Electronic Journal of Statistics
Vol. 0 (0000)
ISSN: 1935-7524
DOI: 10.1214/154957804100000000

Supplementary material for: A Universal Kriging predictor for spatially dependent functional data of a Hilbert Space

Alessandra Menafoglio and Piercesare Secchi

MOX - Department of Mathematics
Politecnico di Milano
Piazza Leonardo da Vinci 32, 20133 Milano, Italy
e-mail: alessandra.menafoglio@polimi.it
e-mail: piercesare.secchi@polimi.it

Matilde Dalla Rosa

Eni S.p.A, Exploration & Production Division
Via Emilia 1, San Donato (MI), Italy
e-mail: matilde.dalla.rosa@eni.com

1. Simulation Study

S.1. Simulation of Non-stationary Functional Processes

The simulation of functional stochastic processes $\{\boldsymbol{\chi_s}, \boldsymbol{s} \in D\}$ of the form (13) can be performed first by simulating a second-order stationary and isotropic residual field $\{\boldsymbol{\delta_s}, \boldsymbol{s} \in D\}$ by direct construction as in (11), then by generating the drift term and finally by summing residual and drift term.

For this Section, the residual fields have been simulated considering the space $H = L^2([0,1])$, endowed with the Fourier orthonormal basis $\{e_j, j \geq 1\}$. Expansion (11) has been truncated to the 7^{th} order for the first dataset of Section 3 and for the 5 collections of datasets analyzed in Subsection S.2; for the second dataset of Section 3, expansion (11) has been truncated to the 25^{th} order, assuming all the coefficients to be zero except for the last seven. The generation of the 7 non-null scalar fields $\{\xi_j(\boldsymbol{s}), \boldsymbol{s} \in D\}$, $1 \leq j \leq 7$, involved in expansion (11) –which in fact determine the structure of spatial dependence of the functional random field– has been performed by means of the geostatistical software ISATIS®. In particular, each scalar field has been independently simulated on a fine grid over the domain $D = [0,2] \times [0,3] \subset \mathbb{R}^2$, according to a gaussian second-order stationary and isotropic distribution; the generating variograms are listed in Table 1. Having obtained the functional residuals over the whole grid by combining the scalar grid realizations, the residual datasets have been finally obtained by sampling uniformly $n = 100$ grid locations.

TABLE 1
Generating variogram models for the scalar fields $\{\xi_j(\boldsymbol{s})\}$, $j = 1, ..., 7$. *Fields* ξ_5, ξ_6, ξ_7 *are generated by the sum of the indicated variogram structures.*

	Structure	(Sill, Range, Nug.)
ξ_1	Exp.	(16, 0.75, 0)
ξ_2	Sph.	(16, 0.75, 0)
ξ_3	Exp.	(16, 1.50, 0)
ξ_4	Sph.	(16, 1.50, 0)
ξ_5	Sph.; Exp.	(8, 0.75, 0); (8, 0.75, 0)
ξ_6	Sph.; Exp.	(8, 0.75, 0); (8, 0.75, 0)
ξ_7	Sph.; Exp.	(12, 1.50, 0); (4, 0.75, 0)

TABLE 2
Coefficients $\beta_{j,l}$, $1 \leq j \leq 7$, $0 \leq l \leq 5$ *of the drift expansion* (S.2) *used for the construction of the complete model and the relative sub-models.*

	$l=0$	$l=1$	$l=2$	$l=3$	$l=4$	$l=5$
$\beta_{1,l}$	1.247	-5.050	4.011	0.389	1.734	1.572
$\beta_{2,l}$	0.979	1.651	-1.531	1.535	0.086	0.710
$\beta_{3,l}$	0.558	4.008	3.096	0.289	1.246	0.502
$\beta_{4,l}$	-0.047	-0.020	0.045	-0.031	0.008	-0.001
$\beta_{5,l}$	0.032	0.022	-0.024	-0.005	-0.008	-0.047
$\beta_{6,l}$	0.029	0.028	0.033	0.046	0.047	-0.0002
$\beta_{7,l}$	0.063	0.042	0.016	0.109	0.057	0.004

For Subsection 3, only stationary datasets –obtained directly from the residuals realizations– have been considered. For Subsection S.2, non-stationary datasets have been built instead. For generating the drift terms, polynomials of degree lower than two have been considered:

$$m_{\boldsymbol{s}}(t) = a_0(t)+a_1(t)x+a_2(t)y+a_3(t)x^2+a_4(t)y^2+a_5(t)xy,\ t \in [0,1], \boldsymbol{s} = (x,y) \in D, \tag{S.1}$$

where a_l are deterministic functional (possibly null) coefficients belonging to L^2. For the construction of a_l, $l = 0, ..., 5$, the same basis with the same truncation as for the residuals has been fixed:

$$a_l(t) = \sum_{j=1}^{7} \beta_{j,l} e_j(t), \quad t \in \mathcal{T}, \tag{S.2}$$

where $\beta_{j,l} \in \mathbb{R}$, $1 \leq j \leq 7$, $0 \leq l \leq 5$, are the deterministic coefficients of the expansion on the Fourier basis. In Table 2 the coefficients $\beta_{j,l}$ relative to the complete model are listed; drift models used in the considered synthetic examples are obtained as sub-models of the complete model as will be specified later on.

S.2. Model Selection Procedure

The first goal of our simulation study is to evaluate the performance of Algorithm 11 in terms of error of model selection (referred to as *model misclassifica-*

tion) –which occurs when the selected model does not coincide with that generating the data– and type of error (in particular over-fitting or under-fitting[1]).

In order to analyze the behavior of the algorithm in different scenarios, pointing out possible tendency to over-fit or under-fit the data, 5 collections of 32 datasets each have been considered.

Data generation follows this scheme. First a set of 32 drift models has been built by considering the complete drift model (S.1) –constructed as previously specified and evaluated in the sampled locations $\boldsymbol{s}_1, ..., \boldsymbol{s}_{100}$– and its 31 sub-models. In particular, for $t \in [0,1]$, $\boldsymbol{s} = (x, y)$, drift model $k = 1, ..., 32$, has been represented through five binary variables $\{\zeta_1^{(k)}, \zeta_2^{(k)}, ..., \zeta_5^{(k)}\}$, by representing:

$$m_{\boldsymbol{s}}^{(k)}(t) = a_0(t) + \zeta_1^{(k)} a_1(t)x + \zeta_2^{(k)} a_2(t)y + \zeta_3^{(k)} a_3(t)x^2 + \zeta_4^{(k)} a_4(t)y^2 + \zeta_5^{(k)} a_5(t)xy, \tag{S.3}$$

and setting $\zeta_l^{(k)} = 1$ if the l-th regressor is included in the sub-model k, $\zeta_l^{(k)} = 0$ otherwise, for $l = 1, ..., 5$. Hence, the sub-models have been ordered according to the bijective relation among the binary numbers $\zeta_5^{(k)}\zeta_4^{(k)}\zeta_3^{(k)}\zeta_2^{(k)}\zeta_1^{(k)}$, $k = 1, ..., 32$, and their decimal representations plus 1 (the constant term is always included): the complete model is thus model $32 = 1 + 31$ ($31 \leftrightarrow 11111$), the spatially constant model is model 1 ($0 \leftrightarrow 00000$), while, for example, sub-model $18 = 1 + 17$ is the model with regressors $\{1, x, xy\}$ ($17 \leftrightarrow 10001$).

Given the set of drift terms, the first collection of 32 functional datasets, $\{\chi_{\vec{\boldsymbol{s}}}^{(k,1)}, k = 1, ..., 32\}$, has been obtained by summing to each drift sub-model $m_{\vec{\boldsymbol{s}}}^{(k)}$ the residuals generated as specified in Subsection S.1 (Figure 1, left panel); the remaining 4 collections $\{\chi_{\vec{\boldsymbol{s}}}^{(k,r)}, k = 1, ..., 32\}$, $r = 2, 3, 4, 5$, have been obtained with the same construction but dividing the residual realization by r (Figure 1, four panels on the right):

$$\chi_{\vec{\boldsymbol{s}}}^{(k,r)} = m_{\vec{\boldsymbol{s}}}^{(k)} + \delta_{\vec{\boldsymbol{s}}}/r,$$

which in fact corresponds to a reduction by a factor $r = 2, ..., 5$ of the variogram sills reported in Table 1.

To illustrate how the datasets collections depend on the amplitude of the stochastic component, consider the drift model 18, which is of the form

$$m_{\boldsymbol{s}}^{(18)}(t) = a_0(t) + a_1(t)x + a_5(t)xy, \tag{S.4}$$

and consider the corresponding non-stationary data $\chi_{\vec{\boldsymbol{s}}}^{(18,r)} = m_{\vec{\boldsymbol{s}}}^{(k)} + \delta_{\vec{\boldsymbol{s}}}/r$, $r = 1, ..., 5$ (Figure 2, upper panels). The increasing importance of the drift term is made already explicit by graphical inspection, but it is even more stressed by the empirical estimate of the semivariogram computed from the data (Figure 2, lower panels). Indeed, for higher levels of residuals amplitude ($r = 1$), the semivariogram is only slightly affected by the drift, presenting an almost

[1] We say that the algorithm finds an over-fitting solution when it selects a drift model including all the generating regressors plus at least one; analogously, we say that the algorithm selects an under-fitting solution when it selects a sub-model of the true model.

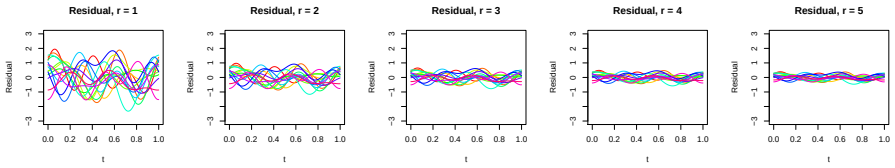


FIGURE 1. *First 15 residual curves for* $r = 1, 2, 3, 4, 5$ *(from left to right). Curves in the panel* r*,* $r = 2, 3, 4, 5$*, are obtained dividing the first panel residuals by* r*, which is the same as considering the same realization of the first residual field but with a lower sill for the generating variogram, namely a sill divided by* r*.*

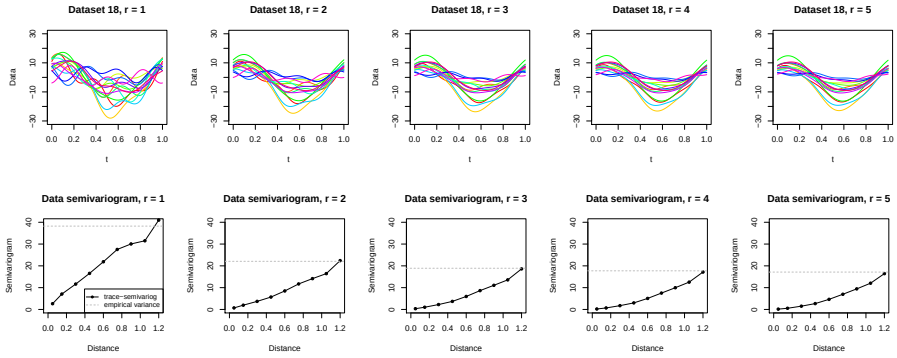


FIGURE 2. *From left to right: First 15 data of dataset* $\chi_{\vec{s}}^{(18,r)}$ *(upper panel) and empirical semivariogram computed from the dataset (lower panel), for* $r = 1, 2, 3, 4, 5$*.*

stationary behavior (e.g., downwards concavity near the origin, presence of a horizontal asymptote for higher distances); on the contrary, for decreasing amplitude of the residuals ($r = 2, 3, 4, 5$), the experimental semivariograms assume a non-stationary aspect (e.g., upwards concavity near the origin, super-quadratic growth for higher distances). This behavior is mainly due to the increasing influence of the drift term on the data, since the empirical semivariogram estimator (20) computed from non-stationary data becomes severely biased when the drift component is predominant with respect to the residual one.

The generated collection of datasets has been used for testing the procedure as follows. For each collection r, $r = 1, ..., 5$, the model selection step has been separately applied to each of the corresponding 32 datasets, considering as candidate models all the 32 polynomials of degree lower than two, namely the complete model (S.1) and all its sub-models, and setting the number of the GLS iterations equal to $M = 5$, which seems sufficient for Algorithm 10 to converge. For each dataset $\chi_{\vec{s}}^{(k,r)}$, $k = 1, ..., 32$, the selected model and, in case of model misclassification, the type of error (over-fitting, under-fitting or none of them)

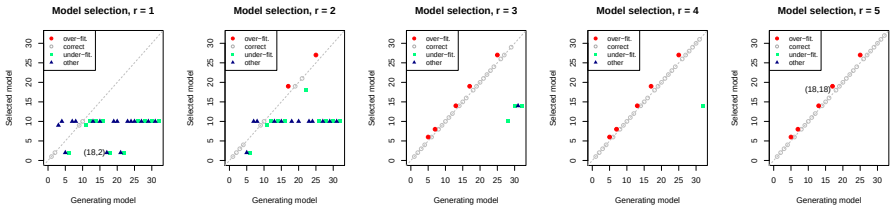


FIGURE 3. *Simulation results for the model selection algorithm. From left to right: results applying Algorithm 11 to the collection of 32 datasets* $\{\chi_{\vec{s}}^{(k,r)}, k = 1, ..., 32\}$ *for* $r = 1, 2, 3, 4, 5$. *The horizontal axis identifies the number of the generating model, the vertical axis the number of the selected model; grey empty dots indicate correct selection, red full dots over-fitting, green square dots under-fitting, blue triangular dots the other cases. The points (18,2) –under-fitting– and (18,18) –correct selection– correspond to the dataset* $\chi_{\vec{s}}^{(18,r)}$, *for* $r = 1, 5$.

has been recorded. Simulation results are shown in Figure 3. It is clear that the number of misclassified models sensibly decreases when the residual amplitude decreases. Indeed, as the r parameter increases, the drift term becomes more significant in the prediction: we thus expect a better performance of Algorithm 11 in cases high signal-to-noise ratio ($r = 3, 4, 5$).

Consider now the behavior of the procedure in terms of over-fitting or under-fitting. Figure 3 shows that in response to a decrease in the residual amplitude ($r = 3, 4, 5$) the behavior moves mainly from under-fitting to correct selection, except for a few cases in which over-fitting occurs (for $r = 5$, only datasets 5, 7, 13, 17, 25 are slightly over-fitted).

What is even more interesting to notice is that in very critical scenarios ($r = 1, 2$), the most common misclassification error is under-fitting (Figure 4, first and second panels). As an example, consider the dataset $\chi_{\vec{s}}^{(18,1)}$ (Figure 4, first panel). Comparing the non-stationary data with the drift curves (Figure 4, second panel), it is evident that the residuals heavily affect the shape of the curves, making them almost indistinguishable from the curves of dataset $\chi_{\vec{s}}^{(2,1)}$ (Figure 4, third panel). Here under-fitting occurs, but, as a matter of fact, drift models 2 and 18 are equally likely for the dataset $\chi_{\vec{s}}^{(18,1)}$: indeed, the larger fluctuations for t in $[0, 0.2]$ presented by dataset 18 –due to the xy component in drift model 18, which is missing in drift model 2– might be due to the residual fluctuation, and thus the simplest model is selected. Hence, in the presence of highly correlated residuals, Algorithm 11 proves to be very parsimonious, which is a very desirable property for a model selection procedure.

S.3. Drift Estimation and Universal Kriging Prediction

Our next goals are the analysis of the performance of Algorithm 10 and the evaluation of Universal Kriging predictions.

Simulations have been performed on the data generated before, focusing on

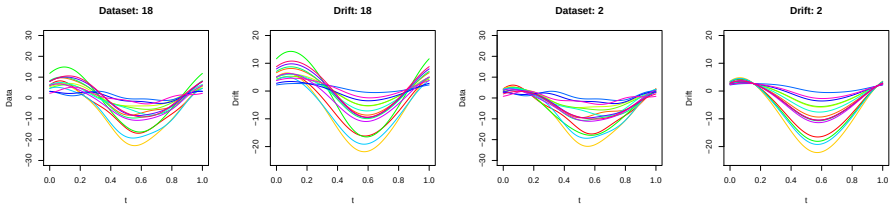


FIGURE 4. *From left to right: first 15 data of dataset* $\chi_{\vec{s}}^{(18,1)}$ *and corresponding drift (first and second panels), first 15 data of dataset* $\chi_{\vec{s}}^{(2,1)}$ *and corresponding drift (third and fourth panels).*

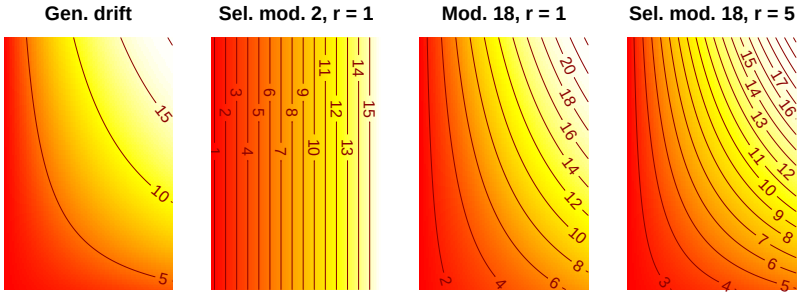
drift model 18 in the presence of residuals $\delta_{\vec{s}}/r$, with $r = 1, 5$ (Figure 1, first and fifth panels). Dataset $\chi_{\vec{s}}^{(18,1)}$ has been used in order to evaluate the effect of under-fitting both on drift estimates and on Universal Kriging predictions; in particular, Algorithm 10 and Universal Kriging prediction have been carried out first having selected the drift model by means of Algorithm 11 and then assuming that the true drift model is known. Dataset $\chi_{\vec{s}}^{(18,5)}$ has been used in order to study the influence of the residuals amplitude on the results, analyzing the performance of the procedures on less noisy data.

In all simulations the number of GLS iterations has been fixed to $M = 5$, which proved to be sufficient for Algorithm 10 to converge. For each of the three cases sketched before, the drift estimation, as well as the Universal Kriging prediction, has been performed over the whole generated grid for every t in $[0, 1]$. Since it is hardly possible to show at once a space-time grid of values –which is 4-dimensional– we consider two kind of visualizations: the functional visualization, obtained by plotting $t \in [0, 1]$ on the horizontal axis, and the value $\chi_{\boldsymbol{s}}(t)$ on the vertical axis for different $\boldsymbol{s} \in D$ –thus ignoring the spatial location– and the space contour representation, obtained by slicing the 4D space-time grid at some fixed t –thus loosing the functional variation–.

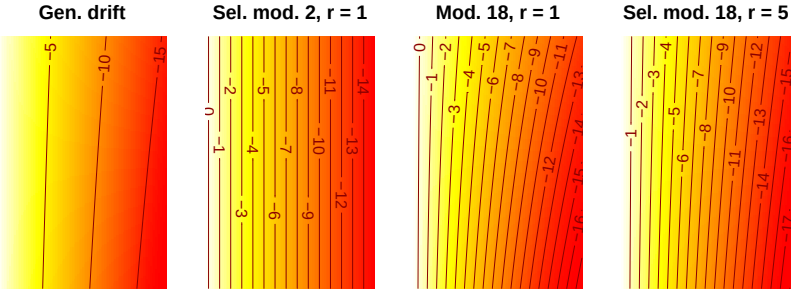
Figure 5 shows the contour plots of the GLS drift estimation, for all the three considered situations, namely for dataset $\chi_{\vec{s}}^{(18,1)}$ with drift model 2 –selected by Algorithm 11– (second panels of Figure 5a and 5b), with drift model 18 (third panels of Figure 5a and 5b) and for dataset $\chi_{\vec{s}}^{(18,5)}$ with drift model 18 (fourth panels of Figure 5a and 5b). Recall that both datasets are characterized by the same generating drift model.

It is clear from Figure 5a that the error in the drift estimate is not negligible in most critical situations, namely for $r = 1$, $t \in [0, 0.2]$ and when model 2 is chosen by Algorithm 11 (Figure 5a second panel; recall Figure 4), and it becomes even more severe where the non-linear behavior of the drift is more apparent, as in the bottom-right part of the spatial domain. However, for higher values of t, the linear model 2 seems appropriate (Figure 5b second panel), although it is more parsimonious than the generating one.

The choice of the drift model has consequences not only on the drift maps, but

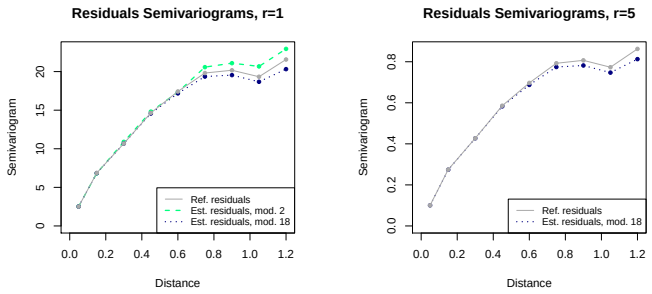


(a) $t = 0.1$



(b) $t = 0.7$

FIGURE 5. *Comparison of true and estimated drift through contour plots, for* $t = 0.1, t = 0.7$. *In each sub-figure, from left to right: generated drift grid, drift estimated with Algorithm 10 from* $\chi_{\vec{s}}^{(18,1)}$ *with model 2 –selected with Algorithm 11–, from* $\chi_{\vec{s}}^{(18,1)}$ *with model 18 and from* $\chi_{\vec{s}}^{(18,5)}$ *with model 18 –selected with Algorithm 11–.*



(a) $r = 1$ (b) $r = 5$

FIGURE 6. *Residuals semivariograms: empirical semivariogram computed from generated residuals (solid grey lines), empirical semivariogram of the estimated residuals with model 18 (dotted blue lines), semivariogram of the residuals estimated with drift model 2 (only in left panel, dashed green line).*

TABLE 3
Cross-validation squared error for $\chi^{(18,1)}_{\vec{s}}$, considering model 2 –selected by Algorithm 11– and model 18.

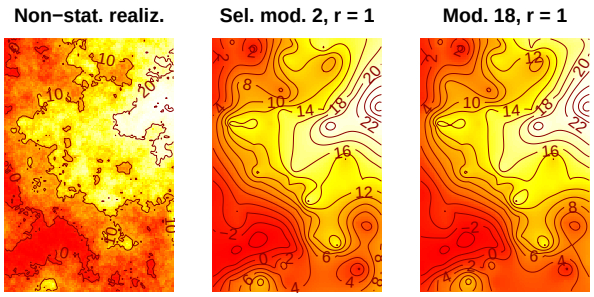
	Selected Model: 2	Correct Model: 18
Median	5.19	5.33
Mean	7.47	7.54

also on the residuals variogram estimate. As a matter of fact, the deterministic variability not captured by under-fitted drifts, is picked up by the corresponding residual variogram, leading to its over-estimation (Figure 6a, green line). Such over-estimation is partially balanced by the downward bias that occurs estimating the variogram from estimated residuals –due to the variance decomposition (29)– which seems not to be very severe in the considered cases (compare gray lines and blue lines in Figure 6a and 6b).

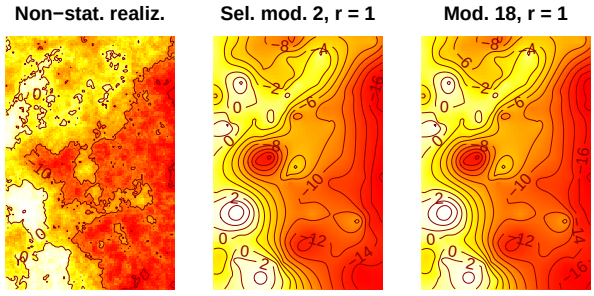
Even though the drift estimate as well as the variogram estimation obtained by selecting model 2 instead of model 18 might not seem very satisfactory, the Universal Kriging prediction appears not to be affected by under-fitting, both for $t = 0.1$ and for $t = 0.7$ (Figure 7a and 7b). Indeed, all the patterns presented by the original grid realization (first panels in Figure 7a and 7b) are well reproduced by both interpolations, with very similar results for models 2 and 18.

Cross-validation results, shown in Table 3, confirm these graphical observations. The statistics relative to the $n = 100$ cross-validation squared errors $\|\chi_{\boldsymbol{s}_i} - \chi^{*k}_{\boldsymbol{s}_i}\|^2$, $i = 1, ..., 100$ and $k = 2, 18$, are very similar, with slightly better results for the selected model –which is not surprisingly, since the optimality criterion in Algorithm 11 is precisely based on cross-validation error–.

Moreover, notice that dataset $\chi^{(18,1)}_{\vec{s}}$ is characterized by a high residuals

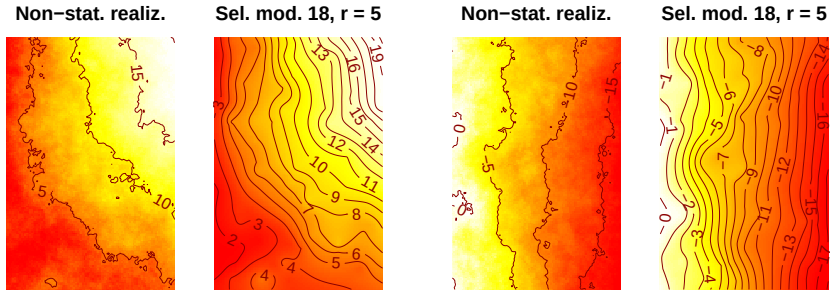


(a) $t = 0.1$



(b) $t = 0.7$

FIGURE 7. *Comparison of simulated grid and UK prediction for* $\chi^{(18,1)}_{\vec{s}}$ *through contour plots, for* $t = 0.1, 0.7$. *In each sub-figure, from left to right: generated grid, UK prediction for with drift model 2 –selected with Algorithm 11– and with drift model 18.*



(a) $t = 0.1$ (b) $t = 0.7$

FIGURE 8. *Comparison of simulated grid and UK prediction for* $\chi_{\vec{s}}^{(18,5)}$ *through contour plots, for* $t = 0.1, t = 0.7$. *In each sub-figure: generated grid (left panel), UK prediction for with drift model 18 –selected with Algorithm 11– (right panel).*

amplitude and thus the prediction turns out to be only slightly drift-driven: therefore, the Universal Kriging prediction proves to be very robust with respect to under-fitting of the drift model. On the contrary, in the presence of low residuals amplitude, as for $r = 5$, the drift term becomes very influential on the data and on the prediction: in such a case, the performance of the model selection algorithm is much more satisfactory than before (Figure 3), as well as the drift estimation (fourth panels of Figure 5a and 5b), and the Universal Kriging prediction appears to be very accurate (Figure 8).

The increasing precision in estimating the drift for decreasing residuals variability is even more apparent by comparing the estimated drift coefficients $\widehat{a}_l$, $l = 0, 1, 5$, computed from dataset $\chi_{\vec{s}}^{(18,1)}$ and $\chi_{\vec{s}}^{(18,5)}$, adopting in both cases drift model 18 (Figure 9). Indeed, the coefficients relative to the case $r = 1$ (Figure 9, upper panels) capture also part of the stochastic variability and are thus much more fluctuating than the reference ones, while the coefficients computed from $\chi_{\vec{s}}^{(18,5)}$ (Figure 9, lower panels) are much more smooth reproducing more precisely the reference ones. Notice that the first situation is particularly critical because both residuals and drift curves are built on the same truncated basis and thus excite the same set of frequencies. The presence of more uncertainty in the estimates for noisier data is confirmed by the curves $\widehat{a}_l \pm 2\sqrt{\Lambda_{ll}}$, $\Lambda = \mathrm{Cov}(\widehat{\boldsymbol{a}}_l)$, $l = 0, 1, 5$, reported in Figure 9, which provide measures of the estimates variability.

Several other scenarios have been considered in the simulation study, obtaining further evidence of the results shown here: the performance of the proposed methodology on simulated data confirmed to be very satisfactory.

Indeed, the combination of Algorithms 10 and 11 leads to a very robust and flexible procedure. On one hand, in the presence of highly correlated data, the adoption of a predictive criterion for selecting the drift model proves to be

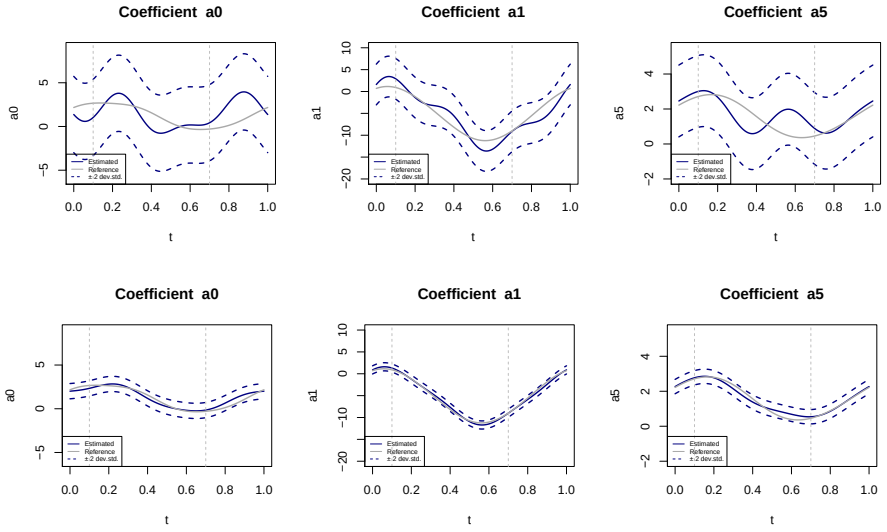


FIGURE 9. *Comparison of the coefficients estimates* $\widehat{a}_l$, $l = 0, 1, 5$, *computed from dataset* $\chi_{\boldsymbol{s}}^{(18,1)}$ *(upper panels) and from dataset* $\chi_{\boldsymbol{s}}^{(18,5)}$ *(lower panels): estimated functional coefficients* $\widehat{a}_l$, $l = 0, 1, 5$ *(solid blue lines), generated coefficients* a_l, $l = 0, 1, 5$ *(solid grey lines). Dashed blue lines correspond to* $\widehat{a}_l \pm 2\sqrt{\Lambda_{ll}}$, $\Lambda = \mathrm{Cov}(\widehat{\boldsymbol{a}}_{\vec{l}})$, $l = 0, 1, 5$. *Vertical dashed grey lines indicate* $t = 0.1$ *and* $t = 0.7$.

appropriate to avoid over-fitting in favor of more parsimonious models; at the same time, in such cases Universal Kriging prediction proves to be very robust to under-fitting. On the other hand, in the presence of less noisy data, the results obtained by Algorithm 11 become more reliable, as well as the drift estimations by Algorithm 10, leading to a very precise prediction.

In any case, the obtained predictions appear very accurate in reproducing all the main patterns presented by the generated realizations, with only a moderate smoothing effect. This is a remarkable result especially given the simplicity of the kriging predictor which involves only global definitions of spatial dependence, besides being linear in the data through scalar coefficients. Indeed, simulations show that the proposed methodology is so flexible that not only global features, but also local structures can be well reproduced by this kind of predictor.