Space-Time Generalized Additive Models

Ali M. Mosammam
Department of Statistics,University of Zanjan, Zanjan, IRAN-a.m.mosammam@znu.ac.ir

Abstract

If data have some temporal/spatial or spatial-temporal dependence and this is not modelled in the systematic part of the classic generalized additive model(GAM), the residuals from that model will also exhibit spatial or spatial temporal dependence, or in other words they will be spatially-temporally correlated. Such dependence would invalidate the theory that produces test statistics in the GAM models; because they were computed assuming independence. In this paper, we develop a space-time generalized additive model to take in to account such correlation. Our simulation results show the advantages of our proposed model.
Keywords: spatial-temporal model; generalized additive model; smoothing; interpolation.

1. Introduction

1 Space-Time Processes

This section is a review of some elementary concepts of spatial-temporal processes. A collection of spatial-temporal random variables

$$\{Y(\boldsymbol{s},t),\ (\boldsymbol{s},t)\in\mathbb{E}\} \tag{1}$$

is called a spatial-temporal *random field*, where the index set $\mathbb{E}$ is often a subset of $\mathbb{R}^d\times\mathbb{R}$. Assume that all first and second-order moments of the random field exist for each $(\boldsymbol{s},t)$. The random field (1) can be decomposed as $Y(\boldsymbol{s},t)=\mu(\boldsymbol{s},t)+\eta(\boldsymbol{s},t), \quad (\boldsymbol{s},t)\in\mathbb{E}$, where the deterministic mean $\mu(\boldsymbol{s},t)=E\left(Y(\boldsymbol{s},t)\right)$ represents drift and the probabilistic term $\eta(\boldsymbol{s},t)$ represents the error term. The spatial-temporal covariance between $Y(\boldsymbol{s}_1,t_1)$ and $Y(\boldsymbol{s}_2,t_2)$ is defined as

$$\sigma(\boldsymbol{s}_1,\boldsymbol{s}_2,t_1,t_2)=\mathrm{cov}\left(Y(\boldsymbol{s}_1,t_1),Y(\boldsymbol{s}_2,t_2)\right)=E\left\{\left(Y(\boldsymbol{s}_1,t_1)-\mu(\boldsymbol{s}_1,t_1)\right)\left(Y(\boldsymbol{s}_2,t_2)-\mu(\boldsymbol{s}_2,t_2)\right)\right\}.$$

If $\mathbb{E}=\mathbb{R}^d\times\mathbb{R}$, the spatial-temporal random field (1) is called *second-order stationary* when

$$\begin{aligned} E\left(Y(\boldsymbol{s},t)\right) &= \mu \qquad \text{for any } (\boldsymbol{s},t)\in\mathbb{E},\ \text{where } \mu \text{ is a constant},\\ \mathrm{cov}\left(Y(\boldsymbol{s},t),Y(\boldsymbol{s}+\boldsymbol{h},t+u)\right) &= \sigma(\boldsymbol{h},u) \qquad \text{for any } (\boldsymbol{s},t),\ (\boldsymbol{s}+\boldsymbol{h},t+u)\in\mathbb{E}.\end{aligned}$$

Furthermore, if $\sigma(\boldsymbol{h},u)$ is a function of only $|\boldsymbol{h}|$ and u, then $\sigma(\boldsymbol{h},u)$ is called an *isotropic* spatial-temporal covariance function. The stationary spatial-temporal random field $Z(\boldsymbol{s},t)$, is said to have a separable spatial-temporal covariance function if $C(\boldsymbol{h},u)=C_S(\boldsymbol{h})C_T(u)$, where $C_S(\boldsymbol{h})$ and $C_T(u)$ are purely spatial and purely temporal covariance functions respectively. Recently several authors have developed approaches to generate positive and negetive-value non-separable models for space-time processes. It is worth citing Cressie and Huang (1999), Gneiting (2002), Stein (2005) and Mosammam (2015). One class of spatially isotropic, fully symmetric, non-separable space-time models proposed by Gneiting (2002) has the form

$$C(\boldsymbol{h},u)=\frac{\sigma^2}{(au^{2\alpha}+1)^{\tau}}\exp\left\{\frac{-c|\boldsymbol{h}|^{2\gamma}}{(au^{2\alpha}+1)^{\beta\gamma}}\right\}. \tag{2}$$

Here a and c are time and space scaling parameters. $\tau>\beta d/2$. A separable covariance function is obtained when $\beta=0$. The parameters α and γ take values in $[0,1]$; the parameter β take values in $[0,1]$.

2 Space-Time Smoothing and Interpolation

Suppose that the spatial-temporal random field (1) does not include any drift. Consider the following *smoothing problem*: find $f \in H$, a reproducing kernel Hilbert space, to minimize

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f(\boldsymbol{s}_i, t_i))^2 + \lambda \|f\|^2, \tag{3}$$

where $\lambda \geq 0$ is called the smoothing parameter and $\|f\|$ is a norm in H defined by a spatial-temporal positive definite kernel $k(\boldsymbol{h}, u)$. The solution to the smoothing problem, called the *representer theorem*, Kimeldorf and Wahba (1971), has a representation of the form $f(\boldsymbol{s}, t) = \sum_{i=1}^{n} b_i k(\boldsymbol{s}_i, t_i, \boldsymbol{s}, t) = \boldsymbol{\tau}^\top \boldsymbol{b} = \boldsymbol{\tau}^\top M^{-1}\boldsymbol{Y}$, where $\boldsymbol{b} = M^{-1}\boldsymbol{Y}$, $M = K + n\lambda I_n$, $\boldsymbol{\tau} = (k((\boldsymbol{s}_1, t_1, \boldsymbol{s}, t), \ldots, k((\boldsymbol{s}_n, t_n, \boldsymbol{s}, t))^\top$ and $K = (k(\boldsymbol{s}_i, t_i, \boldsymbol{s}_j, t_j))$, $i, j = 1, \ldots, n$. Suppose now that our objective function includes a drift term lying in a drift space H_0, assumed to be a finite-dimensional subspace of H of dimension $m \geq 0$. Let $\{\phi_\ell(\boldsymbol{s}_i, t_i)\}_{\ell=1}^{m}$ form a basis of H_0. Let $(H_1, \langle\cdot,\cdot\rangle_1)$ be the orthogonal complement of H_0. Every $f \in H$ has a unique decomposition $f = f_0 + f_1$, where $f_0 \in H_0$ and $f_1 \in H_1$. Consider the smoothing problem: minimize

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f_0(\boldsymbol{s}_i, t_i) - f_1\boldsymbol{s}_i, t_i))^2 + \lambda \|f_1\|^2 \tag{4}$$

over $f_1 \in H_1$ and $f_0 \in H_0$. Assume the points $(\boldsymbol{s}_i, t_i)$, $i = 1, \ldots, n$ are arranged in $\mathbb{E}$ in such a way that the matrix U with the (i, ℓ)-th entry $\phi_\ell(\boldsymbol{s}_i, t_i)$ has full rank m. Straight forward extension of the solution of the problem by Kimeldorf and Wahba (1971) in such a case leads to a general form of the representer

$$f(\boldsymbol{s}, t) = \sum_{\ell=1}^{m} a_\ell \phi_\ell(\boldsymbol{s}, t) + \sum_{i=1}^{n} b_i k(\boldsymbol{s}_i, t_i, \boldsymbol{s}, t) = \boldsymbol{u}^\top \boldsymbol{a} + \boldsymbol{\tau}^\top \boldsymbol{b}, \tag{5}$$

where $\boldsymbol{\tau} = (k(\boldsymbol{s}, t, \boldsymbol{s}_1, t_1), \ldots, k(\boldsymbol{s}, t, \boldsymbol{s}_n, t_n))^\top$, $\boldsymbol{u} = (\phi_1(\boldsymbol{s}, t), \ldots, \phi_m(\boldsymbol{s}, t))^\top$, and the coefficients are given by $\boldsymbol{a} = \left(U^\top M^{-1} U\right)^{-1} U^\top M^{-1}\boldsymbol{Y}$, $\boldsymbol{b} = \left(M^{-1} - M^{-1}U\left(U^\top M^{-1}U\right)^{-1} U^\top M^{-1}\right)\boldsymbol{Y}$ where $M = K + n\lambda I_n$, $K = (k(\boldsymbol{s}_i, t_i, \boldsymbol{s}_j, t_j))$, $i, j = 1, \ldots, n$.
Next we review a general representation of simple kriging which is applied later to spatial-temporal residuals. Given the spatial-temporal random field (1), suppose that the mean of the random field is zero. Consider the problem of predicting $Y(\boldsymbol{s}_0, t_0)$ at new point $(\boldsymbol{s}_0, t_0)$ given the values of $Y(\boldsymbol{s}_i, t_i)$ at sites $(\boldsymbol{s}_i, t_i)$, $i = 1, \ldots, n$. Let $\boldsymbol{Y}(\boldsymbol{0}, \Sigma)$, where $\Sigma = (\sigma(\boldsymbol{s}_i, t_i, \boldsymbol{s}_j, t_j))$ is the covariance matrix between the data sites. We want to predict $Y_0 = Y(\boldsymbol{s}_0, t_0)$. Let $\boldsymbol{Y} = (Y(\boldsymbol{s}_1, t_1), \ldots, Y(\boldsymbol{s}_n, t_n))^\top$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^\top$, $\boldsymbol{\tau}_0 = (\sigma(\boldsymbol{s}_0, t_0, \boldsymbol{s}_1, t_1), \ldots, \sigma(\boldsymbol{s}_0, t_0, \boldsymbol{s}_n, t_n))^\top$. The kriging predictor takes the form $\hat{Y}_0 = \boldsymbol{\alpha}^\top \boldsymbol{Y} = \boldsymbol{\tau}_0^\top \Sigma^{-1}\boldsymbol{Y} = \boldsymbol{\tau}_0^\top \boldsymbol{b}$, where $\boldsymbol{\alpha} = \Sigma^{-1}\boldsymbol{\tau}_0$ and $\boldsymbol{b} = \Sigma^{-1}\boldsymbol{Y}$. The result is exactly equal to that obtained in Section 2, when the covariance σ of the process in this setting is the same as the kernel k in the setting of Section 2 and when $(\boldsymbol{s}_0, t_0)$ here plays the same role as $(\boldsymbol{s}, t)$. The prediction error variance becomes $\sigma^2 - \boldsymbol{\tau}_0^\top \Sigma^{-1}\boldsymbol{\tau}_0$. Then (5) can be rewritten as

$$\hat{Y}(\boldsymbol{s}, t) = \boldsymbol{u}^\top \hat{\boldsymbol{a}} + \boldsymbol{\tau}^\top M^{-1}(\boldsymbol{Y} - \boldsymbol{u}^\top \hat{\boldsymbol{a}}), \tag{6}$$

where $\hat{Y}(\boldsymbol{s}, t)$ is the predicted value at location $(\boldsymbol{s}, t)$, $\hat{\mu}(\boldsymbol{s}, t) = \boldsymbol{u}^\top \hat{\boldsymbol{a}}$ is the fitted drift, $\boldsymbol{u}$ is the vector of predictors, $\hat{\boldsymbol{a}}$ is the vector of estimated drift model coefficients and $\hat{\eta}(\boldsymbol{s}, t) = \boldsymbol{\tau}^\top M^{-1}(\boldsymbol{Y} - \boldsymbol{u}^\top \hat{\boldsymbol{a}})$ is the interpolated residuals where $\boldsymbol{\tau}^\top M^{-1}$ is the vector of simple kriging weights used to interpolate the fitted residuals $\boldsymbol{Y} - \boldsymbol{u}^\top \hat{\boldsymbol{a}}$. This representation suggest a nice interpolation technique, the drift and residuals can be estimated separately and then summed. The advantage of regression-kriging is that we can use any complex regression methods including generalized additive models.

3 Space-Time Generalized Additive Models

A Generalized additive models (GAM) (Hastie and Tibshirani,1990) has the following form $g(\boldsymbol{\mu}_i) = \mathbf{X}_i^*\boldsymbol{\beta} + \sum_{j=1}^{m} \mathbf{f}_j(\mathbf{x}_{ij})$ where $\mathbf{Y}_i$ has some exponential family distribution; $\boldsymbol{\mu}_i \equiv E(\mathbf{Y}_i)$; $\mathbf{X}_i^*$ is the ith row of the model matrix for the strictly parametric model components; $\mathbf{f}_j$ are smooth functions of the covariates $\mathbf{x}_j$. Much standard research has been devoted to fitting GAM with independent data using spline methods Green and Silverman (1994); Wood (2006)). However only limited work has been done for correlated data. If the data have some temporal/spatial or spatial temporal dependence and this is not modelled in the systematic part of the model, the residuals from that model will also exhibit spatial or spatial temporal dependence. Such dependence would invalidate the theory that produces test statistics in the GAM models; because they were computed assuming independence. There are two main options for handling such data: (1) model the spatial/temporal dependence in the systematic part of the model, i.e. by including a smooth of the spatial locations in the model. (2) relax the assumption of independence and estimate the correlation between residuals. In this paper we aim to apply penalized likelihood method to fit GAMs for nonseparable spatio-temporal data. In summary, if we model the spatial temporal dependence between observations then the residuals are more likely to be independent random variables and therefore not violate the assumptions of any classic inferential procedure. Consider the GAM

$$Y(\boldsymbol{s},t)=\sum_{j=1}^{m} f_j(\mathbf{x}_j(\boldsymbol{s},t))+\epsilon(\boldsymbol{s},t), \tag{7}$$

where $\mathbf{x}_j(\boldsymbol{s},t)$ are vectors of covariates and f_j are unknown smooth functions and $\boldsymbol{\epsilon}(\boldsymbol{s},t)$ are error terms with a multivariate normal distribution with mean zero and covariance $\Sigma_{\boldsymbol{\theta}}$. Consider the problem of estimating θ and f based on data $\mathbf{Y}=(Y(\boldsymbol{s}_1,t_1),\ldots,Y(\boldsymbol{s}_n,t_n))^T$ and $\mathbf{x}_j(\boldsymbol{s},t)=(\mathbf{x}_j(\boldsymbol{s}_1,t_1),\ldots,\mathbf{x}_j(\boldsymbol{s}_n,t_n))^T$ $j=1,\cdots,m$. This goal can be achieved by maximizing the penalized log likelihood

$$\ell_p=-\frac{n}{2}\log(2\pi)-\frac{1}{2}\log|\Sigma_{\boldsymbol{\theta}}|-\frac{1}{2}(\mathbf{Y}-\sum_{j=1}^{m}\mathbf{f}_j(\mathbf{x}_j))^T\Sigma_{\boldsymbol{\theta}}^{-1}(\mathbf{Y}-\sum_{j=1}^{m}\mathbf{f}_j(\mathbf{x}_j))-\frac{1}{2}\sum_{j=1}^{m}\lambda_j J_j(\mathbf{f}_j) \tag{8}$$

where $\mathbf{f}_j(\mathbf{x}_j)=[f(\mathbf{x}_j(\boldsymbol{s},t))]^T$, λ_j is the smoothing parameter controlling the tradeoff between the model fit and the smoothness of the regression function, and J_j is a roughness penalty functional.
For simplicity, let us suppose the case of a univariate Gaussian data, with spatial-temporally correlated response variable Y and regressor $\mathbf{x}$. We propose an iterative algorithm which maximizes the penalized log likelihood alternatively with respect to the covariance parameters and the smooth functions:

- Step 1: Start with estimating the drift model using GAM assuming independent errors.
- Step 2: Derive the residuals from an GAM assuming independent errors in Step1.
- Step 3:Calculate the initial value of $\boldsymbol{\theta}$, say $\boldsymbol{\theta}^{(0)}$ by fitting a variogram model to the residuals from an GAM assuming independent errors in Step2.
- Step 4: Then treat $\boldsymbol{\theta}$ as known and select the value of λ and estimate the smooth function f. For fixed $\boldsymbol{\theta}$, maximizing (8) becomes

$$\max_{f(x)}\left\{-\frac{1}{2}(\mathbf{Y}-\mathbf{f}(\mathbf{x}))^T\Sigma_{\boldsymbol{\theta}}^{-1}(\mathbf{Y}-\mathbf{f}(\mathbf{x}))-\frac{1}{2}\lambda J(\mathbf{f})\right\}.$$

It can be shown that for fixed $\boldsymbol{\theta}$ the solution of f to the above maximization problem is a spline if (Green and Silverman 1994). Hence the maximization problem is equivalent to the penalized weighted least squares problem

$$\min_{\boldsymbol{\beta}}\left\{\|\ \Sigma_{\boldsymbol{\theta}}^{-1/2}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})\ \|^2+\lambda\boldsymbol{\beta}^T\mathbf{S}\boldsymbol{\beta}\right\}=\min_{\boldsymbol{\beta}}\left\{\|\ \tilde{\mathbf{Y}}-\tilde{\mathbf{X}}\boldsymbol{\beta}\ \|^2+\lambda\boldsymbol{\beta}^T\mathbf{S}\boldsymbol{\beta}\right\}, \tag{9}$$

where $\Sigma_{\boldsymbol{\theta}}^{-1/2}$ is any square root matrix of $\Sigma_{\boldsymbol{\theta}}^{-1}$ such that $(\Sigma_{\boldsymbol{\theta}}^{-1/2})^T\Sigma_{\boldsymbol{\theta}}^{-1/2}=\Sigma_{\boldsymbol{\theta}}^{-1}$, $\tilde{\mathbf{Y}}=\Sigma_{\boldsymbol{\theta}}^{-1/2}\mathbf{Y}$ and $\tilde{\mathbf{X}}=\Sigma_{\boldsymbol{\theta}}^{-1/2}\mathbf{X}$. The smoothing parameter can be estimated by a number of criteria, e.g. cross validation (CV) or generalized cross validation (GCV). With known $\boldsymbol{\theta}$ and λ, the minimization problem (9) admits a unique solution, i.e., $\hat{\boldsymbol{\beta}}=(\mathbf{X}\Sigma_{\boldsymbol{\theta}}^{-1}\mathbf{X}+\lambda\mathbf{S})^{-1}\mathbf{X}^T\Sigma_{\boldsymbol{\theta}}^{-1}\mathbf{Y}$. Denote the selected λ and the corresponding estimated $\boldsymbol{\theta}$ as $\lambda^{(1)}$ and $\boldsymbol{\beta}^{(1)}$ respectively.

- Step 5: Let $\lambda = \lambda^{(1)}$ and $\boldsymbol{\beta} = \boldsymbol{\beta}^{(1)}$ and derive the residuals of GAM with correlated errors.
- Step 6:Try to find a new estimate of $\boldsymbol{\theta}$ using the residuals of GAM with correlated errors in Step 5. With λ and $\boldsymbol{\beta}$ fixed, maximizing (8) becomes

$$\max_{\boldsymbol{\theta}} \left\{ -\frac{1}{2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T \Sigma_{\boldsymbol{\theta}}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) - \frac{1}{2}\log|\Sigma_{\boldsymbol{\theta}}| \right\}.$$

  Clearly the solution is the MLE of $\boldsymbol{\theta}$ with $\boldsymbol{\beta}^{(1)}$ plugged in. Denote the new estimate as $\boldsymbol{\theta}^{(1)}$.
- Step 7: Stop if $\frac{\|\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(0)}\|}{\|\boldsymbol{\theta}^{(0)}\|} < \epsilon$, say $\epsilon = 10^{-5}$. Otherwise, let $\boldsymbol{\theta}^{(0)} = \boldsymbol{\theta}^{(1)}$ and repeat Steps 4–7.

4 Simulation study

Now, we investigate the empirical performance of the penalized likelihood estimation from the spatio-temporal data through a simulation study. The data are simulated from the model

$$Y(s_1, s_2, t) = f_1(s_1) + f_2(s_2) + b(s_1, s_2, t) + e(s_1, s_2, t), \; t = 1, 2, \cdots, T$$

where $s_1 \in [0, 1]$ and $s_2 \in [0, 1]$ are the spatial coordinates of a data point $\boldsymbol{s} = (s_1, s_2)$;$\mathrm{f}_1(s_1) = 2\sin(\pi s_1)$; $f_2(s_2) = e^{2s_2} - 13.4$ are smooth functions; $\mathbf{b}$ are spatial-temporally correlated errors with distribution $\mathbf{N}(\mathbf{0}, \mathbf{\Sigma})$; and $\mathbf{e}$ are pure measurement errors (producing a nugget effect) with distribution $\mathbf{N}(\mathbf{0}, \eta\mathbf{I})$. s_1 and s_2 are simulated independently from the uniform distribution on $[0, 1]$ with different sample size S. The space time random fields $\boldsymbol{b}$ are independently simulated on the same set of locations and from the same normal distribution $\mathbf{N}(\mathbf{0}, \mathbf{\Sigma})$ where Σ is defined by the covariance with (2), where $\tau = 1, a = c = 0.4, \sigma^2 = 4.5$, $\beta = 0, .9, \gamma = 1, \alpha = 1$. The measurement errors $\mathbf{e}$ are simulated from the normal distribution $\mathbf{N}(\mathbf{0}, \eta\mathbf{I})$ where $\eta = 0.2$. We analyze each simulated dataset in three ways: (1) a GAM assuming independent data; (2) a GAM assuming spatially temporally correlated data by ML estimation; and (3) a GAM assuming spatially temporally correlated data by REML estimation. The GAM with and without correlation structure are fitted to the simulated data by both ML and REML methods. The fitting results, which are based on 100 replicates for $S = 25, T = 20$; $S = 50, T = 10$ and $S = 100, T = 5$ (such that the total sample size is 500), are shown in Tables 1 and 2. As extremely large and extremely small estimates can be produced in estimating the covariance parameters, in addition to the sample means and mean squared error (MSE) we also include the medians and median absolute deviation (MAD) to assess the performance and to provide a more robust summary of simulation results. To evaluate the efficiency of estimation of the smooth functions, we calculated the mean square errors of the fitted functions and the 95 percent CI coverage proportion, which is the proportion of data points that are covered by their 95 percent CIs.
In summary, REML estimation are less stable than ML estimation. On the other hand, REML estimation are less biased than ML estimation. Also, the 95 percent CI coverage proportions from the REML estimation are higher than those from the ML estimation. As the spatial sample size increases and temporal sample size decreases such that the total sample size is fixed, both REML and ML estimation perform better with less bias and smaller variation. It is clear that the 95 percent CI coverage proportions in GAM without correlation structure are much lower while the MSEs are slightly larger than the cases where the spatial temporal correlation is explicitly modeled. Note that for all simulations in this paper the smoothing parameters are estimated by REML method. However, residuals from this model still presented some correlation and so a space time regression kriging model for errors was fitted. The residual were next analyzed for spatio-temporal correlation. Simple kriging was applied to the residuals from space time GAM, and, in order to obtain the final GK prediction, the estimated trend and residuals were combined into one final interpolation using residual space time variogram models parametrized as described in (2). The results show that the space-time regression kriging can explain a significant part of the variation in data. The results of 10–fold cross-validation show that use of spatio-temporal regression-kriging leads to significantly more higher coverage for data.

		Proposed Method				GAM assuming independent data	
		ML		REML			
		coverage	MSE	coverage	MSE	Coverage	MSE
n=25,T=20	Totall	43.91	0.0417	46.494	0.0418	25.974	0.0414
	f_1+f_2	97.48	0.0034	97.88	0.0034	93.96	0.0035
	f_1	89.48	0.0019	90.08	0.0019	79.08	0.0018
	f_2	88.32	0.0021	88.92	0.0021	79.72	0.0023
n=50,T=10	Totall	49.558	0.0397	52.708	0.0398	26.206	0.0389
	f_1+f_2	97.14	0.0041	97.66	0.0041	85.12	0.0046
	f_1	88.68	0.0021	89.48	0.0021	65.72	0.0023
	f_2	89.08	0.0022	89.76	0.0022	69.98	0.0026
n=100,T=5	Totall	56.454	0.0368	58.204	0.0369	26.714	0.0352
	f_1+f_2	97.37	0.0057	97.74	0.0058	74.47	0.0067
	f_1	85.8	0.0029	86.55	0.0029	53.06	0.0034
	f_2	91.11	0.0029	91.98	0.0029	62.83	0.0033

Proposed Method		ML				REML			
		Mean	Median	MSE	mad	Mean	Median	MSE	Mad
n=25,T=20	Nugget	0.452	0.2013	0.9584	0.2531	0.455	0.2070	0.9229	0.2618
	Scale_s	46424275	0.3632	1.034294e+17	0.0546	25563201	0.3702	1.760106e+16	0.0555
	Scale_t	0.944	0.2765	17.9577	0.1551	8419185	0.2927	7.087703e+15	0.1600
	Sill	3.875	3.9845	1.0092	0.4507	714538	4.0992	5.105177e+13	0.5015
n=50,T=10	Nugget	0.321	0.2051	0.437	0.1309	0.2127	0.3	0.453	0.1345
	Scale_s	12685215	0.3610	7.810499e+15	0.0638	0.3659	1826352605	3.277917e+20	0.1343
	Scale_t	0.354	0.2674	0.8189	0.1394	0.2787	559081089	3.072138e+19	0.0678
	Sill	3.865	3.9565	1.0203	0.5353	4.0691	142922780	2.010132e+18	0.4864
n=100,T=5	Nugget	0.1986	0.2053	0.01	0.0897	0.2033	0.2097	0.0101	0.0919
	Scale_s	0.3453	0.3372	0.0075	0.0617	0.3569	0.3475	0.0069	0.0645
	Scale_t	0.2531	0.2673	0.0467	0.1583	0.2766	0.2841	0.0378	0.1632
	Sill	3.7821	3.6827	0.9144	0.5875	3.8684	3.7438	0.828	0.5906

Available Method		ML-non separable				REML-non separable			
		Mean	Median	MSE	Mad	Mean	Median	MSE	Mad
n=25,T=20	Nugget	0.458	0.2161	0.9304	0.2622	0.458	0.2161	0.9304	0.2622
	Scale_s	24555073	0.3621	1.621618e+16	0.0513	24555073	0.3621	1.621618e+16	0.0513
	Scale_t	0.912	0.2687	16.0369	0.1529	0.912	0.2687	16.0369	0.1529
	Sill	3.852	4.0080	1.0364	0.4927	3.852	4.0080	1.0364	0.4927
n=50,T=10	Nugget	0.370	0.2204	0.609	0.1433	0.370	0.2204	0.609	0.1433
	Scale_s	30065394	0.3590	5.232274e+16	0.0698	30065394	0.3590	5.232274e+16	0.0698
	Scale_t	0.432	0.2566	1.3788	0.1448	0.432	0.2566	1.3788	0.1448
	Sill	3.776	3.8949	1.2182	0.5636	3.776	3.8949	1.2182	0.5636
n=100,T=5	Nugget	0.1997	0.2026	0.0102	0.0946	0.1997	0.2026	0.0102	0.0946
	Scale_s	0.3371	0.3312	0.009	0.0666	0.3371	0.3312	0.009	0.0666
	Scale_t	0.2380	0.2525	0.0576	0.2012	0.2380	0.2525	0.0576	0.2012
	Sill	3.7276	3.6092	1.055	0.6005	3.7276	3.6092	1.055	0.6005

References

Cressie, N. A. C. and Huang, H. (1999). Classes of non-separable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, 94:1330–1340.

Gneiting, T. (2002). Non-separable, stationary covariance functions for space-time data. *Journal of the American Statistical Association*, 97(458):590–600.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models. A Roughness Penalty Approach.* Chapman & Hall, London.

Kimeldorf, G. and Wahba, G. (1971). Some results on tchebycheffian spline functions. *Journal of Mathematical Analysis and Applications*, 33(1):82–95.

Mosammam, A. (2015). The reverse dimple in potentially negative-value spacetime covariance models. *Stochastic Environmental Research and Risk Assessment*, 29(2):599–607.

Stein, M. L. (2005). Statistical methods for regular monitoring data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67:667–687.

Wood, S. (2006). *Generalized additive models: an introduction with R.* CRC press.