

Supplement for: “Object oriented spatial analysis of natural concentration levels of chemical species in regional-scale aquifers”

Alessandra Menafoglio^a, Laura Guadagnini^b, Alberto Guadagnini^c, Piercesare Secchi^{a,d}

^a*MOX, Department of Mathematics, Politecnico di Milano*

^b*Department of Civil and Environmental Engineering, Universitat Politècnica de Catalunya*

^c*Department of Civil and Environmental Engineering, Politecnico di Milano*

^d*Center for Analysis, Decisions and Society, Human Technopole*

Abstract

This supplement presents the results of leave-one-out cross-validation for the field case presented in the manuscript.

1. Leave-one-out cross-validation analysis

A leave-one-out cross-validation (LOO-CV) analysis is here performed on the data presented in the manuscript. For the data point \mathcal{X}_{s_i} , $i = 1, \dots, n$, in the sample of smoothed densities, we consider as training set $\mathcal{X}^{(-i)} = \{\mathcal{X}_{s_j}, j \neq i\}$. On this basis, we build the model for the spatial dependence (trace-variogram) and perform kriging prediction, by following the same choices as in Section 5. Results are displayed in Figures 1, 2, 3 and 4. The goodness of the estimated model is confirmed by the sum of squared error (*SSE*) of kriging, defined as

$$SSE(\mathcal{X}_{s_i}) = \|\mathcal{X}_{s_i} - \mathcal{X}_{s_i}^{*(-i)}\|^2.$$

Here, $\mathcal{X}_{s_i}^{*(-i)}$ is the prediction of the i -th observation when this is left out of the training set. The summary statistics of the *SSE*, assessed via LOO-CV, are listed in Table 1. One can notice a consistency between the order of magnitude of the mean *SSE* and that of the trace-variogram (estimated sill: 69.12).

Table 1: Summary statistics for $SSE(\mathcal{X}_{s_i})$.

Min	Q1	Median	Mean	Q3	Max
1.682	22.016	28.871	32.552	37.882	90.186

Figures 1, 2, 3 and 4 also display, for each site, a set of 100 curves (colored lines) simulated by conditional simulation, following the methods described in Section 4 of the manuscript. The depth of the test observation with respect to the simulated sample, $\widehat{SpD}(\mathcal{X}_{s_i}|\{\mathcal{X}_{s_0}|\mathcal{X}\})$, is reported in the titles of the panels. Overall, one can notice a good agreement of the simulated

Table 2: Summary statistics for $\widehat{SpD}(\mathcal{X}_{s_i}|\{\mathcal{X}_{s_0}|\boldsymbol{\mathcal{X}}\})$.

Min	Q1	Median	Mean	Q3	Max
0.099	0.244	0.303	0.333	0.384	0.851

sample with the shape of the test observations, the spatial depths $\widehat{SpD}(\mathcal{X}_{s_i}|\{\mathcal{X}_{s_0}|\boldsymbol{\mathcal{X}}\})$ ranging in $[0.099, 0.851]$. Table 2 reports the summary statistics for $\widehat{SpD}(\mathcal{X}_{s_i}|\{\mathcal{X}_{s_0}|\boldsymbol{\mathcal{X}}\})$.

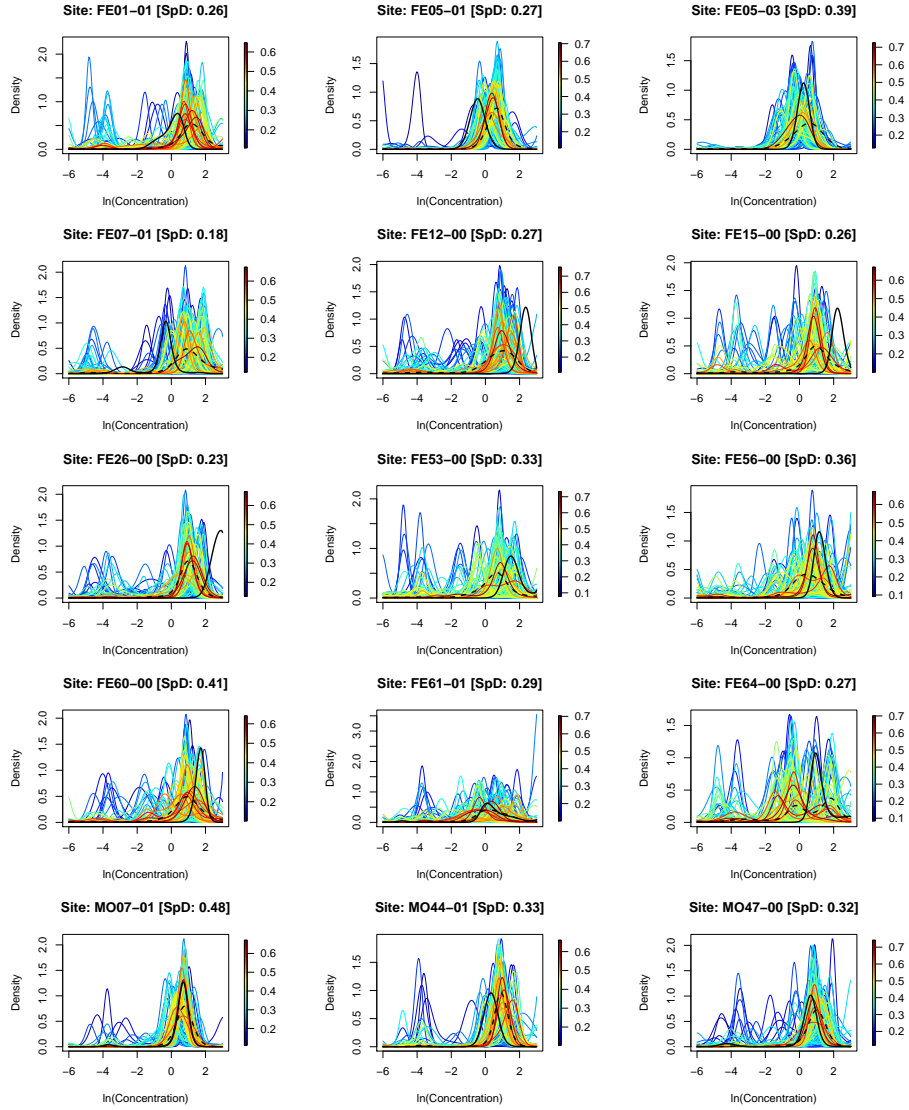


Figure 1: Cross-validation results. In each panel: the solid black curve represents the observation left out of the training set (test data point), the dashed curve represents the kriging prediction, and the solid colored curves represent a set of 100 conditional simulations at the test location; colors are given according to the spatial depth of the simulated curve $\mathcal{X}_{s_0}^{*m}$ with respect to the MC ensemble, i.e., $\widehat{SpD}(\mathcal{X}_{s_i}^{*m} | \{\mathcal{X}_{s_i}\})$.

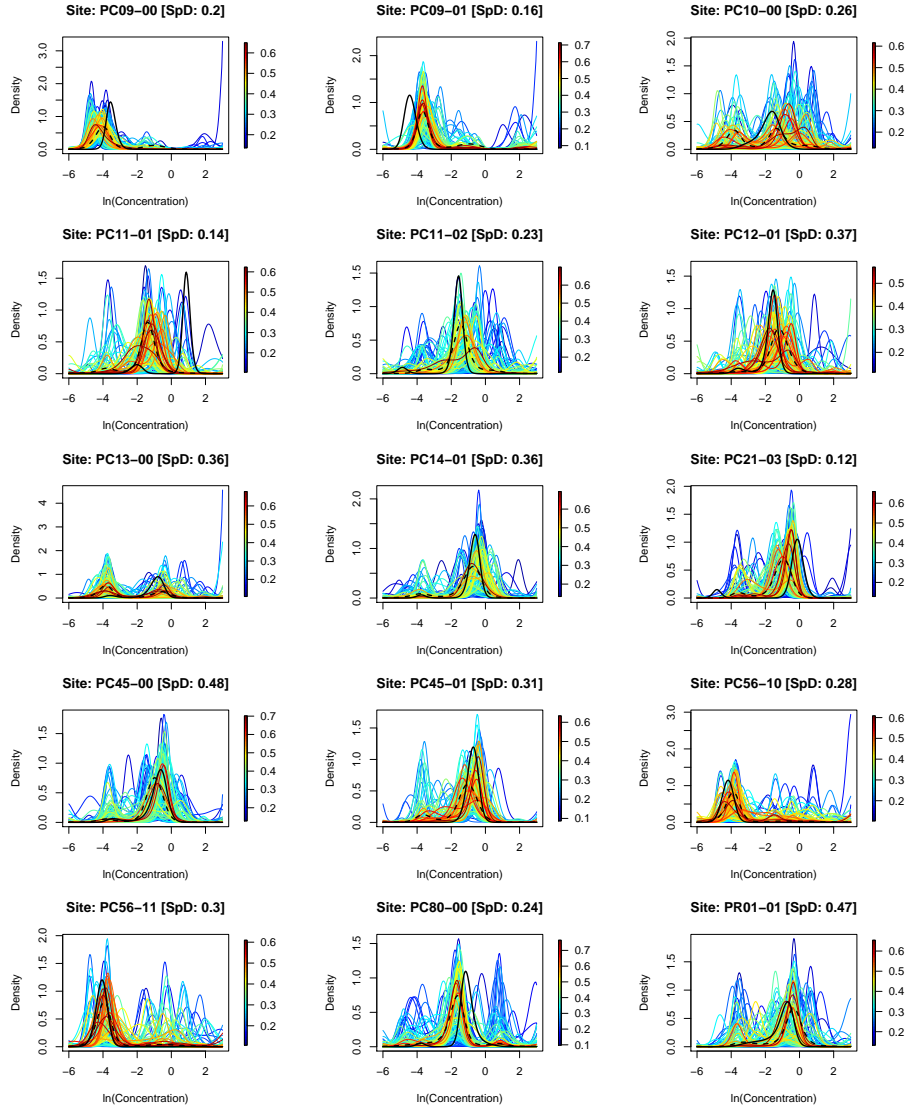


Figure 2: Cross-validation results. In each panel: the solid black curve represents the observation left out of the training set (test data point), the dashed curve represents the kriging prediction, and the solid colored curves represent a set of 100 conditional simulations at the test location; colors are given according to the spatial depth of the simulated curve $\mathcal{X}_{s_0}^{*m}$ with respect to the MC ensemble, i.e., $\widehat{SpD}(\mathcal{X}_{s_i}^{*m} | \{\mathcal{X}_{s_i}\})$.

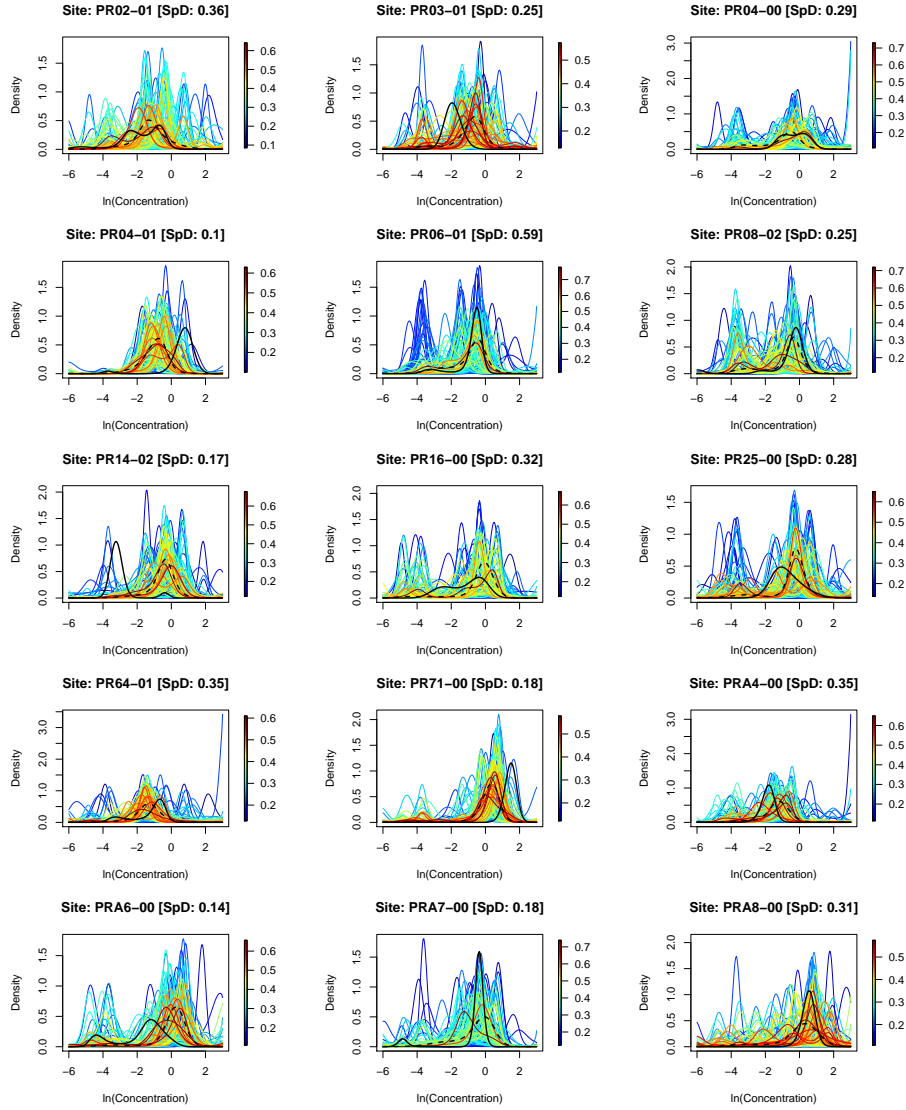


Figure 3: Cross-validation results. In each panel: the solid black curve represents the observation left out of the training set (test data point), the dashed curve represents the kriging prediction, and the solid colored curves represent a set of 100 conditional simulations at the test location; colors are given according to the spatial depth of the simulated curve $\mathcal{X}_{s_0}^{*m}$ with respect to the MC ensemble, i.e., $\widehat{SpD}(\mathcal{X}_{s_i}^{*m} | \{\mathcal{X}_{s_i}\})$.

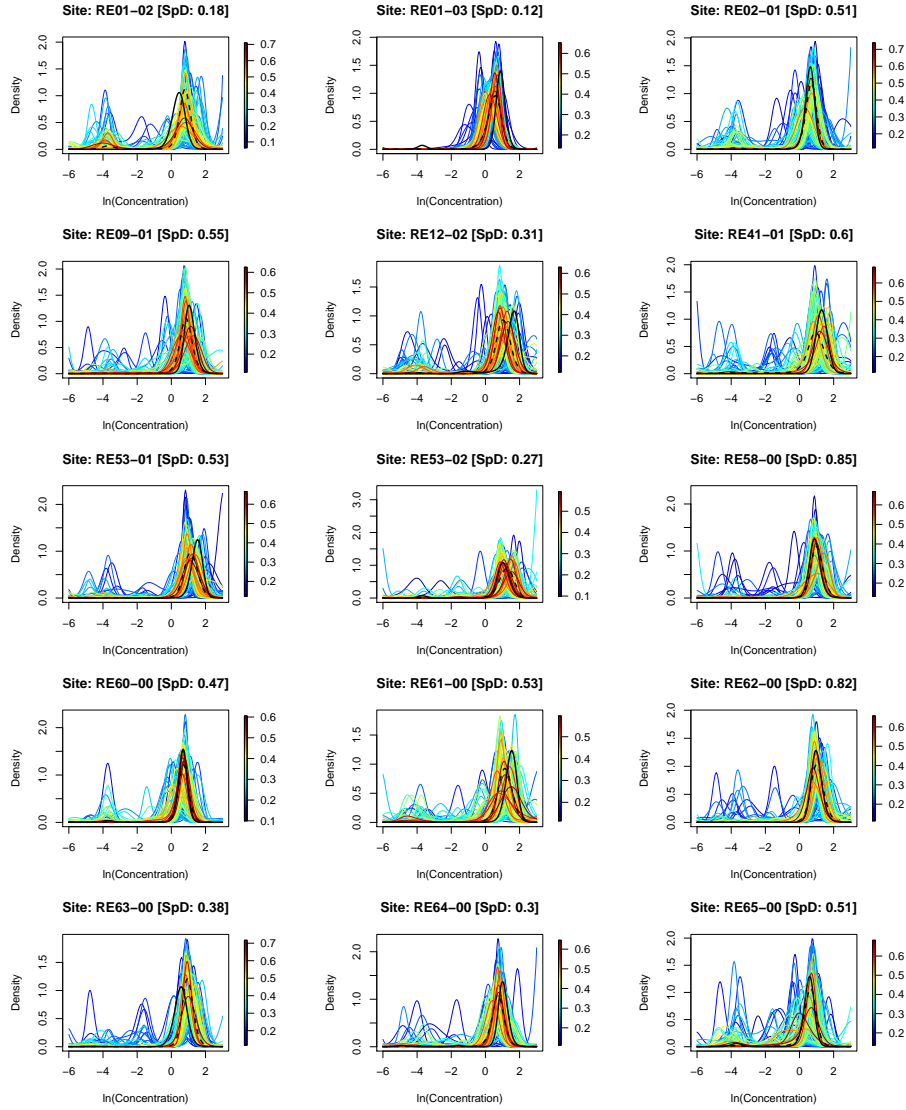


Figure 4: Cross-validation results. In each panel: the solid black curve represents the observation left out of the training set (test data point), the dashed curve represents the kriging prediction, and the solid colored curves represent a set of 100 conditional simulations at the test location; colors are given according to the spatial depth of the simulated curve $\mathcal{X}_{s_0}^{*m}$ with respect to the MC ensemble, i.e., $\widehat{SpD}(\mathcal{X}_{s_i}^{*m} | \{\mathcal{X}_{s_i}\})$.