



Spatial-Temporal Trend Modeling for Ozone Concentration in Tehran City

Seyedeh Somayeh Mousavi and Mohsen Mohammadzadeh*

Tarbiat Modares University

Abstract. Fitting a suitable covariance function for the correlation structure of spatial-temporal data requires de-trending the data. In this article, some potential models for spatial-temporal trend are presented. Eventually the best model will be announced for de-trending tropospheric ozone concentration data for the city of Tehran (Capital city of Iran). By using the selected trend model, some features of the covariance function of de-trended data will be specified.

Keywords. Spatial-temporal random Field; modeling spatial-temporal trend; dynamic spatial linear model; symmetry; separability; ozone concentration.

MSC 2010: 62H11.

1 Introduction

Many environmental and geophysical processes such as atmospheric pollutant concentrations, precipitation fields and surface winds, involve variability over space and time, which are named as spatial-temporal processes. Statistical analysis of spatial-temporal data requires the determination of the correlation structure using the covariance function. This function is usually unknown and should be estimated. Since the trend of the data may cause bias in the covariance function estimation (Cressie, 1993), it is necessary to model the possible trend which may exist in the data, and use the de-trended data for fitting a valid covariance function. However, there are

* Corresponding author

various methods for trend modeling. Certainly, the de-trended data set derived from various trend models induces different covariance functions, which have different effects on the accuracy of data analysis.

In recent researches, more attention has been paid on the modeling of spatial-temporal trend. Cox and Chu (1992) used a generalized linear model approach, by assuming a conditional Weibull distribution for ozone concentrations along with meteorology, to estimate trends in daily maximum ozone levels. Porter et al. (2001) reported on the estimation of trends in ozone concentrations, which were adjusted for meteorological variables at individual monitoring sites. However, Stroud et al. (2001) modeled the mean function at each time-period as a locally weighted mixture of linear regressions. In order to incorporate temporal variation, they allowed the regression coefficients to change through time. Huerta et al. (2004) applied a dynamic linear model for ambient ozone, whilst McMillan et al. (2005) proposed a nearest-neighbor spatial model. Also, Zheng et al. (2007) used a dynamic linear model and generalized additive model to explain ozone trends and Sahu et al. (2007) proposed a space-time model for 8-hour maximum ozone levels to provide input to regulatory activities, detection, evaluation, and analysis of spatial patterns of ozone summaries and temporal trends. However, Fuentes et al. (2008) introduced spatial-temporal models, using covariates that have dynamic coefficients in the space-time model. Finally, Behshad and Mohammadzadeh (2011) evaluated different separability and symmetry tests for spatial-temporal covariance function.

In this paper, we focus on the effect of four linear models in formulating the spatial-temporal trends on the features of covariance structure for cases where the data are spatial-temporal. To summarize, the structure of this article is as follows: in Section 2, four linear models are introduced for modeling the mean structure of a space-time process. In Section 3, we assess the effect of different models on the features of spatial-temporal covariance function. Initially a case study has been carried out on the ozone concentration of the data collected from the city of Tehran, with consideration of the suitable features of the spatial-temporal covariance function. Finally, conclusions are given in Section 4.

2 Spatial-Temporal Trend Modeling

Let $\{Z(\mathbf{s}, t); \mathbf{s} \in R^d, t \in R\}$ denotes a spatial-temporal random field, where \mathbf{s} represents a site in d -dimensional space and t represents time. In general,

a spatial-temporal random field can be decomposed as

$$Z(\mathbf{s}, t) = \mu(\mathbf{s}, t) + \delta(\mathbf{s}, t),$$

where $\mu(\mathbf{s}, t)$ represents the mean surface or spatial-temporal trend, also it corresponds to large scale variations and $\delta(\mathbf{s}, t)$ is a zero mean space-time correlated error that explains the spatial-temporal small scale variations. There are different ways to define a model for the large scale structure, i.e. trend. In this article, four linear space-time models are considered for modeling the spatial-temporal trend.

Suppose $\mathbf{z} = (z(\mathbf{s}_1, t_1), \dots, z(\mathbf{s}_m, t_n))$ is a realization vector of a Gaussian random field observed at spatial locations $\mathbf{s}_1, \dots, \mathbf{s}_m$ and temporal points t_1, \dots, t_n . Assume $\mathbf{Z} \sim N_{mn}(\boldsymbol{\mu}(\mathbf{s}, t), \boldsymbol{\Sigma})$, $\mathbf{Z} = (\mathbf{Z}(\mathbf{s}_1, t_1), \dots, \mathbf{Z}(\mathbf{s}_m, t_n))$, where $\boldsymbol{\Sigma}$ is a $mn \times mn$ positive definite covariance matrix. Also $\mathbf{X}'(\mathbf{s}, t) = (X_1(\mathbf{s}, t), \dots, X_k(\mathbf{s}, t))$ is vector of known functions including the values of covariates and location variables. A simple model is referred to the linear regression model with constant coefficients and independent errors. The development of trend modeling is described systematically in the following section.

2.1 Linear Model with Constant Coefficients and Independent Errors

A simple linear model for spatial-temporal trend is as follows

$$\mu(\mathbf{s}, t) = \beta_0 + \beta_1 X_1(\mathbf{s}, t) + \dots + \beta_k X_k(\mathbf{s}, t), \quad (\mathbf{s}, t) \in R^d \times R, \quad (1)$$

which can be restated as $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, where $\boldsymbol{\mu} = (\mu(\mathbf{s}_1, t_1), \dots, \mu(\mathbf{s}_m, t_n))$, $\mathbf{X} = (\mathbf{x}(\mathbf{s}_i, t_j))$ is a $mn \times (k+1)$ matrix and $\boldsymbol{\beta} = (\beta_0, \dots, \beta_k)$ represents $(k+1) \times 1$ parameter vector specifying the effect of each covariate and locative variable. For the iid errors the ordinary least square estimator of $\boldsymbol{\beta}$ is given by $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}$, where $\mathbf{Z} = (Z(\mathbf{s}_1, t_1), \dots, Z(\mathbf{s}_m, t_n))$.

For Bayesian estimation of $\boldsymbol{\beta}$, it is assumed that $\text{var}(\mathbf{Z}) = \text{var}(\boldsymbol{\delta}) = \boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_{mn}$, where $\mathbf{Z} \sim N_{mn}(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_{mn})$, and \mathbf{I}_{mn} is an $mn \times mn$ identity matrix. Suppose prior distribution for $(\boldsymbol{\beta}, \sigma^2)$ is the multivariate Normal-inverse Gamma, that it is a conjugate distribution and is given by

$$(\boldsymbol{\beta}, \sigma^2) \sim NIG(\boldsymbol{\mu}_\beta, \mathbf{V}_\beta, a, b),$$

which can be expressed as the product of a multivariate Normal distribution for $\boldsymbol{\beta}$ and an inverse Gamma prior for σ^2 given by

$$(\boldsymbol{\beta} | \sigma^2) \sim N(\boldsymbol{\mu}_\beta, \sigma^2 \mathbf{V}_\beta), \quad \sigma^2 \sim IG(a, b).$$

The resulting posterior is also a multivariate Normal-inverse gamma $NIG(\tilde{\boldsymbol{\mu}}_\beta, \tilde{\boldsymbol{\Sigma}}_\beta, \tilde{a}, \tilde{b})$, where

$$\begin{aligned} \tilde{\boldsymbol{\mu}}_\beta &= \tilde{\boldsymbol{\Sigma}}_\beta(\mathbf{X}'\mathbf{z} + \mathbf{V}_\beta^{-1}\boldsymbol{\mu}_\beta), & \tilde{\boldsymbol{\Sigma}}_\beta &= (\mathbf{X}'\mathbf{X} + \mathbf{V}_\beta^{-1})^{-1}, \\ \tilde{a} &= a + \frac{mn}{2}, & \tilde{b} &= b + \frac{\mathbf{z}'\mathbf{z} - \tilde{\boldsymbol{\mu}}'_\beta\tilde{\boldsymbol{\Sigma}}_\beta^{-1}\tilde{\boldsymbol{\mu}}_\beta + \boldsymbol{\mu}'_\beta\mathbf{V}_\beta^{-1}\boldsymbol{\mu}_\beta}{2}. \end{aligned}$$

The conditional posterior distribution of $\boldsymbol{\beta}$ and marginal posterior of σ^2 are given as follows (see Bernardo and Smith, 1994; Ntzoufras, 2009):

$$(\boldsymbol{\beta}|\mathbf{z}, \sigma^2) \sim N_{k+1}(\tilde{\boldsymbol{\mu}}_\beta, \sigma^2\tilde{\boldsymbol{\Sigma}}_\beta), \quad \sigma^2 \sim IG(\tilde{a}, \tilde{b}).$$

Using Gibbs algorithm, the Bayesian estimates of $\boldsymbol{\beta}$ and σ^2 can be derived.

One of the major advantages of the Bayesian approach lies in the ease with subjective information and model/data information are combined. Disadvantage of this model is that the parameters are constant, therefore they do not consider their spatial and temporal dependency. Also it is assumed that the errors are not spatially and temporal dependent.

2.2 Linear Model with Spatial-temporal Coefficients and Independent Errors

If the coefficients of the linear model (1) depends on the locations $(\mathbf{s}_1, t_1), \dots, (\mathbf{s}_m, t_n)$, then the following assumption holds;

$$\mu(\mathbf{s}, t) = \beta_0(\mathbf{s}, t) + \beta_1(\mathbf{s}, t)X_1(\mathbf{s}, t) + \dots + \beta_k(\mathbf{s}, t)X_k(\mathbf{s}, t), \quad (\mathbf{s}, t) \in R^d \times R, \tag{2}$$

which is a new linear model, that can be rewritten as $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, where $\boldsymbol{\mu}$ and \mathbf{X} are similar to previous subsection, but $\boldsymbol{\beta}$ is a $(k + 1) \times mn$ dimensional matrix of regression parameters as follows;

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0(s_1, t_1) & \dots & \beta_0(s_m, t_n) \\ \vdots & \ddots & \vdots \\ \beta_k(s_1, t_1) & \dots & \beta_k(s_m, t_n) \end{bmatrix}$$

For estimating $\boldsymbol{\beta}$, the matrix Normal distribution is supposed with $(k + 1) \times mn$ -dimensional mean matrix \mathbf{M} and $(k + 1)mn \times (k + 1)mn$ covariance matrix $\boldsymbol{\Omega}$, as prior for $\boldsymbol{\beta}$. For ease, $\boldsymbol{\Omega}$ is considered to be the Kronecker product of two smaller matrices, as $\boldsymbol{\Omega} = \boldsymbol{\Sigma}_1 \otimes \boldsymbol{\Sigma}_2$, where $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are

$(k + 1) \times (k + 1)$ - and $mn \times mn$ -dimensional matrices (Rowe, 2003). So the following;

$$\begin{aligned} \pi(\boldsymbol{\beta} | \mathbf{M}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) &= (2\pi)^{-\frac{(k+1)mn}{2}} |\boldsymbol{\Sigma}_1|^{-\frac{(k+1)}{2}} |\boldsymbol{\Sigma}_2|^{-\frac{mn}{2}} \\ &\quad \times \exp \left\{ -\frac{1}{2} \text{tr} \boldsymbol{\Sigma}_1^{-1} (\boldsymbol{\beta} - \mathbf{M}) \boldsymbol{\Sigma}_2^{-1} (\boldsymbol{\beta} - \mathbf{M})' \right\}, \end{aligned}$$

is a prior distribution for $\boldsymbol{\beta}$. If $\boldsymbol{\Sigma}_1 = \mathbf{I}_{(k+1) \times (k+1)}$ and $\boldsymbol{\Sigma}_2 = \mathbf{I}_{mn \times mn}$, then $\boldsymbol{\Omega} = \mathbf{I}$. Therefore rows and columns of the matrix $\boldsymbol{\beta}$ are independent, that is $\boldsymbol{\beta}$ includes $k + 1$ constant parameters $(\beta_0, \dots, \beta_k)$ and can be estimated similar to Subsection 2.1.

Advantage of the model 2 is that the parameters and covariates have spatial and temporal dependency. Its disadvantage is that the errors are independent.

2.3 Linear Model with Spatial-Temporal Errors

Considering model (1), if $\text{var}(\mathbf{Z}) = \text{var}(\boldsymbol{\delta}) = \boldsymbol{\Sigma}$ is known, the generalized least square (GLS) estimator of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}}_{gls} = (\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X}) \mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{Z}.$$

However, when $\boldsymbol{\Sigma}$ is unknown; the method of estimated generalized least square (EGLS) is applied, in which $\hat{\boldsymbol{\Sigma}}$ is used instead of $\boldsymbol{\Sigma}$, so another estimator of $\boldsymbol{\beta}$ would be given by

$$\hat{\boldsymbol{\beta}}_{egls} = (\mathbf{X}' \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{X}) \mathbf{X}' \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{Z}$$

For Bayesian approach, we assume that spatial-temporal covariance matrix $\boldsymbol{\Sigma}$ is separable, that is $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_S \otimes \boldsymbol{\Sigma}_T$, where $\boldsymbol{\Sigma}_S$ is purely a spatial covariance matrix and $\boldsymbol{\Sigma}_T$ is a pure temporal covariance matrix. The conjugated prior distribution for $\boldsymbol{\beta}$ is $N_{k+1}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, where $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ are known. Assuming the random field $Z(\cdot, \cdot)$ is Gaussian, the conditional posterior distribution of $\boldsymbol{\beta}$ is $N_{k+1}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}})$, where

$$\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\Sigma}} (\mathbf{X}' (\boldsymbol{\Sigma}_S^{-1} \otimes \boldsymbol{\Sigma}_T^{-1}) \mathbf{z} + \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\mu}_0), \quad \tilde{\boldsymbol{\Sigma}} = (\mathbf{X}' (\boldsymbol{\Sigma}_S^{-1} \otimes \boldsymbol{\Sigma}_T^{-1}) \mathbf{X} + \boldsymbol{\Sigma}_0^{-1})^{-1}.$$

Disadvantage of this model is that the parameters are constant, therefore their spatial and temporal dependency is not assumed.

2.4 Dynamic Spatial Linear Model

Let the m -dimensional vector $\mathbf{Z}(t) = (Z(\mathbf{s}_1, t), \dots, Z(\mathbf{s}_m, t))$ observed at time point t , ($t = 1, \dots, n$), have multivariate Normal distribution $N_m(\boldsymbol{\mu}(t), \mathbf{V}_t)$. The dynamic spatial linear model for each t is defined by observation and evolution equations as follows:

$$\text{Observation equation : } \quad \boldsymbol{\mu}(t) = \mathbf{F}'_t \boldsymbol{\theta}(t) + \boldsymbol{\epsilon}(t), \quad \boldsymbol{\epsilon}(t) \sim N_m(\mathbf{0}, \mathbf{V}_t), \quad (3)$$

$$\text{Evolution equation : } \quad \boldsymbol{\theta}(t) = \mathbf{G}_t \boldsymbol{\theta}(t-1) + \boldsymbol{\omega}(t), \quad \boldsymbol{\omega}(t) \sim N_q(\mathbf{0}, \mathbf{W}_t), \quad (4)$$

where \mathbf{F}_t is the $q \times m$ design matrix, $\boldsymbol{\theta}(t)$ is the $q \times 1$ state vector, $\boldsymbol{\epsilon}(t)$ is the observational error vector with $m \times m$ covariance matrix \mathbf{V}_t , \mathbf{G}_t is the evolution matrix related to the state vector and $\boldsymbol{\omega}(t)$ is the evolution error vector with $q \times q$ covariance matrix \mathbf{W}_t , also $\boldsymbol{\epsilon}(t)$ and $\boldsymbol{\omega}(t)$ are independent. Therefore at each time (t), the dynamic linear model (DLM) is characterized by a set of known quadruples $\{\mathbf{F}_t, \mathbf{G}_t, \mathbf{V}_t, \mathbf{W}_t\}$. This model is completed with a prior on the initial state vector, $\boldsymbol{\theta}_0 | \mathbf{D}_0 \sim N(\mathbf{m}_0, \mathbf{C}_0)$, where \mathbf{D}_0 denotes the initial information set, and \mathbf{m}_0 and \mathbf{C}_0 are known (West and Harrison, 1997). Assuming that the spatial-temporal observations have cyclical behavior, the state vector can be defined as $\boldsymbol{\theta}(t) = (\boldsymbol{\theta}'_1(t), \boldsymbol{\theta}'_2(t))'$, where $q = r + 2k$, $\boldsymbol{\theta}'_1(t)$ is the $r \times 1$ spatial process and the $2k \times 1$ dimensional vector $\boldsymbol{\theta}'_2(t)$ describes cyclicity of the data with cycle $2k$. Covariates of the spatial process denoted by \mathbf{X} , include variables such as length, width, height and other covariates. Corresponding to the partitioning of $\boldsymbol{\theta}(t)$, design matrix is considered as $\mathbf{F}_t = (\mathbf{X}, \mathbf{F}_1, \dots, \mathbf{F}_k)$, where each of \mathbf{F}_h , $h = 1, \dots, k$, are $m \times 2$ matrices that the entire elements of the first column are 1 and second column are 0. Therefore, the evolution matrix \mathbf{G}_t can be used as a block structure with $\mathbf{G}_t = \text{blockdiag}(\mathbf{I}_{rr}, \mathbf{G}_{t1}, \dots, \mathbf{G}_{tk})$, where each block \mathbf{G}_{th} , $h = 1, \dots, k$, is a 2×2 harmonic matrix of the form

$$\mathbf{G}_{th} = \begin{bmatrix} \cos(2\pi h/p) & \sin(2\pi h/p) \\ -\sin(2\pi h/p) & \cos(2\pi h/p) \end{bmatrix}, \quad h = 1, \dots, k$$

For modeling the spatial dependency of the observations, the covariance matrix $\mathbf{V}_t = \sigma^2 \mathbf{V}_\lambda$ is considered in which $\mathbf{V}_\lambda = \exp(-\mathbf{V}/\lambda)$ and elements of \mathbf{V} are determined by a known spatial correlation function. The evolution variance \mathbf{W}_t can be specified either explicitly or through a discount factor $\alpha \in [0, \infty)$, which defines $\mathbf{W}_t = \alpha \mathbf{P}_t$, where $\mathbf{P}_t = \text{var}(\mathbf{G}_t \boldsymbol{\theta}(t-1) | \mathbf{D}_{t-1})$. A discount factor of $\alpha = 0$ gives a static model, with the same coefficients for

all time periods, whereas $\alpha \rightarrow \infty$ implies coefficients which are independent over time, i.e. no temporal smoothing at all (Stroud et al., 2001).

Using Kalman filtering theorem (West and Harrison, 1997) under models (3) and (4), together with the initial state information $\boldsymbol{\theta}_0 | \mathbf{D}_0 \sim N(\mathbf{m}_0, \mathbf{C}_0)$, for $t = 2, \dots, n$, we have:

- (a) Posterior at $t - 1$: $(\boldsymbol{\theta}(t - 1) | \mathbf{D}_{t-1}) \sim N(\mathbf{m}_{t-1}, \mathbf{C}_{t-1})$, where, $\mathbf{D}_t = \{\mathbf{D}_0, \boldsymbol{\mu}(1), \dots, \boldsymbol{\mu}(t)\}$.
- (b) Prior at t (Evolution): $(\boldsymbol{\theta}(t) | \mathbf{D}_{t-1}) \sim N(\mathbf{a}_t, \mathbf{R}_t)$, where $E(\boldsymbol{\theta}(t) | \mathbf{D}_{t-1}) = \mathbf{a}_t = \mathbf{G}_t \mathbf{m}_{t-1}$ and $V(\boldsymbol{\theta}(t) | \mathbf{D}_{t-1}) = \mathbf{R}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t' + \mathbf{W}_t$.
- (c) Prediction Distribution: $(\boldsymbol{\mu}(t) | \mathbf{D}_{t-1}) \sim N(\mathbf{f}_t, \mathbf{Q}_t)$, where, $E(\boldsymbol{\mu}(t) | \mathbf{D}_{t-1}) = \mathbf{f}_t = \mathbf{F}_t' \mathbf{a}_{t-1}$ and $V(\boldsymbol{\mu}(t) | \mathbf{D}_{t-1}) = \mathbf{Q}_t = \mathbf{F}_t' \mathbf{R}_{t-1} \mathbf{F}_t + \mathbf{V}_t$.
- (d) Posterior at t : $(\boldsymbol{\theta}(t) | \boldsymbol{\mu}(t), \mathbf{D}_{t-1}) \sim N(\mathbf{m}_t, \mathbf{C}_t)$, where $\mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t(\boldsymbol{\mu}(t) - \mathbf{f}_t)$, $\mathbf{C}_t = \mathbf{R}_t - \mathbf{A}_t \mathbf{Q}_t \mathbf{A}_t'$, $\mathbf{A}_t = \mathbf{R}_t \mathbf{F}_t \mathbf{Q}_t^{-1}$.

In this stage, a closed form for $p(\boldsymbol{\mu}_n, \boldsymbol{\theta}_n, \lambda, \sigma^2 | z)$ does not exist which is encountered as a problem, where $\boldsymbol{\mu}_n = (\boldsymbol{\mu}(1), \dots, \boldsymbol{\mu}(n))$ and $\boldsymbol{\theta}_n = (\boldsymbol{\theta}(1), \dots, \boldsymbol{\theta}(n))$. Therefore; the Metropolis-Hastings algorithm is used in order to get the results of posterior and prediction distribution based on DLM.

Considering

$$\boldsymbol{\mu}(t) \sim (\mathbf{0}, \sigma^2 \mathbf{V}_\lambda), \quad \boldsymbol{\theta}(t) \sim (\mathbf{0}, \sigma^2 \mathbf{W}_\lambda),$$

where λ is a non-negative parameter that defines the covariance matrices \mathbf{V}_λ and \mathbf{W}_λ , and σ^2 is a common scale factor. Sample from conditional posterior distributions can be drawn as follows:

a) Sampling from $p(\boldsymbol{\theta}_n, \lambda, \sigma^2 | \boldsymbol{\mu}_n)$: The joint posterior distribution of $\boldsymbol{\theta}_n$, λ and σ^2 is given by

$$p(\boldsymbol{\theta}_n, \lambda, \sigma^2 | \boldsymbol{\mu}_n) = p(\lambda, \sigma^2) p(\boldsymbol{\theta}(0) | \mathbf{D}_0) \prod_{t=1}^n p(\boldsymbol{\mu}(t) | \boldsymbol{\theta}(t), \lambda, \sigma^2) p(\boldsymbol{\theta}(t) | \boldsymbol{\theta}(t-1), \lambda, \sigma^2).$$

Suppose $p(\lambda, \sigma^2) = p(\lambda)p(\sigma^2)$, that is, λ and σ^2 are independent. Using Bayes' theorem and the Markov structure of the model, the joint posterior

can be written as

$$\begin{aligned} p(\boldsymbol{\theta}_n, \lambda, \sigma^2 | \boldsymbol{\mu}_n) &= p(\lambda)p(\sigma^2) \prod_{t=1}^n p(\boldsymbol{\mu}(t) | \boldsymbol{\mu}_{t-1}, \lambda, \sigma^2) \\ &\quad \times p(\boldsymbol{\theta}(n-t) | \boldsymbol{\theta}(n-t+1), \lambda, \sigma^2, \boldsymbol{\mu}_n) \\ &= p(\boldsymbol{\theta}_n | \lambda, \sigma^2, \boldsymbol{\mu}_n) p(\sigma^2 | \lambda, \boldsymbol{\mu}_n) p(\lambda | \boldsymbol{\mu}_n). \end{aligned}$$

The joint posterior distribution of λ and σ^2 can be expressed as

$$\begin{aligned} p(\lambda, \sigma^2 | \boldsymbol{\mu}_n) &\propto p(\lambda)p(\sigma^2)p(\boldsymbol{\mu}(t) | \boldsymbol{\mu}_{t-1}, \lambda, \sigma^2) \\ &\propto p(\lambda)p(\sigma^2) \left(\frac{1}{\sigma^2}\right)^{-\frac{n}{2}} \prod_{t=1}^n |Q_t|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=1}^n e'_t Q_t^{-1} e_t\right\}. \end{aligned}$$

If σ^2 has an inverse Gamma prior distribution with shape parameter α and scale parameter β , then its posterior distribution is also an inverse Gamma with shape parameter $\alpha + \frac{n}{2}$ and scale parameter $\beta + \frac{1}{2} \sum_{t=1}^n e'_t Q_t^{-1} e_t$. Hence, the posterior density of λ can be written as

$$p(\lambda | \boldsymbol{\mu}_n) = \frac{p(\lambda, \sigma^2 | \boldsymbol{\mu}_n)}{p(\sigma^2 | \lambda, \boldsymbol{\mu}_n)} \propto p(\lambda) \prod_{t=1}^n |Q_t|^{-\frac{1}{2}} \left[\beta + \frac{1}{2} \sum_{t=1}^n e'_t Q_t^{-1} e_t\right]^{-(\alpha + \frac{n}{2})}.$$

Then, the posterior density of $\boldsymbol{\theta}_n$ is also given by

$$p(\boldsymbol{\theta}_n | \lambda, \sigma^2, \boldsymbol{\mu}_n) = p(\boldsymbol{\theta}(n) | \lambda, \sigma^2, \boldsymbol{\mu}_n) \prod_{t=1}^n p(\boldsymbol{\theta}(n-t) | \boldsymbol{\theta}(n-t+1), \lambda, \sigma^2, \boldsymbol{\mu}_n),$$

where each term in the product sign is a Normal distribution. A sample of $\boldsymbol{\theta}_n$ can be obtained recursively moving backwards from $\boldsymbol{\theta}(n)$ to $\boldsymbol{\theta}(1)$ (Huerta et al., 2004; Dou, 2008).

In Metropolis-Hasting algorithm, the transition kernel is a mixed distribution for the new state of the chain: $q(\cdot, \cdot)$, the proposal density and $\alpha(\cdot, \cdot)$, the acceptance probability. Since the parameter space is bounded by 0, lognormal distribution is chosen as proposed density. As Moller (2002) notes, the alternative to the random walk Metropolis considers the proposal move to be a random multiple of the current state. From the current state $\lambda^{(j-1)}$, ($j > 1$), the proposed move is given by $\lambda^* = \lambda^{(j-1)} e^Z$, where Z is drawn from a symmetric distribution, such as Normal. In other words, at iteration j , a new λ^* is sampled from this proposal, centered at the previously sampled $\lambda^{(j-1)}$, with a tuning parameter τ^2 as the variance of the

distribution of Z . Gamerman and Lopes (2006) suggests that the acceptance rate, i.e. the ratio of accepted λ^* to the total number of iterations, should be around 50%. So τ^2 is tuned to attain this rate. If the acceptance rate is too high, for example; 70% to 100%, then τ^2 is increased. Similarly, if the acceptance rate is too low, for example, 0 to 20%, then τ^2 is decreased to narrow down the search area for λ^* .

The Metropolis-Hastings algorithm proceeds as follows. Given $\lambda^{(j-1)}$, where $j > 1$,

- Draw λ^* from $LN(\lambda^{(j-1)}, \tau^2)$.
- Compute the acceptance probability $\alpha(\lambda^{(j-1)}, \lambda^*) = \min \left\{ 1, \frac{p(\lambda^* | \boldsymbol{\mu}_n)}{p(\lambda^{(j-1)} | \boldsymbol{\mu}_n)} \times \frac{q(\lambda^{(j)}, \lambda^*)}{q(\lambda^*, \lambda^{(j-1)})} \right\}$.
- Accept λ^* with probability $\alpha(\lambda^{(j-1)}, \lambda^*)$. In other words, sample $u \sim U(0, 1)$ and let $\lambda^{(j)} = \lambda^*$ if $\lambda^* < u$ and $\lambda^{(j)} = \lambda^{(j-1)}$, otherwise.

This algorithm is iteratively performed until it converges. Next, σ^2 is drawn from the inverse Gamma distribution $IG(\alpha + \frac{n}{2}, \beta + \frac{1}{2} \sum_{t=1}^n e_t' Q_t^{-1} e_t)$. We now sample $\boldsymbol{\theta}_n$ given the accepted λ , σ^2 and $\boldsymbol{\mu}_n$ by using the Kalman filter algorithm (Dou, 2008).

b) Sampling from $p(\boldsymbol{\mu}_n | \boldsymbol{\theta}_n, \lambda, \sigma^2)$: Use prior of $\boldsymbol{\mu}(t)$ which has been specified by observation equation, to create proposal distribution. Since the proposal distribution for $\boldsymbol{\mu}(t)$ in observation equation has been $\mathbf{N}_m(\mathbf{F}_t' \boldsymbol{\theta}(t), \mathbf{V}_t)$, by a basic property of the multivariate Normal distribution; it can be transformed to uni-variate Normal distribution.

An advantage of the dynamic spatial linear model is that any linear combination of independent normal DLMS is a normal DLM. The important consequence is that in most practical cases, a DLM can be composed into a linear combination of simple canonical DLMS, the good news being that a modeller only needs to master a very few simple DLMS in order to become proficient at model building. Another advantage of the component structure is robustness.

3 Application to Ozone Data

In this section, the proposed trend models influence on the covariance function features of ozone data will be evaluated. The data set considered in this

study includes the daily averages of ozone concentration in scale parts per billion (ppb), which were measured during year of 2009 at 9 different monitoring stations scattered irregularly in Tehran city. The locations of the monitoring stations are shown in Figure 1. This data are collected by particular and accurate systems by Air Quality Control Company (subsidiary of Tehran's municipality) and Tehran City Directorate of Environment. Since the data are recorded hourly, there are 24 observations in each day and we used their daily means.

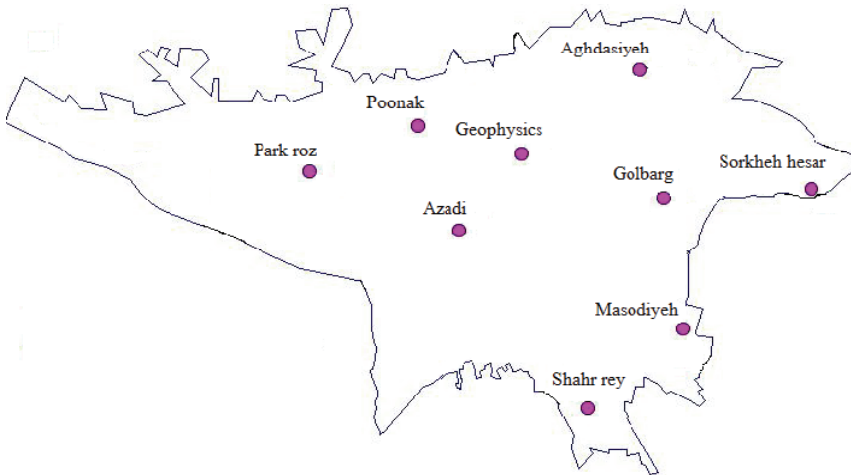


Figure 1. Locations of 9 monitoring stations in the city of Tehran

For considering the normality of ozone data, two most common transformations in the literatures are the square root and the natural logarithm. Histogram and normal QQ plots were plotted in Figure 2 on three measurement scales: original, square root and logarithmic. According to the Figure 2, The distributions of the original data and logarithmic data are not symmetric but the square root transformation converts them to a symmetric distribution. Also using Shapiro-Wilk test for examination of normality, the p-value only the square root transformation is more than 0.05.

Therefore a multivariate Normal distribution for the transformed data is considered. Let $Z(s_i, t_j)$ denotes the square root of observed ozone con-

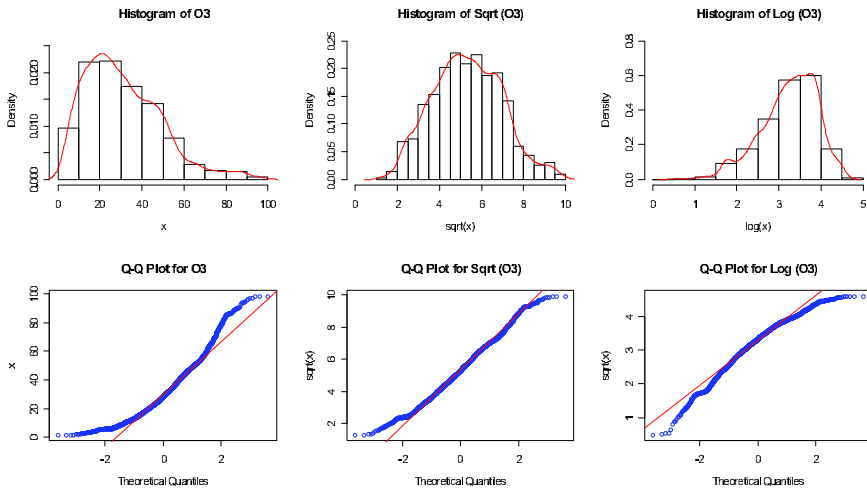


Figure 2. Histograms and normal Q-Q plots for original, square root and logarithmic data

centration, at locations s_i , $i = 1, \dots, 9$ and time $t_j \in \mathbf{T} = \{1, \dots, 365\}$. Ozone is formed by photochemical processes which involve nitrogen oxides (NOX) and volatile organic compounds (VOCs), in other words, Sunlight + VOC + NOX = Ozone. Since accessible covariates were only NO and NO_2 , the trend of the data were modeled using the proposed models in previous section.

Model 1. The trend of the data with two covariates is given by

$$\mu(\mathbf{s}, t) = \beta_0 + \beta_1 NO(\mathbf{s}, t) + \beta_2 NO_2(\mathbf{s}, t).$$

In many applications a Normal prior centered at zero with very large standard error, which is a non-informative prior distribution, is used for β . So the prior distributions $\beta \sim N(\mathbf{0}, 100\mathbf{I})$ and $\frac{1}{\sigma^2} \sim \text{Gamma}(0.01, 0.01)$ are chosen. The parameterization of the $\text{Gamma}(a, b)$ distribution is chosen so that its mean is $a/b = 1$ and its variance is very large. Using Gibbs sampling the estimates of the parameters are obtained as $\hat{\beta}' = (-1.069 \times 10^{-4}, -0.356, 0.109)$, $\hat{\sigma} = 1.0583$ and $R^2 = 0.28$, respectively.

Model 2. In this model β is a $3 \times (365 \times 9)$ dimensional matrix. Choosing multivariate Normal prior $N(\mathbf{0}, 100\mathbf{I})$ for β and inverse Gamma $IG(0.01, 0.01)$ for σ^2 , 5000 values were sampled by Gibbs algorithm, then the β and σ were

estimated, $\hat{\sigma} = 0.3098$. For the fitted model, $R^2 = 0.83$ was obtained, which is much more than the corresponding R^2 value for Model 1, that shows the dependency of the regression parameters to space locations and time points.

Model 3. To consider constant regression parameters and dependent errors, we assume the prior $\beta \sim N(\mathbf{0}, 100\mathbf{I})$, and $\delta(\mathbf{s}, t)$ to have separable spatial-temporal covariance function given by

$$C(\mathbf{h}, u) = \exp\left(-\frac{\|\mathbf{h}\|}{\phi_s}\right) \exp\left(-\frac{t}{\phi_t}\right), \quad \phi_s, \phi_t > 0,$$

where, ϕ_s is the range of the purely spatial covariance and ϕ_t is the range of purely temporal covariance. Usually a Uniform prior $U(a, b)$ is selected for these parameters, where a and b are minimum and maximum values of spatial (or temporal) lags, respectively. Since minimum and maximum values of space lag are 4914 and 24749 meters, by using Lambert transformation, we consider $U(0.051, 0.315)$ for ϕ_s . Since this data set is for 365 days, a Uniform prior $U(1, 364)$ is used for ϕ_t . Bayes estimate of the parameters are obtained as $\hat{\beta}' = (-1.721, -0.261, 0.1272)$, $\hat{\phi}_s = 0.1472$, $\hat{\phi}_t = 182.7008$ and $R^2 = 0.59$. The decrease of R^2 value in this model is likely for assuming that the regression coefficients are fixed.

Model 4. For ozone data, the number of days, stations and regression coefficients are $n = 365$, $m = 9$ and $q = 3$, respectively. In DLM trend model, we consider $\mathbf{G}_t = \mathbf{I}_{3 \times 3}$ and

$$\mathbf{F}'_t = \begin{bmatrix} 1 & NO(s_1, t) & NO_2(s_1, t) \\ \vdots & \vdots & \vdots \\ 1 & NO(s_9, t) & NO_2(s_9, t) \end{bmatrix}, \quad \boldsymbol{\theta}(t) = \begin{bmatrix} \theta_1(t) \\ \theta_2(t) \\ \theta_3(t) \end{bmatrix}$$

where by noting the evolution equation, each member of $\boldsymbol{\theta}(t)$ is simulated from autoregressive model given by

$$\theta_i(t) = \theta_i(t - 1) + \omega_i(t), \quad i = 1, 2, 3.$$

The prior distributions $\boldsymbol{\theta}(\mathbf{0}) | \mathbf{D}_0 \sim \mathbf{N}_3(\mathbf{0}, 100\mathbf{I})$, $\sigma^2 \sim IG(0.01, 0.01)$ and $\lambda \sim U(0.051, 0.315)$ are chosen. Next the MCMC algorithm was run for 10000 iterations. After a burn in period of 5000 iterations, the Bayes estimates of the parameters were obtained as $\hat{\theta}_0(t) = -0.0295$, $\hat{\theta}_1(t) = 1.3067$, $\hat{\theta}_2(t) = 0.1083$, $\hat{\sigma} = 0.158$, $\hat{\lambda} = 0.1584$, and $R^2 = 0.94$. The comparison between the coefficients of determinations for four models, shows that the

DLM is the best model for trend modeling of the ozone concentration in Tehran.

For considering stationarity in correlation structure of de-trended data, were plotted H-scatter plot and accepted stationary covariance function of de-trended data. The Shao and Li nonparametric method is used to test the symmetry and separability of space-time dependency of de-trended data (Shao and Li, 2009). Results of this test for different models summarized in Table 1.

Table 1. Results for the symmetry and separability tests for ozone data, “ + ” signifies the significance of the test

Trend		Test	
Model	R^2	Symmetry	Separability
1	0.28	+	+
2	0.83	+	+
3	0.59	+	+
4	0.94	+	-

Taking care of the coefficients of determinations obtained in this table, are realized that the model 4 is better than other, because it's coefficient of determination is more than other models. Therefore a symmetric, nonseparable and stationary covariance function is the more appropriate model for ozone data. Furthermore differences of the results for symmetry and separability tests among model 4 with previous three models show that each model may have different effect on the features of covariance function of the spatial-temporal data.

4 Conclusion

Since spatial-temporal data depend upon spatial locations and time points of observations, the models which consist constant coefficients or are independent from space and time cannot remove the trend of the data truly. Therefore, estimates of spatial-temporal covariance functions include large bias and these models could not be suitable for modeling the trend of spatial-temporal data. By comparing the coefficients of determinations obtained from de-trended ozone data, the inclusive model of dynamic regression coefficients and dependent spatial-temporal errors, is the most suitable model

for formulating the trend of data. The results of the symmetry and separability test for correlation structure of the de-trended data from different models show that a symmetric and nonseparable spatial-temporal covariance function characterizes the correlation structure of ozone concentration in Tehran.

Acknowledgement

We wish to thank to referees for their valuable comments and suggestions that improved the previous version of this paper. Partial support from Ordered and Spatial Data Center of Excellence of Ferdowsi University of Mashhad is also acknowledged.

References

- Behshad, E. and Mohammadzadeh, M. (2011). Evaluation of tests for separability and symmetry of spatio-temporal covariance function, *Journal of Statistical Research of Iran*, **8**, 1-27.
- Bernardo, J. and Smith, A. (1994). *Bayesian Theory*, Wiley, Chicester, UK.
- Cox, W.M. and Chu, S.H. (1992). Meteorologically adjusted trends in urban areas, a probabilistic approach. *Atmospheric Environment*, **27**, 425– 434.
- Cressie, N. (1993). *Statistics for Spatial Data*, Wiley, New York.
- Dou, Y. (2008). Dynamic Bayesian Models for Modeling Environmental Space-Time Fields, a Thesis Submitted for the Degree of Doctor of Philosophy, the University of British Columbia.
- Fuentes, M., Chen, L. and Davis, J.M. (2008). A class nonseparable and nonstationary spatial-temporal covariance function, *Environmetrics*, **19**, 487-507.
- Gamerman, D., and Lopes, H.F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2nd ed. Chapman and Hall/CRC Press, London.
- Huerta, G., Sanso, B. and Stroud, J.R. (2004). A spatio-temporal model for Mexico City Ozone levels, *Journal of the Royal Statistical Society*, **53**, 231-248.
- McMillan, N., Bortnick, S.M., Irwin, M.E. and Berliner, M. (2005). A hierarchical bayesian model to estimate and forecast Ozone through space and time, *Atmospheric Environment*, **39**, 1373-1382.
- Moller, J. (2002). *Spatial Statistics and Computational Methods*, Springer-Verlag.

- Ntzoufras, I. (2009). *Bayesian Modeling Using Win BUGS*, Department of Statistics, Athens.
- Porter, P.S., Rao, S.T., Zurbenko, I.G., Dunker, A.M. and Wolff, G.T. (2001). Ozone air quality over North America: Part II: An analysis of trend detection and attribution techniques. *Journal of Air and Waste Management Association*, **51**, 283–306.
- Rowe, D.B. (2003). *Multivariate Bayesian Statistics: Models for Source Separation and Signal Unmixing*, CRC Press, New York.
- Sahu, S.K., Gelfand, A.E. and Holland, D.M. (2007). High resolution space-time Ozone modeling for assessing trends, *Journal of American Statistical Society*, **102**, 1221–1234.
- Shao, X. and Li, B. (2009). A tuning parameter free test for properties of space-time covariance function, *Journal of Statistical Planning and Inference*, **13**, 4031-4038.
- Stroud, R.J., Muller, P. and Sanso, B. (2001). Dynamic models spatio-temporal data, *Journal of the Royal Statistical Society, Series B*, **63**, 673-689.
- Zheng, J., Swall, J., Cox, W.M. and Davis, J.M. (2007). Interannual variation in meteorological adjusted Ozone levels in the Eastern United State, *Atmospheric Environment*, **41**, 705-716.
- West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*, Springer-Verlag, New York.

Seyedeh Somayeh Mousavi

Department of Statistics,
Tarbiat Modares University,
Tehran, Iran.

Mohsen Mohammadzadeh

Department of Statistics,
Tarbiat Modares University,
Tehran, Iran.
email: *mohsen_m@modares.ac.ir*