JIRSS (2019)
Vol. 18, No. 02, pp 63-85
DOI:10.29252/jirss.18.2.63

A Skew–Gaussian Spatio–Temporal Process with Non–Stationary Correlation Structure

Zahra Barzegar [1], Firoozeh Rivaz [1], and Majid Jafari Khaledi [2]

[1] Department of Statistics, Faculty of Mathematical Sciences, Shahid Beheshti University, Iran.
[2] Department of Statistics, Faculty of Mathematical Sciences, Tarbiat Modares University, Iran.

Received: 26/04/2018, Revision received: 17/12/2018, Published online: 29/08/2019

Abstract. This paper develops a new class of spatio-temporal process models that can simultaneously capture skewness and non-stationarity. The proposed approach which is based on using the closed skew-normal distribution in the low-rank representation of stochastic processes, has several favorable properties. In particular, it greatly reduces the dimension of the spatio-temporal latent variables and induces flexible correlation structures. Bayesian analysis of the model is implemented through a Gibbs MCMC algorithm which incorporates a version of the Kalman filtering algorithm. All fully conditional posterior distributions have closed forms which show another advantageous property of the proposed model. We demonstrate the efficiency of our model through an extensive simulation study and an application to a real data set comprised of precipitation measurements.

Keywords. Closed-Skew Normal Distribution, Low-Rank Models, Non-Stationarity, Spatio-Temporal Data.

MSC: 60G99; 62M99.

Zahra Barzegar Hargalan (z_barzegar@sbu.ac.ir)
Corresponding Author: Firoozeh Rivaz (f_rivaz@sbu.ac.ir)
Majid Jafari Khaledi (jafari-m@modares.ac.ir)

1 Introduction

Spatio-temporal stochastic processes play a crucial role in the analysis of a variety of problems ranging from meteorological and ocean measurements to environmental pollutants, and disease incidences. It is a common practice to assume that these processes are Gaussian and stationary. While convenient, such assumptions are rarely realistic. More precisely, data often exhibit obvious asymmety in their distributions. In addition, due to local influences in the correlation structure of spatio-temporal processes, there is no justification for assuming stationarity. As a motivating example, we consider the annual total precipitation data (in millimeters) from 92 meteorological stations, geographically distributed across Iran, between 2006 and 2015. The data were prepared based on the monthly precipitation observations that are available online from the Iranian meteorological organization at http://www.irimo.ir. Our explanatory data analysis reveals that the distribution of the data is skewed (See Figure 1). Besides, due to the climatic diversity in Iran, it seems unreasonable to assume spatial covariance structures to be stationary over the spatial scale. The p-values of the stationary test introduced by Bandyopadhyay and Rao (2017) for each time point are displayed in Figure 2, which confirms spatial non-stationarity over time except for the time $t = 7$. Thus, there is a need for a flexible and computationally tractable approach to remove Gaussianity and stationarity assumptions.

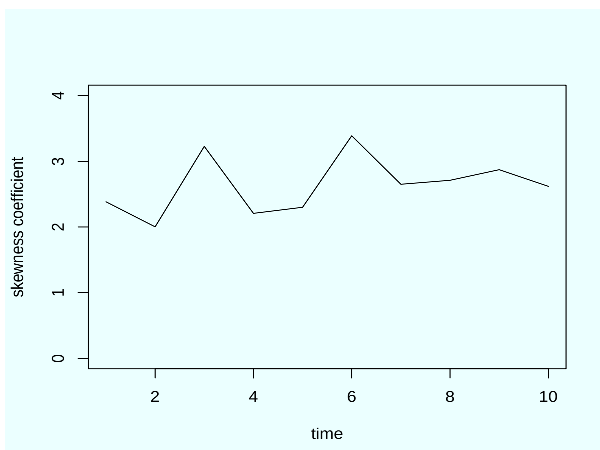

Figure 1: The skewness coefficient of the data versus time.

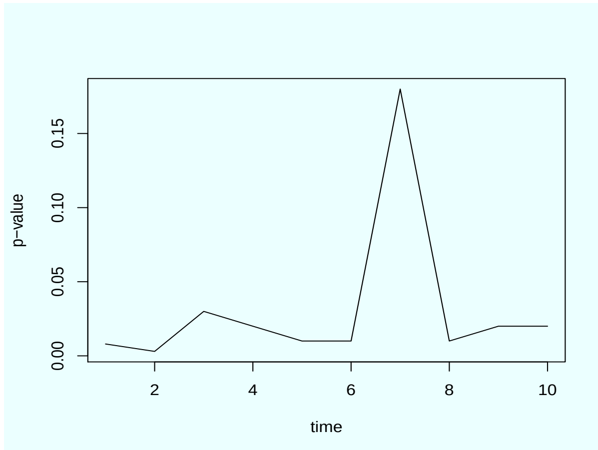

Figure 2: The p-values of the stationarity test for each time point.

Non-stationarity is an important problem in spatial statistics. In recent years, there has been an increasing interest to introduce non-stationary covariance functions, including deformation method (Sampson and Guttorp , 1992), process convolution method (Higdon , 1998; Higdon *et al.* , 1999), piecewise Gaussian processes (Kim *et al.* , 2005), kernel-based methods (Fuentes , 2001, 2002a,b; Nott and Dunsmuir , 2002), basis function models (Nychka *et al.* , 2002) and processes based on stochastic partial differential equations with spatially varying coefficients (Lindgren *et al.* , 2011). In all of these methods, the underlying spatial process is assumed to be Gaussian or transformed Gaussian, which may be unrealistic in practice. Specifically, in the presence of spatial heterogeneity in physical phenomena, the data usually exhibit skewness. Various approaches have been provided in the literature to handle departures from Gaussianity. A common method is to search for the most suitable transformation for the observed data in a family of transformations. De Oliveira *et al.* (1997) proposed a transformed Gaussian spatial model with a transformation from the Box-Cox family. Recently, Xu and Genton (2017) proposed the Tukey g-and-h stochastic process, a flexible class of transformed Gaussian processes. Interestingly, these processes accommodate different levels of skewness and tail heaviness in marginal distributions. Zareifard *et al.* (2018) proposed a convolution of log-Gaussian and Gaussian processes to tackle skewness. Another possible alternative to build skewed spatial processes is based on employing the skew-normal distributions that not only take skewness into account but also share several similar properties with the normal distribution. In this direction, two recently developed approaches have shown great appeal. The first approach seeks a multivariate skew-normal distribution for the finite dimensional distributions of the underlying stationary spatial process. Kim and Mallick (2004) constructed a skew-Gaussian random field using the multivariate skew- normal distribution in Azzalini and Dalla Valle (1996). But, Genton and Zhang (2012) showed some identifiability issues with this model. Since the multivariate skew-normal distribution lacks the closure properties, Allard and Naveau (2007) introduced a skew-Gaussian process based on the multivariate closed skew-normal (CSN) distribution. Rimstad and Omre (2014) provide an extension of the model proposed by Allard and Naveau (2007). Zareifard and Khaledi (2013) applied the multivariate unified skew-normal distribution (Arellano–Valle and Azzalini , 2006) which unifies several plethora of skew-normal models to introduce a skew-Gaussian process. This model extended to the multivariate case by Rivaz (2016). The second approach is based on exploiting the stochastic representation of the skew-normal distribution to create a class of stationary processes that have skewed marginal distributions (Zhang and El-Shaarawi , 2010). All of the aforementioned works are in the spatial context. To the best of our knowledge, there is only one article that introduces skewness into spatio-temporal models. Strictly speaking, Schmidt *et al.* (2017) presented an extension of the spatial skewed model proposed by Zhang and El-Shaarawi (2010) for the spatio-temporal data. However, several questions with regard to this model were posed by Genton and Hering (2017). Besides, the finite-dimensional distributions of the process do not belong to any of the commonly considered families of the multivariate skew-normal distributions.

In this paper, we develop a new approach to construct a non-stationary skew-Gaussian spatio-temporal process. More specifically, we propose to use the closed skew-normal distribution for the random coefficients in a low rank (LR) representation of the spatio-temporal processes. The proposed model includes several widely used classes of spatio-temporal models, for instance, the spatio-temporal random effects model (Cressie *et al.* , 2010) and Gaussian predictive processes (Finley *et al.* , 2012) as special cases. In addition, it not only induces flexible correlation structures but it also reduces the dimension of the spatio-temporal latent variables.

The outline of the article is as follows. In Section 2, we briefly review the definition of the closed skew-normal distribution and some of its properties. In Section 3, we present our skewed extension of the Gaussian LR model with its properties. Details on Bayesian implementation of the model are illustrated in Section 4, while in Sections 5 and 6 we show the efficiency of our model by using simulated data as well as a real data set. Finally, we conclude in Section 7 with some discussions.

2 A Brief Review of the CSN Distribution

We recall the definition and a few key properties of the closed skew-normal distribution, given by Dominguez-Molina *et al.* (2003) and Gonzalez–Farias *et al.* (2004). The multivariate CSN distribution is an extension of the normal density which admits skewness to increase the applicability of it. The CSN distribution inherits many interesting properties of the normal one including being closed under marginalization and conditioning.

A random vector $\mathbf{x} = (x_1, ..., x_n)'$ has a CSN distribution (hereafter it is denoted by $CSN_{n,m}(x; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Gamma}, \boldsymbol{\nu}, \boldsymbol{\Delta})$) if it has the distribution

$$[\Phi_m(0; \boldsymbol{\nu}, \boldsymbol{\Delta} + \boldsymbol{\Gamma\Sigma\Gamma}')]^{-1}\Phi_m(\boldsymbol{\Gamma}(\mathbf{x} - \boldsymbol{\mu}); \boldsymbol{\nu}, \boldsymbol{\Delta})\phi_{\mathbf{n}}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{2.1}$$

where $\boldsymbol{\mu} \in \Re^n$, $\boldsymbol{\Gamma} \in \Re^{m\times n}$, $\boldsymbol{\nu} \in \Re^m$ and $\boldsymbol{\Sigma} \in \Re^{n\times n}$ and $\boldsymbol{\Delta} \in \Re^{m\times m}$ are both covariance matrices. Moreover, $\phi_m(x, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\Phi_m(x, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ are PDF and the cumulative distribution function (CDF) of an m-dimensional normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, respectively. The integer value m defines the skewness dimension. If $\boldsymbol{\Gamma} = 0$, the CSN distribution reduces to an n- variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Additionally, if $m = 1$, the CSN density clearly reduces the proposed skew-normal distribution by Azzalini (2005).

The mean vector and covariance matrix of a CSN distribution are as follows

$$E(\mathbf{X}) = \boldsymbol{\mu} + \boldsymbol{\Sigma\Gamma}'\boldsymbol{\psi}, \tag{2.2}$$

$$Cov(\mathbf{X}) = \boldsymbol{\Sigma} + \boldsymbol{\Sigma\Gamma}'\boldsymbol{\xi\Gamma\Sigma} - \boldsymbol{\Sigma\Gamma}'\boldsymbol{\psi\psi}'\boldsymbol{\Gamma\Sigma}, \tag{2.3}$$

where $\boldsymbol{\psi} = \left.\frac{\left[\nabla_s \boldsymbol{\Phi}_m(\mathbf{s}; \boldsymbol{\nu}, \boldsymbol{\Delta} + \boldsymbol{\Gamma\Sigma\Gamma}')\right]'}{\boldsymbol{\Phi}_m(0; \boldsymbol{\nu}, \boldsymbol{\Delta} + \boldsymbol{\Gamma\Sigma\Gamma}')}\right|_{\mathbf{s}=\mathbf{0}}$ and $\boldsymbol{\xi} = \left.\frac{\nabla_s \nabla'_{\mathbf{s}} \boldsymbol{\Phi}_m(\mathbf{s}; \boldsymbol{\nu}, \boldsymbol{\Delta} + \boldsymbol{\Gamma\Sigma\Gamma}')}{\boldsymbol{\Phi}_m(0; \boldsymbol{\nu}, \boldsymbol{\Delta} + \boldsymbol{\Gamma\Sigma\Gamma}')}\right|_{\mathbf{s}=\mathbf{0}}$ with $\nabla_s = \left(\frac{\partial}{\partial s_1}, ..., \frac{\partial}{\partial s_m}\right)$.

Let $\mathbf{X}$ be distributed as $CSN_{n,m}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Gamma}, \boldsymbol{\nu}, \boldsymbol{\Delta})$. The stochastic representation of $\mathbf{X}$ is

$$\mathbf{X} \overset{d}{=} \boldsymbol{\mu} + (\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Gamma}'\boldsymbol{\Delta}^{-1}\boldsymbol{\Gamma})^{-\frac{1}{2}}\mathbf{V} + \boldsymbol{\Sigma\Gamma}'(\boldsymbol{\Delta} + \boldsymbol{\Gamma\Sigma\Gamma}')^{-1}\mathbf{U}, \tag{2.4}$$

where $\mathbf{V}$ and $\mathbf{U}$ are independent random vectors following $N_n(\mathbf{0}, \mathbf{I})$ and $TN_m(\mathbf{0}; \mathbf{0}, \boldsymbol{\Delta} + \boldsymbol{\Gamma\Sigma\Gamma}')$ distributions, respectively. It must be noted that $TN_m(\mathbf{c}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the normal distribution $N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ truncated below at a point $\mathbf{c}$. In the following, we review several important properties of the CSN distribution that are used in this paper.

Proposition 2.1 *Let $\boldsymbol{X}$ be a random vector with a closed skew-normal distribution $CSN_{n,m}(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{D},\boldsymbol{\nu},\boldsymbol{\Delta})$ and $\boldsymbol{A}$ be a $q\times n$ full rank matrix, then*

$$\boldsymbol{AX}\sim CSN_{q,m}(\boldsymbol{\mu}_{\boldsymbol{A}},\boldsymbol{\Sigma}_{\boldsymbol{A}},\boldsymbol{D}_{\boldsymbol{A}},\boldsymbol{\nu},\boldsymbol{\Delta}_{\boldsymbol{A}}), \tag{2.5}$$

where

$$\boldsymbol{\mu}_{\boldsymbol{A}}=\boldsymbol{A\mu},\quad \boldsymbol{\Sigma}_{\boldsymbol{A}}=\boldsymbol{A\Sigma A}',\quad \boldsymbol{D}_{\boldsymbol{A}}=D\boldsymbol{\Sigma A}'\boldsymbol{\Sigma}_{\boldsymbol{A}}^{-1},$$

and $\boldsymbol{\Delta}_{\boldsymbol{A}}=\boldsymbol{\Delta}+\boldsymbol{D\Sigma D}'-\boldsymbol{D\Sigma A}'\boldsymbol{\Sigma}_{\boldsymbol{A}}^{-1}\boldsymbol{A\Sigma D}'$.

Proposition 2.2 *Let $\boldsymbol{X}$ be a random vector with a closed skew-normal distribution $CSN_{n,m}(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{D},\boldsymbol{\nu},\boldsymbol{\Delta})$ and $\boldsymbol{Z}$ be an n-dimensional normal random vector with mean $\boldsymbol{\mu}_z$ and covariance $\boldsymbol{\Sigma}_z$, and independent of $\boldsymbol{X}$. Then $\boldsymbol{X}+\boldsymbol{Z}$ follows from a closed skew-normal distribution*

$$CSN_{n,m}(\boldsymbol{\mu}_{\boldsymbol{X+Z}},\boldsymbol{\Sigma}_{\boldsymbol{X+Z}},\boldsymbol{D}_{\boldsymbol{X+Z}},\boldsymbol{\upsilon}_{\boldsymbol{X+Z}},\boldsymbol{\Delta}_{\boldsymbol{X+Z}}), \tag{2.6}$$

with $\boldsymbol{\mu}_{\boldsymbol{X+Z}}=\boldsymbol{\mu}+\boldsymbol{\mu}_z$, $\boldsymbol{\Sigma}_{\boldsymbol{X+Z}}=\boldsymbol{\Sigma}+\boldsymbol{\Sigma}_z$, $\boldsymbol{D}_{\boldsymbol{X+Z}}=\boldsymbol{D\Sigma}(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}_z)^{-1}$, $\boldsymbol{\upsilon}_{\boldsymbol{X+Z}}=\boldsymbol{\upsilon}$ *and* $\boldsymbol{\Delta}_{\boldsymbol{X+Z}}=\boldsymbol{\Delta}+(\boldsymbol{D}-\boldsymbol{D}_{\boldsymbol{X+Z}})\boldsymbol{\Sigma D}'$.

Proposition 2.3 *Let $\boldsymbol{X}$ be a random vector with a closed skew-normal distribution $CSN_{n,m}(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{D},\boldsymbol{\nu},\boldsymbol{\Delta})$ and assume that it is partitioned into two components, $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, of dimensions k and $n-k$, respectively, with corresponding partition for $\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{D}$ and $\boldsymbol{\nu}$. Then the conditional distribution of $\boldsymbol{X}_2$ given $\boldsymbol{X}_1=\boldsymbol{x}_1$ is*

$$CSN_{n-k,m}(\boldsymbol{\mu}_2+\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(x_1-\boldsymbol{\mu}_1),\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12},\boldsymbol{D}_2,\boldsymbol{\nu}-\boldsymbol{D}_1x_1,\boldsymbol{\Delta}).$$

3 Proposed Model

Let's assume that the spatio-temporal stochastic process $Y(.,.)=\{Y(\mathbf{s},t);\mathbf{s}\in D\subseteq\mathbb{R}^d,t\in\{1,2,\cdots\}\}$ is observed at locations $\mathbf{s}_{it}$ at time t for $i=1,...,n_t$ and $t=1,2,...,T$. Also, suppose that the sampling model is

$$Y(\mathbf{s}_{it},t)=\mu(\mathbf{s}_{it},t)+\nu(\mathbf{s}_{it},t)+\varepsilon(\mathbf{s}_{it},t), \tag{3.7}$$

where $\mu(\mathbf{s}_{it},t)$ models the large-scale variability and is assumed to be a linear function of p regressors $\mathbf{f}_t'(.)=(f_{t,1}(.),...,f_{t,p}(.))$ with the unknown vector of coefficients $\boldsymbol{\beta}_t\in\Re^p$.

Further, the random effect $\varepsilon(\mathbf{s}_{it}, t)$ is the pure error process with distribution $N(0, \sigma_\varepsilon^2)$ which is considered to be independent of the $\nu(\mathbf{s}_{it}, t)$.

At any fixed time t, we assume an LR structure for the random effect $\nu(\mathbf{s}_{it}, t)$ as

$$\nu(\mathbf{s}_{it}, t) = \mathbf{b}_t^{'}(\mathbf{s}_{it})\boldsymbol{\eta}_t, \tag{3.8}$$

where $\mathbf{b}_t^{'}(\mathbf{s}_{it}) = (b_{1,t}(\mathbf{s}_{it}), ..., b_{r_t,t}(\mathbf{s}_{it}))$ represents a set of r_t ($r_t \ll n_t$) known spatio-temporal basis functions. It is assumed that a first-order Markovian evolution for the random coefficient $\{\boldsymbol{\eta}_t : t = 1, 2, \cdots\}$ which is given as

$$\boldsymbol{\eta}_t = \mathbf{H}_t\boldsymbol{\eta}_{t-1} + \boldsymbol{\omega}_t, \tag{3.9}$$

where $\mathbf{H}_t$ is the $r_t \times r_t$ evolution matrix that measures the dynamic dependence of the process. Also, the innovation vector $\boldsymbol{\omega}_t$ which is independent of $\boldsymbol{\eta}_t$, has an r_t-dimensional zero mean normal distribution with the covariance matrix $\mathbf{W}_t$ that controls the magnitude of the change at time t. Now, we explain our idea for extending the spatio-temporal model (3.7), (3.8) and (3.9) for capturing the skewness. The equation (3.9) can be written in the recursive form

$$\boldsymbol{\eta}_t = (\prod_{k=1}^{t} \mathbf{H}_k)\boldsymbol{\eta}_0 + \sum_{j=1}^{t-1}(\prod_{k=j+1}^{t} \mathbf{H}_k)\boldsymbol{\omega}_j + \boldsymbol{\omega}_t. \tag{3.10}$$

As seen before, the random vector $\boldsymbol{\eta}_t$ is a linear combination of the initial state $\boldsymbol{\eta}_0$ and the error terms $\boldsymbol{\omega}_1, \cdots, \boldsymbol{\omega}_t$. Thus, in order to induce skewness in the equation (3.10), it is enough to consider a CSN distribution for the random vector $\boldsymbol{\eta}_0$. Consequently, with regard to the favorable properties of the CSN distribution, the distribution of the random vector $\boldsymbol{\eta}_t$, as well as the random variable $Y(\mathbf{s}_{it}, t)$ would be CSN. To be more specific, the distribution of the random vector $\boldsymbol{\eta}_0$ is considered as $CSN_{r,m_0}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \mathbf{D}_0, \boldsymbol{\nu}_0, \boldsymbol{\Delta}_0)$. Hereupon, based on propositions (2.1) and (2.2), it is clear that $\boldsymbol{\eta_t}$ has $CSN_{r,m_0}(\boldsymbol{\mu}_{\eta_t}, \boldsymbol{\Sigma}_{\eta_t}, \mathbf{D}_{\eta_t}, \boldsymbol{v}_{\eta_t}, \boldsymbol{\Delta}_{\eta_t})$ with

$$\begin{aligned}
\boldsymbol{\mu}_{\eta_t} &= \mathbf{H}_t\boldsymbol{\mu}_{\eta_{t-1}},\\
\boldsymbol{\Sigma}_{\eta_t} &= \mathbf{H}_t\boldsymbol{\Sigma}_{\eta_{t-1}}\mathbf{H}_t^{'} + \mathbf{W}_t,\\
\mathbf{D}_{\eta_t} &= \mathbf{D}_{\eta_{t-1}}\boldsymbol{\Sigma}_{\eta_{t-1}}\mathbf{H}_t^{'}\boldsymbol{\Sigma}_{\eta_t}^{-1},\\
\boldsymbol{v}_{\eta_t} &= \boldsymbol{v}_{\eta_{t-1}},\\
\boldsymbol{\Delta}_{\eta_t} &= \boldsymbol{\Delta}_{\eta_{t-1}} + \mathbf{D}_{\eta_{t-1}}\boldsymbol{\Sigma}_{\eta_{t-1}}\mathbf{D}_{\eta_{t-1}}^{'} - \mathbf{D}_{\eta_{t-1}}\boldsymbol{\Sigma}_{\eta_{t-1}}\mathbf{H}_t^{'}\boldsymbol{\Sigma}_{\eta_t}^{-1}H_t\boldsymbol{\Sigma}_{\eta_{t-1}}\mathbf{D}_{\eta_{t-1}}^{'}.
\end{aligned}$$

Therefore, the marginal distribution of $Y(\mathbf{s}_{it}, t)$ is $CSN_{1,m_0}(\boldsymbol{\mu}_{(\mathbf{s}_{it},t)}, \boldsymbol{\Sigma}_{(\mathbf{s}_{it},t)}, \mathbf{D}_{(\mathbf{s}_{it},t)}, \boldsymbol{\upsilon}_{(\mathbf{s}_{it},t)}, \boldsymbol{\Delta}_{(\mathbf{s}_{it},t)})$ with the parameters

$$\begin{aligned}
\boldsymbol{\mu}_{(\mathbf{s}_{it},t)} &= \mathbf{f}_t^{'}(s)\boldsymbol{\beta}_t + \mathbf{b}_t^{'}(\mathbf{s}_{it})\boldsymbol{\mu}_{\boldsymbol{\eta}_t},\\
\boldsymbol{\Sigma}_{(\mathbf{s}_{it},t)} &= \mathbf{b}_t^{'}(\mathbf{s}_{it})\boldsymbol{\Sigma}_{\boldsymbol{\eta}_t}\mathbf{b}_t(\mathbf{s}_{it}) + \sigma_\varepsilon^2,\\
\mathbf{D}_{(\mathbf{s}_{it},t)} &= \mathbf{D}_{\boldsymbol{\eta}_t}\boldsymbol{\Sigma}_{\boldsymbol{\eta}_t}\mathbf{b}_t(\mathbf{s}_{it})\boldsymbol{\Sigma}^{-1}_{(\mathbf{s}_{it},t)},\\
\boldsymbol{\upsilon}_{(\mathbf{s}_{it},t)} &= \boldsymbol{\upsilon}_{\boldsymbol{\eta}_t},\\
\boldsymbol{\Delta}_{(\mathbf{s}_{it},t)} &= \boldsymbol{\Delta}_{\boldsymbol{\eta}_t} + \mathbf{D}_{\boldsymbol{\eta}_t}\boldsymbol{\Sigma}_{\boldsymbol{\eta}_t}\mathbf{D}_{\boldsymbol{\eta}_t}^{'} - \mathbf{D}_{\boldsymbol{\eta}_t}\boldsymbol{\Sigma}_{\boldsymbol{\eta}_t}\mathbf{b}_t(\mathbf{s}_{it})\boldsymbol{\Sigma}^{-1}_{(\mathbf{s}_{it},t)}\mathbf{b}_t^{'}(\mathbf{s}_{it})\boldsymbol{\Sigma}_{\boldsymbol{\eta}_t}\mathbf{D}_{\boldsymbol{\eta}_t}^{'}.
\end{aligned}$$

It is worth mentioning that the Gaussianity can be recovered by $\mathbf{D}_0 = 0$. The hierarchical structure of the proposed model, according to the stochastic representation (2.4), is given as

$$\begin{aligned}
Y(\mathbf{s};t)|\boldsymbol{\beta}_t, \boldsymbol{\eta}_t, \sigma_\varepsilon^2 &\sim N(\mathbf{f}_t^{'}(\mathbf{s})\boldsymbol{\beta}_t + \mathbf{b}_t^{'}(\mathbf{s})\boldsymbol{\eta}_t, \sigma_\varepsilon^2); \quad \mathbf{s} = 1, \cdots, n_t, \quad t = 1, ..., T \qquad (3.11)\\
\boldsymbol{\eta}_t|\boldsymbol{\eta}_{t-1}, \mathbf{W}_t &\sim N_r(\mathbf{H}_t\boldsymbol{\eta}_{t-1}, \mathbf{W}_t),\\
\boldsymbol{\eta}_0|\mathbf{U}_0 &\sim N_r(\boldsymbol{\mu}_0 + \boldsymbol{\Sigma}_0\mathbf{D}_0^{'}(\boldsymbol{\Delta}_0 + \mathbf{D}_0\boldsymbol{\Sigma}_0\mathbf{D}_0^{'})^{-1}\mathbf{U}_0, (\boldsymbol{\Sigma}^{-1} + \mathbf{D}_0^{'}\boldsymbol{\Delta}_0\mathbf{D}_0)^{-1}),\\
\mathbf{U}_0 &\sim TN_{m_0}(0; 0, \boldsymbol{\Delta}_0 + \mathbf{D}_0\boldsymbol{\Sigma}_0\mathbf{D}_0^{'}).
\end{aligned}$$

It is worth noting that we can also consider the CSN distribution for the error terms $\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_t$. However, this method unnecessarily increases the complexity of the model. This issue may impact the identifiability of the model parameters. Besides, the skewness dimension will increase as time progresses (as stated before by Naveau *et al.* (2005) in the state-space models). Accordingly, we only consider a CSN distribution for the random effect $\boldsymbol{\eta}_0$. As a result, the skewness dimension is the same as $\boldsymbol{\eta}_0$ for all the time epochs. Furthermore, according to the proposed modeling approach, the mean and covariance functions of the random field $Y(\cdot, \cdot)$ have well-known structures similar to the Gaussian model as

$$\begin{aligned}
E(Y(\mathbf{s}_{it}, t)) &= E(\mu(\mathbf{s}_{it}, t) + \mathbf{b}_t^{'}(\mathbf{s}_{it})\boldsymbol{\eta}_t + \varepsilon(\mathbf{s}_{it}, t)) \qquad (3.12)\\
&= \mu(\mathbf{s}_{it}, t) + \mathbf{b}_t^{'}(\mathbf{s}_{it})E(\boldsymbol{\eta}_t)\\
&= \mu(\mathbf{s}_{it}, t) + \mathbf{b}_t^{'}(\mathbf{s}_{it})\prod_{k=1}^{t}\mathbf{H}_k E(\boldsymbol{\eta}_0),
\end{aligned}$$

and

$$\begin{aligned} Cov(Y(\mathbf{s}_{it}, t), Y(\mathbf{s}_{jt+k}, t+k)) &= \mathbf{b}'_t(\mathbf{s}_{it}) Cov(\boldsymbol{\eta}_t, \boldsymbol{\eta}_{t+k}) \mathbf{b}'_{t+k}(\mathbf{s}_{jt+k})) + \sigma^2_\varepsilon I_{\{k=0, i=j\}} \\ &= \mathbf{b}'_t(\mathbf{s}_{it}) Var(\boldsymbol{\eta}_0) \left(\prod_{l=0}^{k} \mathbf{H}'_{t+l} \right) \mathbf{b}'_{t+k}(\mathbf{s}_{jt+k})) + \sigma^2_\varepsilon I_{\{k=0, i=j\}}, \end{aligned} \tag{3.13}$$

where $E(\boldsymbol{\eta}_0)$ and $Var(\boldsymbol{\eta}_0)$ are determined from the distribution of the initial state $\boldsymbol{\eta}_0$. The expression (3.13) indicates that this modeling strategy can solve the non-stationarity problem in the second-order structure of the random filed.

In the following section, we deal with the Bayesian inference of the hierarchical model (3.11).

4 Bayesian Implementation

The specification of our Bayesian hierarchical model is completed by placing priors on the model parameters. In the absence of prior information, we adopt independent vague normals for the elements of the regression parameters, $\boldsymbol{\beta}_t$, $t = 1, ..., T$, and a vague inverse Gamma prior $IG(a_0, b_0)$ with mean $b_0/(a_0 - 1)$ for σ^2_ε. Additionally, we assume that $\mathbf{W}_t$ is time-invariant and assign an inverse Wishart prior, i.e. $\mathbf{W} \sim IW(\nu_0, \boldsymbol{\Psi})$ where ν_0 denotes the degrees of freedom and $\boldsymbol{\Psi}$ is a scale positive definite matrix. There are several approaches to model the $\mathbf{H}_t$'s. For example, it may be assumed that they are structurally known up to some unknown parameters that change over time (Xu *et al.* , 2015; Ghosh *et al.* , 2010). Here we assume that $\mathbf{H}_t$ is time-invariant and known. However, in the simulation study and applied example we assume that the matrix $\mathbf{H}$ has a known structure up to some unknown parameters.

Our posterior inference is based on a Gibbs MCMC algorithm incorporating the Kalman filter algorithm. Therefore, to carry out the Bayesian inference, we need to compute the full conditional distributions of parameters. For this purpose, we utilize data augmentation to incorporate $\boldsymbol{\eta}_{1:T} = (\boldsymbol{\eta}_1, ..., \boldsymbol{\eta}_T)$. Let $\Upsilon \equiv (\mathbf{Y}'_1, ..., \mathbf{Y}'_T)'$ be the vector of all observations with $\mathbf{Y}'_t = (Y(\mathbf{s}_{1,t}, t), ..., Y(\mathbf{s}_{n_t,t}, t))$. By defining

$$\boldsymbol{\Theta} = (\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_T, \sigma^2_\varepsilon, \mathbf{W}, \boldsymbol{\eta}_{1:T}),$$

the full conditional distributions for each t, $t = 1, 2, ..., T$, are

$$\boldsymbol{\beta}_t|\mathbf{Y}_t, \boldsymbol{\Theta}_{-\boldsymbol{\beta}_t} \sim N((\frac{1}{\sigma_\varepsilon^2}\mathbf{F}_t\mathbf{F}_t' + \boldsymbol{\Sigma}_{\boldsymbol{\beta}_t}^{-1})^{-1}\left(\frac{1}{\sigma_\varepsilon^2}\mathbf{F}_t'(\mathbf{Y}_t - \mathbf{B}_t\boldsymbol{\eta}_t) + \boldsymbol{\Sigma}_{\boldsymbol{\beta}_t}^{-1}\mu_{\boldsymbol{\beta}_t}\right), (\frac{1}{\sigma_\varepsilon^2}\mathbf{F}_t\mathbf{F}_t' + \boldsymbol{\Sigma}_{\boldsymbol{\beta}_t}^{-1})^{-1}),$$

$$\mathbf{W}|\mathbf{Y}_t, \boldsymbol{\Theta}_{-\mathbf{W}} \sim IW(\nu_0 + r + T + 1, \sum_{t=1}^{T}\left(\left(\boldsymbol{\eta}_{t,\phi_t} - \mathbf{H}_t\boldsymbol{\eta}_{t-1,\phi_t}\right)\left(\boldsymbol{\eta}_{t,\phi_t} - \mathbf{H}_t\boldsymbol{\eta}_{t-1,\phi_t}\right)' + \boldsymbol{\Psi}\right)),$$

$$\sigma_\varepsilon^2|\mathbf{Y}_t, \boldsymbol{\Theta}_{-\sigma_\varepsilon^2} \sim IG(a_0 + \sum_{t=1}^{T} n_t/2, b_0 + \frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{n_t}\left(Y(s_i, t) - \mathbf{f}_t'(s_i)\boldsymbol{\beta}_t - \mathbf{b}_t'(s)\boldsymbol{\eta}_t\right)^2),$$

where $\boldsymbol{\Theta}_{-\delta}$ denotes all model parameters except δ, $\mathbb{B}_t = (\mathbf{B}(\mathbf{s}_{1,t}), ..., \mathbf{B}(\mathbf{s}_{n_t,t}))'$ and $\mathbf{F}_t = (\mathbf{f}_t(\mathbf{s}_{1,t}), ..., \mathbf{f}_t(\mathbf{s}_{n_t,t}))'$.

In the following, the inference of $\boldsymbol{\eta}_t$ will be presented using the Kalman filtering method. Let

$$\boldsymbol{\eta}_{t-1}|\mathbb{Y}_{t-1}, \boldsymbol{\Theta}_{-\boldsymbol{\eta}_{t-1}} \sim CSN_{r,m_0}(\hat{\boldsymbol{\mu}}_{t-1}, \hat{\boldsymbol{\Sigma}}_{t-1}, \hat{\mathbf{D}}_{t-1}, \hat{\boldsymbol{\nu}}_{t-1}, \hat{\boldsymbol{\Delta}}_{t-1}) \tag{4.14}$$

where $\mathbb{Y}_t$ denotes the available data until time the point t. Such a distributional assumption is valid due to the distribution of the initial state. On observing $\mathbf{Y}_t$, the posterior of $\boldsymbol{\eta}_t$ will be

$$\boldsymbol{\eta}_t|\mathbb{Y}_t, \boldsymbol{\Theta}_{-\eta_t} \sim CSN_{r,m_0}(\hat{\boldsymbol{\mu}}_t, \hat{\boldsymbol{\Sigma}}_t, \hat{\mathbf{D}}_t, \hat{\boldsymbol{\nu}}_t, \hat{\boldsymbol{\Delta}}_t) \tag{4.15}$$

with

$$\begin{aligned}
\hat{\boldsymbol{\mu}}_t &= \mathbf{H}_t\hat{\boldsymbol{\mu}}_{t-1} + \tilde{\boldsymbol{\Sigma}}_t\mathbf{B}_t'\left(\sigma_\varepsilon^2\mathbf{I} + \mathbf{B}_t\tilde{\boldsymbol{\Sigma}}_t\mathbf{B}_t'\right)^{-1}\mathbf{e}_t,\\
\mathbf{e}_t &= \mathbf{Y}_t - \mathbf{F}_t\boldsymbol{\beta}_t - \mathbf{B}_t\mathbf{H}_t\hat{\boldsymbol{\mu}}_{t-1},\\
\tilde{\boldsymbol{\Sigma}}_t &= \mathbf{H}_t\hat{\boldsymbol{\Sigma}}_{t-1}\mathbf{H}_t' + \mathbf{W}_t,\\
\hat{\boldsymbol{\Sigma}}_t &= \tilde{\boldsymbol{\Sigma}}_t - \tilde{\boldsymbol{\Sigma}}_t\mathbf{B}_t'\left(\sigma_\varepsilon^2\mathbf{I} + \mathbf{B}_t\tilde{\boldsymbol{\Sigma}}_t\mathbf{B}_t'\right)^{-1}\mathbf{B}_t\tilde{\boldsymbol{\Sigma}}_t,\\
\hat{\mathbf{D}}_t &= \hat{\mathbf{D}}_{t-1} - \hat{\boldsymbol{\Sigma}}_{t-1}\mathbf{H}_t'\tilde{\boldsymbol{\Sigma}}_t^{-1},\\
\hat{\boldsymbol{\nu}}_t &= \hat{\boldsymbol{\nu}}_{t-1},\\
\hat{\boldsymbol{\Delta}}_t &= \hat{\boldsymbol{\Delta}}_{t-1} + \left(\hat{\mathbf{D}}_{t-1} - \hat{\mathbf{D}}_t\mathbf{H}_t\right)\hat{\boldsymbol{\Sigma}}_{t-1}\hat{\mathbf{D}}_{t-1}.
\end{aligned}$$

Details of full conditionals and Kalman filtering are presented in the Appendix.

5 Simulation Study

This section presents two simulation examples to evaluate the performance of our proposed non-stationary skew-Gaussian spatio-temporal (NS-SG) model. The first example is to assess the model identifiability whereas the second example is designed to evaluate the impact of misspecification of latent variables distribution on estimation. For this purpose, we generate different datasets from the model (3.11) with 100 spatial sampling points randomly taken on the square $[0, 10] \times [0, 10]$ for each time point t, $t = 1, \ldots, 10$. The basis functions we use are bisquare functions with two resolutions, one with 4 knots and the other one with 16 knots, given as

$$B_{j(l)}(\mathbf{s}) \equiv \begin{cases} \{1 - (\|\mathbf{s} - \mathbf{u}_{j(l)}\|/r_l)^2\}^2 & \|\mathbf{s} - \mathbf{u}_{j(l)}\| \le r_l \\ 0 & \textit{otherwise}, \end{cases} \tag{5.16}$$

where $\{u_{j(l)}\}_j$ are the center points of l_{th} resolution and $\|.\|$ denotes the Euclidean distance between points $\mathbf{s}$ and $\mathbf{u}_{j(l)}$. Additionally, the radius of a bisquare basis function of a particular resolution, r_l, is defined as 1.5 times the shortest distance between the center points of that resolution (Cressie and Johannesson , 2008). A homotopic CSN distribution (Allard and Navaue, 2007) is considered for $\boldsymbol{\eta}_0$ as $CSN(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \alpha\mathbf{I}_r, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ such that $\boldsymbol{\mu}_0$ is a zero vector and $\boldsymbol{\Sigma}_0 = \mathbf{I}_r$. It is worth mentioning that the distribution of $\boldsymbol{\eta}_0$ reduces to the Gaussian distribution as $\alpha \to 0$. Different values of α represent different behaviors of skewness. Additionally, in the data generation setup, we use a simple random walk to generate $\boldsymbol{\eta}_{1:T}$. For priors, we consider $\beta \sim N(0, 10^3)$, $\sigma^2_\varepsilon \sim IG(0.01, 0.01)$, $\mathbf{W} \sim IW(r, I_r)$, and for the skewness parameter α, we choose $N(0, 10^2)$. According to our sensitivity analysis, we see that estimations of parameters are robust in the face of moderate changes in the prior of α. Moreover, to construct a more flexible model, the evolution matrix $\mathbf{H}$ is considered as $\rho\mathbf{I}_r$, where $\mathbf{I}_r$ is an $r \times r$ identity matrix and ρ is an unknown autoregressive parameter with a uniform prior distribution on the interval $(-1, 1)$. For each dataset, the MCMC algorithm was run with a total number of 20,000 iterations based on the NS-SG model. The posterior inferences are based on the last 10,000 iterations. To reduce the correlation between samples after burn-in time, the lag value was taken to be 5.

Example 1

The main objective of this example is to demonstrate model identifiability. To this end, according to the above mechanism with $\mu(\mathbf{s}_{it}; t) = \beta = 2$ and $\sigma^2_\varepsilon = 0.25$, we generated 20 datasets for each value of α by assuming $\alpha \in \{0, 1, 2, 4\}$, representing different behaviors from very weak to strong skewness. To assess the NS-SG model estimation, we computed the bias and mean square error (MSE) which were calculated as

$$Bias(\hat{\theta}) = \bar{\hat{\theta}} - \theta_{True}, \qquad MSE(\hat{\theta}) = \frac{1}{20}\sum_{i=1}^{20}(\hat{\theta}_i - \theta_{True})^2$$

where $\hat{\theta}_i$ is the estimated parameter from the ith simulated dataset and $\bar{\hat{\theta}}$ is the average of $\hat{\theta}_i$ s. Table 1 illustrates the average posterior means, standard deviation (in parenthesis), bias and MSE. As it is evident, all the Bias, MSE and sd are small for different values of α, which clearly indicates that data allow for meaningful inference on the parameters.

Table 1: Posterior mean (standard deviation), Bias and MSE of model parameters under different values for α.

Parameter	True value	$\alpha = 0$			$\alpha = 1$		
		Mean (sd)	Bias	MSE	Mean (sd)	Bias	MSE
α		0.00 (0.06)	0.00	0.00	0.84 (0.17)	-0.16	0.5
β	2	1.89 (0.56)	-0.11	0.33	2.02 (0.45)	0.02	0.21
ρ	1	0.96 (0.07)	-0.04	0.01	1.00 (0.00)	0.00	0.00
σ_ε^2	0.25	0.25 (0.00)	0.00	0.00	0.25 (0.00)	0.00	0.00
Parameter	True value	$\alpha = 2$			$\alpha = 4$		
		Mean (sd)	Bias	MSE	Mean (sd)	Bias	MSE
α		1.8 (0.28)	-0.2	0.12	3.82(0.4)	-0.18	0.2
β	2	1.96 (0.26)	-0.04	0.07	2.02 (0.5)	0.02	0.3
ρ	1	0.99 (0.00)	0.01	0.00	0.97 (0.08)	-0.03	0.00
σ_ε^2	0.25	0.25 (0.00)	0.00	0.00	0.25 (0.00)	0.00	0.00

Example 2

Now we assess the impact of misspecification of latent variables distribution on estimation using several simulated datasets. For this purpose, we select α from $\{0, 1, 4\}$ and generate 20 datasets for each value of α from the previous mechanism with $\mu(\mathbf{s}_{it}, t) = \beta_0 + \beta_1 lat(\mathbf{s}_{it}) + \beta_2 long(\mathbf{s}_{it})$. In all the simulated data, we fix $\beta_0 = 2$, $\beta_1 = 1$ and $\beta_3 = -3$. Then similar to the previous example, we fit the NS-SG model to the data and compare its performance with the Gaussian LR (GLR) model under the same conditions. Table 2 illustrates the posterior mean, bias, MSE and DIC (deviance information criterion) of our model as well as the GLR. The results indicate that the amount of skewness plays a crucial role in estimation and goodness of fit. More precisely, when we go from $\alpha = 1$ to $\alpha = 4$, GLR tends to perform worse in terms of bias and MSE as well as DIC. It should be noted that there is no significant difference between GLR and NS-SG models in estimation of σ_ϵ^2 and ρ. In fact, under both models, bias and MSE of the variance of measurement error and parameter of propagation matrix are noticeably small. However, when $\alpha = 0$, the NS-SG model outperforms the GLR model in parameters estimation (based on bias, MSE and DIC), unexpectedly. According to the posterior mean of skewness parameter ($\alpha = -0.6$), we conclude that most of the simulated datasets have a left skewed distribution. Therefore, our proposed model has shown better performance in comparison to the GLR model. To summarize, the NS-SG model provides a more robust result in comparison to the GLR model.

Table 2: Posterior mean, bias, MSE and DIC of NS-SG and GLR models under different values for α.

Parameter	True value			$\alpha = 0$			
			NS-SG			GLR	
		Mean	Bias	MSE	Mean	Bias	MSE
α		-0.6	-0.60	2.65	-	-	-
β_0	2	2.79	0.79	1.30	5.87	3.87	17.53
β_1	1	0.90	-0.09	0.01	1.03	0.03	0.02
β_2	-3	-3.02	-0.02	0.00	-3.45	-0.45	0.21
ρ	1	0.96	-0.03	0.00	0.92	-0.08	0.00
σ_ε^2	0.25	0.25	0.00	0.00	0.26	0.01	0.00
DIC		**392.34**			1068.22		
Parameter	True value			$\alpha = 1$			
			NS-SG			GLR	
		Mean	Bias	MSE	Mean	Bias	MSE
α		0.81	-0.19	2.80	-	-	-
β_0	2	2.61	0.61	1.15	4.79	2.79	9.11
β_1	1	0.92	-0.08	0.01	1.15	0.15	0.03
β_2	-3	-2.97	0.03	0.00	-3.36	-0.36	0.16
ρ	1	0.97	-0.03	0.00	0.93	-0.07	0.00
σ_ε^2	0.25	0.25	0.00	0.00	0.26	0.01	0.00
DIC		**407.51**			1056.72		
Parameter	True value			$\alpha = 4$			
			NS-SG			GLR	
		Mean	Bias	MSE	Mean	Bias	MSE
α		3.20	-0.80	2.33	-	-	-
β_0	2	2.98	0.98	1.96	5.97	3.97	16.98
β_1	1	0.96	-0.04	0.01	1.08	0.08	0.01
β_2	-3	-3.03	-0.03	0.01	-3.43	-0.43	0.2
ρ	1	0.97	-0.03	0.00	0.92	-0.08	0.00
σ_ε^2	0.25	0.25	0.00	0.00	0.26	0.00	0.00
DIC		**386.55**			1068.56		

6 Application To Modeling of Precipitation Data

Drought due to adverse consequences on socio-economic activities and agricultural productions has become one of the most challenging environmental problems in Iran. Drought is the result of abnormally dry periods that last long enough to cause an imbalance in hydrologic processes (storage and consumption). Due to the geographical location of Iran and the synoptic systems that affect this region, it is clear that one of the important reasons of drought is a significant reduction in annual precipitation. Therefore, accurate knowledge of precipitation levels is a fundamental requirement for understanding and managing climate changes. In this section, we aim to assess the effectiveness of the proposed NS-SG model in prediction of precipitation. We also compare the results with those obtained from the Gaussian LR model on transformed data.

The locations of 92 meteorological stations are shown in Figure 3. As pointed out before, the dataset includes $T = 10$ time points and for each $t = 1, 2, \ldots, T$, $n_t = 92$. Table 3 summarizes the descriptive statistics of the data. As observed, the distributions at each time point showed fairly right-skewed distributions. Since the explanatory analysis of the data showed a significant linear relation between annual total precipitation and elevation, the mean function is assumed to be a first order linear function of elevation (calculated as the elevation (in meters) divided by 100).

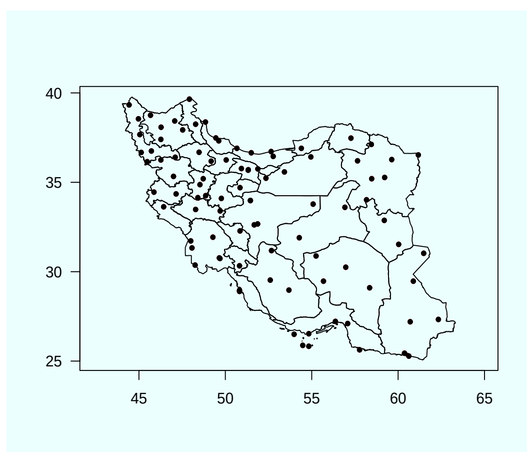

Figure 3: Location of meteorological stations.

Table 3: Descriptive statistics for precipitation data.

Year	Minimum	Maximum	Skewness coefficient
2006	24.83	1929.14	2.42
2007	25.31	1360.81	2.03
2008	14.76	1828.38	3.28
2009	45.75	1321.97	2.25
2010	6.87	1336.17	2.34
2011	31.44	2914.23	3.44
2012	24.02	1926.69	2.69
2013	23.56	1608.94	2.76
2014	6.00	1823.23	2.92
2015	0.00	1772.00	2.66

In the sequel, we need to construct a set of appropriate basis functions. Although in the simulation study, we focused on bisquare basis functions, another choice could be orthogonal basis functions. One of the most familiar orthogonal basis functions is empirical orthogonal functions (EOFs), which are the eigenvectors of the empirical covariance matrix (Cressie and Wikle, 2011). The use of EOFs as basis vectors leads to dramatic dimension reduction and therefore great simplicity in MCMC implementation. The first few EOFs are candidates for basis vectors. For our data, the first seven EOFs capture about 90% of the total variance. To construct a more flexible model, we consider $\mathbf{H} = diag(\rho_1, ..., \rho_r)$ where $\rho_1, ..., \rho_r$ are unknown autoregressive parameters.

In what follows, we evaluate the predictive performance of our NS-SG model in smoothing, filtering and forecasting with the Gaussian LR model. To assess the performance of our model in smoothing and filtering, we create 10 datasets by randomly taking 90% of the available data as training data and the rest as prediction locations. Then, two models will be compared based on the root of mean squared prediction error (RMSPE) and percentage prediction error (PPE) which were calculated as

$$RMSPE = \sqrt{\frac{1}{n_p T} \sum_{t=1}^{T} \sum_{i=1}^{n_{p_t}} \left(Y(\mathbf{s}_{i,t}, t) - \hat{Y}(\mathbf{s}_{i,t}, t)\right)^2},$$

$$PPE = \frac{1}{n_p T} \sum_{t=1}^{T} \sum_{i=1}^{n_{p_t}} 100 \times (Y(\mathbf{s}_{i,t}, t) - \hat{Y}(\mathbf{s}_{i,t}, t))/Y(\mathbf{s}_{i,t}, t),$$

where n_{p_t} is the number of prediction locations in time point t, $n_p = \sum_{t=1}^{T} n_{p_t}$, $\hat{Y}(\mathbf{s}_{i,t}, t)$ is the posterior mean of the predictive distribution of $Y(\mathbf{s}_{i,t}, t)$. It is worth mentioning that, we made a square root transform of the original data (to make them near normal) before fitting the GLR model.

we ran $50,000$ MCMC iterations for each model where the burn-in time was $20,000$ and the lag value was taken to be 10 in order to reduce the correlation between samples after burn-in time. The convergence of the MCMC was verified through the autocorrelations and visual inspection of the trace plots. Evidence reveals no obvious convergence problem. The hyperparameters were fixed to $a_0 = b_0 = 0.01$, $\nu_0 = r$, and $\mathbf{\Psi} = \mathbf{I}_r$. Additionally, we considered a homotopic $CSN_{r,r}(\boldsymbol{\mu}, \mathbf{\Sigma}, \alpha\mathbf{I}_r, \boldsymbol{\mu}, \mathbf{\Sigma})$ for $\boldsymbol{\eta}_0$ in the NS-SG model, where $\boldsymbol{\mu} = (0, ..., 0)^{'}$ and $\mathbf{\Sigma} = diag(100, r)$. Also, a vague normal prior was considered for α. In the case of GLR model, we assign a normal distribution $N_r(\mathbf{0}, 100\mathbf{I}_r)$ to the initial state vector. For the autoregressive parameters, we chose a uniform distribution on the interval $(-1, 1)$. Figures 4 shows the distributions of the RMSPE and PPE across the 10 sets of holdout samples. In general, the NS-SG model performs much better than the GLR model under both RMSPE and PPE criteria. The estimated values of autoregressive parameters are between 0.6 and 0.8. Moreover, the 95% credible interval for α is 0.32 ± 0.15 which confirms the right skewness of the precipitation data.

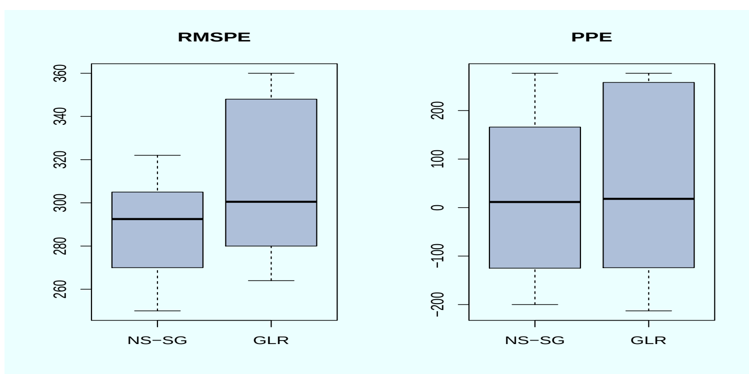

Figure 4: Box plots of the RMSPE and PPE for two models fit to the precipitation data, summarized for each of 10 holdout replicates.

To evaluate the NS-SG model in forecasting, we split the data into two parts: the first 9 years as training data and the last year as testing data. Similar to the above, the hyperparameters were chosen and the MCMC samples were obtained. Table 4 shows the prediction results for the two models on the hold out values. Based on two criteria, the NS-SG model outperforms the transformed GLR model. More precisely, relative RMSPE and PPE are 1.11 and 1.21 for GLR model compared with NS-SG model, respectively.

Table 4: Forecasting results for NS-SG and transformed GLR models.

	NS-SG	GLR
RMSPE	284.50	315.80
PPE	28.00%	33.88%

7 Discussion

In this paper, we proposed a non-stationary skew Gaussian spatio-temporal process based on using the closed skew-normal distribution in the low-rank representation of stochastic processes. Our model not only inherits favorable properties of the low-rank models but also greatly reduces the dimension of the spatio-temporal latent variables. Through simulation studies, we assess the performance of our model in parameters estimation. Also based on a real data example comprised of precipitation measurements, we demonstrated that our model outperforms the transformed Gaussian LR in filtering, smoothing and forecasting. Although, our example does not include many spatial locations, the proposed model can easily be applied to large datasets due to its low-rank structure.

Further extensions of our model can be explored. In many real problems, spatio-temporal datasets exhibit different features in time, such as multimodality and heavy tails as well as skewness. Although there are some remedies in the literature, an approach that addresses these issues in a simple computational way would be very valuable. Another extension of the proposed model could be in the case of areal data such as crime data that is a peak topic in spatio-temporal analysis.

Acknowledgements

The authors are grateful to the editor, associate editor and referees for constructive comments and suggestions which improved the article greatly.

References

Azzalini, A. (2005), The skewnormal distribution and related multivariate families. *Scandinavian Journal of Statistics*, **32**(2), 159–188.

Allard, D. and Naveau, P. (2007), A new spatial skew–normal random field model. *Communications in Statistics-Theory and Methods*, **36**(9), 1821–1834.

Arellano-Valle, R. B. and Azzalini, A. (2006), On the Unification of Families of Skewnormal Distributions. *Scandinavian Journal of Statistics*, **33**(3), 561–574.

Azzalini, A. and Dalla Valle, A. (1996), The multivariate skew–normal distribution. *Biometrika*, **83**(4), 715–726.

Bandyopadhyay, S. and Rao, S. S. (2017), A test for stationarity for irregularly spaced spatial data. J*ournal of the Royal Statistical Society: Series B (Statistical Methodology)*, **79**(1), 95–123.

Cressie, N. and Johannesson, G. (2008), Fixed Rank Kriging for Very Large Spatial Data Sets, *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **70** (1), 209–226 .

Cressie, N. and Wikle, C. K. (2011), *Statistics for spatio–temporal data*. Hoboken.

Cressie, N., Shi, T., and Kang, E. L. (2010), Fixed rank filtering for spatio–temporal data. *Journal of Computational and Graphical Statistics*, 19(3), 724–745.

De Oliveira, V., Kedem, B., and Short, D. A. (1997), Bayesian prediction of transformed Gaussian random fields. *Journal of the American Statistical Association*, **92**(440), 1422–1433.

Dominguez-Molina, J., Gonzalez–Farias, G., and Gupta, A. K. (2003), The multivariate closed skew normal distribution.*Technical Report*, 3–12.

Finley, A. O., Banerjee, S., and Gelfand, A. E. (2012), Bayesian dynamic modeling for large space–time datasets using Gaussian predictive processes. *Journal of geographical systems*, 14(1), 29–47.

Fuentes, M. (2001), A high frequency kriging approach for nonstationary environmental processes. *Environmetrics*, **12**(5), 469–483.

Fuentes, M. (2002a), Interpolation of nonstationary air pollution processes: a spatial spectral approach. *Statistical Modelling*, **2**(4), 281–298.

Fuentes, M. (2002b), Spectral methods for nonstationary spatial processes. *Biometrika*, **89**(1), 197–210.

Genton, M. G. and Hering, A. S. (2017), Comments on: Spatiotemporal models for skewed processes. *Environmetrics*, **28**(6).

Genton, M. G. and Zhang, H. (2012), Identifiability problems in some non–Gaussian spatial random fields. *Chilean Journal of Statistics*, **3**(2), 171–179.

Ghosh, S. K., Bhave, P. V., Davis, J. M., and Lee, H. (2010), Spatio–temporal analysis of total nitrate concentrations using dynamic statistical models. *Journal of the American Statistical Association*, **105**(490), 538–551.

Gonzalez-Farias, G., Dominguez–Molina, A., and Gupta, A. K. (2004), Additive properties of skew normal random vectors. *Journal of Statistical Planning and Inference*, **126** (2), 521–534.

Higdon, D. (1998), A process–convolution approach to modelling temperatures in the North Atlantic Ocean. *Environmental and Ecological Statistics*, **5**(2), 173–190.

Higdon, D., Swall, J., and Kern, J. (1999), Non–stationary spatial modeling. *Bayesian statistics*, **6**(1), 761–768.

Kim, H. M. and Mallick, B. K. (2004), A Bayesian prediction using the skew Gaussian distribution. *Journal of Statistical Planning and Inference*, **120**(1-2), 85–101.

Kim, H. M., Mallick, B. K., and Holmes, C. C. (2005), Analyzing nonstationary spatial data using piecewise Gaussian processes. *Journal of the American Statistical Association*, **100**(470), 653–668.

Lindgren, F., Rue, H., and Lindstrom, J. (2011), An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **73**(4), 423–498.

Naveau, P., Genton, M. G., and Shen, X. (2005), A skewed Kalman filter. *Journal of multivariate Analysis*, **94**(2), 382–400.

Nott, D. J. and Dunsmuir, W. T. (2002), Estimation of nonstationary spatial covariance structure. *Biometrika*, **89**(4), 819–829.

Nychka, D., Wikle, C., and Royle, J. A. (2002), Multiresolution models for nonstationary spatial covariance functions. *Statistical Modelling*, **2**(4), 315–331.

Rimstad, K. and Omre, H. (2014), Skew–Gaussian random fields.*Spatial Statistics* , 10, 43–62.

Rivaz, F. (2016), Optimal network design for Bayesian spatial prediction of multivariate non-Gaussian environmental data. *Journal of Applied Statistics*, **43** (7), 1335–1348.

Sampson, P. D. and Guttorp, P. (1992), Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, **87**(417), 108–119.

Schmidt, A. M., Goncalves, K., and Velozo, P. L. (2017), Spatiotemporal models for skewed processes. *Environmetrics*, **28**(6).

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

Xu, G., Liang, F., and Genton, M. G. (2015), A Bayesian spatio–temporal geostatistical model with an auxiliary lattice for large datasets. *Statistica Sinica*, 61–79.

Zareifard, H., Khaledi, M. J., Rivaz, F., and Vahidi-Asl, M. Q. (2018), Modeling Skewed Spatial Data Using a Convolution of Gaussian and Log–Gaussian Processes. *Bayesian Analysis*, **13**(2), 531–557.

Zareifard, H. and Khaledi, M. J. (2013), Non–Gaussian modeling of spatial data using scale mixing of a unified skew Gaussian process. *Journal of Multivariate Analysis*, 114, 16–28.

Zhang, H. and ElShaarawi, A. (2010), On spatial skew-Gaussian processes and applications. *Environmetrics*, **21**(1), 33–47.

Appendix

This section includes the details of the posterior inference.

$$\begin{aligned}
\pi(\boldsymbol{\beta}_t|\mathbf{Y}_t,\boldsymbol{\Theta}_{-\boldsymbol{\beta}_t}) &\propto \pi(\mathbf{Y}_t|\boldsymbol{\beta}_t,\boldsymbol{\Theta}_{-\boldsymbol{\beta}_t})\pi(\boldsymbol{\beta}_t)\\
&\propto \exp\left(-\frac{1}{2\sigma_\varepsilon^2}\left(\mathbf{Y}_t-\mathbf{F}_t\boldsymbol{\beta}_t-\mathbb{B}_t\boldsymbol{\eta}_t\right)'\left(\mathbf{Y}_t-\mathbf{F}_t\boldsymbol{\beta}_t-\mathbb{B}_t\boldsymbol{\eta}_t\right)\right)\\
&\times \exp\left(-\frac{1}{2}\left(\boldsymbol{\beta}_t-\boldsymbol{\mu}_{\boldsymbol{\beta}_t}\right)'\boldsymbol{\Sigma}_{\boldsymbol{\beta}_t}^{-1}\left(\boldsymbol{\beta}_t-\boldsymbol{\mu}_{\boldsymbol{\beta}_t}\right)\right)\\
&\propto \exp\left(-\frac{1}{2}\left(\boldsymbol{\beta}_t'\left(\frac{\mathbf{F}_t'\mathbf{F}_t}{\sigma_\varepsilon^2}+\boldsymbol{\Sigma}_{\boldsymbol{\beta}_t}^{-1}\right)\boldsymbol{\beta}_t-\boldsymbol{\beta}_t'\left(\mathbf{F}_t'\left(\mathbf{Y}_t-\mathbb{B}_t\boldsymbol{\eta}_t\right)+\boldsymbol{\Sigma}_{\boldsymbol{\beta}_t}^{-1}\boldsymbol{\mu}_{\boldsymbol{\beta}_t}\right)\right)\right).
\end{aligned}$$

$$\boldsymbol{\beta}_t|\mathbf{Y}_t,\boldsymbol{\Theta}_{-\boldsymbol{\beta}_t}\sim N\left(\left(\frac{\mathbf{F}_t'\mathbf{F}_t}{\sigma_\varepsilon^2}+\boldsymbol{\Sigma}_{\boldsymbol{\beta}_t}^{-1}\right)^{-1}\left(\mathbf{F}_t'(\mathbf{Y}_t-\mathbb{B}_t\eta_t)+\boldsymbol{\Sigma}_{\boldsymbol{\beta}_t}^{-1}\boldsymbol{\beta}_0\right),\left(\frac{\mathbf{F}_t'\mathbf{F}_t}{\sigma_\varepsilon^2}+\boldsymbol{\Sigma}_{\boldsymbol{\beta}_t}^{-1}\right)^{-1}\right).$$

$$\begin{aligned}
\pi(\sigma_\varepsilon^2|\mathbb{Y},\boldsymbol{\Theta}_{-\sigma_\varepsilon^2}) &\propto f(\mathbb{Y}|\boldsymbol{\Theta})\pi(\sigma_\varepsilon^2)\\
&\propto \frac{1}{(\sigma_\varepsilon^2)^{\frac{nT}{2}}}\exp\left(-\frac{1}{2\sigma_\varepsilon^2}\sum_{t=1}^{T}\left(\mathbf{Y}_t-\mathbf{F}_t\boldsymbol{\beta}_t-\mathbb{B}_t'\eta_t\right)'\left(\mathbf{Y}_t-\mathbf{F}_t\boldsymbol{\beta}_t-\mathbb{B}_t'\eta_t\right)\right)\\
&\times \frac{1}{(\sigma_\varepsilon^2)^{a_0+1}}\exp\left(-\frac{b_0}{\sigma_\varepsilon^2}\right)\\
&\propto \frac{1}{(\sigma_\varepsilon^2)^{\frac{nT}{2}+a_0+1}}\exp\left(-\frac{1}{\sigma_\varepsilon^2}\left(\frac{1}{2}\sum_{t=1}^{T}\left(\mathbf{Y}_t-\mathbf{F}_t\boldsymbol{\beta}_t-\mathbb{B}_t'\eta_t\right)'\left(\mathbb{Y}_t-\mathbf{F}_t\boldsymbol{\beta}_t-\mathbb{B}_t'\eta_t\right)+b_0\right)\right).
\end{aligned}$$

Therefore,

$$\sigma_\varepsilon^2|\mathbb{Y},\boldsymbol{\Theta}_{-\sigma_\varepsilon^2}\sim IG(a_0+\frac{nT}{2},\frac{1}{2}\sum_{t=1}^{T}\left(\mathbf{Y}_t-\mathbf{F}_t\beta_t-\mathbb{B}_t'\eta_t\right)'\left(\mathbf{Y}_t-\mathbf{F}_t\boldsymbol{\beta}_t-\mathbb{B}_t'\eta_t\right)+b_0).$$

$$\begin{aligned}
\pi(\mathbf{W}|\mathbb{Y},\boldsymbol{\Theta}) &\propto \prod_{t=1}^{T} \pi(\boldsymbol{\eta}_t|\boldsymbol{\eta}_{t-1},\mathbf{W})\pi(\mathbf{W}) \\
&\propto \prod_{t=1}^{T} |\mathbf{W}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\boldsymbol{\eta}_t - \mathbf{H}_t\boldsymbol{\eta}_{t-1}\right)^{'} \mathbf{W}^{-1}\left(\boldsymbol{\eta}_t - \mathbf{H}_t\boldsymbol{\eta}_{t-1}\right)\right) \\
&\times |\boldsymbol{\Psi}|^{\frac{\nu_0}{2}} |\mathbf{W}|^{-\frac{\nu_0+r+1}{2}} \exp\left(-\frac{1}{2}tr\left(\boldsymbol{\Psi}\mathbf{W}^{-1}\right)\right) \\
&\propto \prod_{t=1}^{T} |\mathbf{W}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}tr\left[\left(\boldsymbol{\eta}_t - \mathbf{H}_t\boldsymbol{\eta}_{t-1}\right)^{'} \mathbf{W}^{-1}\left(\boldsymbol{\eta}_t - \mathbf{H}_t\boldsymbol{\eta}_{t-1}\right)\right]\right) \\
&\times |\boldsymbol{\Psi}|^{\frac{\nu_0}{2}} |\mathbf{W}|^{-\frac{\nu_0+r+1}{2}} \exp\left(-\frac{1}{2}tr\left(\boldsymbol{\Psi}\mathbf{W}^{-1}\right)\right) \\
&\propto |\mathbf{W}|^{-\frac{\nu_0+r+T+2}{2}} \exp\left(-\frac{1}{2}tr\sum_{t=1}^{T}\left[\left(\boldsymbol{\eta}_t - \mathbf{H}_t\boldsymbol{\eta}_{t-1}\right)\left(\boldsymbol{\eta}_t - \mathbf{H}_t\boldsymbol{\eta}_{t-1}\right)^{'} \mathbf{W}^{-1} + \boldsymbol{\Psi}\mathbf{W}^{-1}\right]\right) \\
&\propto |\mathbf{W}|^{-\frac{\nu_0+r+T+2}{2}} \exp\left(-\frac{1}{2}tr\left[\sum_{t=1}^{T}\left(\left(\boldsymbol{\eta}_{t,\phi_t} - \mathbf{H}_t\boldsymbol{\eta}_{t-1,\phi_t}\right)\left(\boldsymbol{\eta}_{t,\phi_t} - \mathbf{H}_t\boldsymbol{\eta}_{t-1,\phi_t}\right)^{'} + \boldsymbol{\Psi}\right)\mathbf{W}_t^{-1}\right]\right).
\end{aligned}$$

In what follows, we present the proof of Kalman filter equations. Consider (3.9), we have

$$\begin{aligned}
\left[\boldsymbol{\eta}_t|\mathbb{Y}_{t-1},\boldsymbol{\Theta}_{-\boldsymbol{\eta}_t}\right] &= \left[\mathbf{H}_t\boldsymbol{\eta}_{t-1} + \boldsymbol{\omega}_t|\mathbb{Y}_{t-1},\boldsymbol{\Theta}_{-\boldsymbol{\eta}_t}\right] \\
&= \mathbf{H}_t\left[\boldsymbol{\eta}_{t-1}|\mathbb{Y}_{t-1},\boldsymbol{\Theta}_{-\boldsymbol{\eta}_t}\right] + \boldsymbol{\omega}_t.
\end{aligned}$$

Utilizing Propositions 2.1-2.2 and the above equation, we have

$$\left[\boldsymbol{\eta}_t|\mathbb{Y}_{t-1},\boldsymbol{\Theta}_{-\boldsymbol{\eta}_t}\right] = CSN_{r,m_0}(\mathbf{H}_t\hat{\boldsymbol{\mu}}_{t-1},\tilde{\boldsymbol{\Sigma}}_t,\hat{\mathbf{D}}_t,\hat{\boldsymbol{\nu}}_t,\hat{\boldsymbol{\Delta}}_t),$$

where $\tilde{\boldsymbol{\Sigma}}_t = \mathbf{H}_t\hat{\boldsymbol{\Sigma}}_{t-1}\mathbf{H}_t^{'} + \mathbf{W}_t$, $\hat{\mathbf{D}}_t = \hat{\mathbf{D}}_{t-1}\hat{\boldsymbol{\Sigma}}_{t-1}\mathbf{H}_t^{'}(\mathbf{H}_t\hat{\boldsymbol{\Sigma}}_{t-1}\mathbf{H}_t^{'} + \mathbf{W}_t)^{-1}$, $\hat{\boldsymbol{\nu}}_t = \hat{\boldsymbol{\nu}}_{t-1}$ and $\hat{\boldsymbol{\Delta}}_t = \hat{\boldsymbol{\Delta}}_t + (\hat{\boldsymbol{D}}_{t-1} - \hat{\boldsymbol{D}}_t\mathbf{H}_t)\hat{\boldsymbol{\Sigma}}_{t-1}\hat{\mathbf{D}}_{t-1}$.

On observing $\mathbf{Y}_t$, our objective is to compute the posterior of $\boldsymbol{\eta}_t$, i.e. $\left[\boldsymbol{\eta}_t|\mathbb{Y}_{t-1},\boldsymbol{\Theta}_{-\boldsymbol{\eta}_t}\right]$. To this end, we introduce $\mathbf{e}_t = \mathbf{Y}_t - \mathbf{B}_t[\mathbf{H}_t\hat{\boldsymbol{\mu}}_{t-1} + \boldsymbol{\mu}_{\omega_t}] - \boldsymbol{\mu}_\varepsilon$ as the error in predicting $\mathbf{Y}_t$

from point $t-1$. Following this definition and $\boldsymbol{\mu}_{\omega_t} = \boldsymbol{\mu}_{\varepsilon} = \mathbf{0}$, we have

$$\begin{aligned}[\mathbf{e}_t|\boldsymbol{\eta}_t, \mathbb{Y}_{t-1}, \boldsymbol{\Theta}_{-\eta_t}] &= [\mathbf{B}_t(\boldsymbol{\eta}_t - \mathbf{H}_t\hat{\boldsymbol{\mu}}_{t-1}) + \boldsymbol{\varepsilon}_t|\boldsymbol{\eta}_t, \mathbb{Y}_{t-1}, \boldsymbol{\Theta}_{-\eta_t}] \\ &= [\mathbf{B}_t(\boldsymbol{\eta}_t - \mathbf{H}_t\hat{\boldsymbol{\mu}}_{t-1})_t|\boldsymbol{\eta}_t, \mathbb{Y}_{t-1}, \boldsymbol{\Theta}_{-\eta_t}] + \boldsymbol{\varepsilon}_t.\end{aligned}$$

Then, the normality of $\boldsymbol{\varepsilon}_t$ imply that

$$[\mathbf{e}_t|\boldsymbol{\eta}_t, \mathbb{Y}_{t-1}, \boldsymbol{\Theta}_{-\eta_t}] = N_{n_t}(\mathbf{B}_t(\boldsymbol{\eta}_t - \mathbf{H}_t\hat{\boldsymbol{\mu}}_{t-1}), \sigma_\varepsilon^2\mathbf{I}_{n_t}).$$

To obtain the posterior of $\boldsymbol{\eta}_t$, we link the posterior of it with the

$$[\boldsymbol{\eta}_t|\mathbb{Y}_t, \boldsymbol{\Theta}_{-\eta_t}] = [\boldsymbol{\eta}_t|\mathbb{Y}_{t-1}, \mathbf{Y}_t, \boldsymbol{\Theta}_{-\eta_t}] = [\boldsymbol{\eta}_t|\mathbb{Y}_{t-1}, \mathbf{e}_t].$$

Then applying proposition 2.3 to the distribution of $[\mathbf{e}_t|\boldsymbol{\eta}_t, \mathbb{Y}_{t-1}, \boldsymbol{\Theta}_{-\eta_t}]$ and $[\boldsymbol{\eta}_t|\mathbb{Y}_{t-1}, \boldsymbol{\Theta}_{-\eta_t}, \boldsymbol{\Theta}_{-\eta_t}]$, the distribution of $[\boldsymbol{\eta}_t^T, \mathbf{e}_t^T|\mathbb{Y}_{t-1}, \boldsymbol{\Theta}_{-\eta_t}]^T$ is obtained as

$$CSN_{n+r,m_0}\left(\begin{pmatrix} \mathbf{H}_t\hat{\boldsymbol{\mu}}_{t-1} + \boldsymbol{\mu}_\eta \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \tilde{\boldsymbol{\Sigma}}_t & \tilde{\boldsymbol{\Sigma}}_t\mathbf{B}_t' \\ \mathbf{B}_t\tilde{\boldsymbol{\Sigma}}_t & \mathbf{B}_t\tilde{\boldsymbol{\Sigma}}_t\mathbf{B}_t' + \sigma_\varepsilon^2\mathbf{I} \end{pmatrix}, \begin{pmatrix} \hat{\mathbf{D}}_t \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \hat{\boldsymbol{\nu}}_t \\ \mathbf{0} \end{pmatrix}, \hat{\boldsymbol{\Delta}}_t\right).$$

Again, by employing proposition 2.3, we have

$$[\boldsymbol{\eta}_t|\mathbb{Y}_t, \boldsymbol{\Theta}_{-\eta_t}] = [\boldsymbol{\eta}_t|\mathbf{e}_t, \mathbb{Y}_{t-1}, \boldsymbol{\Theta}_{-\eta_t}] = CSN_{n,m}(\hat{\boldsymbol{\mu}}_t, \hat{\boldsymbol{\Sigma}}_t, \hat{\mathbf{D}}_t, \hat{\boldsymbol{\nu}}_t, \hat{\boldsymbol{\Delta}}_t).$$

with the given components of (4.15).