



Statistical evaluation of natural background levels of groundwater bodies through spatially dependent functional data analysis

Alberto Guadagnini (1), Laura Guadagnini (1), and Alessandra Menafoglio (2)

(1) Politecnico di Milano, Dip. di Ingegneria Civile e Ambientale, Milano, Italy (alberto.guadagnini@polimi.it), (2) Politecnico di Milano, MOX, Department of Mathematics

We illustrate the use of Object-Oriented Kriging and stochastic simulation for the geostatistical characterization of spatially variable Natural Background Levels (NBLs) of a chemical species in a large-scale groundwater body. The approach is motivated by the observation that NBLs tend to display spatial variations due to heterogeneous hydro-geochemical characteristics of the target aquifer and their assessment is critical to identify significant trends of chemical concentrations. NBL estimates typically rely on statistical analysis of concentration samples which can be considered as unaffected by anthropogenic actions. These can be obtained by the application of, e.g., an approach such as pre-selection to temporal series of concentrations and environmental parameters monitored across a network of observation boreholes. Here, we start by considering the ensuing probability density function (PDF) of the chemical species of interest at each observation well as the object of the statistical analysis. We interpret these PDFs through an approach that relies on modeling them as random points in a Bayes Hilbert space and is set in the context of Object Oriented Spatial Statistics. As such, each PDF is seen as an object in the Hilbert space of functional compositions, endowed with the Aitchison geometry, the building-block of our geostatistical analysis being the entire object, rather than a limited number of selected features (e.g., quantiles) of the data. This, in turn, enables us to: (i) perform linear predictions through an (Object-Oriented) Universal Kriging approach; (ii) provide a rigorous uncertainty quantification based on the generation of multiple conditional stochastic simulations, and (iii) assess spatial distributions of the probability of exceedance of reference thresholds. In this context, our approach allows modeling and taking full advantage of these data characteristics upon providing us with the framework to project the entire information content embedded within the data onto a computational grid, to characterize unsampled locations in the system. We exemplify the methodology through the analysis of NBLs of selected chemical species monitored at a network of locations within an alluvial aquifer in Northern Italy. We describe the results of the proposed methodological and theoretical framework and illustrate its potential in comparison to previous studies.