Supplementary material for the paper: Smoothing spatio-temporal data with complex missing data patterns

Eleonora Arnone [1], Laura M. Sangalli [1], and Andrea Vicini [1]

[1] MOX - Dipartimento di Matematica, Politecnico di Milano, Italy

---

Address for correspondence: Laura M. Sangalli, MOX - Dipartimento di Matematica, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy.
E-mail: `laura.sangalli@polimi.it`.
Phone: (+39) 02 2399 4554.
Fax: (+39) 02 2399 4568.

Abstract: Supplemental material.

Key words: functional data with spatial dependence; incomplete and partially observed functional data; nonparametric regression with partial differential equation regularization; smoothing with roughness penalties

1 Proofs of asymptotic results

Proof of Theorem 1. We first prove that the ST-PDE estimator is consistent. Consider the mean square error for the estimator

$$\mathbb{E}(\hat{\mathbf{f}} - \mathbf{f})^2 = b_\nu(\lambda_{S,\nu}, \lambda_{T,\nu}) b_\nu(\lambda_{S,\nu}, \lambda_{T,\nu})^\top + \mathrm{Var}_\nu(\lambda_{S,\nu}, \lambda_{T,\nu})$$

where $b_\nu(\lambda_{S,\nu}, \lambda_{T,\nu})$ and $\mathrm{Var}_\nu(\lambda_{S,\nu}, \lambda_{T,\nu})$ are the bias and the variance of the estimator obtained with ν observations and $\lambda_{S,\nu}$ and $\lambda_{T,\nu}$ as smoothing parameters, respectively. To obtain the consistency, we need $\mathbb{E}(\hat{\mathbf{f}} - \mathbf{f})^2 \to 0$.

For the bias we have:

$$\begin{aligned} b_\nu(\lambda_{S,\nu}, \lambda_{T,\nu}) &= \mathbb{E}(\hat{\mathbf{f}}_\nu) - \mathbf{f} \\ &= \left[\left(B^\top Q B/\nu + \lambda_{S,\nu} P_S + \lambda_{T,\nu} P_T \right)^{-1} B^\top Q B/\nu - I \right] \mathbf{f} \\ &= \left[\left(A_\nu^{-1} + \lambda_{S,\nu} P_S + \lambda_{T,\nu} P_T \right)^{-1} A_\nu^{-1} - I \right] \mathbf{f} \\ &= -\lambda_{S,\nu} A_\nu P_s \mathbf{f} - \lambda_{T,\nu} A_\nu P_T \mathbf{f} + o\left(\lambda_{S,\nu} + \lambda_{T,\nu} \right) \end{aligned}$$

where we have used Taylor expansion and Assumption 1. Analogously, for the variance we have:

$$\begin{aligned}
\mathrm{Var}_\nu\left(\lambda_{S,\nu},\lambda_{T,\nu}\right) &= \frac{\sigma^2}{\nu}\left(B^\top QB/\nu+P\right)^{-1} B^\top QB/\nu\left(B^\top QB/\nu+P\right)^{-1}\\
&= \frac{\sigma^2}{\nu}\left(A_\nu^{-1}+P\right)^{-1} A_\nu^{-1}\left(A_\nu^{-1}+P\right)^{-1}\\
&= \frac{\sigma^2}{\nu}\left(I-A_\nu P\right)^{-1}\left(I-A_\nu P\right)^{-1} A_\nu = \frac{\sigma^2}{\nu}\left(I-A_\nu P\right)^{-2} A_\nu\\
&= \frac{\sigma^2}{\nu}\left(I-\lambda_{S,\nu}A_\nu P_S-\lambda_{T,\nu}A_\nu P_T\right)^{-2} A_\nu\\
&= \frac{\sigma^2}{\nu}A_\nu + o\left(\lambda_{S,\nu}+\lambda_{T,\nu}\right).
\end{aligned}$$

Under the hypothesis that both $\lambda_{S,\nu}$ and $\lambda_{T,\nu}$ goes to zero we thus have that $\mathbb{E}(\hat{\mathbf{f}}-\mathbf{f})^2 \to 0$ as $\nu\to\infty$.

In order to obtain the asymptotic distribution of $\sqrt{\nu}(\hat{\mathbf{f}}-\mathbf{f})$, first of all note that $\mathbb{E}(\sqrt{\nu}(\hat{\mathbf{f}}-\mathbf{f}))$ goes to zero only if $\lambda_{S,\nu}$ and $\lambda_{T,\nu}$ are $o(\nu^{-1/2})$. Then, recall that

$$\begin{aligned}
\hat{\mathbf{f}}_\nu &= \left(B^\top QB/\nu+P\right)^{-1} B^\top Qz/\nu\\
&= \left(B^\top QB/\nu+P\right)^{-1} B^\top Q(W\boldsymbol{\beta}+B\mathbf{f}+\boldsymbol{\epsilon})/\nu.
\end{aligned}$$

Rearranging the above equation and noting that QW has all entries equal to zero we obtain

$$\left(B^\top QB/\nu+P\right)(\hat{\mathbf{f}}_\nu-\mathbf{f})+P\mathbf{f} = B^\top Q\boldsymbol{\epsilon}/\nu.$$

The right side of the equation above can be used as a pivot to derive the asymptotic distribution of the estimator.

□

Proof of Theorem 2. Given $\hat{\mathbf{f}}$, the vector $\hat{\boldsymbol{\beta}}$ is the solution of the score equation

$$\frac{1}{\nu}W^\top(\mathbf{z}-W\hat{\boldsymbol{\beta}}-B\hat{\mathbf{f}})=0.$$

Recalling that $\mathbf{z}=W\boldsymbol{\beta}+\mathbf{f}+\boldsymbol{\epsilon}$, we get

$$\frac{1}{\nu}W^\top W(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})+\frac{1}{\nu}W^\top B(\hat{\mathbf{f}}-\mathbf{f})=\frac{1}{\nu}W^\top\boldsymbol{\epsilon}$$

where $\boldsymbol{\epsilon}$ represents the vector of i.i.d. errors. We thus obtain

$$\hat{\Sigma}_\nu(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})=\frac{1}{\nu}W^\top\boldsymbol{\epsilon}-\frac{1}{\nu}W^\top B(\hat{\mathbf{f}}-\mathbf{f}). \tag{1.1}$$

Recalling that, thanks to Theorem 1, $\hat{\mathbf{f}}$ is a consistent estimator for $\mathbf{f}$, we can use the right side of the equation above as a pivoting quantity to derive the asymptotic normality of $\hat{\boldsymbol{\beta}}$.

In order to prove the consistency of $\hat{\boldsymbol{\beta}}$ it remains to show that $\mathbb{E}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^2 \to 0$. By squaring and taking the expectation of equation (1.1), and recalling that the term $(\hat{\mathbf{f}} - \mathbf{f})$ is independent from $\boldsymbol{\epsilon}$ by construction, we have that $\mathbb{E}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^2$ converges to zero if both $\frac{1}{\nu^2}\mathbb{E}(\boldsymbol{\epsilon}^\top W W^\top \boldsymbol{\epsilon})$ and $\mathbb{E}(\hat{\mathbf{f}} - \mathbf{f})^2$ goes to zero. The first term goes to zero by the central limit theorem, while the second term goes to zero thanks to Theorem 1.

□

2 Test field on C-shaped domain

The test field $f(\mathbf{p}, t)$ on the C-shaped domain is defined as

$$f(x,y,t) = \begin{cases} \cos(t)(q+x) + (y-r)^2 & \text{if} \quad x \geq 0 \wedge y > 0 \\ \cos(2t)(-q-x) + (-y-r)^2 & \text{if} \quad x \geq 0 \wedge y \leq 0 \\ \cos(t)\left(-\arctan\left(\frac{y}{x}\right) r\right) + (\sqrt{x^2+y^2} - r)^2 K(x,y) & \text{if} \quad x < 0 \wedge y > 0 \\ \cos(2t)\left(-\arctan\left(\frac{y}{x}\right) r\right) + (\sqrt{x^2+y^2} - r)^2 K(x,y) & \text{if} \quad x < 0 \wedge y \leq 0 \end{cases}$$

where

$$K(x,y) = \left(\frac{y}{r_0} 1_{|y| \leq r_0 \wedge x > -r} + 1_{|y| > r_0 \vee x \leq -r}\right)^2$$

and $r_0 = 0.1$, $r = 0.5$ and $q = \pi r/2$.

3 Test field on squared domain

The test field $f(\mathbf{p}, t)$ on the squared domain is shown in Figure 12 and is defined as

$$\begin{aligned} f(x,y,t) = &\sin\left(2\pi\left(h(y)x\cos\left(\frac{9}{5}t - 2\right) - y\sin\left(\frac{9}{5}t - 2\right)\right)\right) \\ &\cdot\cos\left(2\pi\left(h(y)x\cos\left(\frac{9}{5}t - 2 + \frac{\pi}{2}\right) + h(x)y\sin\left(\left(\frac{9}{5}t - 2\right)\frac{\pi}{2}\right)\right)\right) \end{aligned}$$

where $h(x) = \frac{1}{2}\sin(5\pi x)e^{-x^2} + 1$.

3.1 Tests for different signal-to-noise ratio

Table 3 and 4 contain the results for simulation studies on the square domain, as in Section 5.5 of the main paper, but with different signal-to-noise ratio, setting the standard deviation σ of the error respectively to $0.20 \cdot$ `range(exact data)` and to $0.01 \cdot$ `range(exact data)`. As for the simulation case considered in Section 5.5 of the main paper, ST-PDE achieves the best results in terms of RMSE, MAE, CRPS and INT in all cases, with a stronger

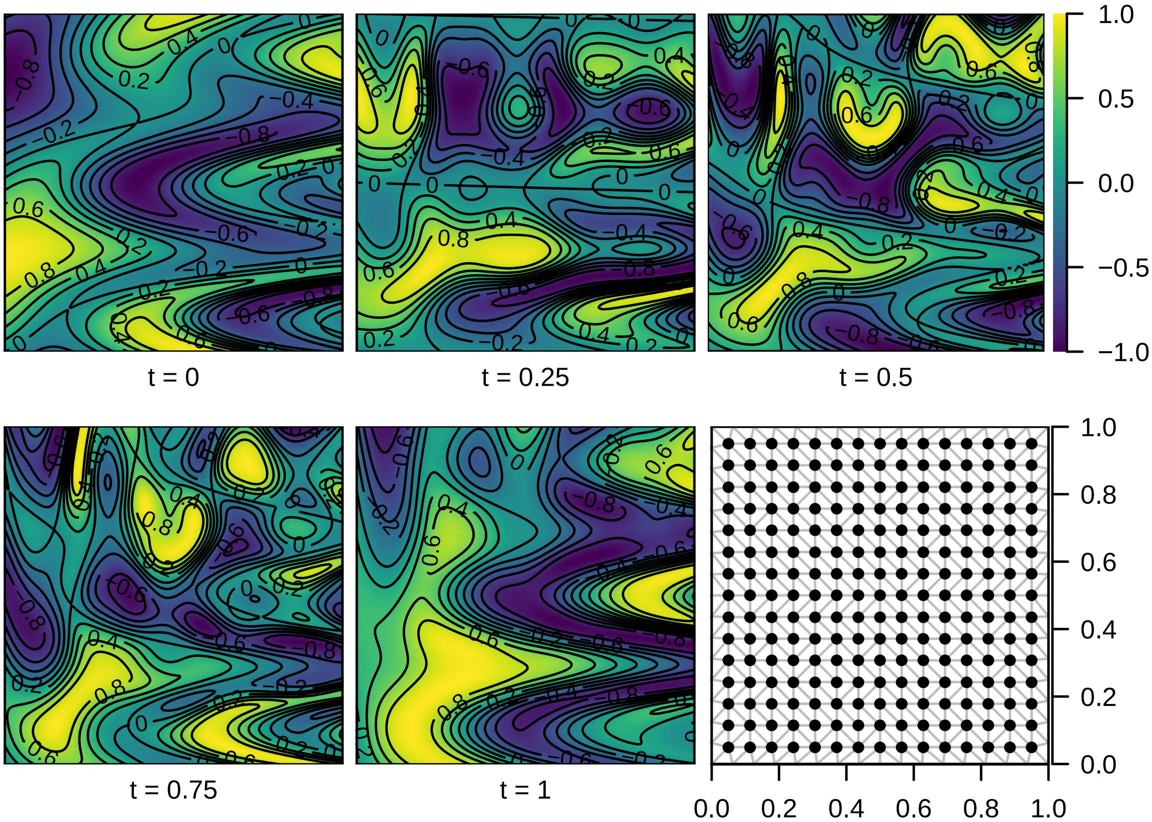

Figure 12: Test field on the c-shaped domain. The field exhibits a complicated behaviors and localized features. The last panel shows the observational spatial grid corresponding to the full (uncensored) design, and the triangulation used for ST-PDE estimate.

advantage over the competing methods for high signal-to-noise ratio. TPS, SOAP and INLA have comparable performances in terms of RMSE and MAE. INLA always achieves the best CVG, with coverage slightly higher then the nominal value 0.95. ST-PDE is the second best method in terms of CVG. However, as already noted in the main paper the standard parametric Wald-type confidence intervals here considered have low coverage for the nonparametric regressors (TPS, SOAP and ST-PDE); this is a well known issue, and several approaches can be followed to construct better confidence intervals, such as bias correction and undersmoothing strategies.

Method	RMSE	MAE	CRPS	INT	CVG
KRIG	0.367 (0.004)	0.539 (0.003)	0.252 (0.003)	7.155 (0.106)	0.483 (0.006)
TPS	0.221 (0.004)	0.416 (0.004)	0.125 (0.002)	1.255 (0.027)	0.861 (0.005)
SOAP	0.228 (0.003)	0.420 (0.004)	0.129 (0.003)	1.429 (0.096)	0.831 (0.013)
DINEOF	0.370 (0.014)	0.539 (0.014)	-	-	-
INLA	0.222 (0.003)	0.416 (0.003)	0.126 (0.001)	1.171 (0.017)	**0.983** (0.002)
ST-PDE	**0.193** (0.005)	**0.389** (0.005)	**0.109** (0.003)	**1.080** (0.054)	0.849 (0.014)

Table 3: Test on square domain, censoring (a), noise standard deviation $\sigma = 0.20 \cdot$ `range(exact data)`. For each competing method we report the median (and inter quartile range), over 20 simulation repetitions, of Root Mean Square Error (RMSE), mean absolute error (MAE), continuous rank probability score (CRPS), interval score (INT), interval coverage (CVG).

Method	RMSE	MAE	CRPS	INT	CVG
KRIG	0.081 (0.004)	0.211 (0.004)	0.041 (0.001)	0.717 (0.014)	0.495 (0.002)
TPS	0.160 (0.001)	0.336 (0.002)	0.089 (0.001)	1.504 (0.019)	0.727 (0.004)
SOAP	0.170 (0.001)	0.346 (0.001)	0.096 (0.001)	1.789 (0.029)	0.683 (0.007)
DINEOF	0.066 (0.021)	0.166 (0.015)	-	-	-
INLA	0.148 (0.001)	0.308 (0.004)	0.077 (0.001)	0.894 (0.014)	**0.955** (0.004)
ST-PDE	**0.038** (0.003)	**0.142** (0.002)	**0.015** (0.001)	**0.196** (0.021)	0.922 (0.009)

Table 4: Test on square domain, censoring (a), noise standard deviation $\sigma = 0.01 \cdot$ `range(exact data)`. Results reported as in Table 3.