Spatio-temporal Methods in Environmental Epidemiology

Alexandra M. Schmidt

Session 12 - Spatio-Temporal Models

Department of Epidemiology, Biostatistics and Occupational Health
McGill University

alexandra.schmidt@mcgill.ca

Applied Bayesian Statistics School 2025

University of Genova, Department of Architecture and Design

03-06, June 2025

McGill

Introduction

- Environmental and ecological processes usually vary across space and time simultaneously
- Spatio-temporal observations are naturally accommodated in the DLM framework

Potential strategies

Some general approaches to incorporating time are as follows:

Approach 1: Treat continuous time as another spatial dimension, e.g. spatio–temporal kriging (Bodnar & Schmid, 2010). There is extra complexity in constructing covariance models compared to purely spatial process modelling (Fuentes, Chen, & Davis, 2008) and possible reductions in the complexity based on time having a natural ordering (unlike space) are not realised

Approach 2: Represent the spatial fields represented as vectors $\mathbf{Z}_t : N_S \times 1$, and combine them across time to get a multivariate time series

Approach 3: Represent the time series as vectors, $\mathbf{Z}_s : 1 \times N_T$, and use multivariate spatial methods e.g. co-kriging.

Approach 4: Build a statistical framework based on deterministic models that describe the evolution of processes over space and time.

Separable Models

- Let ρ be a valid spatial correlation function;
- Let Ω be a $T \times T$ positive definite matrix. Then,

$$C(s, s') = \rho(s, s')\Omega. \tag{1}$$

- Considering stacking the observed values obtained at n spatial locations into a vector, $\mathbf{Y}$, the resultant covariance matrix is given by

$$\Sigma = \mathbf{R} \otimes \Omega, \tag{2}$$

Advantages and Disadvantages of the separable model

- $|\Sigma| = |R|^p |\Omega|^n$ e $\Sigma^{-1} = R^{-1} \otimes \Omega^{-1}$
- $cov(Y_l(s), Y_{l'}(s')) = cov(Y_{l'}(s), Y_l(s')), \forall\, l, l', s$ e s'
- if ρ is symmetric and strictly decreasing, then the spatial range is the same for all the elements of $\mathbf{Y}(.)$
- In a more general setting we could include the structure above into a latent component, say, $\mathbf{v}(s)$, such that

$$\mathbf{Y}(s) = \mathbf{X}(s)\beta + \mathbf{v}(s) + \varepsilon(s), \tag{3}$$

McGill

A Bayesian hierarchical approach

- Le and Zidek (1992) derive a fully Bayesian alternative to kriging in which temporal and spatial modelling are done in a convenient way
- Let $\mathbf{Y}_j = (\mathbf{Y}_j^{(1)'}, \mathbf{Y}_j^{(2)'})$, $j = 1, 2, \cdots, T$, be a q-dimensional random vector where $\mathbf{Y}_j^{(1)}$ and $\mathbf{Y}_j^{(2)}$ are u and g-dimensional vectors, of unobserved (ungauged) and observed (gauged) locations $(q = u + g)$
- Let $\mathbf{f}_j$ be a r-dimensional vector of covariates

$$\begin{aligned} \mathbf{Y}_j \mid \mathbf{f}_j, \mathbf{B}, \Sigma &\sim N_q(\mathbf{B}\mathbf{f}_j, \Sigma) \\ \mathbf{B} \mid \mathbf{B}^0, \Sigma, \mathbf{F} &\sim N_{rq}(\mathbf{B}^0, \Sigma \otimes \mathbf{F}^{-1}) \qquad (4) \\ \Sigma \mid \Psi, m &\sim W^{-1}(\Psi, m) \qquad (5) \end{aligned}$$

- In particular, they consider the following partition of Σ, and using the Bartlett decomposition, they show that

$$\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} \Sigma_{1|2} + \tau\Sigma_{22}\tau' & \tau\Sigma_{22} \\ \Sigma_{22}\tau' & \Sigma_{22} \end{pmatrix}$$

A Bayesian hierarchical approach

- Instead of assigning a prior for Σ they propose a prior for $\Sigma_{1|2}$, τ, and Σ_{22}
- $\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ is the $u \times u$ matrix which represents the residual covariance of $\mathbf{Y}_j^{(1)}$-residuals after optimal linear prediction based on $\mathbf{Y}_j^{(2)}$
- τ is a $u \times g$ matrix representing the slope of the optimal linear predictor $\mathbf{Y}_j^{(1)}$ based on $\mathbf{Y}_j^{(2)}$
- $$\begin{aligned} \Sigma_{22} \mid \Psi, m &\sim W_g^{-1}(\Psi_{22}, m-u) \\ \Sigma_{1|2} \mid \Psi, m &\sim W_u^{-1}(\Psi_{1|2}, m) \\ \tau \mid \Sigma_{1|2}, \Psi &\sim N_{ug}(\eta, \Sigma_{1|2} \otimes \Psi_{22}^{-1}), \end{aligned}$$

  where $(\Psi_{22}, \Psi_{1|2}, \eta)$ denote the decomposition of the prior parameter Ψ, with $\eta = \Psi_{12}\Psi_{22}^{-1}$

McGill

A Bayesian hierarchical approach

- Let $D = ((\mathbf{y}_1^{(2)}, \mathbf{f}_1), \cdots, (\mathbf{y}_n^{(2)}, \mathbf{f}_n))$ be the observed data
- The predictive distributions of future observations at gauged locations $\mathbf{Y}_f^{(2)} \mid D$ and those at the ungauged locations conditioned on the gauged ones, $\mathbf{Y}_f^{(1)} \mid D, \mathbf{Y}_f^{(2)}$ follow multivariate Student-t distributions

Multivariate DLMs

- The multivariate extension of the dynamic linear models leads to the broad classes of multivariate and matrix variate Gaussian DLMs
- For a p-dimensional vector $\mathbf{y}_t = (y_{t1}, \ldots, y_{tp})'$, a multivariate normal DLM can be written as

$$\begin{aligned} \mathbf{y}_t &= \mathbf{F}_t'\theta_t + \varepsilon_t, & \varepsilon_t \sim N(0, \mathbf{V}_t), \qquad (6) \\ \theta_t &= \mathbf{G}_t\theta_{t-1} + \omega_t, & \omega_t \sim N(0, \mathbf{W}_t), \qquad (7) \end{aligned}$$

  where $\theta_0 \sim N(\mathbf{m}_0, \mathbf{C}_0)$.
- Assume that both sequences of observation and evolution covariances matrices $\mathbf{V}_t$ and $\mathbf{W}_t$ are known and of dimensions $(p \times p)$ and $(k \times k)$, respectively.
- Therefore, the model is completely specified by the quadruple $\{\mathbf{F}_t, \mathbf{G}_t, \mathbf{V}_t, \mathbf{W}_t\}$, where $\mathbf{F}_t$s are $(k \times p)$ matrices and $\mathbf{G}_t$s are $(k \times k)$ matrices.
- Analagous to the univariate case, for $t = 1, 2, \ldots,$

$$\begin{aligned} (\theta_t | D_{t-1}) &\sim N(\mathbf{a}_t, \mathbf{R}_t), \\ (\mathbf{y}_t | D_{t-1}) &\sim N(\mathbf{f}_t, \mathbf{Q}_t), \\ (\theta_t | D_t) &\sim N(\mathbf{m}_t, \mathbf{C}_t), \end{aligned}$$

DLMs - Geostatistical set up

- Let $\{y_t(\mathbf{s}), \mathbf{s} \in B \subset \mathbb{R}^d, t \in T\}$ be a spatial random field at *discrete* time t, and usually, $d = 1, 2$ or 3

- If a partial realization of the spatial random field is available at each time t, following (6) we assume a spatial process for each time t

- Let the observed data $y_t(\mathbf{s})$, be a noisy version of the process of interest, $Z_t(\mathbf{s})$, that is, assume

$$
\begin{aligned}
y_t(\mathbf{s}) &= z_t(\mathbf{s}) + v_t(\mathbf{s}) && (8)\\
z_t(\mathbf{s}) &= \mathbf{F}_t'\theta_t + \varepsilon_t^*(\mathbf{s}) && (9)\\
\theta_t &= \mathbf{G}_t\theta_{t-1} + \omega_t, && (10)
\end{aligned}
$$

- $\varepsilon_t^*(\cdot)$ is assumed to follow a zero mean Gaussian process with some covariance matrix Σ_t

- $v_t(\mathbf{s})$ represents measurement error assumed to follow an independent zero mean normal distribution with variance τ^2

- It is assumed further that $\varepsilon_t^*(\mathbf{s})$ are independent across time, and independent of $v_t(\mathbf{s})$ and ω_t, for all t and $\mathbf{s} \in D$

DLMs - Geostatistical set up

- If we substitute equation (9) into (8), we obtain $y_t(\mathbf{s}) = \mathbf{F}_t'\theta_t + \varepsilon_t^*(\mathbf{s}) + v_t(\mathbf{s})$
- If we stack the observations of the process at time t onto a vector $\mathbf{y}_t = (y_t(\mathbf{s}_1), y_t(\mathbf{s}_2), \ldots, y_t(\mathbf{s}_n))'$, and assume a similar structure for ε_t^* and $\mathbf{v}_t$, respectively, we can write $\mathbf{y}_t|\mathbf{F}_t, \theta_t, \mathbf{v}_t, \tau^2 \sim N(\mathbf{F}_t'\theta_t + \mathbf{v}_t, \tau^2\mathbf{I}_n)$
- We can marginalize the distribution of $\mathbf{y}_t$ with respect to $\mathbf{v}_t$
- As $\mathbf{v}_t$ follows a zero mean multivariate normal distribution, then $\mathbf{y}_t|\theta_t, \mathbf{F}_t, \tau^2, \Sigma_t \sim N(\mathbf{F}_t'\theta_t, \Sigma_t + \tau^2\mathbf{I}_n)$
- If spatio-temporal covariates are present in the columns of $\mathbf{F}_t$ the mean of the process is also varying across space for each time t
- If spatio-temporal covariates are not available, then the temporal trend is fixed across space, and ε_t captures, at each location, deviations from this overall temporal structure
- The main issue is the specification of the covariance matrix Σ_t

DLMs - Geostatistical set up

- The simplest alternative to account for spatial correlation at each time t is to model each element of the covariance matrix $\Sigma_t = \Sigma$ $\forall t$, that is, $\Sigma_{ij} = \sigma^2 \rho(d, \phi)$, where $d = ||\mathbf{s}_i - \mathbf{s}_j||$ is the Euclidean distance
  - The parameters in ϕ are typically scalars or low dimensional vectors. E.g.,
    - Exponential $\rho(d, \phi) = \exp\{-d/\phi\}$ or
    - spherical $\rho(d, \phi) = (1 - 1.5(d/\phi) + 0.5(d/\phi)^3)1_{\{d \leq \phi\}}$ or
    - Matérn $\rho(d, \phi) = 2^{1-\phi_2}\Gamma(\phi_2)^{-1}(d/\phi_1)^{\phi_2}\mathcal{K}_{\phi_2}(d/\phi_1)$,

DLMs - Geostatistical set up

- If it is believed that the variance of $\mathbf{v}_t$ changes with time, one alternative is to allow σ^2 to vary smoothly with time, for example, by making $\log \sigma_t^2 = \log \sigma_{t-1}^2 + \varepsilon_{1t}$, with ε_{1t} following a zero mean normal distribution, with possibly unknown but constant variance W_σ, and initial information $\log \sigma_0^2 \sim N(0, b)$, for some known constant b
- This will lead to a covariance matrix $\mathbf{V}_t$ that changes with time. When inference procedure is performed using MCMC methods, the algorithm must be adapted to include steps to sample from the posterior full conditional distributions of the elements of ϕ, and the variances $\sigma_0^2, \sigma_1^2, \ldots, \sigma_T^2$. This can be achieved, for example, through individual Metropolis-Hastings steps for each of these parameters.
- An alternative in the modelling of the covariance matrix is to assign an inverse Wishart prior distribution for Σ_t, which results on a nonstationary (free form) covariance matrix

Spatial interpolation and temporal prediction

- Usually, in spatio-temporal settings the aim is to perform spatial interpolation of the process for unmonitoring locations and temporal prediction
- Assume we want to predict the process for a vector of dimension k, of ungauged locations at observed time $t \in T$, say $\tilde{\mathbf{y}}_t = (y_t(\tilde{\mathbf{s}}_1), \ldots, y_t(\tilde{\mathbf{s}}_k))'$ Assuming that $\mathbf{y} = (y_1(\mathbf{s}_1), \ldots, y_1(\mathbf{s}_n), \ldots, y_T(\mathbf{s}_1), \ldots, y_T(\mathbf{s}_n))'$ represents the vector of the time series observed at n monitoring locations, we need to obtain the predictive posterior distribution, $p(\tilde{\mathbf{y}}_t|\mathbf{y})$, which is given by

$$p(\tilde{\mathbf{y}}_t|\mathbf{y}) = \int_\theta p(\tilde{\mathbf{y}}_t|\mathbf{y}, \Theta)p(\Theta|\mathbf{y})d\Theta, \tag{11}$$

where θ is a parameter vector comprising all the unknowns in the model

McGill

Spatial interpolation and temporal prediction

- Spatial interpolations are obtained by considering the distribution of $(\mathbf{y}_t, \tilde{\mathbf{y}}_t)$ conditional on the parameters, and the initial information θ_0. This distribution is given by

$$\left(\begin{array}{c} \mathbf{y}_t \\ \tilde{\mathbf{y}}_t \end{array} |\theta, \mathbf{V}, \mathbf{W} \right) \sim N\left(\left(\begin{array}{c} \mathbf{F}_t'\theta_t \\ \widetilde{\mathbf{F}}_t'\theta_t \end{array} \right); \left(\begin{array}{cc} \mathbf{V}_y & \mathbf{V}_{y\tilde{y}} \\ \mathbf{V}_{\tilde{y}y} & \mathbf{V}_{\tilde{y}\tilde{y}} \end{array} \right) \right),$$

where $\widetilde{\mathbf{F}}_t$ corresponds to the regression matrix for ungauged locations
Similar notation is used to split the covariance matrix into four blocks. We then have that

$$\tilde{\mathbf{y}}_t|\mathbf{y}_t, \theta \sim N\left(\tilde{\mu}_t + \mathbf{V}_{\tilde{y}y}(\mathbf{V}_y)^{-1}(\mathbf{y}_t - \mu_t), (\mathbf{V}_{\tilde{y}\tilde{y}} - \mathbf{V}_{\tilde{y}y}(\mathbf{V}_y)^{-1}\mathbf{V}_{\tilde{y}y}) \right),$$

where $\mu_t = \mathbf{F}_t'\theta_t$ and $\tilde{\mu}_t = \widetilde{\mathbf{F}}_t'\theta_t$ for all t. Once samples from the posterior distribution of the parameter vector are available, samples from the posterior predictive distribution of $\tilde{\mathbf{y}}_t$ are easily obtained by sampling from the conditional distribution above.

Spatial interpolation and temporal prediction

- Temporal *k* steps ahead prediction at monitoring locations is given by the following posterior predictive distribution

$$\begin{aligned} p(\mathbf{y}_{T+k}|D_T) &= \int \prod_{h=1}^{k} p(\mathbf{y}_{T+h}|\mathbf{F}_{T+h}, \theta_{T+h}, \mathbf{V}) p(\theta_{T+h}|\theta_{T+h-1}, \mathbf{W}) \\ &\times \quad p(\theta|D_T)p(\mathbf{W})p(\mathbf{V})d\tilde{\theta}d\theta d\mathbf{W}d\mathbf{V}, \end{aligned}$$

  where $D_T = \{\mathbf{y}_1, \dots, \mathbf{y}_T\}$, $\theta = (\theta_1, \dots, \theta_T)$ and $\tilde{\theta} = (\theta_{T+1}, \dots, \theta_{T+k})$.
- The integral above can be approximated by

$$p(\mathbf{y}_{T+k}|D_T) \approx \frac{1}{M} \sum_{m=1}^{M} p(\mathbf{y}_{T+k}|\mathbf{F}_{T+k}, \theta_{T+k}^{(m)}, \mathbf{V}^{(m)}).$$

  Here the superscript (m) denotes samples from the posterior of $\theta_{T+1}, \dots, \theta_{T+k}, \mathbf{W}, \mathbf{V}$
- Samples from the distribution above can be obtained by propagating θ_{T+k} following the system equation in (7), and through the samples from the posterior distribution of the parameter vector

McGill

Further extensions

- Huerta, Sansó and Stroud (2004) model ozone data in Mexico city and allow state parameters to vary across space
- Stroud, Müller and Sansó (2001) propose a non-stationary spatiotemporal model. They model the mean function at each time period as a locally weighted mixture of linear regressions and allow the regression coefÂ®cients to change through time
- Gelfand, Banerjee and Gamerman (2005) propose the use of a spatiotemporally varying coefficient form and also consider multivariate spatio-temporal processes through the LMC
- Nobre, Sansó and Schmidt (2011) propose a class of spatio-temporal models based on autoregressive (AR) processes at each location
- Vivar and Ferreira (2009) introduce a class of spatiotemporal models for Gaussian areal data

Further extensions

- Incorporating information through stochastic partial differential equations
  - The DLM framework can be used to accommodate the use of partial differential equations (PDEs) to account for scientific knowledge of the process
  - Broadly speaking, this is done by allowing the state evolution to be defined through a partial differential equation model
  - See Cressie, N. and Wikle, C. K. (2015) *Statistics for spatio-temporal data*. John Wiley & Sons.

A non-normal spatiotemporal model

Environmetrics
The official journal of The International Environmetrics Society
An Association of The International Statistical Institute

Volume 28, Issue 6
September 2017
e2411

SPECIAL INVITED ARTICLE

Spatiotemporal models for skewed processes

Alexandra M. Schmidt, Kelly C. M. Gonçalves, Patrícia L. Velozo

First published: 15 September 2016 | https://doi.org/10.1002/env.2411 | Citations: 13

Read the full text >

PDF TOOLS SHARE

References Related Information

Abstract

In the analysis of most spatiotemporal processes in environmental studies, observations present skewed distributions. Usually, a single transformation of the data is used to approximate normality, and stationary Gaussian processes are assumed to model the transformed data. The choice of transformation is key for spatial interpolation and temporal prediction. We propose a spatiotemporal model for skewed data that does not require the use of data transformation. The process is decomposed as the sum of a purely temporal structure with two independent components that are considered to be partial realizations from independent spatial Gaussian processes, for each time t. The model has an asymmetry parameter that might vary with location and time, and if this is

Recommended

Comments on: Spatiotemporal models for skewed processes

Marc G. Genton, Amanda S. Hering

Environmetrics

Discussion of paper: "Spatial-temporal models for skewed processes"

Gabriel Huerta, Jonathan Stroud

Environmetrics

McGill

A standard approach

In the analysis of most spatio-temporal processes in environmental studies, observations present skewed distributions, with a heavy right or left tail. Commonly we assume that

- a transformation of the response variable follows a GP (the same one at all sampling locations)
- the process is stationary and isotropic $\rightarrow$ distribution is unchanged when the origin of the index set is translated, and the process is invariant under rotation about the origin

McGill

Transformed Gaussian Random Fields
De Oliveira et. al (1997)

Let $\{Z(s), s \in D\}$, $D \subset \mathbb{R}^2$ be the random field of interest. They propose to model

$$\{Y(s) = g_\lambda(Z(s)),\ s \in D\}$$

where each $g_\lambda(\cdot)$ is a nonlinear monotone transformation, $g'_\lambda(\cdot)$ exists and is continuous in $\Lambda \times \mathbb{R} \rightarrow$ a possible family is the Box-Cox family of power transformations

- Pros
  - inference procedure is performed under a single framework $\rightarrow \lambda$ is jointly estimated with other model parameters
  - spatial interpolation is performed integrating out the posterior distribution of the parameters
- Cons
  - How to describe the estimated spatial correlation? And what about the association between covariates and the outcome?

Effect of data transformation

- Wallin and Bolin (2015) point out that a transformation $Y(\mathbf{s}) = g(Z(\mathbf{s}))$ of the spatial process may induce dependence between the mean and covariance structures of $Z(\mathbf{s})$
- E.g.: $Y(\mathbf{s}) = \sqrt{Z(\mathbf{s})}$, such that

$$Y(\mathbf{s}) = \mathbf{X}(\mathbf{s})'\beta + S(\mathbf{s}) + \varepsilon(\mathbf{s}),$$

then Wallin and Bolin (2015) show that

$$\begin{aligned} E(Z(\mathbf{s})) &= C_S(0) + 2\left(\mathbf{X}(\mathbf{s})'\beta\right)^2 \\ C(Z(\mathbf{s}), Z(\mathbf{s}')) &= 2C_S(\mathbf{s}-\mathbf{s}')^2 + 4(\mathbf{X}(\mathbf{s})\beta)(\mathbf{X}(\mathbf{s}')\beta)C_S(\mathbf{s}-\mathbf{s}') \end{aligned}$$

- Covariance structure depends on the mean structure $\Rightarrow$ nonstationary $\Rightarrow$ more complicated when we assume more complex models (nonlinear mean, nonstationary covariance...)

Should we model the process in its original scale?

Motivation

Monthly mean temperature in the South of Brazil from January 2002 until December 2012
N=21 Locations and T=115 months

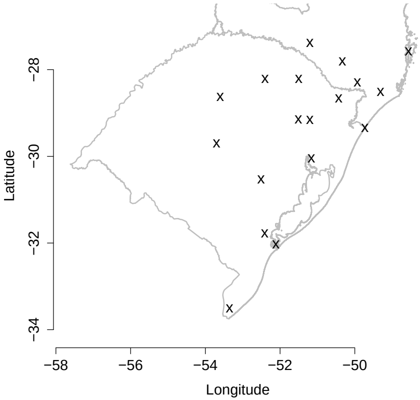
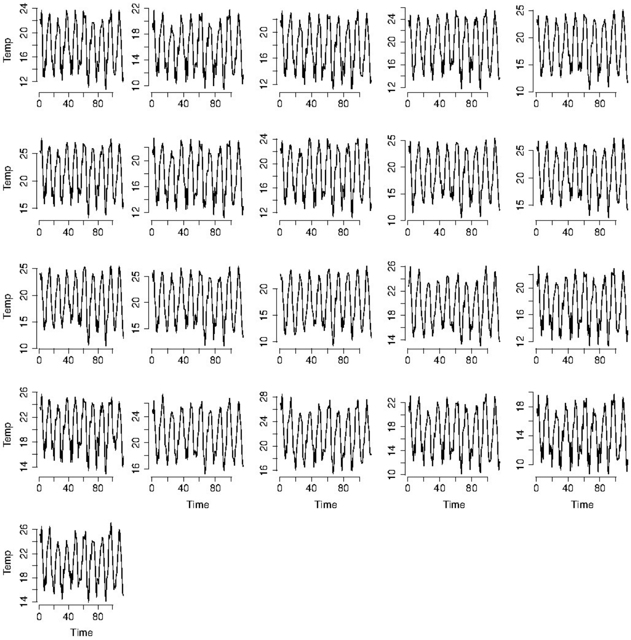


Figure: Temperature monitoring locations in the South of Brazil.

Motivation



A standard approach

Multivariate Dynamic linear model (MDLM)

$$
\begin{aligned}
\mathbf{y}_t &= \mathbf{X}_t \theta_t + \beta\, altitude + v_t, \quad v_t \sim N(0, \Sigma) \\
\theta_t &= \mathbf{G} \theta_{t-1} + \omega_t, \quad \omega_t \sim N(0, \mathbf{W}) \\
\theta_0 &\sim N(\mathbf{m}_0, \mathbf{C}_0),
\end{aligned}
$$

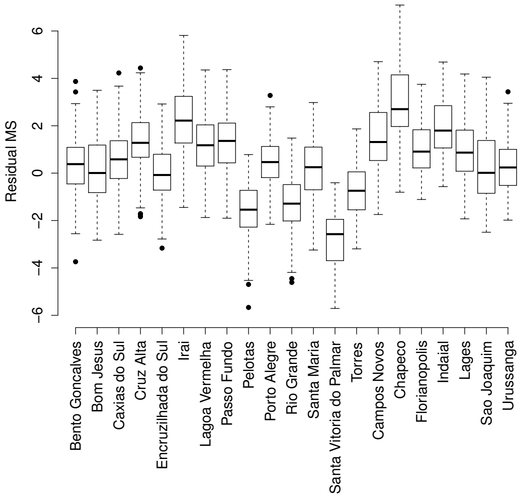
with $\theta_t = (\theta_{t1}, \theta_{t2}, \theta_{t3})'$, $\boldsymbol{X_t} = (1, 1, 0)$,

$$
\boldsymbol{G} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(2\pi/12) & \sin(2\pi/12) \\ 0 & -\sin(2\pi/12) & \cos(2\pi/12) \end{pmatrix},
$$

and $\Sigma_{ij} = \sigma^2 \exp(-\frac{1}{\phi} d_{ij})$, where d_{ij} is the Euclidean distance between locations s_i and s_j

Motivation

- Box-plots of the residuals after fitting the MDLM



McGill

Univariate Skew-normal distribution

A continuous random variable is said to follow a skew-normal distribution with parameter of asymmetry $\alpha \in \mathbb{R}$, if its pdf is given by

$$f(h) \quad = \quad 2\phi(h)\Phi(\alpha h), \quad h \in \mathbb{R}$$

where $\phi(u)$ and $\Phi(u)$ are, respectively, the pdf and cdf of the standard normal distribution evaluated at u.

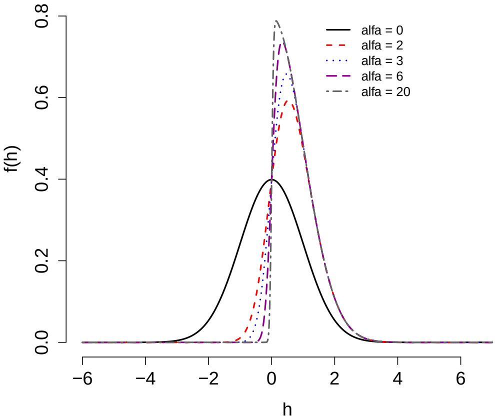


Figure: PDF of the skew-normal for different values of $\alpha \in \{0, 2, 3, 6, 20\}$.

Univariate Skew-normal distribution

- Parameter of asymmetry can be rewritten as a function of $\delta \in (-1, 1)$ such that

$$\alpha = \frac{\delta}{\sqrt{1-\delta^2}} \Longrightarrow \delta = \frac{\alpha}{\sqrt{1+\alpha^2}}$$

-

$$H = \frac{\alpha|\eta| + e}{\sqrt{1+\alpha^2}} = \delta|\eta| + (1-\delta^2)^{1/2} e,$$

where $H \sim SN(\alpha)$, $\eta, e \overset{iid}{\sim} N(0,1)$ and $|\eta|$ is the absolute value of η .

- Study absolute value of δ:
  - Sign of δ indicates positive or negative asymmetry
  - δ close to 0, smaller degree of asymmetry
  - the closer it is to 1, the greater the asymmetry

Multivariate skew-normal distribution

Definition (Azzalini & Dalla Valle, 1996)

A random vector **H** follows a multivariate skew normal distribution if each element of **H** can be written as

$$H_l = \delta_l |\eta_0| + (1-\delta_l^2)^{1/2} e_l, \quad l = 1, \cdots, L$$

where $\eta_0 \sim N(0,1)$, $\boldsymbol{e} = (e_1, \cdots, e_L)$ is a random vector following a multivariate normal distribution with correlation matrix **M**, and with standard normal marginals, the elements of **e** are independent of η_0

Some references that use this distribution: Kim & Mallick(2004) and Frühwirth-Schnatter & Pyne(2010).

Multivariate skew-normal distribution

- Some care must be taken when using the multivariate skew-normal of Azzalini & Dalla Valle (1996) for spatial observations
  - Zhang & El-Shaarawi (2010) mention that, when the asymmetry is high, the spatial correlation between 2 locations gets close to 1 regardless of the distance between these two locations
  - When there is a single realization of the spatial process, Genton & Zhang (2012) mention that the parameters are not well identified, even if the number of locations is high. They also show that the model proposed by Zhang & El-Shaarawi (2010) avoids this problem

Proposed model by Zhang & El-Shaarawi (2010)

Let $Z(\boldsymbol{s})$ be a process defined in a continuous region, $\boldsymbol{s} \in \mathbb{G} \subset \mathbb{R}^p$.
Assume

$$Z(\boldsymbol{s}) = m(\boldsymbol{s}) + \sigma|\eta(\boldsymbol{s})| + e(\boldsymbol{s})$$

where $\sigma \in \mathbb{R}$ and $\eta(\boldsymbol{s})$ is independent of $e(\boldsymbol{s}+\boldsymbol{r})$, $\forall\{\boldsymbol{s}, \boldsymbol{s}+\boldsymbol{r} \in \mathbb{G}\}$,
$\{\eta(\boldsymbol{s}) \sim GP(0, 1, \rho_\eta(d)), \boldsymbol{s} \in \mathbb{G}\}$,
$\{e(\boldsymbol{s}) \sim GP(0, 1, \tau I(d=0) + V\rho(d)), \boldsymbol{s} \in \mathbb{G}\}$.

Aim: Extend the Zhang and El-Shaarawi (2010) model to handle a spatio-temporal process

McGill

Proposed Model

Let $\{Z_t(\boldsymbol{s}); \boldsymbol{s} \in \mathbb{G}; t = 1, 2, \ldots\}$ be a stochastic process in discrete time t and location $\boldsymbol{s} \in \mathbb{G}$, with $\mathbb{G} \subset \mathbb{R}^p$, $p = 1$, 2, or 3. Let

$$Z_t(\boldsymbol{s}) \quad = \quad m_t(\boldsymbol{s}) + \sigma(\boldsymbol{s})|\eta_t(\boldsymbol{s})| + \sqrt{V_t}\omega_t(\boldsymbol{s}) + \sqrt{\tau}\varepsilon_t(\boldsymbol{s})$$

$$\begin{aligned} m_t(\boldsymbol{s}) &= \boldsymbol{X}_t(\boldsymbol{s})\theta_t + \boldsymbol{A}_t(\boldsymbol{s})\gamma, \\ \theta_t &= \boldsymbol{G}_t\theta_{t-1} + \boldsymbol{e}_t, \quad \boldsymbol{e}_t \sim N_K(0, \boldsymbol{W}) \end{aligned}$$

McGill

Proposed Model

Let $\{Z_t(\boldsymbol{s}); \boldsymbol{s} \in \mathbb{G}; t = 1, 2, \ldots\}$ be a stochastic process in discrete time t and location $\boldsymbol{s} \in \mathbb{G}$, with $\mathbb{G} \subset \mathbb{R}^p$, $p = 1$, 2, or 3. Let

$$Z_t(\boldsymbol{s}) \quad = \quad m_t(\boldsymbol{s}) + \sigma(\boldsymbol{s})|\eta_t(\boldsymbol{s})| + \sqrt{V_t}\omega_t(\boldsymbol{s}) + \sqrt{\tau}\varepsilon_t(\boldsymbol{s})$$

- $\sigma(\boldsymbol{s}) \in \mathbb{R}$
- $\sigma(\boldsymbol{s}) = 0 \Rightarrow$ GP model
- *A priori*, we assume

$$\sigma(\boldsymbol{s}) \mid \mu_\sigma, V_\sigma \sim^{iid} N(\mu_\sigma, V_\sigma)$$

- $\mu_\sigma \sim N(m_{00}, C_{00})$, and $V_\sigma \sim \exp(C_v)$
- Then, marginally, the prior for $\sigma(\mathbf{s})$ has

$$E(\sigma(\mathbf{s})) = m_{00} \qquad \text{and} \qquad V(\sigma(\mathbf{s})) = 1/C_v + C_{00}$$

$$Cov(\sigma(\mathbf{s}), \sigma(\mathbf{s}')) = C_{00} \qquad \text{and} \qquad \rho_\sigma = Corr(\sigma(\mathbf{s}), \sigma(\mathbf{s}')) = \frac{C_{00}}{1/C_v + C_{00}}$$

- Fix C_v according to its 1-to-1 relationship with ρ_σ, as $C_v = \frac{\rho_\sigma}{C_{00}(1-\rho_\sigma)}$

McGill

Proposed Model

Let $\{Z_t(\boldsymbol{s}); \boldsymbol{s} \in \mathbb{G}; t = 1, 2, \ldots\}$ be a stochastic process in discrete time t and location $\boldsymbol{s} \in \mathbb{G}$, with $\mathbb{G} \subset \mathbb{R}^p$, $p = 1$, 2, or 3. Let

$$Z_t(\boldsymbol{s}) = m_t(\boldsymbol{s}) + \sigma(\boldsymbol{s})|\eta_t(\boldsymbol{s})| + \sqrt{V_t}\omega_t(\boldsymbol{s}) + \sqrt{\tau}\varepsilon_t(\boldsymbol{s})$$

- $V, \tau > 0$ are scale parameters
- η_t and ω_t are independent, zero mean GPs, with variance equals 1 and spatial correlation functions ρ_η and ρ, respectively
- $\varepsilon_t(\boldsymbol{s}) \sim N(0, 1)$ i.i.d.

Proposed Model

Let $\{Z_t(\boldsymbol{s}); \boldsymbol{s} \in \mathbb{G}; t = 1, 2, \ldots\}$ be a stochastic process in discrete time t and location $\boldsymbol{s} \in \mathbb{G}$, with $\mathbb{G} \subset \mathbb{R}^p$, $p = 1$, 2, or 3. Let

$$Z_t(\boldsymbol{s}) \quad = \quad m_t(\boldsymbol{s}) + \sigma(\boldsymbol{s})|\eta_t(\boldsymbol{s})| + \sqrt{V_t}\omega_t(\boldsymbol{s}) + \sqrt{\tau}\varepsilon_t(\boldsymbol{s})$$

$$\log V_t = \log V_{t-1} + e_t^v, \quad e_t^v \sim N(0, V_v)$$

McGill

Proposed Model - Properties

- If $\boldsymbol{X}_{0t} \sim N(0, \Sigma_0)$ and $\mathbf{X}_t \sim N(0, \Sigma_1)$, independent, and $Z_t(\boldsymbol{s}) = \delta|X_{0t}(\boldsymbol{s})| + (1-\delta^2)^{1/2} X_t(\boldsymbol{s})$, with $\delta \in [-1, 1]$, let $\mathbf{W} = |\boldsymbol{X}_{0t}|$ then
$$f(\boldsymbol{z}_t) = \int_{\mathbb{R}_+^n} \phi_L(\boldsymbol{z}_t - \delta \boldsymbol{w}_t; (1-\delta^2)\Sigma_1) f_{w_t}(\boldsymbol{w}_t) d\boldsymbol{w}_t,$$
does not follow a multivariate skew normal distribution (Genton & Zhang, 2012)

- It can be shown that, for each time t and location $\boldsymbol{s}$,
$$Y_t(\boldsymbol{s}) = \frac{Z_t(\boldsymbol{s}) - m_t(\boldsymbol{s})}{\sqrt{\sigma(\boldsymbol{s})^2 + V_t + \tau}} \text{ has a pdf of the form}$$
$$2\phi(y)\Phi(\alpha y), \text{ i.e. } Y_t(\boldsymbol{s}) \sim SN\left(\frac{\sigma(\boldsymbol{s})}{\sqrt{V_t + \tau}}\right).$$

Resultant Covariance Structure

Let $W_t(\boldsymbol{s}) = \sigma(\boldsymbol{s})|\eta_t(\boldsymbol{s})| + \sqrt{V_t}\omega_t(\boldsymbol{s}) + \sqrt{\tau}\varepsilon_t(\boldsymbol{s})$, $\forall \boldsymbol{s} \in \mathbb{G}$, then

If $\sigma(\boldsymbol{s}) = \sigma$, $\forall \boldsymbol{s}$, and $V_t = V$, $\forall t$

$$
\begin{aligned}
&\text{cov}(W_t(\boldsymbol{s}), W_{t'}(\boldsymbol{s}')) \\
&= \begin{cases} \frac{2}{\pi}\sigma^2\left(\sqrt{1-\rho_\eta^2(d)} + \rho_\eta(d)\arcsin(\rho_\eta(d)) - 1\right) + V\rho(d) + \tau I(d), & t = t' \\ 0, & t \neq t' \end{cases}
\end{aligned}
$$

$d = ||\boldsymbol{s} - \boldsymbol{s}'||$, and $I(d)$ is an indicator function, such that $I(d) = 1$, if $d = 0$.

- Same parameter for asymmetry $\alpha = \frac{\sigma}{V+\tau}$
- Isotropic covariance structure

Resultant Covariance Structure

Let $W_t(\boldsymbol{s}) = \sigma(\boldsymbol{s})|\eta_t(\boldsymbol{s})| + \sqrt{V_t}\omega_t(\boldsymbol{s}) + \sqrt{\tau}\varepsilon_t(\boldsymbol{s})$, $\forall \boldsymbol{s} \in \mathbb{G}$, then

If $\sigma(\boldsymbol{s})$, and $V_t = V$, $\forall t$

$$\begin{aligned}&\text{cov}(W_t(\boldsymbol{s}), W_{t'}(\boldsymbol{s}'))\\ &= \begin{cases} \frac{2}{\pi}\sigma(\boldsymbol{s})\sigma(\boldsymbol{s'})\left(\sqrt{1-\rho_\eta^2(d)} + \rho_\eta(d)\text{arcsin}(\rho_\eta(d)) - 1\right) + V\rho(d) + \tau I(d), & t = t' \\ 0, & t \neq t' \end{cases}\end{aligned}$$

$d = ||\boldsymbol{s} - \boldsymbol{s}'||$, and $I(d)$ is an indicator function, such that $I(d) = 1$, if $d = 0$.

- Parameter for asymmetry $\alpha(\boldsymbol{s}) = \frac{\sigma(\boldsymbol{s})}{V+\tau}$
- Nonstationary (in space) covariance structure

McGill

Resultant Covariance Structure

Let $W_t(\boldsymbol{s}) = \sigma(\boldsymbol{s})|\eta_t(\boldsymbol{s})| + \sqrt{V_t}\omega_t(\boldsymbol{s}) + \sqrt{\tau}\varepsilon_t(\boldsymbol{s})$, $\forall \boldsymbol{s} \in \mathbb{G}$, then

If $\sigma(\boldsymbol{s})$ and V_t

$$\operatorname{cov}(W_t(\boldsymbol{s}), W_{t'}(\boldsymbol{s}'))$$
$$= \begin{cases} \frac{2}{\pi}\sigma(\boldsymbol{s})\sigma(\boldsymbol{s'})\left(\sqrt{1-\rho_\eta^2(d)} + \rho_\eta(d)\arcsin(\rho_\eta(d)) - 1\right) + V_t\rho(d) + \tau I(d), & t = t' \\ 0, & t \neq t' \end{cases}$$

$d = ||\boldsymbol{s} - \boldsymbol{s}'||$, and $I(d)$ is an indicator function, such that $I(d) = 1$, if $d = 0$.

- Parameter for asymmetry $\alpha_t(\boldsymbol{s}) = \frac{\sigma(\boldsymbol{s})}{V_t + \tau}$
- Nonstationary covariance structure

McGill

For observed times t, spatial interpolation can be performed by considering

$$(\mathbf{z}_t^u \mid \mathbf{z}_t^m, \varphi) \sim N\left(\delta_t^u + \boldsymbol{\Psi}^{um}\left(\boldsymbol{\Omega}^m\right)^{-1}(\boldsymbol{z}_t^m - \delta_t^m); \boldsymbol{\Omega}^u - \boldsymbol{\Psi}^{um}\left(\boldsymbol{\Omega}^m\right)^{-1}\boldsymbol{\Psi}^{mu}\right),$$

where $\delta_t^u = \mathbf{X}_t^u\theta_t + \boldsymbol{A}^u\gamma + \sigma^u|\eta_t^u|$, and $\delta_t^m = \mathbf{X}_t^m\theta_t + \boldsymbol{A}^m\gamma + \sigma^m|\eta_t^m|$, for all t, and

$$p(\mathbf{z}_t^u \mid \mathbf{z}) \quad = \quad \int_{\varphi^*} p(\mathbf{z}_t^u \mid \mathbf{z}_t^m, \varphi^*) p(\eta_t^u \mid \eta_t^m, \mathbf{z}_t^m, \varphi)\, p(\varphi \mid \mathbf{z}_t^m) d\varphi^*$$

Temporal prediction k steps ahead, for the monitoring locations, can be obtained from

$$p(\mathbf{z}_{T+k} \mid \mathbf{z}) = \int_{\varphi^*} \left[\prod_{j=1}^{k} p(\mathbf{z}_{T+j} \mid \mathbf{X}_{T+j}, \theta_{T+j}, \eta_{T+j}, V_{T+j}, \gamma, \tau, \boldsymbol{\Sigma}, \phi)\right.$$
$$\left. p(\theta_{T+j} \mid \theta_{T+j-1}, W)\, p(V_{T+j} \mid V_{T+j-1}, V_v)\, p(\eta_{T+j} \mid \phi_\eta)\right] p(\varphi \mid \boldsymbol{z}) d\varphi^*$$

We have $T = 115$, $L = 21$, $\theta_t = (\theta_{t1}, \theta_{t2}, \theta_{t3})'$, $\boldsymbol{X_t} = (1, 1, 0)$ and $\boldsymbol{G} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(2\pi/12) & \sin(2\pi/12) \\ 0 & -\sin(2\pi/12) & \cos(2\pi/12) \end{pmatrix}$, observed altitude is included as a covariate in **A**.

McGill

Fitted models

Fitted Model	Parameter Specification		
	V_t	$\sigma(\boldsymbol{s})$	ϕ_η, ϕ
MVT	V_t	$\sigma(\boldsymbol{s})$	ϕ_η, ϕ
MA1-$\rho_\sigma = 0.99$	$V_t = V$	$\sigma(\boldsymbol{s})$	ϕ_η, ϕ
MA1-$\rho_\sigma = 0.95$	$V_t = V$	$\sigma(\boldsymbol{s})$	ϕ_η, ϕ
MA1-$\rho_\sigma = 0.90$	$V_t = V$	$\sigma(\boldsymbol{s})$	ϕ_η, ϕ
MA1-$\rho_\sigma = 0.33$	$V_t = V$	$\sigma(\boldsymbol{s})$	ϕ_η, ϕ
MA2-$\rho_\sigma = 0.99$	$V_t = V$	$\sigma(\boldsymbol{s})$	$\phi_\eta = \phi$
MA2-$\rho_\sigma = 0.95$	$V_t = V$	$\sigma(\boldsymbol{s})$	$\phi_\eta = \phi$
MA2-$\rho_\sigma = 0.90$	$V_t = V$	$\sigma(\boldsymbol{s})$	$\phi_\eta = \phi$
MA2-$\rho_\sigma = 0.33$	$V_t = V$	$\sigma(\boldsymbol{s})$	$\phi_\eta = \phi$
MA3	$V_t = V$	$\sigma(\boldsymbol{s}) = \sigma$	ϕ_η, ϕ
MA4	$V_t = V$	$\sigma(\boldsymbol{s}) = \sigma$	$\phi_\eta = \phi$
MS	$V_t = V$	$\sigma(\boldsymbol{s}) = 0$	ϕ
MSVT	V_t	$\sigma(\boldsymbol{s}) = 0$	ϕ
MSL	$V_t = V$	$\sigma(\boldsymbol{s}) = 0$	ϕ
MSR	$V_t = V$	$\sigma(\boldsymbol{s}) = 0$	ϕ

Model comparison is performed through

- Posterior predictive loss (Gelfand & Ghosh, 1998)
- Continuous Ranked Probability Score (CRPS) (Gneiting and Raftery, 2007)
- MSE and MAE
- We removed the last 5 observations from all sites for predictive purposes

Fitted Model	**Parameter Specification**			**Within sample**					
	V_t	$\sigma(\boldsymbol{s})$	ϕ_η, ϕ	CRPS	MSE	MAE	D	G	P
MVT	V_t	$\sigma(\boldsymbol{s}_l)$	ϕ_η, ϕ	0.834	1.774	1.023	7562	4286	10837
MA1-$\rho_\sigma = 0.99$	$V_t = V$	$\sigma(\boldsymbol{s})$	ϕ_η, ϕ	0.741	1.341	0.921	4696	3234	6158
MA1-$\rho_\sigma = 0.95$	$V_t = V$	$\sigma(\boldsymbol{s})$	ϕ_η, ϕ	0.744	1.341	0.920	4792	3239	6345
MA1-$\rho_\sigma = 0.90$	$V_t = V$	$\sigma(\boldsymbol{s})$	ϕ_η, ϕ	0.747	1.347	0.922	4858	3249	6467
MA1-$\rho_\sigma = 0.33$	$V_t = V$	$\sigma(\boldsymbol{s})$	ϕ_η, ϕ	0.745	1.344	0.921	4855	3240	6471
MA2-$\rho_\sigma = 0.99$	$V_t = V$	$\sigma(\boldsymbol{s}_l)$	$\phi_\eta = \phi$	0.862	1.598	1.009	7701	3853	11548
MA2-$\rho_\sigma = 0.95$	$V_t = V$	$\sigma(\boldsymbol{s}_l)$	$\phi_\eta = \phi$	0.867	1.588	1.006	7998	3828	12168
MA2-$\rho_\sigma = 0.90$	$V_t = V$	$\sigma(\boldsymbol{s}_l)$	$\phi_\eta = \phi$	0.878	1.600	1.008	8275	3835	12714
MA2-$\rho_\sigma = 0.33$	$V_t = V$	$\sigma(\boldsymbol{s}_l)$	$\phi_\eta = \phi$	0.879	1.588	1.007	8610	3834	13387
MA3	$V_t = V$	$\sigma(\boldsymbol{s}_l) = \sigma$	ϕ_η, ϕ	1.193	3.506	1.478	14011	8448	19574
MA4	$V_t = V$	$\sigma(\boldsymbol{s}_l) = \sigma$	$\phi_\eta = \phi$	1.212	3.503	1.478	14899	8448	21351
MS	$V_t = V$	$\sigma(\boldsymbol{s}_l) = 0$	ϕ	1.208	3.550	1.489	14934	8578	21290
MSVT	V_t	$\sigma(\boldsymbol{s}_l) = 0$	ϕ	1.163	3.622	1.503	12709	8738	16679
MSL	$V_t = V$	$\sigma(\boldsymbol{s}_l) = 0$	ϕ	1.341	3.599	1.527	18260	8661	27858
MSR	$V_t = V$	$\sigma(\boldsymbol{s}_l) = 0$	ϕ	1.266	3.450	1.480	16136	8320	23952

Model comparison is performed through

- Posterior predictive loss (Gelfand & Ghosh, 1998)
- Continuous Ranked Probability Score (CRPS) (Gneiting and Raftery, 2007)
- MSE and MAE
- We removed the last 5 observations from all sites for predictive purposes

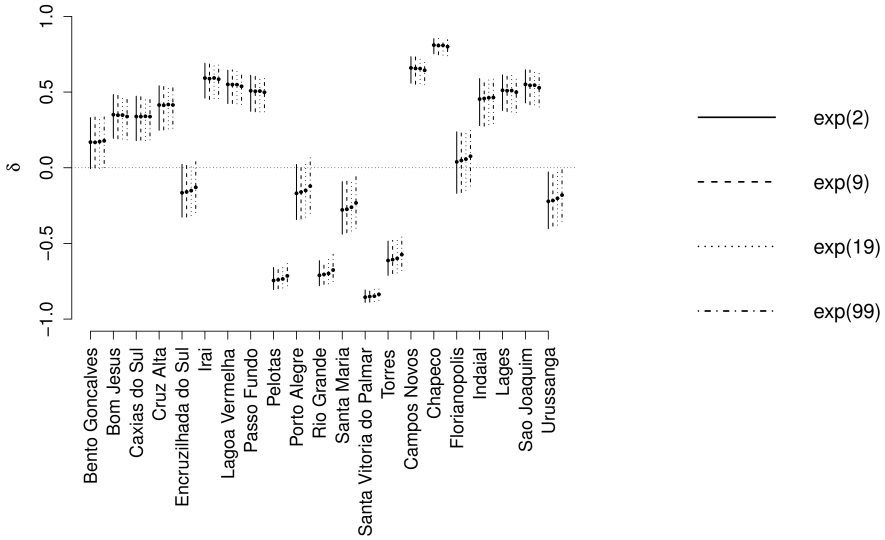
Fitted Model	**Parameter Specification**			**Out of sample**					
	V_t	$\sigma(\boldsymbol{s})$	ϕ_η, ϕ	CRPS	MSE	MAE	D	G	P
MVT	V_t	$\sigma(\boldsymbol{s}_l)$	ϕ_η, ϕ	1.335	1.408	0.955	8629	366	16981
MA1-$\rho_\sigma = 0.99$	$V_t = V$	$\sigma(\boldsymbol{s})$	ϕ_η, ϕ	0.885	1.662	1.047	591	422	760
MA1-$\rho_\sigma = 0.95$	$V_t = V$	$\sigma(\boldsymbol{s})$	ϕ_η, ϕ	0.885	1.608	1.028	602	408	796
MA1-$\rho_\sigma = 0.90$	$V_t = V$	$\sigma(\boldsymbol{s})$	ϕ_η, ϕ	0.876	1.517	0.992	605	392	818
MA1-$\rho_\sigma = 0.33$	$V_t = V$	$\sigma(\boldsymbol{s})$	ϕ_η, ϕ	0.880	1.571	1.011	606	396	816
MA2-$\rho_\sigma = 0.99$	$V_t = V$	$\sigma(\boldsymbol{s}_l)$	$\phi_\eta = \phi$	0.915	1.351	0.948	858	344	1371
MA2-$\rho_\sigma = 0.95$	$V_t = V$	$\sigma(\boldsymbol{s}_l)$	$\phi_\eta = \phi$	0.930	1.347	0.947	901	341	1461
MA2-$\rho_\sigma = 0.90$	$V_t = V$	$\sigma(\boldsymbol{s}_l)$	$\phi_\eta = \phi$	0.944	1.322	0.936	931	341	1521
MA2-$\rho_\sigma = 0.33$	$V_t = V$	$\sigma(\boldsymbol{s}_l)$	$\phi_\eta = \phi$	0.964	1.322	0.935	980	336	1625
MA3	$V_t = V$	$\sigma(\boldsymbol{s}_l) = \sigma$	ϕ_η, ϕ	1.173	2.918	1.393	1458	747	2169
MA4	$V_t = V$	$\sigma(\boldsymbol{s}_l) = \sigma$	$\phi_\eta = \phi$	1.203	2.942	1.399	1558	738	2378
MS	$V_t = V$	$\sigma(\boldsymbol{s}_l) = 0$	ϕ	1.199	2.874	1.385	1561	743	2379
MSVT	V_t	$\sigma(\boldsymbol{s}_l) = 0$	ϕ	1.158	3.077	1.444	1099	771	1426
MSL	$V_t = V$	$\sigma(\boldsymbol{s}_l) = 0$	ϕ	1.284	3.138	1.431	1837	794	2880
MSR	$V_t = V$	$\sigma(\boldsymbol{s}_l) = 0$	ϕ	1.228	2.895	1.371	1643	740	2547

Data Analysis - Mean Temperature in the South of Brazil

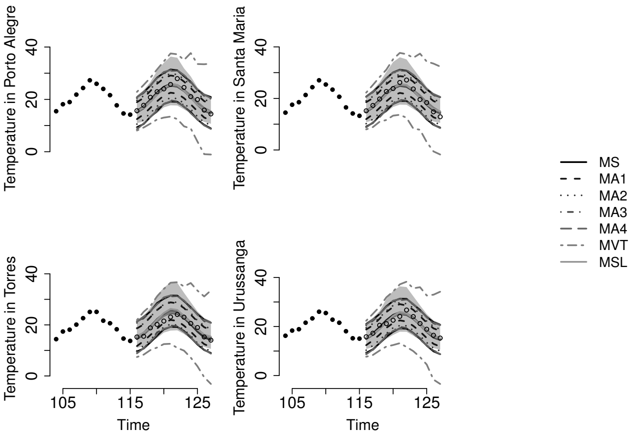
Parameter	Prior specification for V_σ			
	$V_\sigma \sim \exp(99)$	$V_\sigma \sim \exp(19)$	$V_\sigma \sim \exp(9)$	$V_\sigma \sim \exp(2)$
μ_σ	0.17 (-0.17;0.49)	0.16 (-0.27;0.56)	0.12 (-0.33; 0.57)	-0.01 (-0.521;0.327)
V_σ	0.29 (0.22;0.37)	0.61 (0.42;0.86)	0.79 (0.49;1.18)	0.8 (0.45;1.41)

Table: Posterior summary (mean and 95% credible intervals in brackets) of the hyperparameters in the prior specification of $\sigma(\mathbf{s})$, under model MA1.

Data Analysis - Mean Temperature in the South of Brazil



Data Analysis - Mean Temperature in the South of Brazil



McGill

Data Analysis - Mean Temperature in the South of Brazil

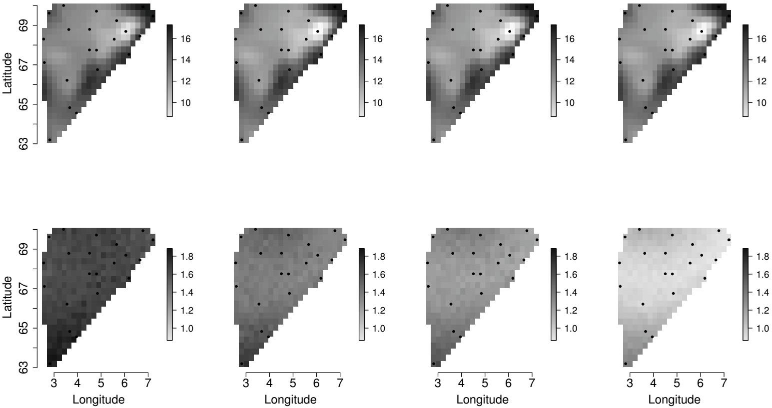


Figure: Surfaces of the posterior mean (first row) and point-wise standard deviation of the predictive distribution for unmonitored locations under model MA1 for all different prior specifications for V_σ, from left to right $V_\sigma \sim \exp(2)$, $V_\sigma \sim \exp(9)$, $V_\sigma \sim \exp(19)$, and $V_\sigma \sim \exp(99)$

Discussion

- We extend the proposed spatial skew-normal model of Zhang and El-Sharaawi (2010) to accommodate temporal observations at each location
- Marginally, the resultant distribution of the process is skew-normal with different skewness parameter for each location and/or time
- Particular cases of the model encompass the normal distribution, and a skew-normal distribution with the same skewness parameter for all locations and instants in time