new formation and dissolution of condensed water vapor. The fluctuation $Z_t(s)$ is not just the symmetric random noise, but is driven by a hidden selection process $W_t(s)$, which represents certain meteorological conditions, known as weather fronts such as temperature and moisture. Although the distributions of their elements, $Z_t(s)$ and $W_t(s)$, are both symmetric, when they are correlated, the distribution of $X_t(s)$ becomes skewed. This representation explains that our model is suitable for data that are censored and skewed. Therefore, the proposed SPG can potentially simulate realistic precipitation observations.

2.4 Inference and computational issues

The inference of model (2) is straightforward as it belongs to the generalized Tobit model (McDonald and Moffitt, 1980). We show in the Appendix that the log-likelihood, $\ell(\boldsymbol{\theta}|\mathbf{y}_1,\ldots,\mathbf{y}_T)$, can be written as

$$\begin{aligned}\ell(\boldsymbol{\theta}|\mathbf{y}_1,\ldots,\mathbf{y}_T) = \\ \textstyle\sum_{t=2}^{T}\sum_{s=1}^{N} \mathbb{1}_{\{y_t(s)>0\}}\left[\log\left\{f_{\mathcal{SST}}\left(\frac{y_t(s)-\boldsymbol{\beta}_s'\mathbf{y}_{t-1}}{b_0+b_1\bar{y}_{t-1}}\right)\right\} - \log\left\{b_0+b_1\bar{y}_{t-1}\right\}\right] \\ +\textstyle\sum_{t=2}^{T}\sum_{s=1}^{N} \mathbb{1}_{\{y_t(s)=0\}}\log\left\{F_{\mathcal{SST}}\left(\frac{u_t(s)-\boldsymbol{\beta}_s'\mathbf{y}_{t-1}}{b_0+b_1\bar{y}_{t-1}}\right)\right\},\end{aligned} \quad (6)$$

where $\boldsymbol{\theta}$ is the vector of parameters to be estimated, $\mathbb{1}_{\{y_t(s)>0\}}$ is an indicator function that takes a value of 1 when $y_t(s) > 0$ and 0 otherwise, the cutoffs $u_t(s)$ are chosen to be the precision limit of the measuring device, $f_{\mathcal{SST}}(\cdot;\alpha,\nu)$ is the scaled skew-t density function, and $F_{\mathcal{SST}}(\cdot;\alpha,\nu)$ is the distribution function.

To quantify the spatial dependence and make the estimation of the autoregressive parameters computationally feasible, we further parameterize the $N \times N$ matrix B using the idea in Sigrist et al. (2012). Since the external wind vector in their models is not available in our study, we assume B to be a anisotropic covariance matrix and independent of t. For our application, there are only 8 locations within small flat area of $1km^2$, it is reasonable to assume spatial stationarity. Specifically, we assume that $\beta_{ij} = \phi\exp\{-(\mathbf{s}_i-\mathbf{s}_j)^T\Sigma^{-1}(\mathbf{s}_i-\mathbf{s}_j)\}$ and

$$\Sigma^{-1} = \frac{1}{\rho}\begin{bmatrix} \cos(\omega) & \sin(\omega) \\ -c\sin(\omega) & c\cos(\omega) \end{bmatrix}^T \begin{bmatrix} \cos(\omega) & \sin(\omega) \\ -c\sin(\omega) & c\cos(\omega) \end{bmatrix},$$

where $c > 0$ controls the degree of anisotropy, $\omega \in [0, \pi/2]$ is the angle of rotation, $\phi > 0$ represents the variance, and $\rho > 0$ is a scaling parameter representing the spatial range. Based on the structure, we can account for the spatial dependence and assume that faraway sites are less correlated. Furthermore, to make sure the process is temporally stationary, the largest eigenvalue of matrix B should be less than 1. Here, we use a sufficient condition for B by constraining $\phi/\rho < \frac{1}{N \max_{ij}(d_{ij})}$. Note that the condition only holds if B is a valid covariance matrix.

The final unknown parameters in Equation (6) are $\boldsymbol{\theta} = (\phi, \rho, c, \omega, b_0, b_1, \alpha, \nu)^T$. Although we reduce the number of unknown parameters to eight, the optimization of the likelihood that can only be achieved numerically is still potentially unstable. Thus, we consider several different numerical optimization methods: two derivative-free algorithms (COBYLA and Nelder-Mead) and two derivative-based algorithms (BFGS and CG). Many literatures point out that the `optim` function in R is not numerically stable, especially when a re-parameterization exists (Mullen, 2014; Nash and Varadhan, 2011; Nash, 2014a). Therefore, we employ the functions in two recently developed R packages, `nloptr` (Johnson, 2014) and `Rcgmin` (Nash, 2014b), as a substitution of `optim`. To make sure that the optimization reaches the global maximum, we use different optimization algorithms with multiple sets of initial values until we get the same optimized values. The `sn` package (Azzalini, 2011) was used to evaluate $f_{\mathcal{SST}}(\cdot)$ and $F_{\mathcal{SST}}(\cdot)$. We also notice that the estimation of the degree of freedom ν is typically not numerically stable as in other similar problems, especially when ν is not integers. Here, we suggest to evaluate the likelihood over a sequence of values of ν. In our simulation study, we choose $\nu = 3$ to represent heavy tail, $\nu = 7$ to represent moderate tail, and $\nu = 20$ to represent light tail, because a value of ν smaller than 3 leads to an infinite mean or variance, and a very large ν is too close to the Gaussian case and significantly increases the computational burden. In the application, we tried different ν starting from 3 and the best ν can be selected according to the maximized likelihood function.

3 Simulation studies

We designed a simulation study to validate the inference procedure introduced in Section 2.4 and compare to the original censored VAR models proposed by Koch and Naveau (2015) using the Gaussian dynamics. We generate multiple synthetic datasets from the censored VAR model (2) at the eight locations shown in Figure 1, where $Z_t(s)$ follows a skew-t distribution. By choosing different skewness parameter α and degree of freedom ν, the generated data can approximate the model in Koch and Naveau (2015) as a special case by letting $\alpha = 0$ and ν be a large value. In the simulation, we set $\nu = 20$ to be large and the two models with $\nu = 3, 7$ for comparison. In addition, we set the skewness parameter, $\alpha = 5$ to represent the skewed data and set $(T, N, \phi, \rho, c, \omega, b_1, b_2) = (10,000, 3, 1/3, 1, 1.5, 1.5, 0.5, 0.5)$. The sample sizes, T and N, are the same as the application in Koch and Naveau (2015) for comparison reason.

We fit the models with skew-t and Gaussian dynamics to the synthetic datasets. The summary of estimated values based on 50 independent realizations are shown in Table S1 of the Supplementary Materials. With different dynamics, the optimized common factors, such as b_0 and b_1, are different as well. The results show that Gaussian dynamics tends to overestimate the b_0 and b_1 so that the overestimated variance will compensate the light tail of Gaussian distribution.

Based on the estimated model parameters, we can calculate and visualize some key statistics of two SWGs and compare their differences statistically. We draw quantile-quantile (QQ) plots between the synthetic data and 50 parametric bootstrap samples for two scenarios, $(\nu, \alpha) = (20, 0)$ and $(\nu, \alpha) = (3, 5)$. The results are shown in Figure 2. Not surprisingly, the Gaussian model cannot reproduce the heavy-tailed and right-skewed behavior that results from a large skewness parameter and a small degree of freedom. In both cases, our model with the skew-t errors successfully reproduces these statistical properties.

Incorrectly assuming a Gaussian model for the data also affects the estimation of other important statistical properties of rainfall. As we mentioned, Gaussian model typically overestimates b_0 and b_1 of heavy-tailed data in order to produce a high rainfall intensity. However, as we explained in Section 2.3, a large b_0 and b_1 also leads to a low dry probability. Therefore, failing to correctly specify the degree of freedom will underestimate the dry probability as well, as shown in Figure 3, where the distribution of the dry probability is obtained from the bootstrap samples. Therefore, only relying on the heteroscedastic standard deviations, without considering a right-skewed and heavy-tailed error term, is not enough to capture the statistical properties of high-frequency rainfall patterns and the stochastic simulations will not be realistic.

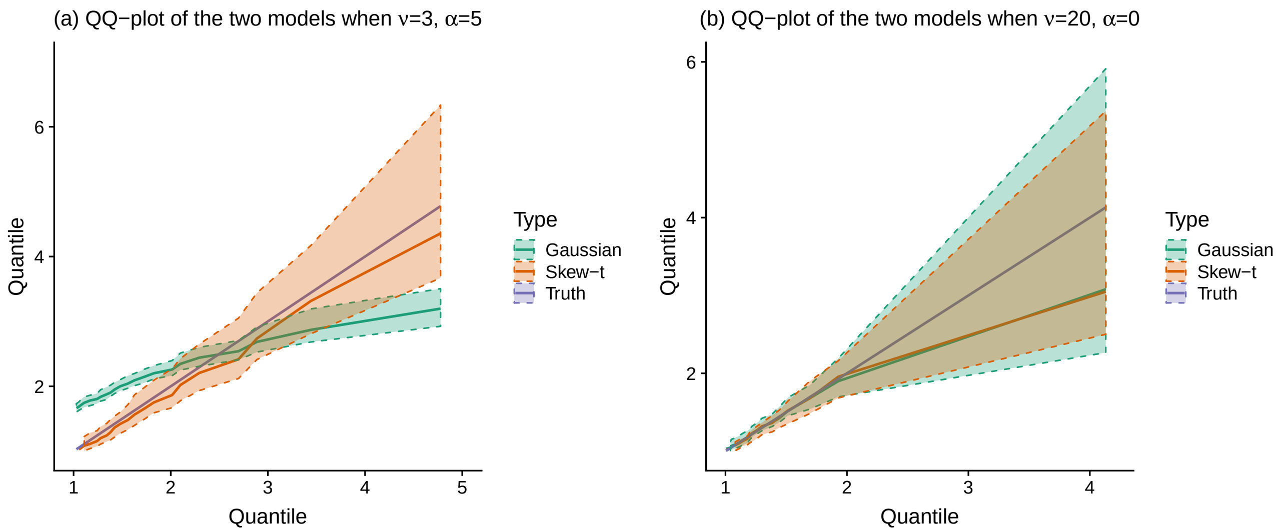

Figure 2: Quantile-quantile (QQ) plot between the synthetic data (purple curve) and 50 parametric bootstrap samples using (a) a skew-t error (orange region) and (b) a Gaussian error (green region). The solid lines are the median curves and the shaded areas are the 95% confidence intervals.

Although the main point of a SWG is to reproduce key statistics of the weather as we have shown, we also include some comparisons related to prediction performance in the Supplementary

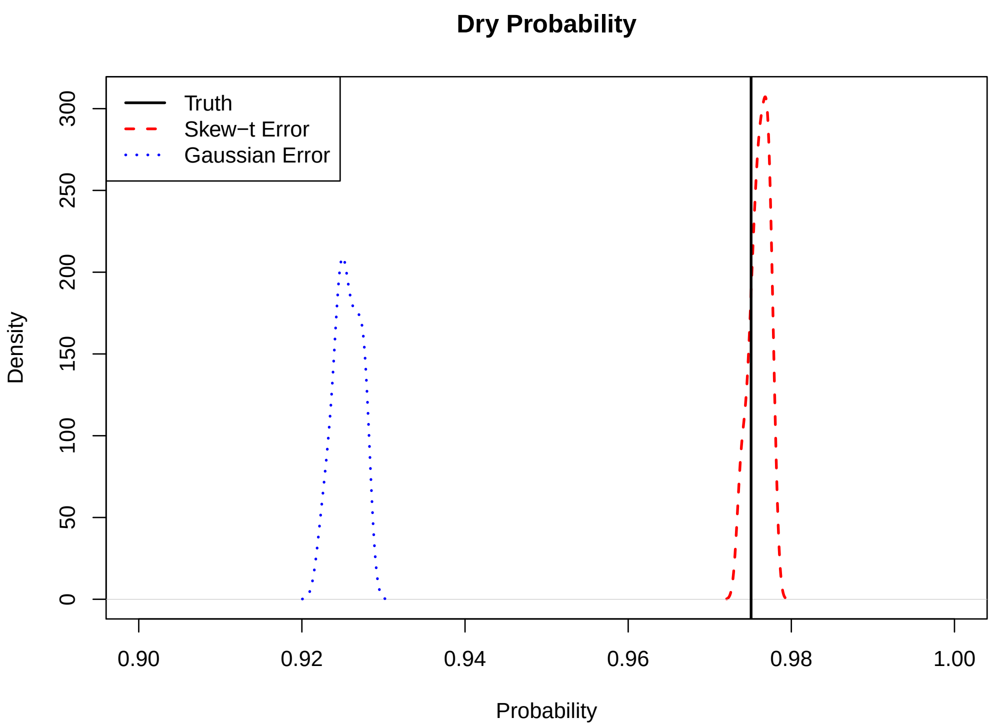

Figure 3: The dry probabilities from the synthetic dataset and 50 parametric bootstrap samples using skew-t and Gaussian models. The black line represents the truth from the synthetic data; the red dashed and blue dotted curves are the empirical density of the dry probability from 50 parametric bootstrap samples using Gaussian and skew-t models, respectively.

Materials. The results show that when ν is large and $\alpha = 0$, two models produce similar results. In contrast, when α is positive and ν is small, the root mean square error of our SWG outperforms the Gaussian model, which indicates that Gaussian model becomes less reliable in the prediction of highly skewed and heavy-tailed precipitation data.

4 Application to Lausanne precipitation data

The motivating data, as shown in Figure 1, were collected by GAIA Lab, Institute of Earth Surface Dynamics (IDYST), the University of Lausanne in 2016 using eight Pluvimate acoustic rain gauges (Collister and Mattey, 2008). The detailed description of the Pluvimate rain gauges can be found in Benoit et al. (2018). This new instrument can measure precipitations at a $0.01mm$ and 30 second resolution with the standard rain rate unit, mm/hr. From April 4th to April 14th, 2016, a total number of 230,400 observations were recorded by the network of eight rain gauges located within a radius of $1km$, as shown in Figure S1 in the Supplementary Materials.

Table 1 and Figure S2 in the Supplementary Materials show the distribution of the collected rainfall data. Most of the observations (92.4%) were zeros, corresponding to no rain or dry events. Most of the rain intensities are below $5mm/hr$, while at some periods there are heavy rainfall with the rain rate around $10mm/hr$ We also observe that the highest rain rate reached more than $30mm/hr$. The QQ-plot also shows that these rainfall data are zero-inflated, right-skewed, and heavy-tailed and completely deviate from Gaussian distribution.

Table 1: Distribution of the collected rainfall data.

Rain rate (mm/hr)	$=0$	$(0,1.2]$	$(1.2,5]$	$(5,10]$	$(10,33.6]$
Percentage	92.40%	5.55%	1.84%	.19%	.02%

The cutoffs are chosen to be the lowest positive rate $u_t(s) = 1.2mm/hr$ and other parameters are fitted by the methods mentioned in Section 2.4. The fitted results are shown in Table 2, along with the results of the Gaussian model for comparison purposes. We can see that the maximum likelihood estimators (MLEs) of both the degree of freedom ν and the skewness parameter α are small. Similar to the simulation study, the values of b_0 and b_1 estimated from the Gaussian model are larger than those estimated from the skew-t model.

Table 2: Maximum likelihood estimators (MLEs) from the censored VAR model with skew-t and Gaussian errors.

Parameters	Skew-t Errors		Gaussian Errors	
	Est	95%CI	Est	95%CI
ϕ	6.362	$(5.359, 7.553)$	6.227	$(5.158, 7.296)$
ρ	1.334	$(1.323, 1.344)$	1.364	$(1.355, 1.373)$
c	1.504	$(0.543, 2.465)$	1.671	$(0.712, 2.640)$
ω	1.748	$(0.723, 2.773)$	1.323	$(0.483, 2.163)$
b_0	.441	$(.405, .477)$	1.190	$(0.912, 1.468)$
b_1	.341	$(.330, .352)$	0.459	$(.448, .471)$
α	.028	$(.004, .053)$		
ν	4			

4.1 Conditional rainfall generation

From the fitted model, we can generate 50 parametric bootstrap samples of the high-frequency rainfall. Each sample can be viewed as a synthetic realization of the 30-second rainfall data. To examine the similarity of the simulated samples to the real data, we use a one-step-ahead conditional simulation, which generates data conditional on observations one step in the past. The main goal here is to reproduce the statistical properties of the real data, we use a QQ plot shown in Figure 4, to compare the real data with the bootstrap samples from the two models. From the results, both models can reproduce the low-intensity rainfall well statistically. However, in terms of the heavy rainfall (defined as having an intensity larger than $10mm/hr$), the 95% confidence band of the Gaussian model significantly underestimates the precipitation, whereas the skew-t model captures the heavy tail very well.

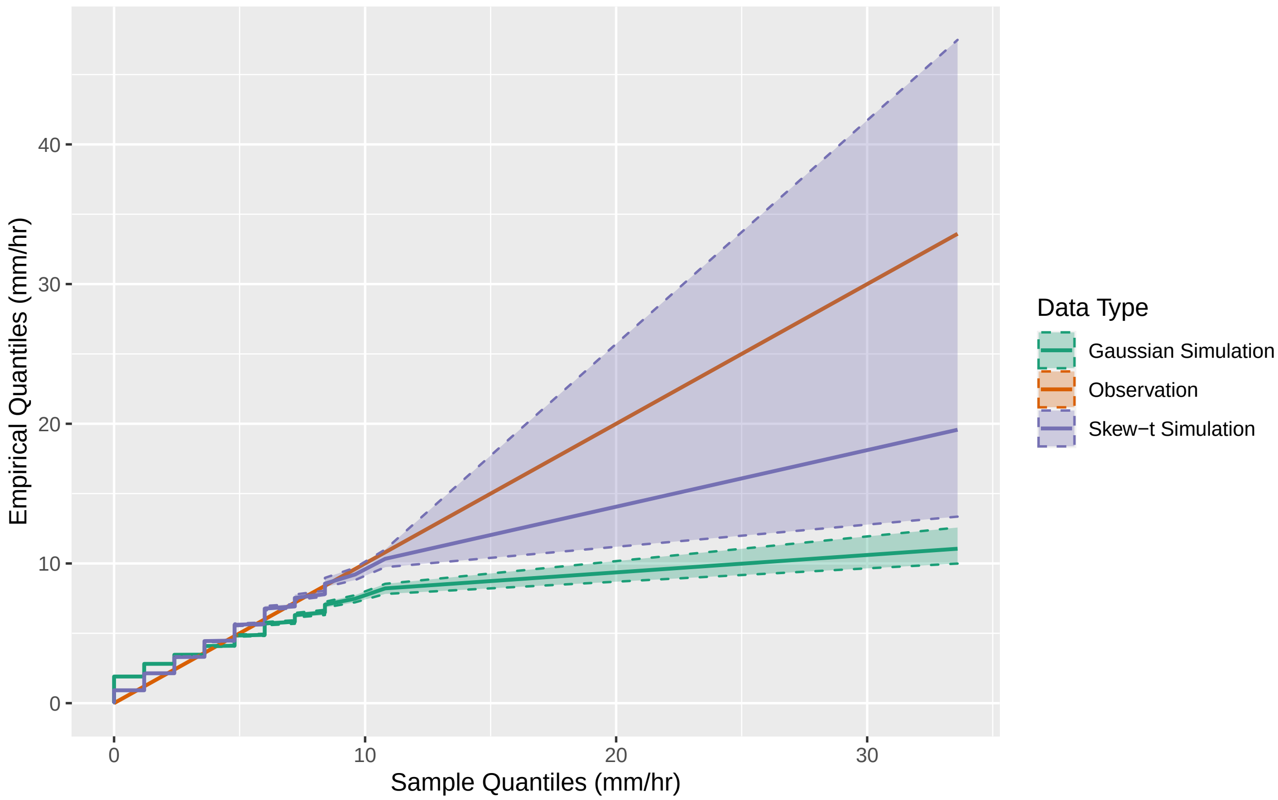

Figure 4: Quantile-quantile (QQ) plot between the observed rainfall (purple curve) and 50 parametric bootstrap samples using a skew-t error (orange region) and a Gaussian error (green region). The solid lines are the median curves and the shaded areas are the 95% confidence intervals.

The spatio-temporal patterns of rainfall occurrence are assessed by the rain concurrences and dry probability. Figure 5 shows the histogram of simultaneously rainy locations for both the observed and simulated rainfall data, from which we can see the spatial dependences among multisite rainfall occurrences. The simulated histograms combines all of the 50 parametric bootstrap samples. The eight locations are usually all dry. In this case, the simulated data slightly overestimate the counts. In contrast, when at least one location is rainy, the simulated data have fewer counts than real observations, which means the spatial correlation is slightly underestimated. Overall, the rain concurrences are reproduced reasonably well by our model.

The temporal dependence of rainfall occurrence can be reflected by the dry/wet spell length

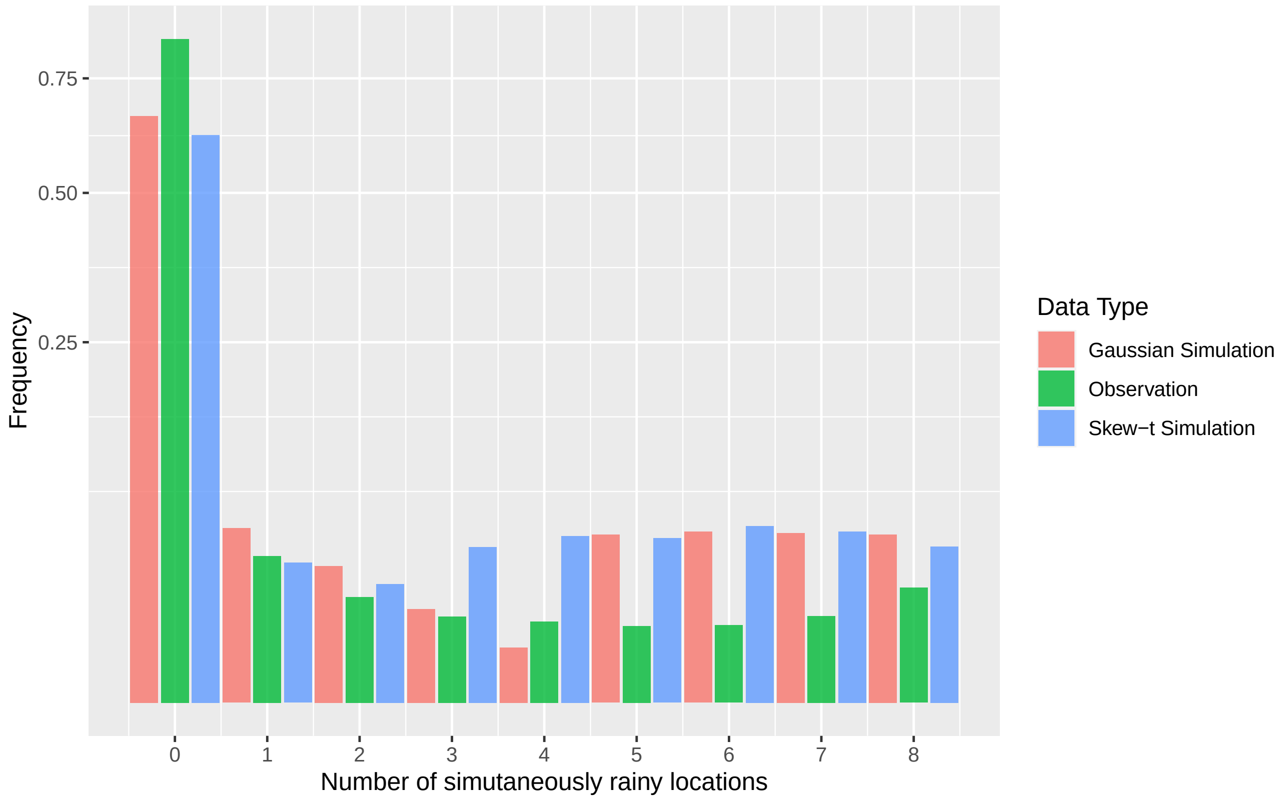

Figure 5: Histogram showing the simultaneously rainy locations. Each bar represents how many corresponding events happen in time. The y-axis is shown in square-root scale.

and conditional dry/wet probability. The dry/wet spell length is less important for the high-frequency data of only ten days and thus we only show the conditional probabilities. Table 3 displays the four different conditional probabilities. Overall, the simulated data can reproduce the conditional dry/wet probability well. However, the simulated data tend to underestimate the probability of transition between wet and dry. This is an expected result of the hard thresholding effect. Nevertheless, it does not affect much on the overall distribution of the rainfall intensity. If it is crucial to accurately reproduce the dry or wet probability, a separate model for rainfall occurrences might be a better choice.

Compared to the Gaussian model, the major advantage of the skew-t model is that it can

Table 3: Conditional probabilities from the observed rainfall data and the mean ($\pm 1.96\times$standard deviation) of the conditional probabilities from 50 parametric bootstrap samples generated using a skew-t and Gaussian model.

	Wet|Wet	Dry|Wet	Wet|Dry	Dry|Dry
Skew-t Dynamics	.976($\pm$.0045)	.024($\pm$.0045)	.024($\pm$.0045)	.976($\pm$.0045)
Gaussian Dynamics	.974($\pm$.0062)	.026($\pm$.0062)	.026($\pm$.0062)	.974($\pm$.0062)
Observation	.971	.029	.029	.971

reproduce both dry events and heavy rainfall. However, since the Gaussian and skew-t models we considered use the same spatio-temporal modeling strategy, i.e., the VAR model with a time-varying threshold, the spatio-temporal properties of the rainfall occurrences do not show much difference.

4.2 Short-term rainfall prediction

A SWG is not meant for prediction but generate multiple realizations with similar statistical properties of the real data. However, we can provide some predictive results by the mean or quantiles of multiple realizations based on the conditional distribution of our SWG. We give an example here to illustrate the performance of our model in one-step-ahead prediction, compared to a multivariate LSTM model as a popular deep learning-based model for time series prediction. We use the last day of rainfall ($5,760$ points) at each location as the testing set and other data as the training set. We re-fit our model using the training data and fit the LTSM model using one hidden layer of 100 LSTM units with rectified linear unit (ReLU) activation, Adam optimization algorithm, and mean squared error (MSE) as the loss function. The LSTM model is run on Keras with 100 epoch in Python.

As a precipitation generator, our model can simultaneously generate 50 one-stead-ahead realizations based on the conditional distribution we assume, then we use the mean of 50 realizations as our predictor in the comparison. We also compute the 95% prediction band by the quantiles of the 50 realizations to show the uncertainties. In terms of the MSE, the result of our model (MSE= 0.099) outperforms the LSTM (MSE= 0.126). To visualize the results, in Figure 6, we show one heavy rain event in the evening of April 13 and 95% prediction band from our model as there is almost no rain at other time. The results show that the 95% prediction band covers the observed rainfalls. Also, the prediction of LSTM is more variable with a few obvious peaks far from the observed values. Instead, our predictor is more stable since it collects the mean of 50 realizations.

From the results, two main differences between LSTM predictor and our stochastic precipitation generator are noticed, even though both of the models provide good prediction. Firstly, a SWG can serve as a predictor via generating multiple independent realizations from the conditional distribution, which can be also used to quantify the uncertainties. But LSTM provides a deterministic result when the model parameter is fixed. Secondly, as many other deep learning methods, the prediction of LSTM is done via optimizing certain loss function, e.g. MSE, thus the results are harder to interpret than our model-based prediction. For example, the predicted values from LSTM model can be negative values. Although we can treat the negative value as zero by the ReLU activation function in the output layer, we lose some meaningful interpretations of the rainfall properties.

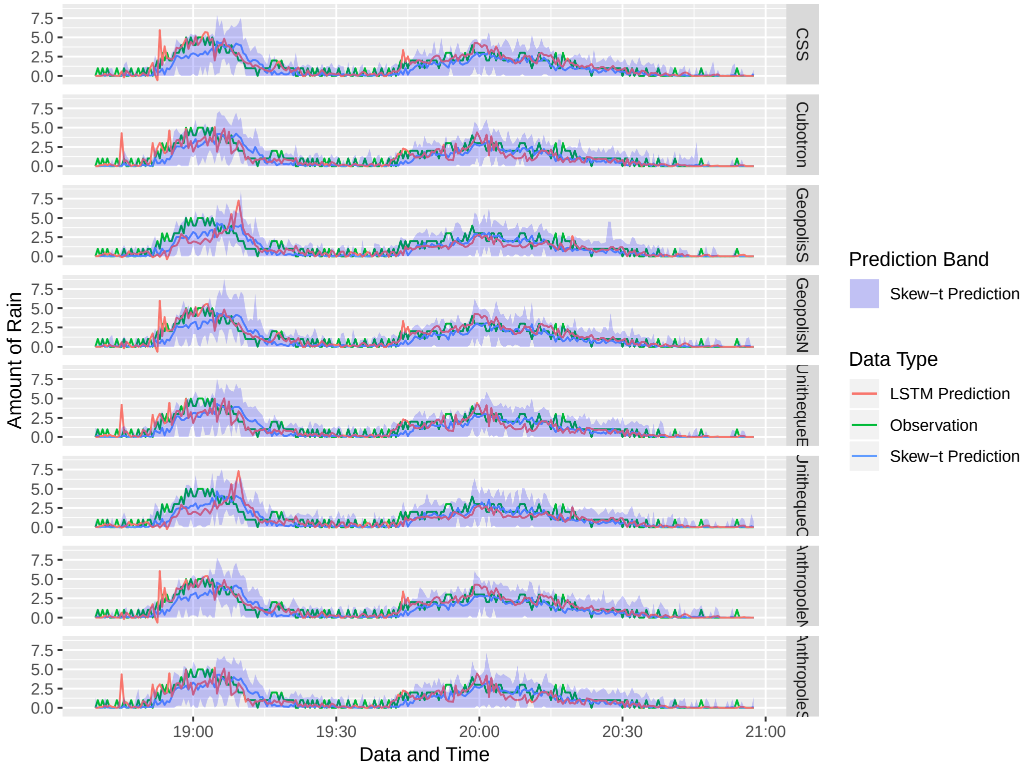

Figure 6: The performance of predicted rain rate (mm/hr) (with uncertainties) during a heavy rainfall period on April 13, 2016. The blue line and blue band are the mean prediction and 95% prediction band from our model, respectively. The red line is the prediction from LSTM model. The green line is the last day of rainfall observation as the truth.

5 Discussion

In this study, we proposed a new SWG for high frequency rainfall that is capable of reproducing large quantities of zeros and extreme values. This model utilizes skew-t random variables with a censoring mechanism to simultaneously drive both the occurrence and intensity of rainfall. By applying a VAR model for spatio-temporal dependence, the inference can be achieved simply by maximizing the likelihood function.

We applied the SWG to a rarely assessed and fine-scale precipitation dataset collected by an acoustic rain gauges every 30 seconds and spatially around 10-100 meters. The results show that the SPG can generate high-intensity rainfall better than the baseline model with Gaussian errors. We also show that our SWG can be used for prediction purposes and its performance is comparable to LSTM, a modern deep learning method. Compared to the LSTM, our SWG can provide uncertainties and is more interpretable.

Although we only consider the VAR framework with a linear spatio-temporal dependence, the model can be generalized by considering other spatio-temporal dependences. One situation was discussed in Tadayon and Khaledi (2015), where a Bayesian hierarchical model was used with only skew-normal error and purely spatial dependence. Another possibility is to use a censored multivariate skew-symmetric distribution. However, the associated inference might be challenging, since likelihood functions do not have a closed form as in our model. The main reason is that skew-symmetric random variables do not retain all of the desired properties when they become Gaussian or student-t random variables, e.g., they are not closed under convolution or conditioning. Thus, some existing results under certain assumptions, such as additive models, or solutions derived from conditional distributions, such as the Kriging predictor, would no longer hold in skew-symmetric cases.

In addition, the seasonality can be considered by taking spatially and temporally varying cutoffs. We use constant cutoffs due to our data are both spatially and temporally stationary regarding the seasonality, as our dataset is observed in an $1km \times 1km$ area on the University of Lausanne campus and an 11-day period from 4th to 14th April, 2016. However, when the seasonality is observed, we can estimate the cutoff $u_t(s)$ based on the method in Sun and Stein (2015), where the estimated cutoffs $\hat{u}_t(s)$ are chosen to be the empirical quantile corresponding to the dry probabilities which is further fitted by logistic regression with harmonic terms.

We parameterize the autoregression matrix, which reduces the number of parameters while taking spatial information into account. Although such a parameterization fits our purpose well, some limitations exist. First, we selected a covariance function to construct the autoregression matrix, which implies that the single-site temporal correlation is homogenous across many locations. Second, the correlation matrix only allows for a positive correlation among locations, whereas an autoregression matrix could also have negative correlations. Last, we use a sufficient condition for the parameters in the spatial covariance function to satisfy the temporally stationary condition. However, it is more favorable to have a necessary and sufficient condition to ensure the stationarity. Thus if only less sites are assessed it is better to estimate the autoregression matrix in element-wise.

6 Acknowledgement

This research was supported by King Abdullah University of Science and Technology (KAUST), Office of Sponsored Research (OSR) under Award No: OSR-2019-CRG7-3800. The dataset used in this study is provided by GAIA Lab, Institute of Earth Surface Dynamics (IDYST), the University of Lausanne. We acknowledge their efforts for collecting the data.

SUPPLEMENTARY MATERIALS

Figures and tables: The supplementary materials consist of additional figures and tables.

Data and source codes: The data and source codes in this work is available on the following repository: https://github.com/aleksada/MultisiteHighFreqPG.

References

Ailliot, P., Allard, D., Monbet, V., and Naveau, P. (2015). Stochastic weather generators: an overview of weather type models. *Journal de la Société Française de Statistique*, 156(1):101–

113.

Ailliot, P., Thompson, C., and Thomson, P. (2009). Space-time modelling of precipitation by using a hidden Markov model and censored Gaussian distributions. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 58(3):405–426.

Allard, D. (2012). Modeling spatial and spatio-temporal non Gaussian processes. In *Advances and Challenges in Space-time Modelling of Natural Events*, pages 141–164. Springer.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, pages 171–178.

Azzalini, A. (2011). R package sn: The skew-normal and skew-t distributions (version 0.4-17). *URL http://azzalini. stat. unipd. it/SN*, 20.

Azzalini, A. (2013). *The Skew-Normal and Related Families*, volume 3. Cambridge University Press.

Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):579–602.

Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):367–389.

Azzalini, A. and Regoli, G. (2012). Some properties of skew-symmetric distributions. *Annals of the Institute of Statistical Mathematics*, 64(4):857–879.

Baxevani, A. and Lennartsson, J. (2015). A spatiotemporal precipitation generator based on a censored latent Gaussian field. *Water Resources Research*, 51(6):4338–4358.

Bell, T. L. (1987). A space-time stochastic model of rainfall for satellite remote-sensing studies. *Journal of Geophysical Research: Atmospheres*, 92(D8):9631–9643.

Benoit, L., Allard, D., and Mariethoz, G. (2018). Stochastic rainfall modeling at sub-kilometer scale. *Water Resources Research*, 54(6):4108–4130.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Boschert, S. and Rosen, R. (2016). Digital twin–the simulation aspect. In *Mechatronic Futures*, pages 59–74. Springer.

Branco, M. D. and Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*, 79(1):99–113.

Chan, S., Kendon, E., Roberts, N., Fowler, H., and Blenkinsop, S. (2016). The characteristics of summer sub-hourly rainfall over the southern uk in a high-resolution convective permitting model. *Environmental Research Letters*, 11(9):094024.

Collister, C. and Mattey, D. (2008). Controls on water drop volume at speleothem drip sites: An experimental study. *Journal of Hydrology*, 358(3-4):259–267.

Durbán, M. and Glasbey, C. (2001). Weather modelling using a multivariate latent Gaussian model. *Agricultural and Forest Meteorology*, 109(3):187–201.

El Saddik, A. (2018). Digital twins: The convergence of multimedia technologies. *IEEE MultiMedia*, 25(2):87–92.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007.

Falconer, G. and Mitchell, S. (2012). Smart city framework. *Cisco Internet Business Solutions Group (IBSG)*, pages 1–11.

Flecher, C., Naveau, P., Allard, D., and Brisson, N. (2010). A stochastic daily weather generator for skewed data. *Water Resources Research*, 46(7).

Gabriel, K. and Neumann, J. (1962). A Markov chain model for daily rainfall occurrence at Tel Aviv. *Quarterly Journal of the Royal Meteorological Society*, 88(375):90–95.

Genton, M. G. (2004). *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. CRC Press.

Gilchrist, A. (2016). *Industry 4.0: The Industrial Internet of Things*. Springer.

Glasbey, C. and Nevison, I. (1997). Rainfall modelling using a latent Gaussian variable. In *Modelling Longitudinal and Spatially Correlated Data*, pages 233–242. Springer.

Hamilton, J. D. (1994). *Time Series Analysis*, volume 2. Princeton University Press.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of Economic and Social Measurement, Volume 5, number 4*, pages 475–492. NBER.

Johnson, S. (2014). The nlopt nonlinear-optimization package [software].

Katz, R. W. (1977). Precipitation as a chain-dependent process. *Journal of Applied Meteorology*, 16(7):671–676.

Kleiber, W., Katz, R., and Rajagopalan, B. (2012). Daily spatiotemporal precipitation simulation using latent and transformed gaussian processes. *Water Resources Research*, 48:W01523.

Koch, E. and Naveau, P. (2015). A frailty-contagion model for multi-site hourly precipitation driven by atmospheric covariates. *Advances in Water Resources*, 78:145–154.

Mary, B., Beaudoin, N., Brisson, N., and Launay, M. (2009). *Conceptual Basis, Formalisations and Parameterization of the STICS Crop Model*. Quae.

McDonald, J. F. and Moffitt, R. A. (1980). The uses of tobit analysis. *The Review of Economics and Statistics*, pages 318–321.

Mullen, K. M. (2014). Continuous global optimization in r. *Journal of Statistical Software*, 060(i06).

Nash, J. C. (2014a). On Best Practice Optimization Methods in R. *Journal of Statistical Software*, 60(i02).

Nash, J. C. (2014b). *Rcgmin: Conjugate Gradient Minimization of Nonlinear Functions*. R package version 2013-2.21.

Nash, J. C. and Varadhan, R. (2011). Unifying Optimization Algorithms to Aid Software System Users: optimx for R. *Journal of Statistical Software*, 43(i09).

Parra, L., Sendra, S., Lloret, J., and Bosch, I. (2015). Development of a conductivity sensor for monitoring groundwater resources to optimize water management in smart city environments. *Sensors*, 15(9):20990–21015.

Rasmussen, P. (2013). Multisite precipitation generation using a latent autoregressive model. *Water Resources Research*, 49(4):1845–1857.

Richardson, C. W. (1981). Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research*, 17(1):182–190.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-k., and Woo, W.-c. (2015). Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'15, pages 802–810, Cambridge, MA, USA. MIT Press.

Shi, X., Gao, Z., Lausen, L., Wang, H., Yeung, D.-Y., Wong, W.-k., and WOO, W.-c. (2017). Deep learning for precipitation nowcasting: A benchmark and a new model. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems 30*, pages 5617–5627. Curran Associates, Inc.

Sigrist, F., Künsch, H. R., and Stahel, W. A. (2012). A dynamic nonstationary spatio-temporal model for short term prediction of precipitation. *The Annals of Applied Statistics*, pages 1452–1477.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, pages 1–48.

Söderberg, R., Wärmefjord, K., Carlson, J. S., and Lindkvist, L. (2017). Toward a digital twin for real-time geometry assurance in individualized production. *CIRP Annals*, 66(1):137–140.

Srikanthan, R. and McMahon, T. (2001). Stochastic generation of annual, monthly and daily climate data: A review. *Hydrology and Earth System Sciences Discussions*, 5(4):653–670.

Sun, Y. and Stein, M. L. (2015). A stochastic space-time model for intermittent precipitation occurrences. *The Annals of Applied Statistics*, 9(4):2110–2132.

Tadayon, V. and Khaledi, M. J. (2015). Bayesian analysis of skew gaussian spatial models based on censored data. *Communications in Statistics-Simulation and Computation*, 44(9):2431–2441.

Taylor, J. W., McSharry, P. E., and Buizza, R. (2009). Wind power density forecasting using ensemble predictions and time series models. *IEEE Transactions on Energy Conversion*, 24(3):775.

Wilks, D. S. and Wilby, R. L. (1999). The weather generation game: a review of stochastic weather models. *Progress in Physical Geography*, 23(3):329–357.

Xu, G. and Genton, M. G. (2017). Tukey g-and-h random fields. *Journal of the American Statistical Association*, 112(519):1236–1249.

7 Appendix

7.1 Proof of the equivalence between Equation (2) and Equations (3)-(4)

Equation (3) can be derived directly from Equation (2) by letting $X_t(s) = \boldsymbol{\beta}_s' \mathbf{Y}_{t-1}(s) + \varepsilon_t(s)$. Further, denote Equation (3) by $\mathbf{Y}_t = g(\mathbf{X}_t)$, then $X_t(s) = \boldsymbol{\beta}_s' \mathbf{Y}_{t-1} + \varepsilon_t(s) = \boldsymbol{\beta}_s' g(\mathbf{X}_{t-1}) + \varepsilon_t(s)$. We note that each element $\varepsilon_t(s)$ follows skew-t distribution with mean zero and variance σ_t, which is linearly associated with a standard skew-t random variable for any t and s. Therefore, using the conditioning representation of the standard skew-t random variable (Azzalini, 2013), we have $\delta_t(s) = \begin{cases} V_t(s), & W_t(s) > 0, \\ -V_t(s), & W_t(s) \leq 0, \end{cases}$ where $V_t(s)$ and $W_t(s)$ are correlated standard t random variables. Since $\varepsilon_t(s)$ is a linear transformation of $\delta_t(s)$, we have $\varepsilon_t(s) =$

$c_t(s) + \begin{cases} Z_t(s), & W_t(s) > 0, \\ -Z_t(s), & W_t(s) \leq 0, \end{cases}$ with a fixed value $c_t(s)$. Therefore, we can obtain (4) by combining terms. $\square$

7.2 Proof of the likelihood function in Equation (5)

The proof is an extension of the related methods in Koch and Naveau (2015), as we generalize the error term to a non-Gaussian distribution.

Denote the density function of $\{Y_t(s); t = 1, \ldots, T, s = 1, \ldots, N\}$ by $h(Y_t(s); t = 1, \ldots, T, s = 1, \ldots, N)$. Due to the Markov property of the VAR(1) process, the density can be written as $h(Y_t(s); t = 1, \ldots, T, s = 1, \ldots, N) = \prod_{t=2}^{T} h(\mathbf{Y}_t|\mathbf{y}_{t-1})$. Since $z_t(s)$ are mutually independent for any t and s, $u_t(s)$ is given, and $\sigma_t(s)$ only depends on $\mathbf{Y}_{t-1}$, we can conclude that, conditional on $\mathbf{Y}_{t-1}$, $Y_t(s) = (\boldsymbol{\beta}'_s\mathbf{Y}_{t-1} + \varepsilon_t(s))\mathbb{1}_{\{\boldsymbol{\beta}'_s\mathbf{Y}_{t-1}+\varepsilon_t(s)>u_t(s)\}}$ as a function of $z_t(s)$ are mutually independent for any t and s. Therefore, we have $h(Y_t(s); t = 1, \ldots, T, s = 1, \ldots, N) = \prod_{t=2}^{T}\prod_{s=1}^{N} h(Y_t(s)|\mathbf{y}_{t-1})$.

Due to the censoring mechanism, the log-density of $Y_t(s)$, ℓ, is a mixture of discrete part ℓ_d and continuous part ℓ_c. When $y_t(s) > 0$, $Y_t(s) = \boldsymbol{\beta}'_s\mathbf{y}_{t-1} + (b_0 + b_1\bar{Y}_{t-1})Z_t(s)$, and $Z_t(s)$ follows a skew-t distribution with density $f_{\mathcal{SST}}(x)$. By changing a variable, in this case, we have

$$\log\{h(Y_t(s)|\mathbf{y}_{t-1})\} = \left[\log\left\{f_{\mathcal{SST}}\left(\frac{y_t(s) - \boldsymbol{\beta}'_s\mathbf{Y}_{t-1}}{b_0 + b_1\bar{Y}_{t-1}}\right)\right\} - \log\left\{b_0 + b_1\bar{Y}_{t-1}\right\}\right].$$

When $y_t(s) = 0$, $\mathbb{P}\{y_t(s) = 0|\mathbf{y}_{t-1}\} = \mathbb{P}\{y_t(s) \leq u_t(s)|\mathbf{y}_{t-1}\}$. By writing $Y_t(s)$ as a function of $Z_t(s)$, we can get that

$$\log\mathbb{P}\{y_t(s) = 0|\mathbf{y}_{t-1}\} = \log\left\{F_{\mathcal{SST}}\left(\frac{u_t(s) - \boldsymbol{\beta}'_s\mathbf{y}_{t-1}}{b_0 + b_1\bar{Y}_{t-1}}\right)\right\}$$

Therefore, we have

$$
\begin{aligned}
&\log\{h(Y_t(s)|\mathbf{y}_{t-1}\} \\
&= \sum_{t=2}^{T}\sum_{s=1}^{N} \ell_c \mathbb{1}_{\{y_t(s)>0\}} + \ell_d \mathbb{1}_{\{y_t(s)=0\}} \\
&= \sum_{t=2}^{T}\sum_{s=1}^{N} \mathbb{1}_{\{y_t(s)>0\}} \left[\log\left\{f_{\mathcal{SST}}\left(\frac{y_t(s)-\boldsymbol{\beta}_s'\mathbf{y}_{t-1}}{b_0+b_1\bar{Y}_{t-1}}\right)\right\} - \log\left\{b_0+b_1\bar{Y}_{t-1}\right\}\right] \\
&\quad + \sum_{t=2}^{T}\sum_{s=1}^{N} \mathbb{1}_{\{y_t(s)=0\}} \log\left\{F_{\mathcal{SST}}\left(\frac{u_t(s)-\boldsymbol{\beta}_s'\mathbf{y}_{t-1}}{b_0+b_1\bar{Y}_{t-1}}\right)\right\}. \quad \square
\end{aligned}
$$