Elsevier Editorial System(tm) for Water Research

Manuscript Draft

Manuscript Number: WR44505R1

Title: Geostatistical multimodel approach for the assessment of the spatial distribution of natural background concentrations in large-scale groundwater bodies

Article Type: Research Paper

Keywords: Natural background levels; groundwater quality; chemical status; multimodel analyses; contaminated aquifers

Corresponding Author: Dr. Laura Guadagnini,

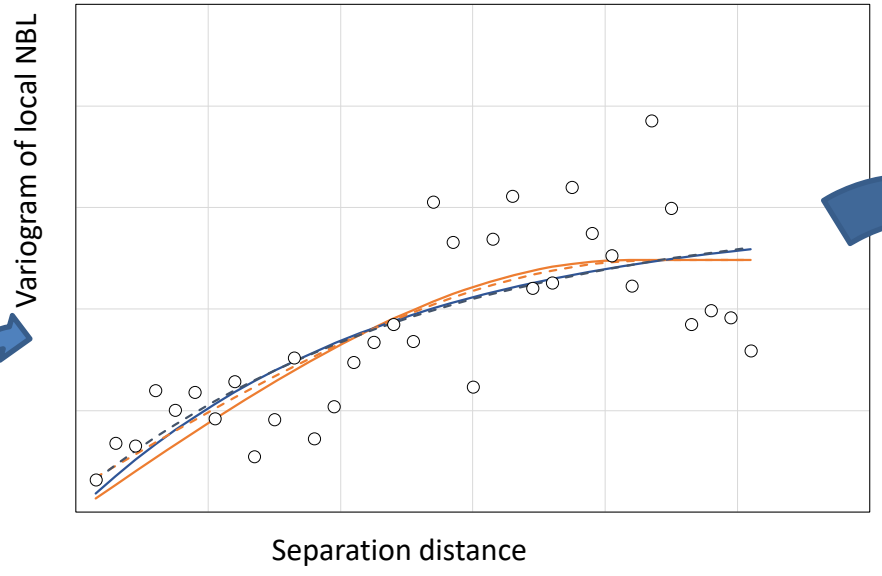
Corresponding Author's Institution: Politecnico di Milano

First Author: Antonio Molinari, PhD

Order of Authors: Antonio Molinari, PhD; Laura Guadagnini; Marco Marcaccio; Alberto Guadagnini, PhD

Abstract: Quantification of the (spatially distributed) natural contributions to the chemical signature of groundwater resources is an emerging issue in the context of competitive groundwater uses as well as water regulation and management frameworks. Here, we illustrate a geostatistically-based approach for the characterization of spatially variable Natural Background Levels (NBLs) of target chemical species in large-scale groundwater bodies yielding evaluations of local probabilities of exceedance of a given threshold concentration. The approach is exemplified by considering three selected groundwater bodies and focusing on the evaluation of NBLs of ammonium and arsenic, as detected from extensive time series of concentrations collected at monitoring boreholes. Our study is motivated by the observation that reliance on a unique NBL value as representative of the natural geochemical signature of a reservoir can mask the occurrence of localized areas linked to diverse strengths of geogenic contributions to the groundwater status. We start from the application of the typical Pre-Selection (PS) methodology to the scale of each observation borehole to identify local estimates of NBL values. The latter are subsequently subject to geostatistical analysis to obtain estimates of their spatial distribution and the associated uncertainty. A multimodel framework is employed to interpret available data. The impact of alternative variogram models on the resulting spatial distributions of NBLs is assessed through probabilistic weights based on model identification criteria. Our findings highlight that assessing possible impacts of anthropogenic activities on groundwater environments with the aim of designing targeted solutions to restore a good groundwater quality status should consider a probabilistic description of the spatial distribution of NBLs. The latter is useful to provide enhanced information upon which one can then build decision-making protocols embedding the quantification of the associated uncertainty.

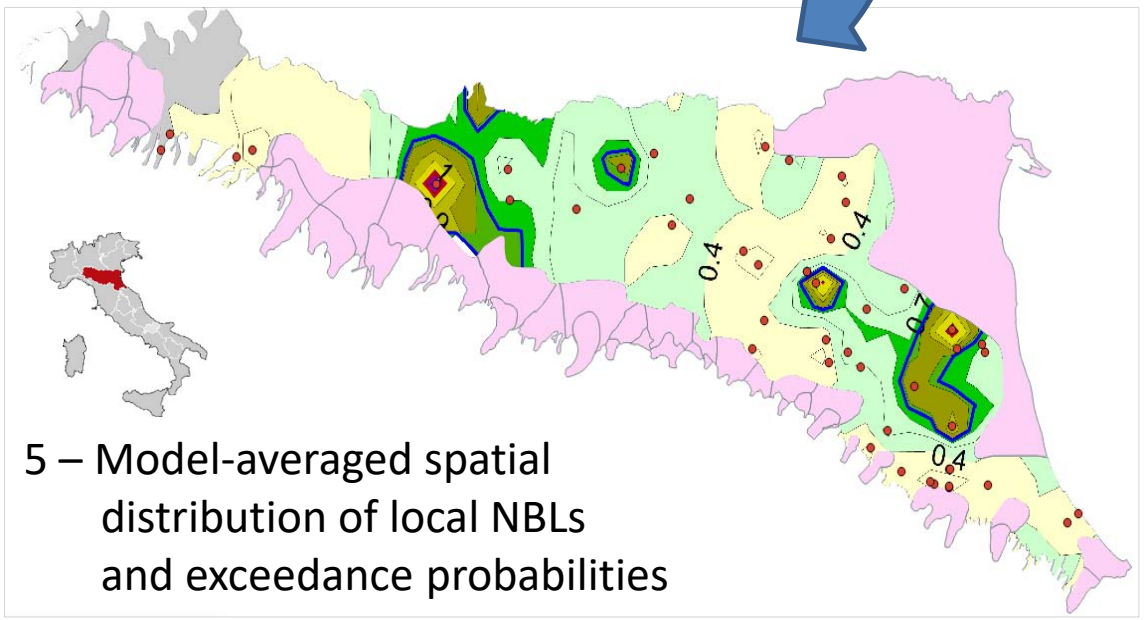
4 – Multi-model variogram assessment



3 - Geospatial analysis

2 - Assessment of local NBL for each monitoring well

1 - Monitored concentrations



5 – Model-averaged spatial distribution of local NBLs and exceedance probabilities

Geostatistical multimodel approach for the assessment of the spatial distribution of natural background concentrations in large-scale groundwater bodies

A. Molinari[a], L. Guadagnini[a,*], M. Marcaccio[b], A. Guadagnini[a,c]

[a]Politecnico di Milano, Dipartimento di Ingegneria Civile e Ambientale, Piazza L. Da Vinci 32, 20133 Milano, Italy

[b]Arpae Emilia-Romagna, Direzione Tecnica, Largo Caduti del Lavoro 6, 40122 Bologna, Italy

[c]University of Arizona, Department of Hydrology and Atmospheric Sciences, 85721 Tucson, AZ, USA

*Corresponding author: laura.guadagnini@polimi.it

Abstract

Quantification of the (spatially distributed) natural contributions to the chemical signature of groundwater resources is an emerging issue in the context of competitive groundwater uses as well as water regulation and management frameworks. Here, we illustrate a geostatistically-based approach for the characterization of spatially variable Natural Background Levels (NBLs) of target chemical species in large-scale groundwater bodies yielding evaluations of local probabilities of exceedance of a given threshold concentration. The approach is exemplified by considering three selected groundwater bodies and focusing on the evaluation of NBLs of ammonium and arsenic, as detected from extensive time series of concentrations collected at monitoring boreholes. Our study is motivated by the observation that reliance on a unique NBL value as representative of the natural geochemical signature of a reservoir can mask the occurrence of localized areas linked to diverse strengths of geogenic contributions to the groundwater status. We start from the application of the typical Pre-Selection (PS) methodology to the scale of each observation borehole to identify local estimates of NBL values. The latter are subsequently subject to geostatistical analysis to obtain estimates of their spatial distribution and the associated uncertainty. A multimodel framework is employed to interpret available data. The impact of alternative variogram models on the resulting spatial distributions of NBLs is assessed through probabilistic weights based on model identification criteria. Our findings highlight that assessing possible impacts of anthropogenic activities on groundwater environments with the aim of designing targeted solutions to restore a good groundwater quality status should consider a probabilistic description of the spatial distribution of NBLs. The latter is useful to provide enhanced information upon which one can then build decision-making protocols embedding the quantification of the associated uncertainty.

Keywords: Natural background levels; groundwater quality; chemical status; multimodel analyses; contaminated aquifers

1. Introduction

Modern society is characterized by an ever-increasing competitive use of groundwater resources, these being subject to many anthropogenic stresses (e.g., domestic use, irrigation and farming activities, industrial operations). To assist evaluation of the resilience of groundwater resources and the soil-water environment serving local communities, several studies are targeted to the analysis of water quality deterioration (e.g., Liu et al., 2017; Zlatanović et al., 2017; Heibati et al., 2017) or of water footprint characteristics (e.g., Qian et al., 2018). When dealing with the assessment of the qualitative status of a target groundwater body as a result of, e.g., in-place monitoring activities, it is not uncommon to evidence areas where detected chemical concentrations attain large values. In some instances, the latter can be directly or indirectly associated with the petrographical composition of the investigated aquifer (e.g., Hinsby and Condesso de Melo, 2006) or with site-specific characteristics such as the occurrence of organic matter (e.g., vegetal matter or peats) which can enhance release of chemical species to groundwater (e.g., Redman et al., 2002; Molinari et al., 2013 and references therein). These elements can in turn yield high natural levels of metals, such as Arsenic, even in crops (e.g., maize or rice) intended for human consumption (e.g., Kumarathilaka et al., 2018). Quantification of the actual (spatially distributed) natural (or geogenic) contributions to chemical concentration is an emerging issue causing increasingly pressing concerns in the context of competitive use of groundwater resources, water regulation and management frameworks at national and European levels, with implications in several industrial activities. Misclassifications of areas where sampled concentrations attain large values as a consequence of geogenic contributions yielding marked Natural Background Levels (NBLs) can have important socio-economic implications related to public health and risk assessment issues. Inaccurate risk assessment analyses can therefore yield an improper classification of the chemical status of an investigated aquifer which might lead to setting unrealistic remediation goals.

Characterization of the actual natural signature of groundwater bodies is a main theme of the EU Water Framework Directive (WFD 2000/60/EC, article 17). A key component required for the

reversal of identified marked and sustained upward trends of contaminants is the proper estimation of NBLs of aquifer bodies. Main aspects related to the definition of NBLs are illustrated in article 2.5 of the GroundWater Daughter Directive (GWDD 2006/118/EC). The latter has been recently amended by Directive 2014/80/EC stating that "the monitoring strategy and interpretation of the data should take account of the fact that flow conditions and groundwater chemistry vary laterally and vertically". With reference to these concepts, and prior to the enactment of Directive 2014/80/EC, Molinari et al. (2012) observe that (*a*) NBLs tend to increase with the average depth of a water body and (*b*) whenever possible, NBLs should be estimated via robust experimental characterization of the geochemical system and modeling studies performed, e.g., through state-of-the-art multicomponent reactive transport approaches.

Statistical analysis of monitored data represents the typical approach employed for NBL estimations (Edmunds et al., 2003; Wendland et al., 2005; Panno et al., 2006; Walter, 2008; Kim et al., 2015). In this context, the EU research project BRIDGE (2007), Background cRiteria for the IDentification of Groundwater thrEsholds, proposes a methodology termed as Pre-Selection (PS). The latter is based on the identification of pristine groundwater samples across an available set of sampled data, as representative of the natural population of the resident concentration. As a result of this procedure, a unique (or bulk) NBL value is estimated and associated with the examined subsurface reservoir, implying that all concentrations exceeding that level should be ascribed to anthropogenic sources. The typical signature of a given chemical species in groundwater may be defined through a range of concentrations rather than a single value (Reimann and Garrett, 2005; Hinsby et al., 2008). This is related to the interaction and feedback between diverse natural, atmospheric, geological, chemical and biological processes taking place in both the vadose and saturated zone during groundwater infiltration and circulation (Edmunds et al., 2003; Wendland et al., 2005; Panno et al., 2006; European Commission, 2009). These concepts are not completely embedded in current regulatory frameworks which requires an estimate of only one threshold value, considered as uniform across a given water body and against which anthropogenic contaminations

should be assessed (Reimann and Garrett, 2005). Otherwise, NBLs can attain markedly different local values, for instance because of the occurrence of diverse petrographic provinces or redox conditions within the same groundwater body, especially in large-scale reservoirs (with areal extent of, e.g., thousands of square kilometers). Hence, the common practice of evaluating the chemical status relying on a single NBL value cannot be considered as realistic and might lead to severe over- or under-estimation of the typical natural signature.

Ducci et al. (2016) and Dalla Libera et al. (2017) recognize that the spatial distribution of NBLs should reflect the heterogeneity of the investigated groundwater body. Critical assumptions associated with these studies are (*a*) the reliance on a unique model employed to interpret experimental variograms, (*b*) the incomplete quantification of the uncertainty of model parameters and estimated concentrations, and (*c*) the lack of a direct estimation of local NBLs associated with each monitoring well upon which exceedance probability maps can be conditioned. Yet, it is well documented that estimated values and uncertainty analyses relying on a single (conceptual and/or mathematical) model can lead to statistical bias or underestimation of the overall uncertainty linked to the system behavior due to undersampling of the space of possible descriptive models. These aspects can be seamlessly embedded within a Maximum Likelihood framework and subsequent reliance on Model Quality criteria to consider uncertainty in the mathematical model depicting the system as well as in its parameters (e.g., Carrera and Neuman, 1986; Ye et al., 2004, Bianchi Janetti et al., 2012 and references therein; Gimeno et al., 2017).

Here, our key objective is to illustrate an approach for the estimation of local NBLs at the borehole scale through the application of a geostatistically-based methodology to yield exceedance probability maps. We accomplish this objective by relying on Maximum Likelihood and formal model identification criteria to take into account uncertainty stemming from multiple and competing models which can be employed to interpret sample variograms of local values of NBLs. To the best of our knowledge, this approach stands as one of the first applications targeted at the evaluation of NBL spatial maps by including quantification of uncertainty associated with the variogram model

employed to interpret the spatial distribution of local NBLs and probability exceedance concentration maps, the latter being usually developed without the evaluation of local NBLs (e.g., Ungaro et al., 2008; Ayotte et al., 2006; Liu et al., 2004; Gaus et al., 2003).

2. Materials and methods

2.1. Study areas

Following the application of WFD 2000/60/EC, a total of 144 groundwater bodies have been delimited within the Emilia-Romagna Region, Italy (Regione Emilia-Romagna, 2010). These are part of the Po Basin fill, which is a syntectonic sedimentary wedge (Ricci Lucchi, 1984) forming the infill of the Pliocene-Pleistocene fore-deep.

The information acquired from sedimentological and hydrogeological analyses has led to the identification of three main hydrogeological complexes, i.e., Apennines alluvial fans, Apennine alluvial plain and alluvial and deltaic Po plain. The system is characterized by a multilayered confined or semiconfined configuration where gravel is gradually replaced by sand deposits in the northern part of the plain, the thickness of fine deposits increasing towards the north portion of the plain (Regione Emilia-Romagna, 2010; Farina et al., 2014). Additional information regarding hydrogeological settings of the study region are available in Molinari et al. (2012) and Farina et al. (2014). An upper confined portion and a lower confined portion have been distinguished within this multilayered system. Our study is focused on the three largest groundwater bodies identified. Two of these are located in the upper confined segment of the aquifer system of the Po Basin fill, the remaining one representing a deep confined water body. Figure 1a depicts limits and planar extent of the two upper confined groundwater bodies, respectively indicated as 0610 and 0630, while the limits and planar extent of the deep confined water body, termed as 2700, are depicted in Figure 1b.

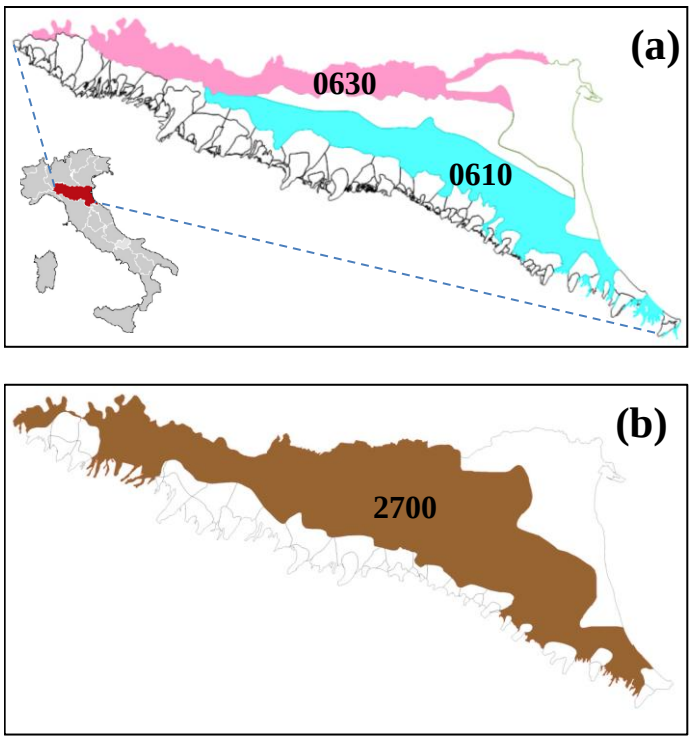


Fig. 1. Planar extent of the investigated water bodies in (a) the upper confined portion and (b) the lower confined portion of the aquifer system of the Po Basin fill; light blue = water body 0610, pink = water body 0630, brown = water body 2700.

Table 1 lists the average depth and thickness as well as the planar area of the three groundwater bodies analyzed.

Table 1

Extension and main characteristic length scales of the three groundwater bodies investigated.

Groundwater body	Average thickness (m)	Average depth (m)	Areal extent (km^2)
0610	130	75	2928
0630	110	65	1995
2700	180	200	6934

Each of these groundwater bodies is subject to various levels of anthropogenic stresses because of diverse and competing uses of the subsurface, in terms of water consumption and withdrawals for agricultural and industrial purposes (Farina et al., 2014). Anthropogenic pollution tends to decrease going from superficial reservoirs to the deepest water body (i.e., 2700) which mostly receives diluted concentrations of chemical species from recharging areas located at some distance.

We consider these three systems because of their significant planar extent whose representative scale is of the order of hundreds of kilometers and where the relevance of considering regional-scale heterogeneous distributions of NBL values for the assessment of groundwater quality is markedly evident.

2.2. Available dataset

We ground our analyses on the time series of concentrations collected at monitoring stations associated with records of about 20 years of observations. These recordings (*a*) extend between 1987 and 2008 (albeit not continuously for some wells), (*b*) have been taken at a six-month interval, and (*c*) constitute a unique data-base that we employ in the context of our investigation. The chemical species considered in this study are ammonium (NH_4) and arsenic (As), which represent critical elements for the achievement of a good chemical status for all three water bodies analyzed, according to Italian Regulation (D. Lgs. 30/09, i.e., Decreto Legislativo n. 30, 16 March 2009) and GWDD 2006/118/CE. As described in Molinari et al. (2012), to which we refer for further details, all concentration data have been subjected to a preliminary exploratory statistical analysis that identified ammonium and arsenic as critical species of concern.

Table 2 lists the number of monitoring stations and the total number of samples collected within the 20-year long record of observations at locations included in the extensive network of observation wells managed by the "Agenzia Regionale per la Prevenzione e l'Ambiente dell'Emilia-Romagna" (ARPAE - Regional Agency for Environmental Protection, Emilia-Romagna).

Table 2

Number of monitoring stations and total number of samples available

		number of samples	
Groundwater body	monitoring stations	As	NH_4
0610	90	1968	2230
0630	75	1692	1917
2700	55	1201	1383

183
184
185 *2.3 Data analysis*

186 *2.3.1 NBL estimation*

187 Within the framework of the EU research project BRIDGE (2007), the Pre-Selection (PS)
188 methodology has been developed for the assessment of the overall geochemical signature of large-
189 scale aquifer systems under data scarcity. The methodology relies on the statistical analysis of the
190 information collected across a monitoring network and is based on the selection of samples that
191 meet certain criteria and can be considered unaffected by anthropogenic influence. Typically
192 adopted criteria for the exclusion of influenced samples are associated with the following
193 conditions: (*a*) chloride concentrations > 1000 mg/L, as indicator of salinity; (*b*) nitrates (NO_3)
194 concentrations > 10 mg/L, as indicator of human influence caused by, e.g., fertilizers; and (*c*)
195 ammonium (NH_4) concentrations > 0.5 mg/L, as indicator of human impact under reducing
196 conditions. Additional criteria, such as redox conditions, dissolved oxygen, sulfate concentration,
197 can be considered for sample exclusion, as pointed out by Hinsby and Condesso de Melo (2006)
198 and Hinsby et al. (2008) to maximize the possibility of grouping homogeneous data.

199 Samples exhibiting markers of anthropogenic contamination (e.g., nitrates or pesticides)
200 larger than a given value have been removed from the original data bank and the residual set is used
201 to estimate the median value for the remaining concentrations of the target chemical species at each
202 monitoring well. The NBL value is then evaluated in terms of a selected percentile of the medians
203 associated with each monitoring well within the investigated water body. The 90^{th}, 95^{th}, or 97.7^{th}
204 percentile are typically considered, depending on the degree of knowledge of the hydrogeochemical
205 system. Wendland et al. (2005) propose to consider the 90^{th} percentile of all of the calculated
206 medians stemming from each monitoring well in the investigated reservoir as an estimate of NBL
207 for the whole water body. Hereinafter, we refer to this quantity as NBL_{90}^{PS}.

As already stated, limitations inherent to the application of this procedure include: (*a*) all the information associated with temporal variability of concentrations are shadowed, and (*b*) only one NBL value is estimated for the whole aquifer body, without the possibility to assess any kind of spatial variability across the system. With the aim of embedding within the analysis spatial and temporal information linked to the scale of observation boreholes, we structure our study through the following main steps:

1. perform sample selection for temporal records at each observation borehole following typically adopted exclusion criteria, as illustrated above and indicated in the original BRIDGE methodology;
2. evaluate a local NBL of the selected chemical species at each observation well as the 90th percentile of concentration values retained at step 1;
3. perform a multimodel geostatistical analysis of the results from step 2, to (*a*) construct and interpret empirical variograms of NBLs, (*b*) project local NBL values onto a computational grid through Kriging and evaluate the associated variance, and (*c*) assess probabilities of exceeding given threshold concentrations, considered as environmental performance metrics characterizing the chemical status of the investigated system (see Section 2.3.2).

We emphasize that this approach directly imbues, as a result of step 2, the effects of the monitored temporal variability of concentrations. As a result of step 3, spatial distributions of direct local NBLs estimated for each observation well can be obtained, together with an appraisal of the associated uncertainty, as reflected in the multimodel theoretical framework considered.

2.3.2 Spatial distribution of local NBLs

Following the approach described in Section 2.3.1 (step 3) we appraise the spatial distribution of local NBLs within the target groundwater bodies through a geostatistical approach, framed in the context of a Bayesian multimodel analysis. The study is performed according to the following steps: (*a*) characterization of the spatial correlation structure of the variable by means of experimental variograms; (*b*) selection of a set of alternative / competing theoretical variogram

models and estimation of their parameters (including their uncertainty) through Maximum Likelihood (ML, see Section 2.3.3); (*c*) evaluation (through appropriate discrimination criteria and posterior model weights, see Section 2.3.3) of the relative benefit associated with any of the models considered to interpret available data; (*d*) projection of sample local NBLs onto a computational grid via Kriging by relying on all of the calibrated models; (*e*) assessment of multimodel mean and variance of local NBLs at each grid node (see Section 2.3.4); and (*e*) evaluation of the probability of locally exceeding a given value, i.e., the uniform NBL value obtained through the PS procedure on the regional scale.

A base 10 logarithmic transformation is applied to NBL concentration data to map these onto the unbounded support comprising both positive and negative values. Omnidirectional experimental variograms are assessed on the basis of the results of a preliminary variogram analysis. Due to spatial arrangement of available sample points, Kriging estimates and variance are calculated on a uniform grid, with spacing equal to 5 km. Geostatistical analyses has been performed through the well-known and widely tested Stanford Geostatistical Modeling Software (SGeMS; Remy et al., 2009).

We calibrate each of the models selected to interpret the evaluated sample variograms through ML parameter estimation and apply model identification criteria to rank the tested models in terms of posterior probabilistic weights. The latter are then used to weigh results associated with each of the selected models and calculate multimodel mean and variance.

2.3.3 Maximum Likelihood (ML) parameter estimation and model quality criteria

Let N be the number of available observations of a model output Y collected in vector $\mathbf{Y}^* = [Y_1^*, ... Y_N^*]$. Note that in our application these coincide with values of (log-transformed) NBLs (see also Section 3). The covariance matrix of measurement errors, $\mathbf{B}_Y$, is here considered to be diagonal with non-zero terms equal to the observation error variance, σ_i^2 (Carrera and Neuman, 1986). Denoting by $\hat{\mathbf{Y}} = [\hat{Y}_1, ... \hat{Y}_N]$ the vector of model predictions at locations where data are

available, the ML estimate $\hat{\mathbf{x}}$ of the vector of the M uncertain model parameters can be obtained by minimizing with respect to $\mathbf{X}$ the negative log likelihood criterion:

$$NLL = \sum_{i=1}^{N} \frac{J_i}{\sigma_i^2} + \ln\left|\mathbf{B}_Y\right| + N\ln(2\pi) \quad (1)$$

where $J_i = \left(Y_i^* - \hat{Y}_i\right)^2$. Criterion (1) includes the weighted least square criterion (Carrera and Neuman, 1986; Bianchi Janetti et al., 2012 and references therein). Here, minimization of (1) is achieved using the iterative Levenberg-Marquardt algorithm as embedded in the well documented computational framework PEST (Doherty, 2002).

Alternative (competing) models which can be used to interpret available system states can be ranked by various criteria (e.g., Neuman, 2003; Ye et al., 2004, 2008; Riva et al., 2011; Bianchi Janetti et al., 2012; Ciriello et al., 2015 and references therein), including:

$$AIC = NLL + 2M \quad (2)$$

$$AIC_c = NLL + 2M + \frac{2M(M+1)}{N-M-1} \quad (3)$$

$$KIC = NLL + M\ln\left(\frac{N}{2\pi}\right) - \ln\left|\mathbf{Q}\right| \quad (4)$$

Here, **Q** represents the Cramer-Rao lower-bound approximation for the covariance matrix of the parameter estimates, i.e., the inverse expected Fisher information matrix, which renders a quantitative appraisal of the quality of parameter estimates and of the information content carried by data about model parameters (see, e.g., Ye et al., 2008 for details). The Akaike information criterion, AIC, is due to Akaike (1974), AIC_c to Hurvich and Tsai (1989) and KIC to Kashyap (1982). The lowest value of a given model identification criterion indicates the most favored model (according to the criterion itself) at the expense of the remaining models. Note that *KIC* tends to (*a*) penalize models proportionally to the number of their parameters, through the quantity $M \cdot \ln(N/2\pi)$ (Ye et al., 2008; Hernandez et al., 2006; Riva et al., 2011) and (*b*) favor models with smaller expected information content per observation, when considering models with equal

parameter numbers, minimum *NLL* values and prior probability of parameters linked to such a minimum (Ye et al., 2008). In light of these observations, we base the analyses presented in this study on *KIC* (4).

The discrimination criteria (2)-(4) can also be considered to assign posterior probability weights quantifying uncertainty associated with each of the tested models. The posterior probability linked to model M_k (k = 1, ..., N_M, N_M being the number of interpreting models considered) is evaluated as (Ye et al., 2008):

$$p(M_k \mid \mathbf{Y}^*) = \frac{\exp\left(-\frac{1}{2}\Delta IC_k\right) p(M_k)}{\sum_{i=1}^{N_M} \exp\left(-\frac{1}{2}\Delta IC_i\right) p(M_i)} \tag{5}$$

Here, $\Delta IC_k = IC_k - IC_{min}$, IC_k being either *AIC* (2), AIC_c (3) or *KIC* (4) and $IC_{min} = \min\{IC_k\}$ its minimum value calculated across the range of models examined; $p(M_k)$ is the prior probability associated with each model. One can set $p(M_k) = 1/N_M$. In case no prior information is available, all models being then characterized by the same prior probability.

Grounding our study on model identification criteria and the ensuing posterior probabilities (5) enables one to rank the models analyzed through their posterior probabilities and discriminate among them in a relative sense.

2.3.4 Multimodel Mean and Variance

We consider a collection **M** of K mutually exclusive variogram models, M_k, upon which lead statistics, such as mean and variance/covariance, of NBL values are computed through Kriging at the nodes of a selected computational grid covering a given aquifer body. The models are uncertain, each of them being assigned the same prior probability $p(M_k)$. Variogram model M_k is employed in a Kriging framework to yield the mean (expectation) $E(\mathbf{Y} \mid \mathbf{D}, M_k)$ and the covariance $Cov(\mathbf{Y} \mid \mathbf{D}, M_k)$ of a vector $\mathbf{Y}$ of random (log-transformed) NBL values, conditional on the prior data vector **D**. The entries of the latter are evaluated according to the procedure illustrated in

Section 2.3.1. Averaging across the moments provided by all K variogram models renders the following (Bayesian-averaged) lead moments (Draper, 1995; Hoeting et al., 1999):

$$E(\mathbf{Y}\,|\,\mathbf{D}) = \sum_{k=1}^{K} E(\mathbf{Y}\,|\,\mathbf{D}, M_k)\, p(M_k\,|\,\mathbf{D}) \tag{6}$$

$$\begin{aligned} Cov(\mathbf{Y}\,|\,\mathbf{D}) = & \sum_{k=1}^{K} Cov(\mathbf{Y}\,|\,\mathbf{D}, M_k)\, p(M_k\,|\,\mathbf{D}) \\ & + \sum_{k=1}^{K} \left[E(\mathbf{Y}\,|\,\mathbf{D}, M_k) - E(\mathbf{Y}\,|\,\mathbf{D}) \right] \\ & \cdot \left[E(\mathbf{Y}\,|\,\mathbf{D}, M_k) - E(\mathbf{Y}\,|\,\mathbf{D}) \right]^T p(M_k\,|\,\mathbf{D}) \end{aligned} \tag{7}$$

T denoting transpose. The conditional covariance $Cov(\mathbf{Y}\,|\,\mathbf{D})$, resulting from Bayesian model averaging (BMA), is the sum of a within- and between-model contribution. Posterior model probabilities, $p(M_k\,|\,\mathbf{D})$, are calculated according to (5) and weigh the contribution of model M_k to BMA moments. For the purpose of our demonstration, we evaluate $E(\mathbf{Y}\,|\,\mathbf{D}, M_k)$ and $Cov(\mathbf{Y}\,|\,\mathbf{D}, M_k)$ through Kriging performed upon relying on variogram model M_k characterized through ML parameter estimates.

3. Results and discussion

Prior to the application of the PS procedure, the correct attribution of each monitoring station to a target groundwater body has been assessed on the basis of the technical characteristics of the boreholes (e.g., position of the filters) and the available hydrogeological information (e.g., the depth of permeable layers). Note that the application of the exclusion criteria described in Section 2.3.1 has been performed by disregarding NH_4 because the collected sample cores provide evidence of natural occurrence of paleo-peats (Amorosi et al., 1996; Cremonini et al., 2008) consistent with documented increasing NH_4 concentrations with depth in the three reservoirs investigated (see also Molinari et al., 2012).

Exclusion of samples associated with anthropogenic influence (according to the criteria listed above) is followed by identification and removal of outliers from the remaining data set,

yielding the temporal record subsequently employed for local NBL estimation at each control point. Outliers are here identified as high concentration values that lie outside an interval centered around the sample mean of the data remaining after pre-selection and of width equal to three times the associated standard deviation. Excluding these records from the analyses is consistent with the observation that the bulk local environmental behavior is not significantly influenced by isolated (in time) and significantly large concentrations recorded, for instance, as a result of unusual processes taking place only in a particular year (and linked to the occurrence of, e.g., extreme rains that can cause large infiltration enhancing remobilization and diffusion of natural compounds), as opposed to a typical environmental baseline observed across the entire monitoring period.

Application of PS to the time series of each monitoring well is inevitably linked to a reduction of the number of records to be then subject to statistical analyses. We then ground our NBL calculations solely on monitoring wells which, following data selection, exhibit a time series with more than five records, characterizing an active monitoring period spanning at least 3 years (note that samples are collected on a six-month basis). Note that reliance on this kind of analysis is warranted for aquifers where a significant number of monitoring points is available with (*a*) a reliable and extensive temporal record of observations and (*b*) an appropriate degree of spatial coverage of the system. Table 3 lists the number of observation wells effectively employed for each water body after data selection.

Table 3

Number of observation wells available for spatial analysis of NBLs after data selection.

groundwater body	monitoring stations Ammonium	monitoring stations Arsenic
0610	51	50
0630	62	60
2700	47	45

When compared against Table 2, these data reveal that the upper confined water body 0610 suffers from a significant reduction (about 50%) of the total number of monitoring points due to

exclusion of data evidencing possible anthropogenic impact on the water body. Otherwise, only a limited reduction of observation wells is observed in the case of the other two water bodies (with a reduction of 20% and 16%, respectively for water body 0630 and 2700). With reference to anthropogenic pressures, these results are consistent with the observation that (*a*) the water body 2700 is a deep reservoir and hence subject to more limited anthropogenic stresses than the other two bodies, which constitute upper confined reservoirs, (*b*) groundwater body 0630 is located in the proximity of the Po river, that can act as a source of recharge of the reservoir, thus reducing the impact of stresses caused, for instance, by pumping activities, and (*c*) reservoir 0610 is subject to a enhanced anthropogenic impact due to exploitation associated with intensive industrial/agricultural activities.

3.1. Ammonium

Values of local NBL_{90} obtained at each monitoring well through the methodology described in Section 2.3 display a high degree of spatial variability. These range from 0.828 to 20.835 mg/L (i.e., -0.08 to 1.319 in logarithmic scale) within groundwater body 0610, while ranging from 0.025 to 14.406 mg/L (i.e., -1.602 to 1.159 in logarithmic scale) and from 0.025 to 30.362 mg/L (i.e., -1.602 to 1.482 in logarithmic scale), respectively within groundwater body 0630 and 2700. Values of NBL_{90}^{PS} estimated through the standard PS procedure at the regional scale (i.e., one value for each reservoir) for the three water bodies analyzed (Molinari et al., 2012) are listed in Table 4 together with the corresponding logarithmic-transformed values. We observe that all estimates exceed EU Drinking Water Standard for ammonium.

Table 4

Values for NH_4 NBL_{90} estimated via the original PS procedure (NBL_{90}^{PS}). EU Drinking Water Standard for NH_4 concentration is 0.5 mg/L.

groundwater body	NBL_{90}^{PS} (mg/L)	Log_{10} NBL_{90}^{PS}
0610	4.6	0.66
0630	5.2	0.71
2700	12.0	1.08

For the purpose of our application, classical isotropic Spherical or Exponential variogram models, with or without nugget, are considered. Table 5 lists the results of variogram model calibration analysis. For each groundwater body, these include the estimated set of model parameters (i.e., nugget, range, and sill) based on minimization of *NLL* (1) and the associated posterior probabilities calculated through (4) and (5). As stated, the analyses have been performed by considering logarithmic transformation of values of NBL_{90} (termed as $NH_4Log_{10}\,NBL_{90}$).

Table 5

Estimated parameters of variogram models of $NH_4Log_{10}\,NBL_{90}$ values and associated posterior probability (*p*) based on *KIC* (4). Here, Sph = Spherical model, Exp = Exponential model; n = nugget, a = range (practical range for Exponential model), c = sill; (*) NBL_{90} values are given in (mg/L).

	Groundwater body 0610				Groundwater body 0630				Groundwater body 2700			
Model	n (*)	a (km)	c (*)	*p* (%)	n (*)	a (km)	c (*)	*p* (%)	n (*)	a (km)	c (*)	*p* (%)
Sph	-	18.7	0.106	52	-	83.5	0.496	40	-	81.7	0.485	6
Exp	-	22.9	0.107	48	-	138.2	0.581	49	-	100.4	0.509	34
Sph	-	-	-	-	0.045	87.9	0.452	4.2	0.1	101.0	0.401	5
Exp	-	-	-	-	0.031	159.6	0.574	6.6	0.1	208.9	0.538	55

The spatial variogram structure of $NH_4Log_{10}\,NBL_{90}$ within groundwater body 0610 has been interpreted by a Spherical and an Exponential model without nugget effect (introducing a nugget effect in the model calibration procedure is associated with near-zero estimated nugget and posterior weight and is therefore disregarded). Both variogram models exhibit similar values of range (practical range in the case of Exponential model) and sill, with nearly coinciding posterior probabilities.

Four theoretical models, i.e., Spherical and Exponential models with and without nugget, are considered for groundwater bodies 0630 and 2700. The contribution of the nugget to the total variance is less than 10% in the case of groundwater body 0630. Otherwise, the estimated nugget

effect for groundwater body 2700 is equal to 15.7% and 20% of the total variance, respectively for the Exponential and Spherical model with nugget, suggesting the occurrence of a significant degree of variability between sample pairs at short distances.

Estimated values of correlation scale (i.e., the range) are lowest for water body 0610, a finding that might be related to the spatial arrangement of sampling boreholes across the system or to the observation that state variables, such as given quantiles of concentration time series observed at multiple wells, are not necessarily characterized by a strong degree of spatial correlation, due to the dynamics associated with the key processes driving chemical migration in the system. Otherwise, a consistent spatial persistence of correlation structure of ammonium NBL_{90} is observed in water bodies 0630 and 2700, where spatial variations of kriged estimates are then expected to be quite smooth. The largest posterior probability values for water body 0630 are associated with Spherical and Exponential models without nugget. On the other hand, the Exponential variogram model (with or without nugget) is unambiguously favored in water body 2700.

As described in Section 2.3, the calibrated variogram models are then employed to obtain spatial distributions of Kriging estimates of log-transformed values of NBL_{90} and their associated variance across each of the water bodies considered. For the purpose of this study we rely on Ordinary Kriging, other flavors of Kriging being compatible with our approach. Kriging estimates and variances obtained for each model are then weighted on the basis of posterior probability values according to (6) and (7), to yield spatial distributions of estimated mean and variance of NH_4Log_{10} NBL_{90} based on the complete set of tested models and fully including information on model uncertainty.

As an example of the results obtained by considering individual variogram models, spatial distributions of kriged values for local NBLs of ammonium in water body 2700 with the four models listed in Table 5 are depicted in Appendix A. The spatial maps of kriged values of NH_4Log_{10} NBL_{90} for groundwater bodies 0610, 0630 and 2700 and resulting from our multimodel analysis are respectively depicted in Figures 2, 3 and 4 together with the probability of exceeding

424 the uniform regional (log-transformed) NBL_{90}^{PS} value calculated via the original PS method (see
425 Table 4). For completeness, the corresponding spatial distributions of Kriging variance for the three
426 water bodies investigated are depicted in Appendix B. Exceedance probabilities are calculated by
427 assuming a Gaussian distribution of log-transformed NBL_{90} with local mean and variance rendered
428 by the multimodel analysis. Areas characterized by estimated mean values of $NH_4Log_{10}NBL_{90}$
429 larger than their uniformly distributed counterpart based on the original PS method are also
430 demarcated in Figures 2a, 3a, and 4a.

431 Spatial patterns of mean Log_{10} NBL_{90} display a marked degree of variability across the
432 systems, suggesting that relying on a single (uniform) value can shadow the proper representation
433 of the actual NBL distribution within a given reservoir. For example, one can observe the
434 occurrence of areas where Kriging-based NBL_{90} exhibit values larger than NBL_{90}^{PS}. Otherwise, one
435 can also observe the occurrence of regions where estimated average NBLs are lower than NBL_{90}^{PS},
436 yet still larger than the EU drinking water standard set for ammonium. These observations reinforce
437 the idea that the use of a single NBL value as representative of the whole reservoir may lead to
438 misleading conclusions with reference to the natural behavior of the water body. For example, one
439 might argue that values larger than NBL_{90}^{PS} be associated with external causes (e.g., anthropogenic
440 activities) while they could be linked to specific and local hydrogeochemical natural processes.
441 Likewise, one should also consider that it is possible that concentration values in some areas be
442 lower than NBL_{90}^{PS}, yet larger than their regulatory-based counterpart (i.e., 0.5 mg/L in the case of
443 NH_4), as a result of natural processes rather than anthropogenically induced pollution phenomena.
444 These observations are also supported by the exceedance probability maps depicted in Figures 2b,
445 3b, and 4b. These suggest that there is non-negligible probability that NBL_{90}^{PS} be exceeded over a
446 large portion of the system, reinforcing the concept that the emergence of areas associated with a
447 geogenic origin of a target pollutant is masked when a single NBL_{90} value is taken as representative
448 of the whole reservoir. These sets of results represent an element upon which one could derive

information for the design of additional investigations. These could be directed, for example, to reduce uncertainty (as quantified, e.g., by large values of Kriging variance) in critical areas and/or to support the findings of the geostatistical analyses through the joint use of other types of information, including, e.g., (*a*) site-specific mineralogy, (*b*) the extent of petrographic provinces, or (*c*) the attainment of local natural hydrochemical equilibria linked to water-rock interactions with the aim of providing a complete picture of the natural signature of the system behavior. An analysis of this kind is outside the scope of the current study and will be the subject of future investigations.

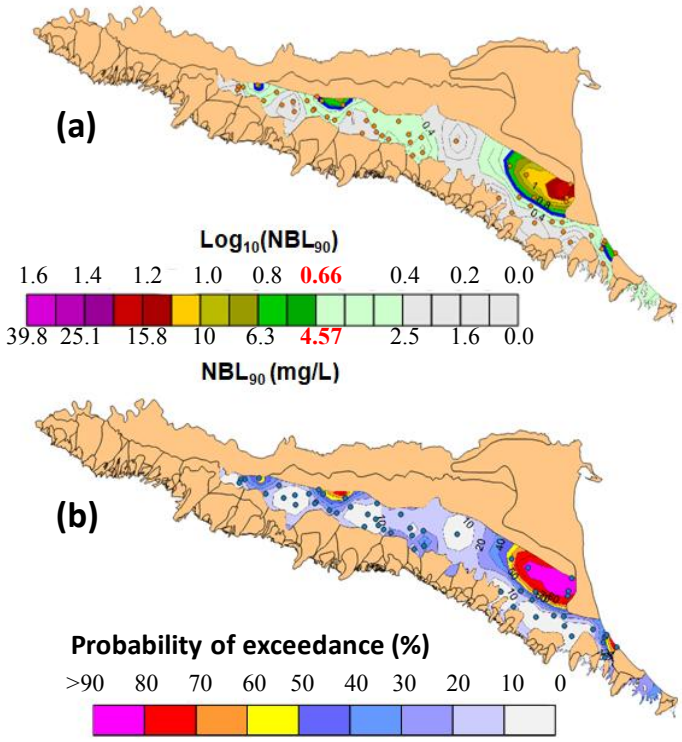


Fig. 2. Groundwater body 0610: spatial maps of (a) kriged $NH_4Log_{10}NBL_{90}$ values and (b) probability of exceedance of the uniform regional (log-transformed) NBL_{90}^{PS} value calculated via the original PS method. The value $NH_4Log_{10}\,NBL_{90}^{PS}$ = 0.66 (red font on the color scale; corresponding to NBL_{90}^{PS} = 4.6 mg/L) is denoted by the thick blue isoline.

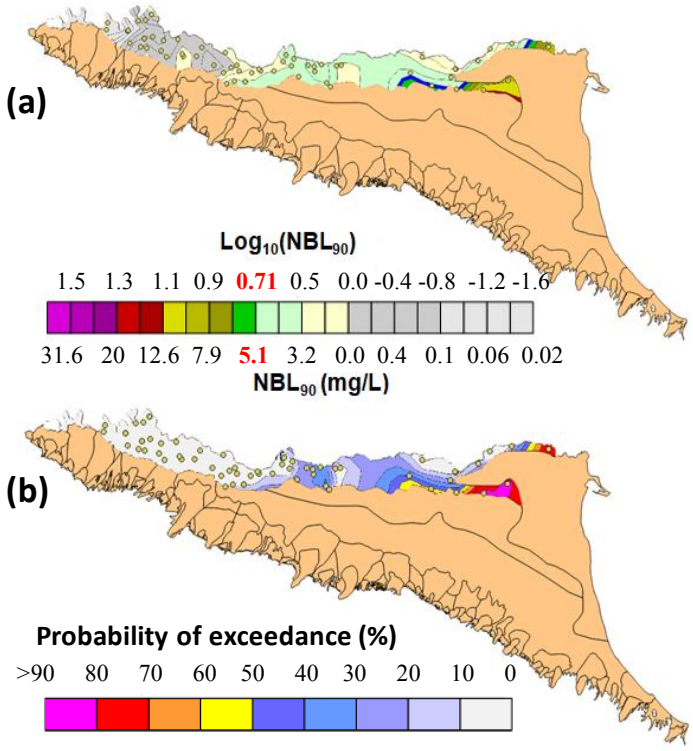


Fig. 3. Groundwater body 0630: spatial maps of (a) kriged $NH_4Log_{10}NBL_{90}$ values and (b) probability of exceedance of the uniform regional (log-transformed) NBL_{90}^{PS} value calculated via the original PS method. The value $NH_4Log_{10}\,NBL_{90}^{PS}$ = 0.71 (red font on the color scale; corresponding to NBL_{90}^{PS} = 5.12 mg/L) is denoted by the thick blue isoline.

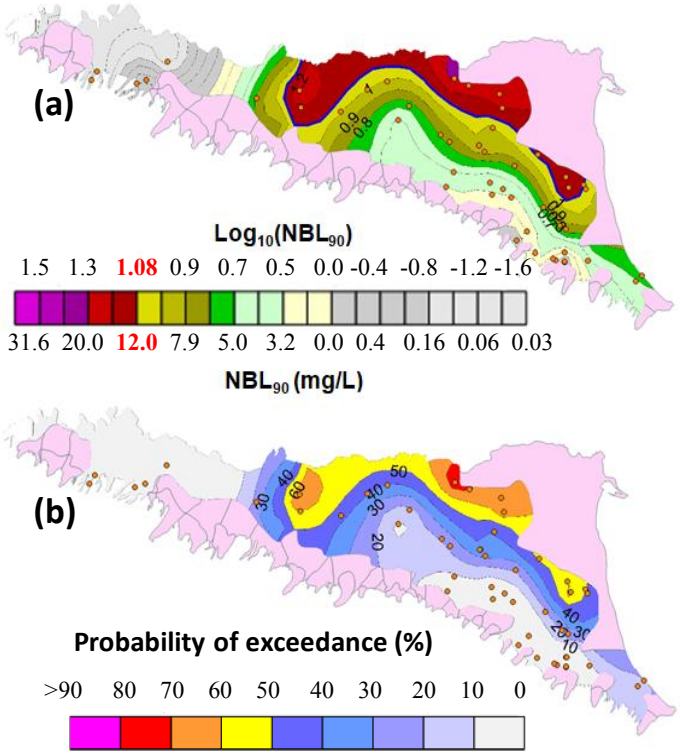


Fig. 4. Groundwater body 2700: spatial maps of (a) kriged $NH_4Log_{10}NBL_{90}$ values and (b) probability of exceedance of the uniform regional (log-transformed) NBL_{90}^{PS} value calculated via the original PS method. The value $NH_4Log_{10}\,NBL_{90}^{PS}$ = 1.08 (red font on the color scale; corresponding to NBL_{90}^{PS} = 12.0 mg/L) is denoted by the thick blue isoline.

3.2 Arsenic

Values of local NBL_{90} obtained for arsenic (As) through the methodology described in Section 2.3 display a considerable degree of spatial variability. These range from 1 to 120.4 µg/L (from 0 to 2.08 in logarithmic scale) within groundwater body 0610, and from 1 to 49.8 µg/L (from 0 to 1.7 in logarithmic scale) and from 0.5 to 70 µg/L (from -0.3 to 1.84 in logarithmic scale), respectively within reservoirs 0630 and 2700.

Table 6

Values for As NBL_{90} estimated via the original PS procedure (NBL_{90}^{PS}). EU Drinking Water Standard for As concentration is 10 µg/L.

groundwater body	NBL_{90}^{PS} (µg/L)	$Log_{10}\ NBL_{90}^{PS}$
0610	33	1.52
0630	4	0.60
2700	6	0.77

Estimated values of NBL_{90}^{PS} for the three investigated water bodies (Molinari et al., 2012) are listed in Table 6 together with the corresponding log-transformed counterparts. We note that EU Drinking Water Standard for As (i.e., 10 µg/L) is exceeded only in the case of the 0610 groundwater body.

Table 7 lists the results of the variogram model calibration analyses including, for each groundwater body, the estimated set of parameters based on minimization of *NLL* (1) and the associated posterior probabilities calculated through (4) and (5).

Table 7

Estimated parameters of variogram models of $AsLog_{10}$ NBL_{90} values and associated posterior probability (*p*) based on *KIC* (4). Here, PN = Pure Nugget, Sph = Spherical model, Exp = Exponential model; n = nugget, a = range (practical range for Exponential model), c = sill; (*) NBL_{90} values are given in (µg/L).

	Groundwater body 0610				Groundwater body 0630				Groundwater body 2700			
Model	n (*)	a (km)	c (*)	*p* (%)	n (*)	a (km)	c (*)	*p* (%)	n (*)	a (km)	c (*)	*p* (%)
PN	0.33	-	-	100	0.252	-	-	0.002	0.238	-	-	6.7
Sph	-	-	-	-	-	12.78	0.256	0.016	-	11.57	0.243	31.6
Exp	-	-	-	-	-	14.57	0.257	0.030	-	11.57	0.243	31.6
Sph	-	-	-	-	0.132	19.80	0.125	99.77	0.07	14.42	0.172	15.0
Exp	-	-	-	-	10^{-5}	14.58	0.257	0.177	0.07	14.42	0.172	15.0

We base our results on log-transformed values of local NBL_{90} (termed as $AsLog_{10}$ NBL_{90}) estimated for each control borehole. The data observed for water body 0610 are characterized by a lack of spatial correlation structure, a pure nugget effect being the only model of choice with the ability to interpret the available data. This result is consistent with the low estimated values for the range of the variogram models employed to characterize the spatial correlation structure of ammonium NBL_{90} in the same water body (see Table 5). As observed in the case of ammonium, these results might be related to (*a*) the lack of sampling points separated by sufficiently small length scales or (*b*) the nature of the variable analyzed, which might or might not display significant spatial correlation structure. As a consequence, kriged local NBL values coincide with the mean of

514 the available data, the associated variance being equal to the nugget. For these reasons, spatial maps
515 of $AsNBL_{90}$ for water body 0610 are not further analyzed.

516 The set of theoretical models employed to interpret the sample spatial correlation structure
517 of $AsNBL_{90}$ across groundwater bodies 0630 and 2700 comprises a pure nugget variogram model,
518 and spherical or exponential models with and without nugget. In the case of water body 0630,
519 model discrimination criteria assign the largest probabilistic weight (very close to 100%) to the
520 spherical model, which is also the model characterized by the largest contribution of the nugget to
521 the total variance (close to 50%). The other tested models are characterized by essentially negligible
522 weights. In the case of water body 2700, our results indicate that all models tested are associated
523 with a non-negligible probabilistic weight, the highest scores being equally assigned to the spherical
524 and exponential models without nugget. It is noted that addition of a nugget effect contributes to
525 about 28% of the total variance of the variogram models analyzed. Comparison of estimated
526 variogram model parameters obtained for arsenic and ammonium shows that variogram ranges for
527 As are significantly lower than those linked to NH_4. This suggests that the spatial distribution of
528 values of $AsNBL_{90}$ may be driven by phenomena occurring at a more localized scale than in the
529 case of NH_4.

530 We apply Ordinary Kriging to obtain Kriging estimates and variances of $AsNBL_{90}$ for each
531 variogram model. These are then weighted via posterior probability values according to (6) and (7),
532 to yield spatial distributions of estimated mean and variance of $AsLog_{10}NBL_{90}$ grounded on the
533 complete set of tested models. As an example of the results obtained by considering individual
534 variogram models, spatial distributions of kriged values for local NBLs of arsenic in water body
535 2700 with the four models listed in Table 5 are depicted in Appendix C.

536 Kriged values of $AsLog_{10}NBL_{90}$ obtained from our multimodel analysis and associated with
537 groundwater bodies 0630 and 2700 are respectively depicted in Figures 5 and 6 together with the
538 probability of exceeding the corresponding uniform regional (log-transformed) NBL_{90}^{PS} value (see
539 Table 6). Exceedance probabilities are calculated following the procedure outlined in Section 3.1.

Areas characterized by estimated mean values of $AsLog_{10}NBL_{90}$ larger than their uniformly distributed counterpart based on the original PS method are also demarcated in Figures 5a and 6a. The corresponding spatial distributions of Kriging variance are depicted in Appendix D.

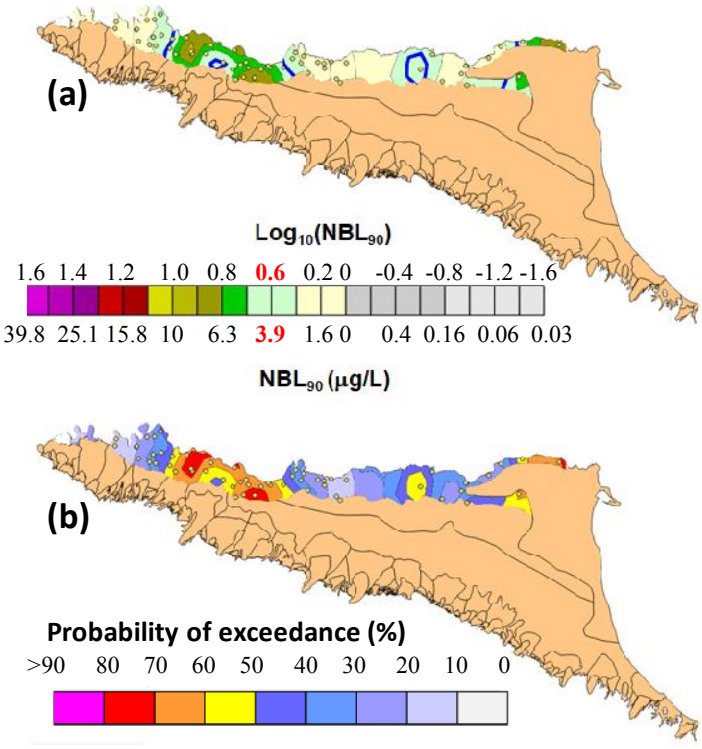


Fig. 5. Groundwater body 0630: spatial maps of (a) kriged $AsLog_{10}NBL_{90}$ values and (b) probability of exceedance of the uniform regional (log-transformed) NBL_{90}^{PS} value calculated via the original PS method. The value $AsLog_{10}\,NBL_{90}^{PS} = 0.60$ (red font on the color scale; corresponding to $NBL_{90}^{PS} = 4$ µg/L) is denoted by the thick blue isoline.

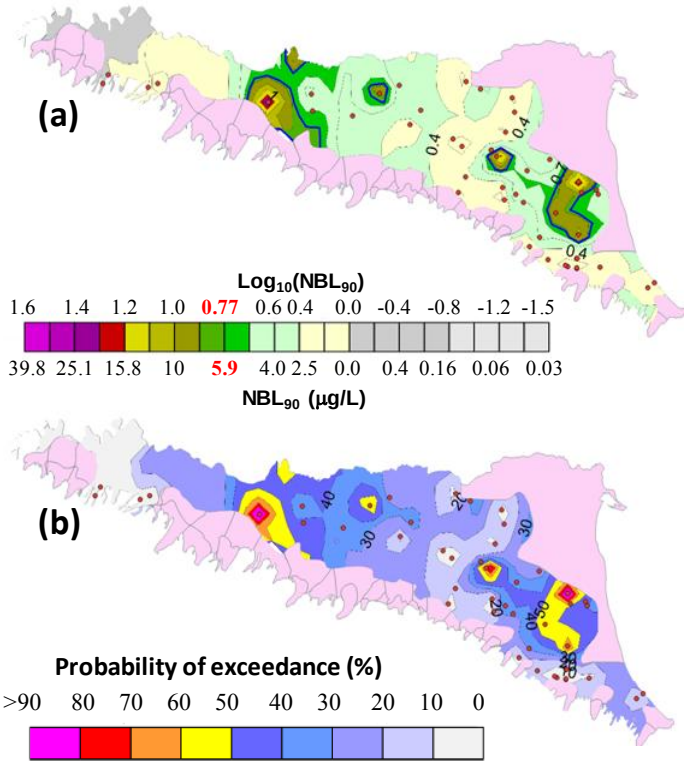


Fig. 6. Groundwater body 2700: spatial maps of (a) kriged $AsLog_{10}NBL_{90}$ values and (b) probability of exceedance of the uniform regional (log-transformed) NBL_{90}^{PS} value calculated via the original PS method. The value $AsLog_{10}\,NBL_{90}^{PS}$ = 0.77 µg/L (red font on the color scale; corresponding to NBL_{90}^{PS} = 6 µg/L) is denoted by the thick blue isoline.

Similar to ammonium, we observe that considering a single NBL_{90} value as representative of the whole water body can lead to underestimating/overestimating the actual natural signature which can locally occur across the reservoir. These findings are markedly relevant for aquifers 0630 and 2700 where arsenic NBL_{90}^{PS} values are lower than the EU drinking water standard, suggesting the potential occurrence of localized areas characterized by local NBL values larger than the regulation limit (10 µg/L). These observations are strengthened and quantified by the resulting exceedance probabilities maps of arsenic (Fig. 5b and 6b). The latter could be consistent with the occurrence of localized high natural arsenic content associated with the occurrence of vegetal matter, specific solid fractions and redox conditions, as evidenced by Molinari et al. (2013, 2015).

4. Conclusions

Our work leads to the following major conclusions:

1. Relying on the estimate of a single (or bulk) NBL value in large-scale groundwater bodies taken to represent the overall (median) behavior of the reservoir, tends to (*a*) mask the actual distribution of local NBLs and (*b*) overestimate (or underestimate) natural background concentrations within localized portions of the system. This might lead to misleading conclusions within areas where values larger than the bulk NBL could be inappropriately associated with external causes (e.g., anthropogenic activities) while they could otherwise be linked to specific and localized natural processes. We show that probabilities of exceedance of EU drinking water standard for hazardous species, such as arsenic, can vary quite significantly across a groundwater body. These elements evidence limitations associated with current directives relying on a unique NBL value taken as representative of the bulk behavior of an aquifer system. Such an approach can be especially critical in large-scale groundwater systems (i.e., with planar extent of the order of thousands of square kilometers) of the kind we analyze in this study. Considering our strategy, which is based on the application of the typical Pre-Selection (PS) methodology to the scale of each observation borehole where temporal records of monitored concentrations are available, yields local estimates of NBL values. These can then be embedded in a geostatistical analysis to characterize their spatial variability.
2. By applying our modified PS approach, we estimate local scale NBLs, evaluated via the information available at a given observation borehole. The interpretive multimodel approach employed allows embedding explicitly uncertainties linked to the choice of the theoretical model selected to describe the degree of spatial correlation (as embedded in the variogram model) of such local scale NBL values. Relying on a single interpretive variogram model can lead to an incomplete quantification of uncertainty, thus suggesting the need for considering the contributions of a set of candidate models to underpin the incomplete knowledge of the system behavior.

3. Spatial estimates of local NBLs and the ensuing uncertainty can serve as a basis to assist (*i*) the demarcation of point and diffuse sources of potentially contaminating areas defining key starting points and clean-up goals for remediation actions; and (*ii*) the management of monitoring networks to optimize information content from areas associated with high uncertainty or high probability of exceedance of environmental limits/standards.
4. Future European and national regulations related to NBLs assessment in large scale water bodies should foresee the possibility of adopting spatial maps grounded on local NBLs (based on temporal data sets). Otherwise, a unique value of NBL could be considered as representative of the whole system for small scale reservoirs (planar extension lower than hundreds of square kilometers) where limited amount of monitoring stations is typically available and hampers a robust assessment of NBL spatial maps (with the associated uncertainty).

Acknowledgments

The research has been performed in the framework of an agreement between the Politecnico di Milano (Department of Civil and Environmental Engineering) and ARPAE - Regional Agency for Prevention, Environment and Energy of Emilia-Romagna who provided environmental data employed within the present study. Funding from ARPAE and the Emilia - Romagna Region is gratefully acknowledged. The EU and MIUR are also acknowledged for funding, in the frame of the collaborative international Consortium (WE-NEED) financed under the ERA-NET WaterWorks2014 Cofunded Call. This ERA-NET is an integral part of the 2015 Joint Activities developed by the Water Challenges for a Changing World Joint Programme Initiative (Water JPI). Data for the geostatistical analyses are available at the following link: https://data.mendeley.com/datasets/km6fgx227j/draft?a=3bdb20a5-adc7-429e-b1d0-3a12bd4202d8.

621 **References**

622 Akaike, H., 1974. A new look at statistical model identification. IEEE Trans. Autom. Control. AC
623 19, 716-723.

624 Amorosi, A., Farina, M., Severi, P., Preti, D., Caporale, L., Di Dio, G., 1996. Genetically related
625 alluvial deposits across active fault zones: an example of alluvial fan-terrace correlation
626 from the upper Quaternary of the southern Po Basin, Italy. Sediment. Geol. 102, 275-295.

627 Ayotte, J.D., Nolan, B.T., Nuckols, J.R., Cantor, K.P., Robinson, G.R. Baris, D., Hayes, L.,
628 Karagas, M., Bress, W., Silverman, D.T., Lubin, J.H., 2006. Modeling the Probability of
629 Arsenic in Groundwater in New England as a Tool for Exposure Assessment. Environ. Sci.
630 Technol 40(11), 3578-3585.

631 Bianchi Janetti, E., Dror, I., Guadagnini, A., Riva, M., Berkowitz, B., 2012. Estimation of single-
632 metal and competitive sorpion isotherm through maximum likelihood and model quality
633 criteria. Soil Sci. Soc. Am. J. 76(4), 1229-1245. doi:10.2136/sssaj2012.0010.

634 BRIDGE - Background cRiteria for the IDentification of Groundwater Thresholds 2007. http://nfp-
635 at.eionet.europa.eu/irc/eionet-circle/bridge/info/data/en/index.htm. (Accessed 25 August
636 2018); or https://cordis.europa.eu/result/rcn/51965_en.html (accessed 25 August 2018).

637 Carrera, J., Neuman, S.P., 1986. Estimation of aquifer parameters under transient and steady state
638 conditions: I. Maximum likelihood method incorporating prior information. Water Resour.
639 Res. 22(2), 199-210.

640 Ciriello, V., Edery, Y., Guadagnini, A., Berkowitz, B., 2015. Multimodel framework for
641 characterization of transport in porous media. Water Resour. Res. 51(5), 3384-3402,
642 doi:10.1002/2015WR017047.

643 Cremonini, S., Etiope, G., Italiano, F., Martinelli G., 2008. Evidence of possible enhanced peat
644 burning by deep-origin methane in the Po River delta Plain (Italy). J. Geol. 116, 401-413.

645 Dalla Libera, N., Fabbri, P., Mason, L., Piccinini, L., Pola, M., 2017. Geostatistics as a tool to
646 improve the Natural Background Level definition: an application in groundwater. Sci. Total
647 Environ. 598, 330-340.
648 Decreto Legislativo n. 30 del 16 marzo 2009 (Legislation Decree n. 30, 16 March, 2009).
649 Application of the Directive 2006/118/CE, related to the protection of groundwater
650 resources from pollution and deterioration (Attuazione della direttiva 2006/118/CE, relativa
651 alla protezione delle acque sotterranee dall'inquinamento e dal deterioramento). Gazzetta
652 Ufficiale n. 79 of 4 April 2009 (in Italian).
653 Directive 2000/60/EC - Water Framework Directive (WFD). Directive of the European Parliament
654 and of the Council of 23 October 2000 establishing a framework for Community action in
655 the field of water policy, OJ L327, 22 Dec 2000, pp 1-73.
656 Directive 2006/118/EC, GroundWater Daughter Directive (GWDD). Directive of the European
657 Parliament and of the Council of 12 December 2006 on the protection of groundwater
658 against pollution and deterioration, OJ L372, 27 Dec 2006, pp 19-31.
659 Directive 2014/80/EU amending Annex II to Directive 2006/118/EC of the European Parliament
660 and of the Council on the Protection of Groundwater Against Pollution and Deterioration,
661 OJ L182, 21 June 2014. pp. 52-55.
662 Doherty, J., 2002. PEST: model independent parameter estimation. User manual. Watermark
663 Numerical Computing, Corinda, Queensland, Australia.
664 Draper, D., 1995. Assessment and propagation of model uncertainty. J. R. Stat. Soc. Series B57, 45-
665 97.
666 Ducci, D., Condesso de Melo, M.T., Preziosi, E., Sellerino, M., Parrone, D., Ribeiro, L., 2016.
667 Combining natural background levels (NBLs) assessment with indicator kriging analysis to
668 improve groundwater quality data interpretation and management. Sci. Total Environ. 569-
669 570, 569-584. doi:10.1016/j.scitotenv.2016.06.184.

Edmunds W.M., Shand, P., Hart, P., Ward, R.S., 2003. The natural (baseline) quality of groundwater: a UK pilot study. Sci. Total Environ. 310, 25-35. doi:10.1016/S0048-9697(02)00620-4.

European Commission. Guidance on groundwater status and trend assessment, guidance document no 18. Technical Report 2009, ISBN 978-92-79-11374-1 European Communities, Luxembourg, 2009.

Farina, M., Marcaccio, M., Zavatti, A., 2014. Experiences and perspectives to monitoring of groundwater resources: the contribution of Emilia Romagna (Esperienze e prospettive nel monitoraggio delle acque sotterranee: il contributo dell'Emilia-Romagna). Bologna, Pitagora Editrice, 528 pp. (in Italian).

Gaus, I., Kinniburgh, D.G., Talbot, J.C., Webster, R., 2003. Geostatistical analysis of arsenic concentration in groundwater in Bangladesh using disjunctive kriging. Environ. Geol. 44, 939-948. doi:10.1007/s00254-003-0837-7.

Gimeno, P., Marcé, R., Bosch, Ll., Comas, J., Corominas, Ll. 2017. Incorporating model uncertainty into the evaluation of interventions to reduce microcontaminant loads in rivers. Water Res. 124, 415-424. https://doi.org/10.1016/j.watres.2017.07.036.

Heibati, M., Stedmon, C.A., Stenroth, K., Rauch, S., Toljander, J., Säve-Söderbergh, M., Murphy, K.R., 2017. Assessment of drinking water quality at the tap using fluorescence spectroscopy. Water Res. 125, 1-10, https://doi.org/10.1016/j.watres.2017.08.020

Hernandez, A.F., Neuman, S.P., Guadagnini, A., Carrera, J., 2006. Inverse stochastic moment analysis of steady state flow in randomly heterogeneous media, Water Resour. Res. 42, W05425, doi:10.1029/2005WR004449.

Hinsby, K., Condesso de Melo, M.T., 2006. Application and evaluation of a proposed methodology for derivation of groundwater threshold values-a case study summary report. In Deliverable D22 of the EU project ‘‘BRIDGE’’ 2006,. http://nfp-at.eionet.europa.eu/Public/irc/eionet-circle/bridge/library?l=/deliverables/d22_final_reppdf/_EN_1.0_&a=d.

696 Hinsby, K., Condesso de Melo, M.T., Dahl, M., 2008. European case studies supporting the
697 derivation of natural background levels and groundwater threshold values for the protection
698 of dependent ecosystems and human health. Sci. Total Environ. 401(1-3), 1-20.
699 Hoeting, J., Madigan, D., Raferty, A., Volinksy, C., 1999. Bayesian model averaging: A tutorial.
700 Statitical Science 14, 382-417.
701 Hurvich, C.M., Tsai, C.L., 1989. Regression and time series model selection in small sample.
702 Biometrika 76(2), 99-104.
703 Kashyap, R.L., 1982. Optimal choice of AR and MA parts in autoregressive moving average
704 models. IEEE Trans. Pattern Anal. Mach. Intel. PAMI 4(2), 99-104.
705 Kim, K.H., Yun, S.T., Kim, H.K., Kim, J.W., 2015. Determination of natural backgrounds and
706 thresholds of nitrate in South Korean groundwater using model-based statistical approaches.
707 J. Geochem. Explor. 148, 196-205.
708 Kumarathilaka, P., Seneweera, S., Meharg,A., Bundschuh, J. 2018. Arsenic speciation dynamics in
709 paddy rice soil-water environment: sources, physico-chemical, and biological factors - a
710 review. Water Res. Available online 21 April 2018.
711 https://doi.org/10.1016/j.watres.2018.04.034.
712 Liu, C.W., Jang, C.S., Liao, C.M., 2004. Evaluation of arsenic contamination potential using
713 indicator kriging in the Yun-Lin aquifer (Taiwan). Sci. Total Environ. 321, 173-188.
714 doi:10.1016/j.scitotenv.2003.09.002
715 Liu, G., Tao, Y., Zhang, Y., Lut, M., Knibbe, W.J., van der Wielen, P., Liu, W., Medema, G., van
716 der Meer, W., 2017. Hotspots for selected metal elements and microbes accumulation and
717 the corresponding water quality deterioration potential in an unchlorinated drinking water
718 distribution system. Water Res. 124, 435-445.
719 Molinari, A., Guadagnini, L., Marcaccio, M., Guadagnini, A., 2012. Natural background levels and
720 threshold values of chemical species in three large-scale groundwater bodies in Northern
721 Italy. Sci. Total Environ. 425, 9-19. doi:10.1016/j.scitotenv.2012.03.015.

722 Molinari, A., Guadagnini, L., Marcaccio, M., Straface, S., Sanchez-Vila, X., Guadagnini, A., 2013.
723 Arsenic release from deep natural solid matrices under experimentally controlled redox
724 conditions. Sci. Total Environ. 444, 231-240. doi: 10.1016/j.scitotenv.2012.11.093
725 Molinari, A., Guadagnini, L., Marcaccio, M., Guadagnini, A., 2015. Arsenic fractioning in natural
726 solid matrices sampled in a deep groundwater body. Geoderma 247, 88-96.
727 https://doi.org/10.1016/j.geoderma.2015.02.011.
728 Neuman, S.P., 2003. Maximum likelihood Bayesian averaging of alternative conceptual-
729 mathematical models. Stoch. Env. Res. Risk A. 17(5), 291-305. doi:10D1007/s00477-003-
730 0151-7.
731 Panno, S.V., Kelly, W.R., Martinsek, A.T., Hackley, K.C., 2006. Estimating background and
732 threshold nitrate concentrations using probability graphs. Groundwater 44(5), 697-709.
733 Qian, Y., Dong, H., Geng, Y., Zhong, S., Tian, X., Yu, Y., Chen, Y., Moss, D.A., 2018. Water
734 footprint characteristic of less developed water-rich regions: Case of Yunnan, China. Water
735 Res., *in press*, Accepted Manuscript, https://doi.org/10.1016/j.watres.2018.03.075.
736 Redman, A.D., Macalady, D., Ahmann, D., 2002. Natural organic matter affects arsenic speciation
737 and sorption onto hematite. Environ. Sci. Technol. 36, 2889-2896.
738 Regione Emilia-Romagna, 2010. Council Decree (Delibera di Giunta) n. 350 of 8/02/2010,
739 Approval of the activities of the Emilia Romagna Region related to the implementation of
740 Directive 2000/60/CE aiming at the design and adoption of the Management Plans of the
741 hydrographic districs Padano, Appennino settentrionale and Appennino centrale
742 (Approvazione delle attività della Regione Emilia-Romagna riguardanti l'implementazione
743 della Direttiva 2000/60/CE ai fini della redazione ed adozione dei Piani di Gestione dei
744 Distretti idrografici Padano, Appennino settentrionale e Appennino centrale).
745 http://ambiente.regione.emilia-romagna.it/acque/temi/piani%20di%20gestione (In Italian,
746 Accessed 25 March 2018).

747 Reimann, C., Garret, R.G., 2005. Geochemical background: concept and reality. Sci. Total Environ.
748 350, 12-27. https://doi.org/10.1016/j.scitotenv.2005.01.047.
749 Remy, N., Boucher, A., Wu, J., 2009. Applied Geostatistics with SGeMS: A User's Guide 1st
750 Edition, Cambridge University Press, New York. ISBN: 978-0-521-51414-9.
751 Ricci Lucchi, F., 1984. Flysh, molassa, clastic deposits: traditional and innovative approaches to the
752 analysis of north Apeninic basins (Flysch, molassa, cunei clastici: tradizione e nuovi
753 approcci nell'analisi dei bacini orogenici dell'Appennino settentrionale). Cento Anni di
754 Geologia Italiana. Volume Giubilare 1° centenario Soc. Geol. Ital., 279-295.
755 Riva, M., Panzeri, M., Guadagnini, A., Neuman, S.P., 2011. Role of model selection criteria in
756 geostatistical inverse estimation of statistical data- and model- parameters. Water Resour.
757 Res. 47, W07502. doi:10.1029/2011WR010480.
758 Ungaro, F., Ragazzi, F., Cappellin, R., Giandon, P., 2008. Arsenic concentration in the soils of the
759 Brenta Plain (Northern Italy): Mapping the probability of exceeding contamination
760 thresholds. J. Geochem. Explor. 96, 117-131. doi:10.1016/j.gexplo.2007.03.006.
761 Walter, T., 2008. Determining natural background values with probability plots. EU Groundwater
762 Policy Developments Conference, UNESCO, Paris, France, 13-15 Nov 2008.
763 Wendland, F., Hannappel, S., Kunkel, R., Schenk, R., Voigt, H.J., Wolter, R., 2005. A procedure to
764 define natural groundwater conditions of groundwater bodies in Germany. Water Sci.
765 Technol. 51(3-4), 249-257.
766 Ye, M., Neuman, S.P., Meyer, P.D., 2004. Maximum likelihood Bayesian averaging of spatial
767 variability models in unsaturated fractured tuff. Water Resour. Res. 40(5), W051131-
768 W0511317.
769 Ye, M., Meyer, P.D., Neuman, S.P., 2008. On model selection criteria in multimodel analysis.
770 Water Resour. Res. 44, W03428.

771 Zlatanović, Lj., van der Hoek, J.P., Vreeburg, J.H.G., 2017. An experimental study on the influence
772 of water stagnation and temperature change on water quality in a full-scale domestic
773 drinking water system. Water Res. 123, 761-772.

Appendix A. Supplementary data

Spatial distributions of kriged values for local NBLs of ammonium in water body 2700 with the four models listed in Table 5 (the value $NH_4Log_{10}\,NBL_{90}^{PS}$ = 1.08 (red font on the color scale; corresponding to NBL_{90}^{PS} = 12.0 mg/L) is denoted by the thick blue isoline.

Appendix B. Supplementary data

Spatial distributions of Kriging variance obtained for ammonium for water bodies 0610, 0630 and 2700 as a result of the multimodel analysis.

Appendix C. Supplementary data

Spatial distributions of kriged values for local NBLs of arsenic in water body 2700 with the four models listed in Table 5. The value $AsLog_{10}\,NBL_{90}^{PS}$ = 0.77 µg/L (red font on the color scale; corresponding to NBL_{90}^{PS} = 6 µg/L) is denoted by the thick blue isoline.

Appendix D. Supplementary data

Spatial distributions of Kriging variance obtained for arsenic for water bodies 0630 and 2700 as a result of the multimodel analysis.

Geostatistical multimodel approach for the assessment of the spatial distribution of natural background concentrations in large-scale groundwater bodies

A. Molinari[a], L. Guadagnini[a,*], M. Marcaccio[b], A. Guadagnini[a,c]

[a]Politecnico di Milano, Dipartimento di Ingegneria Civile e Ambientale, Piazza L. Da Vinci 32, 20133 Milano, Italy

[b]Arpae Emilia-Romagna, Direzione Tecnica, Largo Caduti del Lavoro 6, 40122 Bologna, Italy

[c]University of Arizona, Department of Hydrology and Atmospheric Sciences, 85721 Tucson, AZ, USA

*Corresponding author: laura.guadagnini@polimi.it

Abstract

Quantification of the (spatially distributed) natural contributions to the chemical signature of groundwater resources is an emerging issue in the context of competitive groundwater uses as well as water regulation and management frameworks. Here, we illustrate a geostatistically-based approach for the characterization of spatially variable Natural Background Levels (NBLs) of target chemical species in large-scale groundwater bodies yielding evaluations of local probabilities of exceedance of a given threshold concentration. The approach is exemplified by considering three selected groundwater bodies and focusing on the evaluation of NBLs of ammonium and arsenic, as detected from extensive time series of concentrations collected at monitoring boreholes. Our study is motivated by the observation that reliance on a unique NBL value as representative of the natural geochemical signature of a reservoir can mask the occurrence of localized areas linked to diverse strengths of geogenic contributions to the groundwater status. We start from the application of the typical Pre-Selection (PS) methodology to the scale of each observation borehole to identify local estimates of NBL values. The latter are subsequently subject to geostatistical analysis to obtain estimates of their spatial distribution and the associated uncertainty. A multimodel framework is employed to interpret available data. The impact of alternative variogram models on the resulting spatial distributions of NBLs is assessed through probabilistic weights based on model identification criteria. Our findings highlight that assessing possible impacts of anthropogenic activities on groundwater environments with the aim of designing targeted solutions to restore a good groundwater quality status should consider a probabilistic description of the spatial distribution of NBLs. The latter is useful to provide enhanced information upon which one can then build decision-making protocols embedding the quantification of the associated uncertainty.

Keywords: Natural background levels; groundwater quality; chemical status; multimodel analyses; contaminated aquifers

1. Introduction

Modern society is characterized by an ever-increasing competitive use of groundwater resources, these being subject to many anthropogenic stresses (e.g., domestic use, irrigation and farming activities, industrial operations). To assist evaluation of the resilience of groundwater resources and the soil-water environment serving local communities, several studies are targeted to the analysis of water quality deterioration (e.g., Liu et al., 2017; Zlatanović et al., 2017; Heibati et al., 2017) or of water footprint characteristics (e.g., Qian et al., 2018). When dealing with the assessment of the qualitative status of a target groundwater body as a result of, e.g., in-place monitoring activities, it is not uncommon to evidence areas where detected chemical concentrations attain large values. In some instances, the latter can be directly or indirectly associated with the petrographical composition of the investigated aquifer (e.g., Hinsby and Condesso de Melo, 2006) or with site-specific characteristics such as the occurrence of organic matter (e.g., vegetal matter or peats) which can enhance release of chemical species to groundwater (e.g., Redman et al., 2002; Molinari et al., 2013 and references therein). These elements can in turn yield high natural levels of metals, such as Arsenic, even in crops (e.g., maize or rice) intended for human consumption (e.g., Kumarathilaka et al., 2018). Quantification of the actual (spatially distributed) natural (or geogenic) contributions to chemical concentration is an emerging issue causing increasingly pressing concerns in the context of competitive use of groundwater resources, water regulation and management frameworks at national and European levels, with implications in several industrial activities. Misclassifications of areas where sampled concentrations attain large values as a consequence of geogenic contributions yielding marked Natural Background Levels (NBLs) can have important socio-economic implications related to public health and risk assessment issues. Inaccurate risk assessment analyses can therefore yield an improper classification of the chemical status of an investigated aquifer which might lead to setting unrealistic remediation goals.

Characterization of the actual natural signature of groundwater bodies is a main theme of the EU Water Framework Directive (WFD 2000/60/EC, article 17). A key component required for the

reversal of identified marked and sustained upward trends of contaminants is the proper estimation of NBLs of aquifer bodies. Main aspects related to the definition of NBLs are illustrated in article 2.5 of the GroundWater Daughter Directive (GWDD 2006/118/EC). The latter has been recently amended by Directive 2014/80/EC stating that "the monitoring strategy and interpretation of the data should take account of the fact that flow conditions and groundwater chemistry vary laterally and vertically". With reference to these concepts, and prior to the enactment of Directive 2014/80/EC, Molinari et al. (2012) observe that (*a*) NBLs tend to increase with the average depth of a water body and (*b*) whenever possible, NBLs should be estimated via robust experimental characterization of the geochemical system and modeling studies performed, e.g., through state-of-the-art multicomponent reactive transport approaches.

Statistical analysis of monitored data represents the typical approach employed for NBL estimations (Edmunds et al., 2003; Wendland et al., 2005; Panno et al., 2006; Walter, 2008; Kim et al., 2015). In this context, the EU research project BRIDGE (2007), Background cRiteria for the IDentification of Groundwater thrEsholds, proposes a methodology termed as Pre-Selection (PS). The latter is based on the identification of pristine groundwater samples across an available set of sampled data, as representative of the natural population of the resident concentration. As a result of this procedure, a unique (or bulk) NBL value is estimated and associated with the examined subsurface reservoir, implying that all concentrations exceeding that level should be ascribed to anthropogenic sources. The typical signature of a given chemical species in groundwater may be defined through a range of concentrations rather than a single value (Reimann and Garrett, 2005; Hinsby et al., 2008). This is related to the interaction and feedback between diverse natural, atmospheric, geological, chemical and biological processes taking place in both the vadose and saturated zone during groundwater infiltration and circulation (Edmunds et al., 2003; Wendland et al., 2005; Panno et al., 2006; European Commission, 2009). These concepts are not completely embedded in current regulatory frameworks which requires an estimate of only one threshold value, considered as uniform across a given water body and against which anthropogenic contaminations

should be assessed (Reimann and Garrett, 2005). Otherwise, NBLs can attain markedly different local values, for instance because of the occurrence of diverse petrographic provinces or redox conditions within the same groundwater body, especially in large-scale reservoirs (with areal extent of, e.g., thousands of square kilometers). Hence, the common practice of evaluating the chemical status relying on a single NBL value cannot be considered as realistic and might lead to severe over- or under-estimation of the typical natural signature.

Ducci et al. (2016) and Dalla Libera et al. (2017) recognize that the spatial distribution of NBLs should reflect the heterogeneity of the investigated groundwater body. Critical assumptions associated with these studies are (*a*) the reliance on a unique model employed to interpret experimental variograms, (*b*) the incomplete quantification of the uncertainty of model parameters and estimated concentrations, and (*c*) the lack of a direct estimation of local NBLs associated with each monitoring well upon which exceedance probability maps can be conditioned. Yet, it is well documented that estimated values and uncertainty analyses relying on a single (conceptual and/or mathematical) model can lead to statistical bias or underestimation of the overall uncertainty linked to the system behavior due to undersampling of the space of possible descriptive models. These aspects can be seamlessly embedded within a Maximum Likelihood framework and subsequent reliance on Model Quality criteria to consider uncertainty in the mathematical model depicting the system as well as in its parameters (e.g., Carrera and Neuman, 1986; Ye et al., 2004, Bianchi Janetti et al., 2012 and references therein; Gimeno et al., 2017).

Here, our key objective is to illustrate an approach for the estimation of local NBLs at the borehole scale through the application of a geostatistically-based methodology to yield exceedance probability maps. We accomplish this objective by relying on Maximum Likelihood and formal model identification criteria to take into account uncertainty stemming from multiple and competing models which can be employed to interpret sample variograms of local values of NBLs. To the best of our knowledge, this approach stands as one of the first applications targeted at the evaluation of NBL spatial maps by including quantification of uncertainty associated with the variogram model

employed to interpret the spatial distribution of local NBLs and probability exceedance concentration maps, the latter being usually developed without the evaluation of local NBLs (e.g., Ungaro et al., 2008; Ayotte et al., 2006; Liu et al., 2004; Gaus et al., 2003).

2. Materials and methods

2.1. Study areas

Following the application of WFD 2000/60/EC, a total of 144 groundwater bodies have been delimited within the Emilia-Romagna Region, Italy (Regione Emilia-Romagna, 2010). These are part of the Po Basin fill, which is a syntectonic sedimentary wedge (Ricci Lucchi, 1984) forming the infill of the Pliocene-Pleistocene fore-deep.

The information acquired from sedimentological and hydrogeological analyses has led to the identification of three main hydrogeological complexes, i.e., Apennines alluvial fans, Apennine alluvial plain and alluvial and deltaic Po plain. The system is characterized by a multilayered confined or semiconfined configuration where gravel is gradually replaced by sand deposits in the northern part of the plain, the thickness of fine deposits increasing towards the north portion of the plain (Regione Emilia-Romagna, 2010; Farina et al., 2014). Additional information regarding hydrogeological settings of the study region are available in Molinari et al. (2012) and Farina et al. (2014). An upper confined portion and a lower confined portion have been distinguished within this multilayered system. Our study is focused on the three largest groundwater bodies identified. Two of these are located in the upper confined segment of the aquifer system of the Po Basin fill, the remaining one representing a deep confined water body. Figure 1a depicts limits and planar extent of the two upper confined groundwater bodies, respectively indicated as 0610 and 0630, while the limits and planar extent of the deep confined water body, termed as 2700, are depicted in Figure 1b.

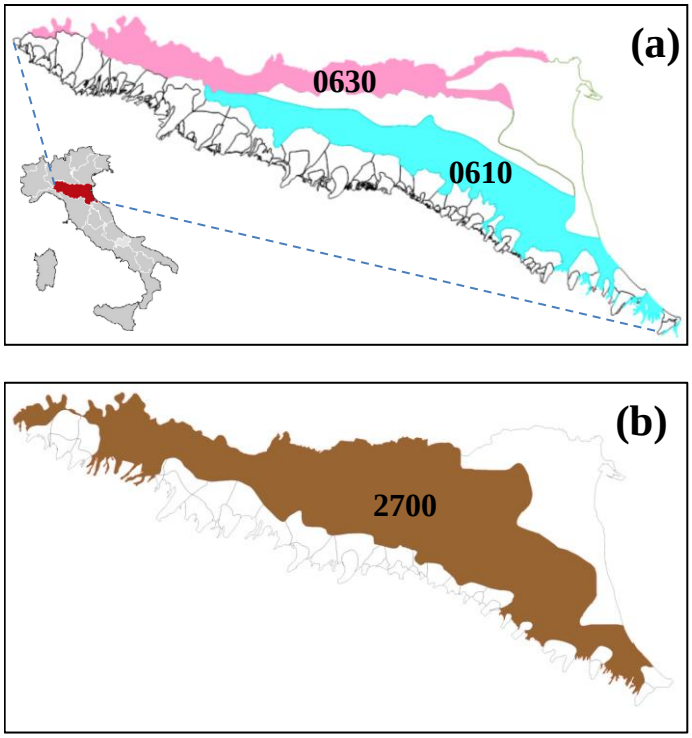


Fig. 1. Planar extent of the investigated water bodies in (a) the upper confined portion and (b) the lower confined portion of the aquifer system of the Po Basin fill; light blue = water body 0610, pink = water body 0630, brown = water body 2700.

Table 1 lists the average depth and thickness as well as the planar area of the three groundwater bodies analyzed.

Table 1

Extension and main characteristic length scales of the three groundwater bodies investigated.

Groundwater body	Average thickness (m)	Average depth (m)	Areal extent (km^2)
0610	130	75	2928
0630	110	65	1995
2700	180	200	6934

Each of these groundwater bodies is subject to various levels of anthropogenic stresses because of diverse and competing uses of the subsurface, in terms of water consumption and withdrawals for agricultural and industrial purposes (Farina et al., 2014). Anthropogenic pollution tends to decrease going from superficial reservoirs to the deepest water body (i.e., 2700) which mostly receives diluted concentrations of chemical species from recharging areas located at some distance.

We consider these three systems because of their significant planar extent whose representative scale is of the order of hundreds of kilometers and where the relevance of considering regional-scale heterogeneous distributions of NBL values for the assessment of groundwater quality is markedly evident.

2.2. Available dataset

We ground our analyses on the time series of concentrations collected at monitoring stations associated with records of about 20 years of observations. These recordings (*a*) extend between 1987 and 2008 (albeit not continuously for some wells), (*b*) have been taken at a six-month interval, and (*c*) constitute a unique data-base that we employ in the context of our investigation. The chemical species considered in this study are ammonium (NH_4) and arsenic (As), which represent critical elements for the achievement of a good chemical status for all three water bodies analyzed, according to Italian Regulation (D. Lgs. 30/09, i.e., Decreto Legislativo n. 30, 16 March 2009) and GWDD 2006/118/CE. As described in Molinari et al. (2012), to which we refer for further details, all concentration data have been subjected to a preliminary exploratory statistical analysis that identified ammonium and arsenic as critical species of concern.

Table 2 lists the number of monitoring stations and the total number of samples collected within the 20-year long record of observations at locations included in the extensive network of observation wells managed by the "Agenzia Regionale per la Prevenzione e l'Ambiente dell'Emilia-Romagna" (ARPAE - Regional Agency for Environmental Protection, Emilia-Romagna).

Table 2

Number of monitoring stations and total number of samples available

		number of samples	
Groundwater body	monitoring stations	As	NH_4
0610	90	1968	2230
0630	75	1692	1917
2700	55	1201	1383

183
184
185 *2.3 Data analysis*

186 *2.3.1 NBL estimation*

187 Within the framework of the EU research project BRIDGE (2007), the Pre-Selection (PS)
188 methodology has been developed for the assessment of the overall geochemical signature of large-
189 scale aquifer systems under data scarcity. The methodology relies on the statistical analysis of the
190 information collected across a monitoring network and is based on the selection of samples that
191 meet certain criteria and can be considered unaffected by anthropogenic influence. Typically
192 adopted criteria for the exclusion of influenced samples are associated with the following
193 conditions: (*a*) chloride concentrations > 1000 mg/L, as indicator of salinity; (*b*) nitrates (NO_3)
194 concentrations > 10 mg/L, as indicator of human influence caused by, e.g., fertilizers; and (*c*)
195 ammonium (NH_4) concentrations > 0.5 mg/L, as indicator of human impact under reducing
196 conditions. Additional criteria, such as redox conditions, dissolved oxygen, sulfate concentration,
197 can be considered for sample exclusion, as pointed out by Hinsby and Condesso de Melo (2006)
198 and Hinsby et al. (2008) to maximize the possibility of grouping homogeneous data.

199 Samples exhibiting markers of anthropogenic contamination (e.g., nitrates or pesticides)
200 larger than a given value have been removed from the original data bank and the residual set is used
201 to estimate the median value for the remaining concentrations of the target chemical species at each
202 monitoring well. The NBL value is then evaluated in terms of a selected percentile of the medians
203 associated with each monitoring well within the investigated water body. The 90^{th}, 95^{th}, or 97.7^{th}
204 percentile are typically considered, depending on the degree of knowledge of the hydrogeochemical
205 system. Wendland et al. (2005) propose to consider the 90^{th} percentile of all of the calculated
206 medians stemming from each monitoring well in the investigated reservoir as an estimate of NBL
207 for the whole water body. Hereinafter, we refer to this quantity as NBL_{90}^{PS}.

As already stated, limitations inherent to the application of this procedure include: (*a*) all the information associated with temporal variability of concentrations are shadowed, and (*b*) only one NBL value is estimated for the whole aquifer body, without the possibility to assess any kind of spatial variability across the system. With the aim of embedding within the analysis spatial and temporal information linked to the scale of observation boreholes, we structure our study through the following main steps:

1. perform sample selection for temporal records at each observation borehole following typically adopted exclusion criteria, as illustrated above and indicated in the original BRIDGE methodology;
2. evaluate a local NBL of the selected chemical species at each observation well as the 90^{th} percentile of concentration values retained at step 1;
3. perform a multimodel geostatistical analysis of the results from step 2, to (*a*) construct and interpret empirical variograms of NBLs, (*b*) project local NBL values onto a computational grid through Kriging and evaluate the associated variance, and (*c*) assess probabilities of exceeding given threshold concentrations, considered as environmental performance metrics characterizing the chemical status of the investigated system (see Section 2.3.2).

We emphasize that this approach directly imbues, as a result of step 2, the effects of the monitored temporal variability of concentrations. As a result of step 3, spatial distributions of direct local NBLs estimated for each observation well can be obtained, together with an appraisal of the associated uncertainty, as reflected in the multimodel theoretical framework considered.

2.3.2 Spatial distribution of local NBLs

Following the approach described in Section 2.3.1 (step 3) we appraise the spatial distribution of local NBLs within the target groundwater bodies through a geostatistical approach, framed in the context of a Bayesian multimodel analysis. The study is performed according to the following steps: (*a*) characterization of the spatial correlation structure of the variable by means of experimental variograms; (*b*) selection of a set of alternative / competing theoretical variogram

234 models and estimation of their parameters (including their uncertainty) through Maximum
235 Likelihood (ML, see Section 2.3.3); (*c*) evaluation (through appropriate discrimination criteria and
236 posterior model weights, see Section 2.3.3) of the relative benefit associated with any of the models
237 considered to interpret available data; (*d*) projection of sample local NBLs onto a computational
238 grid via Kriging by relying on all of the calibrated models; (*e*) assessment of multimodel mean and
239 variance of local NBLs at each grid node (see Section 2.3.4); and (*e*) evaluation of the probability of
240 locally exceeding a given value, i.e., the uniform NBL value obtained through the PS procedure on
241 the regional scale.

242 A base 10 logarithmic transformation is applied to NBL concentration data to map these
243 onto the unbounded support comprising both positive and negative values. Omnidirectional
244 experimental variograms are assessed on the basis of the results of a preliminary variogram
245 analysis. Due to spatial arrangement of available sample points, Kriging estimates and variance are
246 calculated on a uniform grid, with spacing equal to 5 km. Geostatistical analyses has been
247 performed through the well-known and widely tested Stanford Geostatistical Modeling Software
248 (SGeMS; Remy et al., 2009).

249 We calibrate each of the models selected to interpret the evaluated sample variograms
250 through ML parameter estimation and apply model identification criteria to rank the tested models
251 in terms of posterior probabilistic weights. The latter are then used to weigh results associated with
252 each of the selected models and calculate multimodel mean and variance.

253 *2.3.3 Maximum Likelihood (ML) parameter estimation and model quality criteria*

254 Let N be the number of available observations of a model output Y collected in vector
255 $\mathbf{Y}^* = \left[Y_1^*, ...Y_N^*\right]$. Note that in our application these coincide with values of (log-transformed) NBLs
256 (see also Section 3). The covariance matrix of measurement errors, $\mathbf{B}_Y$, is here considered to be
257 diagonal with non-zero terms equal to the observation error variance, σ_i^2 (Carrera and Neuman,
258 1986). Denoting by $\hat{\mathbf{Y}} = \left[\hat{Y}_1, ...\hat{Y}_N\right]$ the vector of model predictions at locations where data are

available, the ML estimate $\hat{\mathbf{x}}$ of the vector of the M uncertain model parameters can be obtained by minimizing with respect to $\mathbf{X}$ the negative log likelihood criterion:

$$NLL = \sum_{i=1}^{N} \frac{J_i}{\sigma_i^2} + \ln\left|\mathbf{B}_Y\right| + N \ln\left(2\pi\right) \tag{1}$$

where $J_i = \left(Y_i^* - \hat{Y}_i\right)^2$. Criterion (1) includes the weighted least square criterion (Carrera and Neuman, 1986; Bianchi Janetti et al., 2012 and references therein). Here, minimization of (1) is achieved using the iterative Levenberg-Marquardt algorithm as embedded in the well documented computational framework PEST (Doherty, 2002).

Alternative (competing) models which can be used to interpret available system states can be ranked by various criteria (e.g., Neuman, 2003; Ye et al., 2004, 2008; Riva et al., 2011; Bianchi Janetti et al., 2012; Ciriello et al., 2015 and references therein), including:

$$AIC = NLL + 2M \tag{2}$$

$$AIC_c = NLL + 2M + \frac{2M\left(M+1\right)}{N-M-1} \tag{3}$$

$$KIC = NLL + M \ln\left(\frac{N}{2\pi}\right) - \ln\left|\mathbf{Q}\right| \tag{4}$$

Here, $\mathbf{Q}$ represents the Cramer-Rao lower-bound approximation for the covariance matrix of the parameter estimates, i.e., the inverse expected Fisher information matrix, which renders a quantitative appraisal of the quality of parameter estimates and of the information content carried by data about model parameters (see, e.g., Ye et al., 2008 for details). The Akaike information criterion, AIC, is due to Akaike (1974), AIC_c to Hurvich and Tsai (1989) and KIC to Kashyap (1982). The lowest value of a given model identification criterion indicates the most favored model (according to the criterion itself) at the expense of the remaining models. Note that *KIC* tends to (*a*) penalize models proportionally to the number of their parameters, through the quantity $M \cdot \ln\left(N/2\pi\right)$ (Ye et al., 2008; Hernandez et al., 2006; Riva et al., 2011) and (*b*) favor models with smaller expected information content per observation, when considering models with equal

parameter numbers, minimum *NLL* values and prior probability of parameters linked to such a minimum (Ye et al., 2008). In light of these observations, we base the analyses presented in this study on *KIC* (4).

The discrimination criteria (2)-(4) can also be considered to assign posterior probability weights quantifying uncertainty associated with each of the tested models. The posterior probability linked to model M_k (k = 1, ..., N_M, N_M being the number of interpreting models considered) is evaluated as (Ye et al., 2008):

$$p(M_k \mid \mathbf{Y}^*) = \frac{\exp\left(-\frac{1}{2}\Delta IC_k\right) p(M_k)}{\sum_{i=1}^{N_M} \exp\left(-\frac{1}{2}\Delta IC_i\right) p(M_i)} \tag{5}$$

Here, $\Delta IC_k = IC_k - IC_{min}$, IC_k being either *AIC* (2), AIC_c (3) or *KIC* (4) and $IC_{min} = \min\{IC_k\}$ its minimum value calculated across the range of models examined; $p(M_k)$ is the prior probability associated with each model. One can set $p(M_k) = 1/N_M$. In case no prior information is available, all models being then characterized by the same prior probability.

Grounding our study on model identification criteria and the ensuing posterior probabilities (5) enables one to rank the models analyzed through their posterior probabilities and discriminate among them in a relative sense.

2.3.4 Multimodel Mean and Variance

We consider a collection **M** of K mutually exclusive variogram models, M_k, upon which lead statistics, such as mean and variance/covariance, of NBL values are computed through Kriging at the nodes of a selected computational grid covering a given aquifer body. The models are uncertain, each of them being assigned the same prior probability $p(M_k)$. Variogram model M_k is employed in a Kriging framework to yield the mean (expectation) $E(\mathbf{Y} \mid \mathbf{D}, M_k)$ and the covariance $Cov(\mathbf{Y} \mid \mathbf{D}, M_k)$ of a vector $\mathbf{Y}$ of random (log-transformed) NBL values, conditional on the prior data vector **D**. The entries of the latter are evaluated according to the procedure illustrated in

Section 2.3.1. Averaging across the moments provided by all K variogram models renders the following (Bayesian-averaged) lead moments (Draper, 1995; Hoeting et al., 1999):

$$E(\mathbf{Y}|\mathbf{D}) = \sum_{k=1}^{K} E(\mathbf{Y}|\mathbf{D}, M_k)\, p(M_k|\mathbf{D}) \tag{6}$$

$$\begin{aligned} Cov(\mathbf{Y}|\mathbf{D}) = & \sum_{k=1}^{K} Cov(\mathbf{Y}|\mathbf{D}, M_k)\, p(M_k|\mathbf{D}) \\ & + \sum_{k=1}^{K} \left[E(\mathbf{Y}|\mathbf{D}, M_k) - E(\mathbf{Y}|\mathbf{D})\right] \\ & \cdot \left[E(\mathbf{Y}|\mathbf{D}, M_k) - E(\mathbf{Y}|\mathbf{D})\right]^T p(M_k|\mathbf{D}) \end{aligned} \tag{7}$$

T denoting transpose. The conditional covariance $Cov(\mathbf{Y}|\mathbf{D})$, resulting from Bayesian model averaging (BMA), is the sum of a within- and between-model contribution. Posterior model probabilities, $p(M_k|\mathbf{D})$, are calculated according to (5) and weigh the contribution of model M_k to BMA moments. For the purpose of our demonstration, we evaluate $E(\mathbf{Y}|\mathbf{D}, M_k)$ and $Cov(\mathbf{Y}|\mathbf{D}, M_k)$ through Kriging performed upon relying on variogram model M_k characterized through ML parameter estimates.

3. Results and discussion

Prior to the application of the PS procedure, the correct attribution of each monitoring station to a target groundwater body has been assessed on the basis of the technical characteristics of the boreholes (e.g., position of the filters) and the available hydrogeological information (e.g., the depth of permeable layers). Note that the application of the exclusion criteria described in Section 2.3.1 has been performed by disregarding NH_4 because the collected sample cores provide evidence of natural occurrence of paleo-peats (Amorosi et al., 1996; Cremonini et al., 2008) consistent with documented increasing NH_4 concentrations with depth in the three reservoirs investigated (see also Molinari et al., 2012).

Exclusion of samples associated with anthropogenic influence (according to the criteria listed above) is followed by identification and removal of outliers from the remaining data set,

yielding the temporal record subsequently employed for local NBL estimation at each control point. Outliers are here identified as high concentration values that lie outside an interval centered around the sample mean of the data remaining after pre-selection and of width equal to three times the associated standard deviation. Excluding these records from the analyses is consistent with the observation that the bulk local environmental behavior is not significantly influenced by isolated (in time) and significantly large concentrations recorded, for instance, as a result of unusual processes taking place only in a particular year (and linked to the occurrence of, e.g., extreme rains that can cause large infiltration enhancing remobilization and diffusion of natural compounds), as opposed to a typical environmental baseline observed across the entire monitoring period.

Application of PS to the time series of each monitoring well is inevitably linked to a reduction of the number of records to be then subject to statistical analyses. We then ground our NBL calculations solely on monitoring wells which, following data selection, exhibit a time series with more than five records, characterizing an active monitoring period spanning at least 3 years (note that samples are collected on a six-month basis). Note that reliance on this kind of analysis is warranted for aquifers where a significant number of monitoring points is available with (*a*) a reliable and extensive temporal record of observations and (*b*) an appropriate degree of spatial coverage of the system. Table 3 lists the number of observation wells effectively employed for each water body after data selection.

Table 3

Number of observation wells available for spatial analysis of NBLs after data selection.

groundwater body	monitoring stations Ammonium	monitoring stations Arsenic
0610	51	50
0630	62	60
2700	47	45

When compared against Table 2, these data reveal that the upper confined water body 0610 suffers from a significant reduction (about 50%) of the total number of monitoring points due to

exclusion of data evidencing possible anthropogenic impact on the water body. Otherwise, only a limited reduction of observation wells is observed in the case of the other two water bodies (with a reduction of 20% and 16%, respectively for water body 0630 and 2700). With reference to anthropogenic pressures, these results are consistent with the observation that (*a*) the water body 2700 is a deep reservoir and hence subject to more limited anthropogenic stresses than the other two bodies, which constitute upper confined reservoirs, (*b*) groundwater body 0630 is located in the proximity of the Po river, that can act as a source of recharge of the reservoir, thus reducing the impact of stresses caused, for instance, by pumping activities, and (*c*) reservoir 0610 is subject to a enhanced anthropogenic impact due to exploitation associated with intensive industrial/agricultural activities.

3.1. Ammonium

Values of local NBL_{90} obtained at each monitoring well through the methodology described in Section 2.3 display a high degree of spatial variability. These range from 0.828 to 20.835 mg/L (i.e., -0.08 to 1.319 in logarithmic scale) within groundwater body 0610, while ranging from 0.025 to 14.406 mg/L (i.e., -1.602 to 1.159 in logarithmic scale) and from 0.025 to 30.362 mg/L (i.e., -1.602 to 1.482 in logarithmic scale), respectively within groundwater body 0630 and 2700. Values of NBL_{90}^{PS} estimated through the standard PS procedure at the regional scale (i.e., one value for each reservoir) for the three water bodies analyzed (Molinari et al., 2012) are listed in Table 4 together with the corresponding logarithmic-transformed values. We observe that all estimates exceed EU Drinking Water Standard for ammonium.

Table 4

Values for NH_4 NBL_{90} estimated via the original PS procedure (NBL_{90}^{PS}). EU Drinking Water Standard for NH_4 concentration is 0.5 mg/L.

groundwater body	NBL_{90}^{PS} (mg/L)	Log_{10} NBL_{90}^{PS}
0610	4.6	0.66
0630	5.2	0.71
2700	12.0	1.08

For the purpose of our application, classical isotropic Spherical or Exponential variogram models, with or without nugget, are considered. Table 5 lists the results of variogram model calibration analysis. For each groundwater body, these include the estimated set of model parameters (i.e., nugget, range, and sill) based on minimization of *NLL* (1) and the associated posterior probabilities calculated through (4) and (5). As stated, the analyses have been performed by considering logarithmic transformation of values of NBL_{90} (termed as $NH_4Log_{10}\,NBL_{90}$).

Table 5

Estimated parameters of variogram models of $NH_4Log_{10}\,NBL_{90}$ values and associated posterior probability (*p*) based on *KIC* (4). Here, Sph = Spherical model, Exp = Exponential model; n = nugget, a = range (practical range for Exponential model), c = sill; (*) NBL_{90} values are given in (mg/L).

	Groundwater body 0610				Groundwater body 0630				Groundwater body 2700			
Model	n (*)	a (km)	c (*)	*p* (%)	n (*)	a (km)	c (*)	*p* (%)	n (*)	a (km)	c (*)	*p* (%)
Sph	-	18.7	0.106	52	-	83.5	0.496	40	-	81.7	0.485	6
Exp	-	22.9	0.107	48	-	138.2	0.581	49	-	100.4	0.509	34
Sph	-	-	-	-	0.045	87.9	0.452	4.2	0.1	101.0	0.401	5
Exp	-	-	-	-	0.031	159.6	0.574	6.6	0.1	208.9	0.538	55

The spatial variogram structure of $NH_4Log_{10}\,NBL_{90}$ within groundwater body 0610 has been interpreted by a Spherical and an Exponential model without nugget effect (introducing a nugget effect in the model calibration procedure is associated with near-zero estimated nugget and posterior weight and is therefore disregarded). Both variogram models exhibit similar values of range (practical range in the case of Exponential model) and sill, with nearly coinciding posterior probabilities.

Four theoretical models, i.e., Spherical and Exponential models with and without nugget, are considered for groundwater bodies 0630 and 2700. The contribution of the nugget to the total variance is less than 10% in the case of groundwater body 0630. Otherwise, the estimated nugget

effect for groundwater body 2700 is equal to 15.7% and 20% of the total variance, respectively for the Exponential and Spherical model with nugget, suggesting the occurrence of a significant degree of variability between sample pairs at short distances.

Estimated values of correlation scale (i.e., the range) are lowest for water body 0610, a finding that might be related to the spatial arrangement of sampling boreholes across the system or to the observation that state variables, such as given quantiles of concentration time series observed at multiple wells, are not necessarily characterized by a strong degree of spatial correlation, due to the dynamics associated with the key processes driving chemical migration in the system. Otherwise, a consistent spatial persistence of correlation structure of ammonium NBL_{90} is observed in water bodies 0630 and 2700, where spatial variations of kriged estimates are then expected to be quite smooth. The largest posterior probability values for water body 0630 are associated with Spherical and Exponential models without nugget. On the other hand, the Exponential variogram model (with or without nugget) is unambiguously favored in water body 2700.

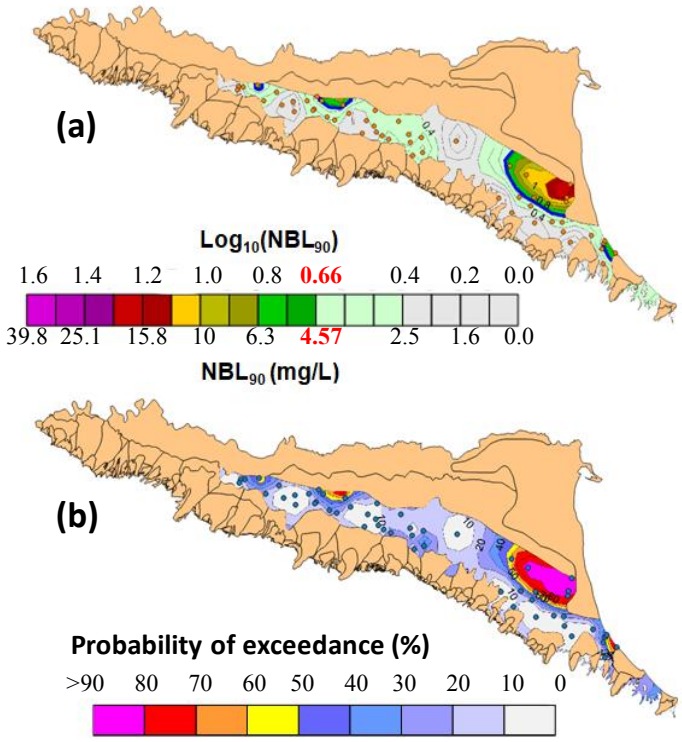
As described in Section 2.3, the calibrated variogram models are then employed to obtain spatial distributions of Kriging estimates of log-transformed values of NBL_{90} and their associated variance across each of the water bodies considered. For the purpose of this study we rely on Ordinary Kriging, other flavors of Kriging being compatible with our approach. Kriging estimates and variances obtained for each model are then weighted on the basis of posterior probability values according to (6) and (7), to yield spatial distributions of estimated mean and variance of NH_4Log_{10} NBL_{90} based on the complete set of tested models and fully including information on model uncertainty.

As an example of the results obtained by considering individual variogram models, spatial distributions of kriged values for local NBLs of ammonium in water body 2700 with the four models listed in Table 5 are depicted in Appendix A. The spatial maps of kriged values of NH_4Log_{10} NBL_{90} for groundwater bodies 0610, 0630 and 2700 and resulting from our multimodel analysis are respectively depicted in Figures 2, 3 and 4 together with the probability of exceeding

424 the uniform regional (log-transformed) NBL_{90}^{PS} value calculated via the original PS method (see
425 Table 4). For completeness, the corresponding spatial distributions of Kriging variance for the three
426 water bodies investigated are depicted in Appendix B. Exceedance probabilities are calculated by
427 assuming a Gaussian distribution of log-transformed NBL_{90} with local mean and variance rendered
428 by the multimodel analysis. Areas characterized by estimated mean values of $NH_4Log_{10}NBL_{90}$
429 larger than their uniformly distributed counterpart based on the original PS method are also
430 demarcated in Figures 2a, 3a, and 4a.

431 Spatial patterns of mean Log_{10} NBL_{90} display a marked degree of variability across the
432 systems, suggesting that relying on a single (uniform) value can shadow the proper representation
433 of the actual NBL distribution within a given reservoir. For example, one can observe the
434 occurrence of areas where Kriging-based NBL_{90} exhibit values larger than NBL_{90}^{PS}. Otherwise, one
435 can also observe the occurrence of regions where estimated average NBLs are lower than NBL_{90}^{PS},
436 yet still larger than the EU drinking water standard set for ammonium. These observations reinforce
437 the idea that the use of a single NBL value as representative of the whole reservoir may lead to
438 misleading conclusions with reference to the natural behavior of the water body. For example, one
439 might argue that values larger than NBL_{90}^{PS} be associated with external causes (e.g., anthropogenic
440 activities) while they could be linked to specific and local hydrogeochemical natural processes.
441 Likewise, one should also consider that it is possible that concentration values in some areas be
442 lower than NBL_{90}^{PS}, yet larger than their regulatory-based counterpart (i.e., 0.5 mg/L in the case of
443 NH_4), as a result of natural processes rather than anthropogenically induced pollution phenomena.
444 These observations are also supported by the exceedance probability maps depicted in Figures 2b,
445 3b, and 4b. These suggest that there is non-negligible probability that NBL_{90}^{PS} be exceeded over a
446 large portion of the system, reinforcing the concept that the emergence of areas associated with a
447 geogenic origin of a target pollutant is masked when a single NBL_{90} value is taken as representative
448 of the whole reservoir. These sets of results represent an element upon which one could derive

449 information for the design of additional investigations. These could be directed, for example, to
450 reduce uncertainty (as quantified, e.g., by large values of Kriging variance) in critical areas and/or
451 to support the findings of the geostatistical analyses through the joint use of other types of
452 information, including, e.g., (*a*) site-specific mineralogy, (*b*) the extent of petrographic provinces,
453 or (*c*) the attainment of local natural hydrochemical equilibria linked to water-rock interactions with
454 the aim of providing a complete picture of the natural signature of the system behavior. An analysis
455 of this kind is outside the scope of the current study and will be the subject of future investigations.
456



457
458
459 Fig. 2. Groundwater body 0610: spatial maps of (a) kriged $NH_4Log_{10}NBL_{90}$ values and (b)
460 probability of exceedance of the uniform regional (log-transformed) NBL_{90}^{PS} value calculated via the
461 original PS method. The value $NH_4Log_{10}\,NBL_{90}^{PS}$ = 0.66 (red font on the color scale; corresponding to
462 NBL_{90}^{PS} = 4.6 mg/L) is denoted by the thick blue isoline.
463

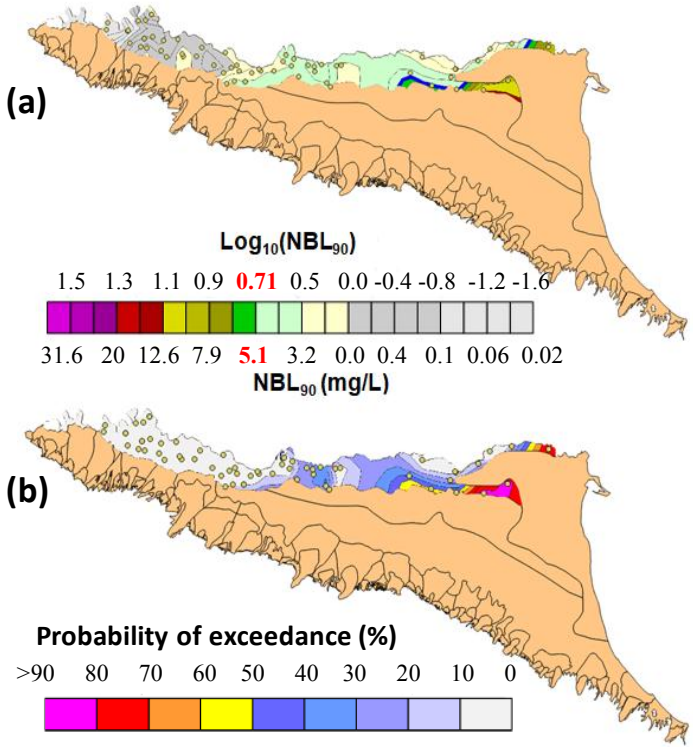


Fig. 3. Groundwater body 0630: spatial maps of (a) kriged $NH_4Log_{10}NBL_{90}$ values and (b) probability of exceedance of the uniform regional (log-transformed) NBL_{90}^{PS} value calculated via the original PS method. The value $NH_4Log_{10}\,NBL_{90}^{PS}$ = 0.71 (red font on the color scale; corresponding to NBL_{90}^{PS} = 5.12 mg/L) is denoted by the thick blue isoline.

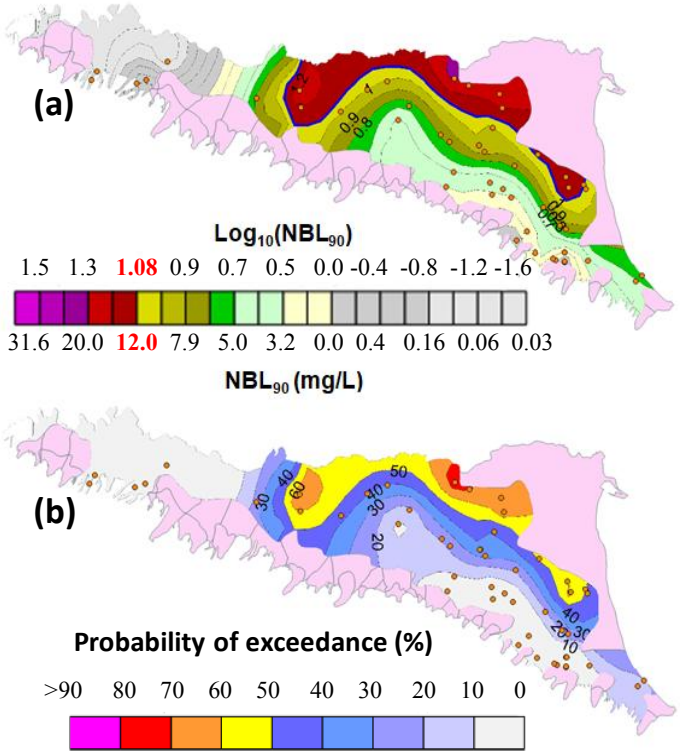


Fig. 4. Groundwater body 2700: spatial maps of (a) kriged $NH_4Log_{10}NBL_{90}$ values and (b) probability of exceedance of the uniform regional (log-transformed) NBL_{90}^{PS} value calculated via the original PS method. The value $NH_4Log_{10}\,NBL_{90}^{PS}$ = 1.08 (red font on the color scale; corresponding to NBL_{90}^{PS} = 12.0 mg/L) is denoted by the thick blue isoline.

3.2 Arsenic

Values of local NBL_{90} obtained for arsenic (As) through the methodology described in Section 2.3 display a considerable degree of spatial variability. These range from 1 to 120.4 µg/L (from 0 to 2.08 in logarithmic scale) within groundwater body 0610, and from 1 to 49.8 µg/L (from 0 to 1.7 in logarithmic scale) and from 0.5 to 70 µg/L (from -0.3 to 1.84 in logarithmic scale), respectively within reservoirs 0630 and 2700.

Table 6

Values for As NBL_{90} estimated via the original PS procedure (NBL_{90}^{PS}). EU Drinking Water Standard for As concentration is 10 µg/L.

groundwater body	NBL_{90}^{PS} (µg/L)	$Log_{10}\ NBL_{90}^{PS}$
0610	33	1.52
0630	4	0.60
2700	6	0.77

Estimated values of NBL_{90}^{PS} for the three investigated water bodies (Molinari et al., 2012) are listed in Table 6 together with the corresponding log-transformed counterparts. We note that EU Drinking Water Standard for As (i.e., 10 µg/L) is exceeded only in the case of the 0610 groundwater body.

Table 7 lists the results of the variogram model calibration analyses including, for each groundwater body, the estimated set of parameters based on minimization of *NLL* (1) and the associated posterior probabilities calculated through (4) and (5).

Table 7

Estimated parameters of variogram models of $AsLog_{10}$ NBL_{90} values and associated posterior probability (*p*) based on *KIC* (4). Here, PN = Pure Nugget, Sph = Spherical model, Exp = Exponential model; n = nugget, a = range (practical range for Exponential model), c = sill; (*) NBL_{90} values are given in (µg/L).

	Groundwater body 0610				Groundwater body 0630				Groundwater body 2700			
Model	n (*)	a (km)	c (*)	*p* (%)	n (*)	a (km)	c (*)	*p* (%)	n (*)	a (km)	c (*)	*p* (%)
PN	0.33	-	-	100	0.252	-	-	0.002	0.238	-	-	6.7
Sph	-	-	-	-	-	12.78	0.256	0.016	-	11.57	0.243	31.6
Exp	-	-	-	-	-	14.57	0.257	0.030	-	11.57	0.243	31.6
Sph	-	-	-	-	0.132	19.80	0.125	99.77	0.07	14.42	0.172	15.0
Exp	-	-	-	-	10^{-5}	14.58	0.257	0.177	0.07	14.42	0.172	15.0

We base our results on log-transformed values of local NBL_{90} (termed as $AsLog_{10}$ NBL_{90}) estimated for each control borehole. The data observed for water body 0610 are characterized by a lack of spatial correlation structure, a pure nugget effect being the only model of choice with the ability to interpret the available data. This result is consistent with the low estimated values for the range of the variogram models employed to characterize the spatial correlation structure of ammonium NBL_{90} in the same water body (see Table 5). As observed in the case of ammonium, these results might be related to (*a*) the lack of sampling points separated by sufficiently small length scales or (*b*) the nature of the variable analyzed, which might or might not display significant spatial correlation structure. As a consequence, kriged local NBL values coincide with the mean of

514 the available data, the associated variance being equal to the nugget. For these reasons, spatial maps
515 of $AsNBL_{90}$ for water body 0610 are not further analyzed.

516 The set of theoretical models employed to interpret the sample spatial correlation structure
517 of $AsNBL_{90}$ across groundwater bodies 0630 and 2700 comprises a pure nugget variogram model,
518 and spherical or exponential models with and without nugget. In the case of water body 0630,
519 model discrimination criteria assign the largest probabilistic weight (very close to 100%) to the
520 spherical model, which is also the model characterized by the largest contribution of the nugget to
521 the total variance (close to 50%). The other tested models are characterized by essentially negligible
522 weights. In the case of water body 2700, our results indicate that all models tested are associated
523 with a non-negligible probabilistic weight, the highest scores being equally assigned to the spherical
524 and exponential models without nugget. It is noted that addition of a nugget effect contributes to
525 about 28% of the total variance of the variogram models analyzed. Comparison of estimated
526 variogram model parameters obtained for arsenic and ammonium shows that variogram ranges for
527 As are significantly lower than those linked to NH_4. This suggests that the spatial distribution of
528 values of $AsNBL_{90}$ may be driven by phenomena occurring at a more localized scale than in the
529 case of NH_4.

530 We apply Ordinary Kriging to obtain Kriging estimates and variances of $AsNBL_{90}$ for each
531 variogram model. These are then weighted via posterior probability values according to (6) and (7),
532 to yield spatial distributions of estimated mean and variance of $AsLog_{10}NBL_{90}$ grounded on the
533 complete set of tested models. As an example of the results obtained by considering individual
534 variogram models, spatial distributions of kriged values for local NBLs of arsenic in water body
535 2700 with the four models listed in Table 5 are depicted in Appendix C.

536 Kriged values of $AsLog_{10}NBL_{90}$ obtained from our multimodel analysis and associated with
537 groundwater bodies 0630 and 2700 are respectively depicted in Figures 5 and 6 together with the
538 probability of exceeding the corresponding uniform regional (log-transformed) NBL_{90}^{PS} value (see
539 Table 6). Exceedance probabilities are calculated following the procedure outlined in Section 3.1.

Areas characterized by estimated mean values of $AsLog_{10}NBL_{90}$ larger than their uniformly distributed counterpart based on the original PS method are also demarcated in Figures 5a and 6a. The corresponding spatial distributions of Kriging variance are depicted in Appendix D.

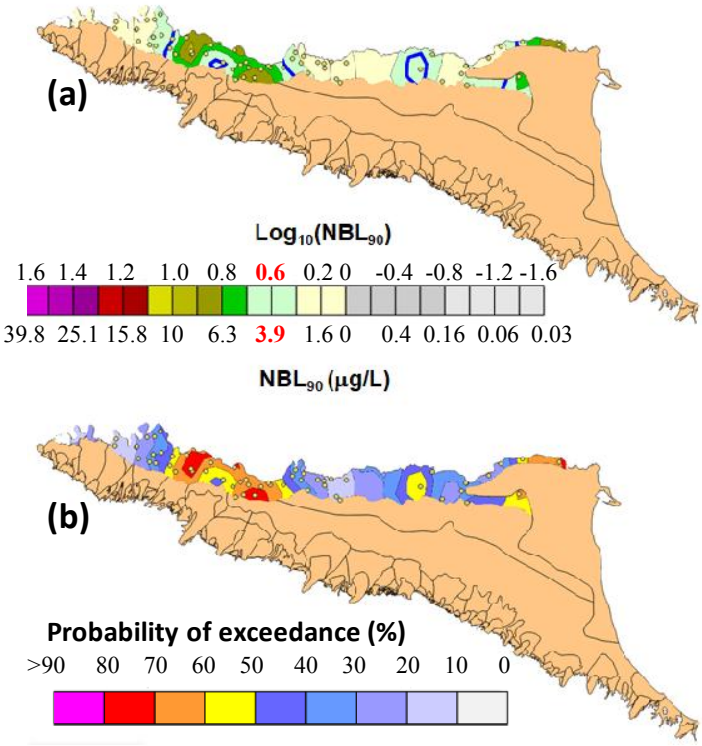


Fig. 5. Groundwater body 0630: spatial maps of (a) kriged $AsLog_{10}NBL_{90}$ values and (b) probability of exceedance of the uniform regional (log-transformed) NBL_{90}^{PS} value calculated via the original PS method. The value $AsLog_{10}\,NBL_{90}^{PS} = 0.60$ (red font on the color scale; corresponding to $NBL_{90}^{PS} = 4$ µg/L) is denoted by the thick blue isoline.

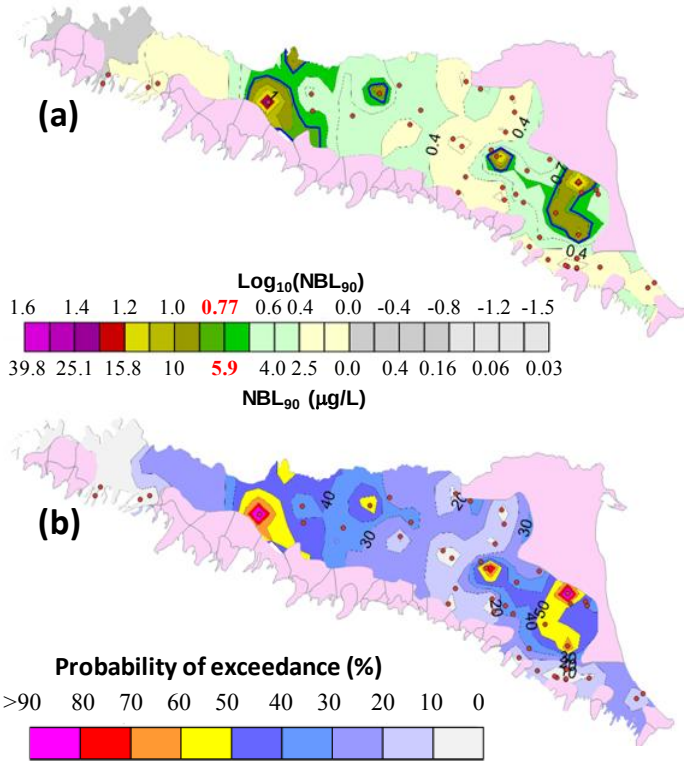


Fig. 6. Groundwater body 2700: spatial maps of (a) kriged $AsLog_{10}NBL_{90}$ values and (b) probability of exceedance of the uniform regional (log-transformed) NBL_{90}^{PS} value calculated via the original PS method. The value $AsLog_{10}\,NBL_{90}^{PS}$ = 0.77 µg/L (red font on the color scale; corresponding to NBL_{90}^{PS} = 6 µg/L) is denoted by the thick blue isoline.

Similar to ammonium, we observe that considering a single NBL_{90} value as representative of the whole water body can lead to underestimating/overestimating the actual natural signature which can locally occur across the reservoir. These findings are markedly relevant for aquifers 0630 and 2700 where arsenic NBL_{90}^{PS} values are lower than the EU drinking water standard, suggesting the potential occurrence of localized areas characterized by local NBL values larger than the regulation limit (10 µg/L). These observations are strengthened and quantified by the resulting exceedance probabilities maps of arsenic (Fig. 5b and 6b). The latter could be consistent with the occurrence of localized high natural arsenic content associated with the occurrence of vegetal matter, specific solid fractions and redox conditions, as evidenced by Molinari et al. (2013, 2015).

4. Conclusions

Our work leads to the following major conclusions:

1. Relying on the estimate of a single (or bulk) NBL value in large-scale groundwater bodies taken to represent the overall (median) behavior of the reservoir, tends to (*a*) mask the actual distribution of local NBLs and (*b*) overestimate (or underestimate) natural background concentrations within localized portions of the system. This might lead to misleading conclusions within areas where values larger than the bulk NBL could be inappropriately associated with external causes (e.g., anthropogenic activities) while they could otherwise be linked to specific and localized natural processes. We show that probabilities of exceedance of EU drinking water standard for hazardous species, such as arsenic, can vary quite significantly across a groundwater body. These elements evidence limitations associated with current directives relying on a unique NBL value taken as representative of the bulk behavior of an aquifer system. Such an approach can be especially critical in large-scale groundwater systems (i.e., with planar extent of the order of thousands of square kilometers) of the kind we analyze in this study. Considering our strategy, which is based on the application of the typical Pre-Selection (PS) methodology to the scale of each observation borehole where temporal records of monitored concentrations are available, yields local estimates of NBL values. These can then be embedded in a geostatistical analysis to characterize their spatial variability.
2. By applying our modified PS approach, we estimate local scale NBLs, evaluated via the information available at a given observation borehole. The interpretive multimodel approach employed allows embedding explicitly uncertainties linked to the choice of the theoretical model selected to describe the degree of spatial correlation (as embedded in the variogram model) of such local scale NBL values. Relying on a single interpretive variogram model can lead to an incomplete quantification of uncertainty, thus suggesting the need for considering the contributions of a set of candidate models to underpin the incomplete knowledge of the system behavior.

3. Spatial estimates of local NBLs and the ensuing uncertainty can serve as a basis to assist (*i*) the demarcation of point and diffuse sources of potentially contaminating areas defining key starting points and clean-up goals for remediation actions; and (*ii*) the management of monitoring networks to optimize information content from areas associated with high uncertainty or high probability of exceedance of environmental limits/standards.
4. Future European and national regulations related to NBLs assessment in large scale water bodies should foresee the possibility of adopting spatial maps grounded on local NBLs (based on temporal data sets). Otherwise, a unique value of NBL could be considered as representative of the whole system for small scale reservoirs (planar extension lower than hundreds of square kilometers) where limited amount of monitoring stations is typically available and hampers a robust assessment of NBL spatial maps (with the associated uncertainty).

Acknowledgments

The research has been performed in the framework of an agreement between the Politecnico di Milano (Department of Civil and Environmental Engineering) and ARPAE - Regional Agency for Prevention, Environment and Energy of Emilia-Romagna who provided environmental data employed within the present study. Funding from ARPAE and the Emilia - Romagna Region is gratefully acknowledged. The EU and MIUR are also acknowledged for funding, in the frame of the collaborative international Consortium (WE-NEED) financed under the ERA-NET WaterWorks2014 Cofunded Call. This ERA-NET is an integral part of the 2015 Joint Activities developed by the Water Challenges for a Changing World Joint Programme Initiative (Water JPI).

Data for the geostatistical analyses are available at the following link: https://data.mendeley.com/datasets/km6fgx227j/draft?a=3bdb20a5-adc7-429e-b1d0-3a12bd4202d8.

621 **References**

622 Akaike, H., 1974. A new look at statistical model identification. IEEE Trans. Autom. Control. AC
623 19, 716-723.

624 Amorosi, A., Farina, M., Severi, P., Preti, D., Caporale, L., Di Dio, G., 1996. Genetically related
625 alluvial deposits across active fault zones: an example of alluvial fan-terrace correlation
626 from the upper Quaternary of the southern Po Basin, Italy. Sediment. Geol. 102, 275-295.

627 Ayotte, J.D., Nolan, B.T., Nuckols, J.R., Cantor, K.P., Robinson, G.R. Baris, D., Hayes, L.,
628 Karagas, M., Bress, W., Silverman, D.T., Lubin, J.H., 2006. Modeling the Probability of
629 Arsenic in Groundwater in New England as a Tool for Exposure Assessment. Environ. Sci.
630 Technol 40(11), 3578-3585.

631 Bianchi Janetti, E., Dror, I., Guadagnini, A., Riva, M., Berkowitz, B., 2012. Estimation of single-
632 metal and competitive sorpion isotherm through maximum likelihood and model quality
633 criteria. Soil Sci. Soc. Am. J. 76(4), 1229-1245. doi:10.2136/sssaj2012.0010.

634 BRIDGE - Background cRiteria for the IDentification of Groundwater Thresholds 2007. http://nfp-
635 at.eionet.europa.eu/irc/eionet-circle/bridge/info/data/en/index.htm. (Accessed 25 August
636 2018); or https://cordis.europa.eu/result/rcn/51965_en.html (accessed 25 August 2018).

637 Carrera, J., Neuman, S.P., 1986. Estimation of aquifer parameters under transient and steady state
638 conditions: I. Maximum likelihood method incorporating prior information. Water Resour.
639 Res. 22(2), 199-210.

640 Ciriello, V., Edery, Y., Guadagnini, A., Berkowitz, B., 2015. Multimodel framework for
641 characterization of transport in porous media. Water Resour. Res. 51(5), 3384-3402,
642 doi:10.1002/2015WR017047.

643 Cremonini, S., Etiope, G., Italiano, F., Martinelli G., 2008. Evidence of possible enhanced peat
644 burning by deep-origin methane in the Po River delta Plain (Italy). J. Geol. 116, 401-413.

645 Dalla Libera, N., Fabbri, P., Mason, L., Piccinini, L., Pola, M., 2017. Geostatistics as a tool to
646 improve the Natural Background Level definition: an application in groundwater. Sci. Total
647 Environ. 598, 330-340.
648 Decreto Legislativo n. 30 del 16 marzo 2009 (Legislation Decree n. 30, 16 March, 2009).
649 Application of the Directive 2006/118/CE, related to the protection of groundwater
650 resources from pollution and deterioration (Attuazione della direttiva 2006/118/CE, relativa
651 alla protezione delle acque sotterranee dall'inquinamento e dal deterioramento). Gazzetta
652 Ufficiale n. 79 of 4 April 2009 (in Italian).
653 Directive 2000/60/EC - Water Framework Directive (WFD). Directive of the European Parliament
654 and of the Council of 23 October 2000 establishing a framework for Community action in
655 the field of water policy, OJ L327, 22 Dec 2000, pp 1-73.
656 Directive 2006/118/EC, GroundWater Daughter Directive (GWDD). Directive of the European
657 Parliament and of the Council of 12 December 2006 on the protection of groundwater
658 against pollution and deterioration, OJ L372, 27 Dec 2006, pp 19-31.
659 Directive 2014/80/EU amending Annex II to Directive 2006/118/EC of the European Parliament
660 and of the Council on the Protection of Groundwater Against Pollution and Deterioration,
661 OJ L182, 21 June 2014. pp. 52-55.
662 Doherty, J., 2002. PEST: model independent parameter estimation. User manual. Watermark
663 Numerical Computing, Corinda, Queensland, Australia.
664 Draper, D., 1995. Assessment and propagation of model uncertainty. J. R. Stat. Soc. Series B57, 45-
665 97.
666 Ducci, D., Condesso de Melo, M.T., Preziosi, E., Sellerino, M., Parrone, D., Ribeiro, L., 2016.
667 Combining natural background levels (NBLs) assessment with indicator kriging analysis to
668 improve groundwater quality data interpretation and management. Sci. Total Environ. 569-
669 570, 569-584. doi:10.1016/j.scitotenv.2016.06.184.

Edmunds W.M., Shand, P., Hart, P., Ward, R.S., 2003. The natural (baseline) quality of groundwater: a UK pilot study. Sci. Total Environ. 310, 25-35. doi:10.1016/S0048-9697(02)00620-4.

European Commission. Guidance on groundwater status and trend assessment, guidance document no 18. Technical Report 2009, ISBN 978-92-79-11374-1 European Communities, Luxembourg, 2009.

Farina, M., Marcaccio, M., Zavatti, A., 2014. Experiences and perspectives to monitoring of groundwater resources: the contribution of Emilia Romagna (Esperienze e prospettive nel monitoraggio delle acque sotterranee: il contributo dell'Emilia-Romagna). Bologna, Pitagora Editrice, 528 pp. (in Italian).

Gaus, I., Kinniburgh, D.G., Talbot, J.C., Webster, R., 2003. Geostatistical analysis of arsenic concentration in groundwater in Bangladesh using disjunctive kriging. Environ. Geol. 44, 939-948. doi:10.1007/s00254-003-0837-7.

Gimeno, P., Marcé, R., Bosch, Ll., Comas, J., Corominas, Ll. 2017. Incorporating model uncertainty into the evaluation of interventions to reduce microcontaminant loads in rivers. Water Res. 124, 415-424. https://doi.org/10.1016/j.watres.2017.07.036.

Heibati, M., Stedmon, C.A., Stenroth, K., Rauch, S., Toljander, J., Säve-Söderbergh, M., Murphy, K.R., 2017. Assessment of drinking water quality at the tap using fluorescence spectroscopy. Water Res. 125, 1-10, https://doi.org/10.1016/j.watres.2017.08.020

Hernandez, A.F., Neuman, S.P., Guadagnini, A., Carrera, J., 2006. Inverse stochastic moment analysis of steady state flow in randomly heterogeneous media, Water Resour. Res. 42, W05425, doi:10.1029/2005WR004449.

Hinsby, K., Condesso de Melo, M.T., 2006. Application and evaluation of a proposed methodology for derivation of groundwater threshold values-a case study summary report. In Deliverable D22 of the EU project ‘‘BRIDGE’’ 2006,. http://nfp-at.eionet.europa.eu/Public/irc/eionet-circle/bridge/library?l=/deliverables/d22_final_reppdf/_EN_1.0_&a=d.

696 Hinsby, K., Condesso de Melo, M.T., Dahl, M., 2008. European case studies supporting the
697 derivation of natural background levels and groundwater threshold values for the protection
698 of dependent ecosystems and human health. Sci. Total Environ. 401(1-3), 1-20.
699 Hoeting, J., Madigan, D., Raferty, A., Volinksy, C., 1999. Bayesian model averaging: A tutorial.
700 Statitical Science 14, 382-417.
701 Hurvich, C.M., Tsai, C.L., 1989. Regression and time series model selection in small sample.
702 Biometrika 76(2), 99-104.
703 Kashyap, R.L., 1982. Optimal choice of AR and MA parts in autoregressive moving average
704 models. IEEE Trans. Pattern Anal. Mach. Intel. PAMI 4(2), 99-104.
705 Kim, K.H., Yun, S.T., Kim, H.K., Kim, J.W., 2015. Determination of natural backgrounds and
706 thresholds of nitrate in South Korean groundwater using model-based statistical approaches.
707 J. Geochem. Explor. 148, 196-205.
708 Kumarathilaka, P., Seneweera, S., Meharg,A., Bundschuh, J. 2018. Arsenic speciation dynamics in
709 paddy rice soil-water environment: sources, physico-chemical, and biological factors - a
710 review. Water Res. Available online 21 April 2018.
711 https://doi.org/10.1016/j.watres.2018.04.034.
712 Liu, C.W., Jang, C.S., Liao, C.M., 2004. Evaluation of arsenic contamination potential using
713 indicator kriging in the Yun-Lin aquifer (Taiwan). Sci. Total Environ. 321, 173-188.
714 doi:10.1016/j.scitotenv.2003.09.002
715 Liu, G., Tao, Y., Zhang, Y., Lut, M., Knibbe, W.J., van der Wielen, P., Liu, W., Medema, G., van
716 der Meer, W., 2017. Hotspots for selected metal elements and microbes accumulation and
717 the corresponding water quality deterioration potential in an unchlorinated drinking water
718 distribution system. Water Res. 124, 435-445.
719 Molinari, A., Guadagnini, L., Marcaccio, M., Guadagnini, A., 2012. Natural background levels and
720 threshold values of chemical species in three large-scale groundwater bodies in Northern
721 Italy. Sci. Total Environ. 425, 9-19. doi:10.1016/j.scitotenv.2012.03.015.

722 Molinari, A., Guadagnini, L., Marcaccio, M., Straface, S., Sanchez-Vila, X., Guadagnini, A., 2013.
723 Arsenic release from deep natural solid matrices under experimentally controlled redox
724 conditions. Sci. Total Environ. 444, 231-240. doi: 10.1016/j.scitotenv.2012.11.093
725 Molinari, A., Guadagnini, L., Marcaccio, M., Guadagnini, A., 2015. Arsenic fractioning in natural
726 solid matrices sampled in a deep groundwater body. Geoderma 247, 88-96.
727 https://doi.org/10.1016/j.geoderma.2015.02.011.
728 Neuman, S.P., 2003. Maximum likelihood Bayesian averaging of alternative conceptual-
729 mathematical models. Stoch. Env. Res. Risk A. 17(5), 291-305. doi:10D1007/s00477-003-
730 0151-7.
731 Panno, S.V., Kelly, W.R., Martinsek, A.T., Hackley, K.C., 2006. Estimating background and
732 threshold nitrate concentrations using probability graphs. Groundwater 44(5), 697-709.
733 Qian, Y., Dong, H., Geng, Y., Zhong, S., Tian, X., Yu, Y., Chen, Y., Moss, D.A., 2018. Water
734 footprint characteristic of less developed water-rich regions: Case of Yunnan, China. Water
735 Res., *in press*, Accepted Manuscript, https://doi.org/10.1016/j.watres.2018.03.075.
736 Redman, A.D., Macalady, D., Ahmann, D., 2002. Natural organic matter affects arsenic speciation
737 and sorption onto hematite. Environ. Sci. Technol. 36, 2889-2896.
738 Regione Emilia-Romagna, 2010. Council Decree (Delibera di Giunta) n. 350 of 8/02/2010,
739 Approval of the activities of the Emilia Romagna Region related to the implementation of
740 Directive 2000/60/CE aiming at the design and adoption of the Management Plans of the
741 hydrographic districs Padano, Appennino settentrionale and Appennino centrale
742 (Approvazione delle attività della Regione Emilia-Romagna riguardanti l'implementazione
743 della Direttiva 2000/60/CE ai fini della redazione ed adozione dei Piani di Gestione dei
744 Distretti idrografici Padano, Appennino settentrionale e Appennino centrale).
745 http://ambiente.regione.emilia-romagna.it/acque/temi/piani%20di%20gestione (In Italian,
746 Accessed 25 March 2018).

747 Reimann, C., Garret, R.G., 2005. Geochemical background: concept and reality. Sci. Total Environ.
748 350, 12-27. https://doi.org/10.1016/j.scitotenv.2005.01.047.
749 Remy, N., Boucher, A., Wu, J., 2009. Applied Geostatistics with SGeMS: A User's Guide 1st
750 Edition, Cambridge University Press, New York. ISBN: 978-0-521-51414-9.
751 Ricci Lucchi, F., 1984. Flysh, molassa, clastic deposits: traditional and innovative approaches to the
752 analysis of north Apeninic basins (Flysch, molassa, cunei clastici: tradizione e nuovi
753 approcci nell'analisi dei bacini orogenici dell'Appennino settentrionale). Cento Anni di
754 Geologia Italiana. Volume Giubilare 1° centenario Soc. Geol. Ital., 279-295.
755 Riva, M., Panzeri, M., Guadagnini, A., Neuman, S.P., 2011. Role of model selection criteria in
756 geostatistical inverse estimation of statistical data- and model- parameters. Water Resour.
757 Res. 47, W07502. doi:10.1029/2011WR010480.
758 Ungaro, F., Ragazzi, F., Cappellin, R., Giandon, P., 2008. Arsenic concentration in the soils of the
759 Brenta Plain (Northern Italy): Mapping the probability of exceeding contamination
760 thresholds. J. Geochem. Explor. 96, 117-131. doi:10.1016/j.gexplo.2007.03.006.
761 Walter, T., 2008. Determining natural background values with probability plots. EU Groundwater
762 Policy Developments Conference, UNESCO, Paris, France, 13-15 Nov 2008.
763 Wendland, F., Hannappel, S., Kunkel, R., Schenk, R., Voigt, H.J., Wolter, R., 2005. A procedure to
764 define natural groundwater conditions of groundwater bodies in Germany. Water Sci.
765 Technol. 51(3-4), 249-257.
766 Ye, M., Neuman, S.P., Meyer, P.D., 2004. Maximum likelihood Bayesian averaging of spatial
767 variability models in unsaturated fractured tuff. Water Resour. Res. 40(5), W051131-
768 W0511317.
769 Ye, M., Meyer, P.D., Neuman, S.P., 2008. On model selection criteria in multimodel analysis.
770 Water Resour. Res. 44, W03428.

771 Zlatanović, Lj., van der Hoek, J.P., Vreeburg, J.H.G., 2017. An experimental study on the influence
772 of water stagnation and temperature change on water quality in a full-scale domestic
773 drinking water system. Water Res. 123, 761-772.

Appendix A. Supplementary data

Spatial distributions of kriged values for local NBLs of ammonium in water body 2700 with the four models listed in Table 5 (the value $NH_4Log_{10}\,NBL_{90}^{PS}$ = 1.08 (red font on the color scale; corresponding to NBL_{90}^{PS} = 12.0 mg/L) is denoted by the thick blue isoline.

Appendix B. Supplementary data

Spatial distributions of Kriging variance obtained for ammonium for water bodies 0610, 0630 and 2700 as a result of the multimodel analysis.

Appendix C. Supplementary data

Spatial distributions of kriged values for local NBLs of arsenic in water body 2700 with the four models listed in Table 5. The value $AsLog_{10}\,NBL_{90}^{PS}$ = 0.77 µg/L (red font on the color scale; corresponding to NBL_{90}^{PS} = 6 µg/L) is denoted by the thick blue isoline.

Appendix D. Supplementary data

Spatial distributions of Kriging variance obtained for arsenic for water bodies 0630 and 2700 as a result of the multimodel analysis.

Table 1

Extension and main characteristic length scales of the three groundwater bodies investigated.

Groundwater body	Average thickness (m)	Average depth (m)	Areal extent (km^2)
0610	130	75	2928
0630	110	65	1995
2700	180	200	6934

Table 2

Number of monitoring stations and total number of samples available

		number of samples	
Groundwater body	monitoring stations	As	NH_4
0610	90	1968	2230
0630	75	1692	1917
2700	55	1201	1383

Table 3

Number of observation wells available for spatial analysis of NBLs after data selection.

groundwater body	monitoring stations Ammonium	monitoring stations Arsenic
0610	51	50
0630	62	60
2700	47	45

Table 4

Values for NH_4 NBL_{90} estimated via the original PS procedure (NBL_{90}^{PS}). EU Drinking Water Standard for NH_4 concentration is 0.5 mg/L.

groundwater body	NBL_{90}^{PS} (mg/L)	Log_{10} NBL_{90}^{PS}
0610	4.6	0.66
0630	5.2	0.71
2700	12.0	1.08

Table 5

Estimated parameters of variogram models of NH_4Log_{10} NBL_{90} values and associated posterior probability (*p*) based on *KIC* (4). Here, Sph = Spherical model, Exp = Exponential model; n = nugget, a = range (practical range for Exponential model), c = sill; (*) NBL_{90} values are given in (mg/L).

	Groundwater body 0610				Groundwater body 0630				Groundwater body 2700			
Model	n (*)	a (km)	c (*)	*p* (%)	n (*)	a (km)	c (*)	*p* (%)	n (*)	a (km)	c (*)	*p* (%)
Sph	-	18.7	0.106	52	-	83.5	0.496	40	-	81.7	0.485	6
Exp	-	22.9	0.107	48	-	138.2	0.581	49	-	100.4	0.509	34
Sph	-	-	-	-	0.045	87.9	0.452	4.2	0.1	101.0	0.401	5
Exp	-	-	-	-	0.031	159.6	0.574	6.6	0.1	208.9	0.538	55

Table 6

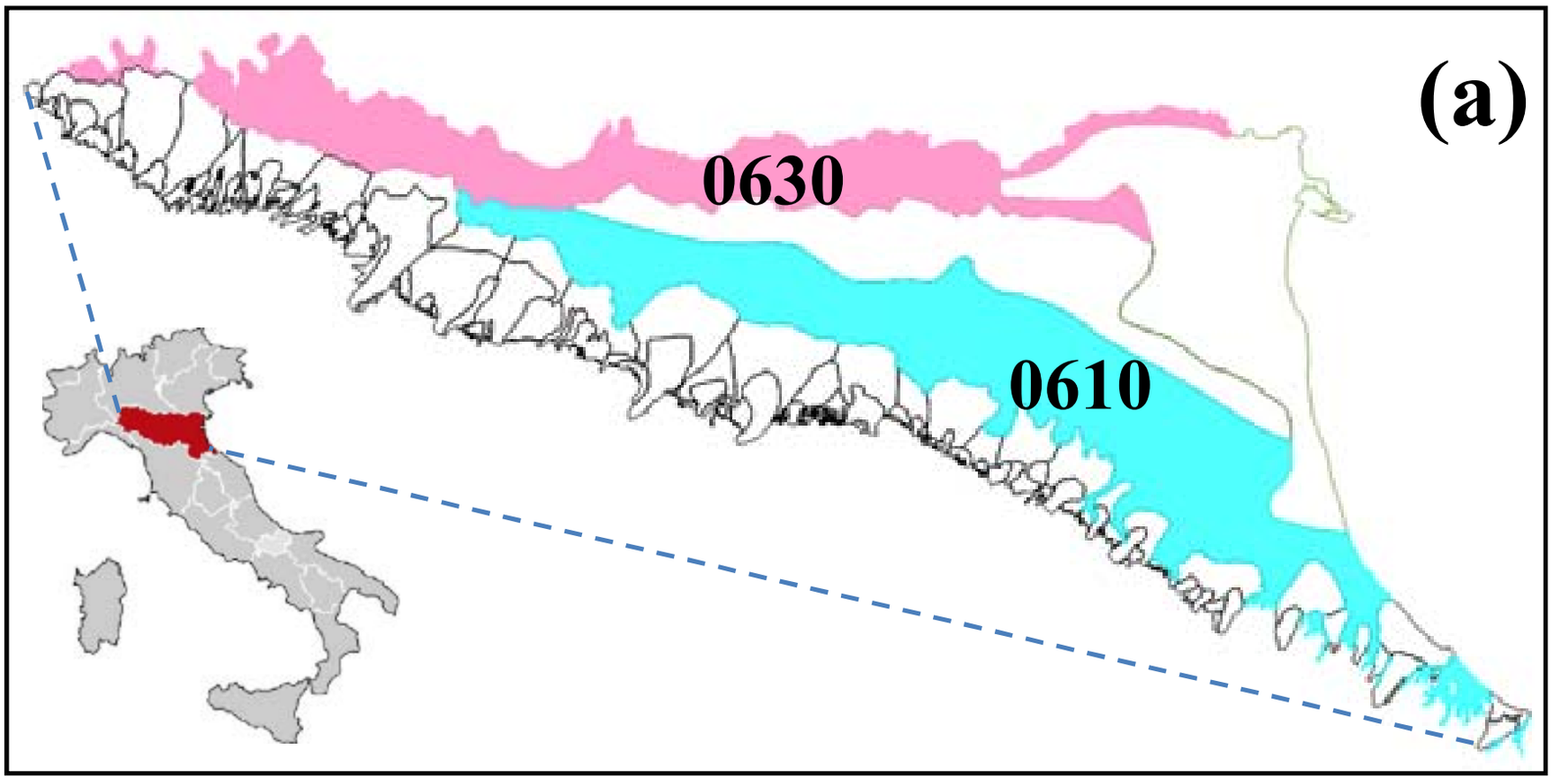
Values for As NBL_{90} estimated via the original PS procedure (NBL_{90}^{PS}). EU Drinking Water Standard for As concentration is 10 µg/L.

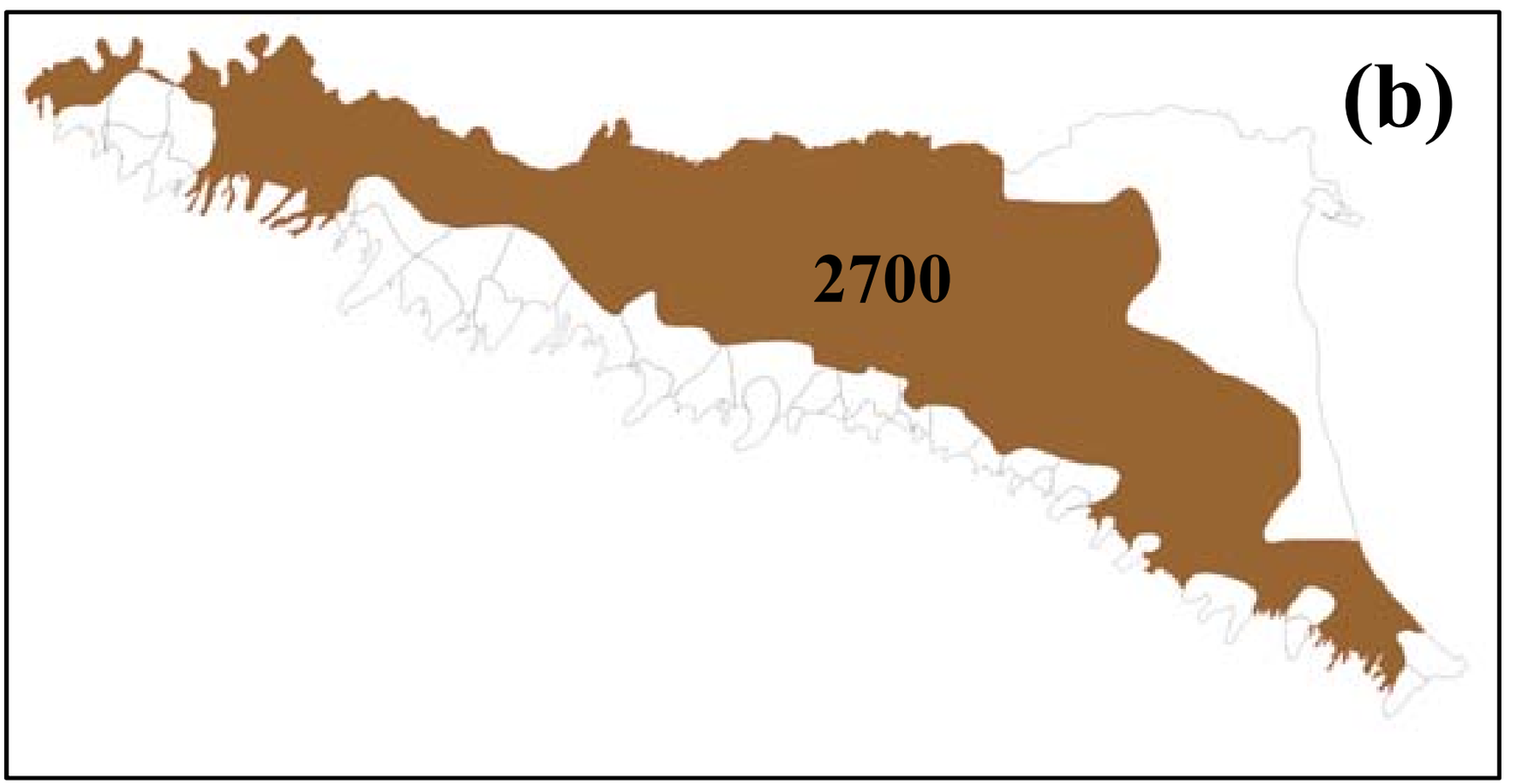
groundwater body	NBL_{90}^{PS} (µg/L)	Log_{10} NBL_{90}^{PS}
0610	33	1.52
0630	4	0.60
2700	6	0.77

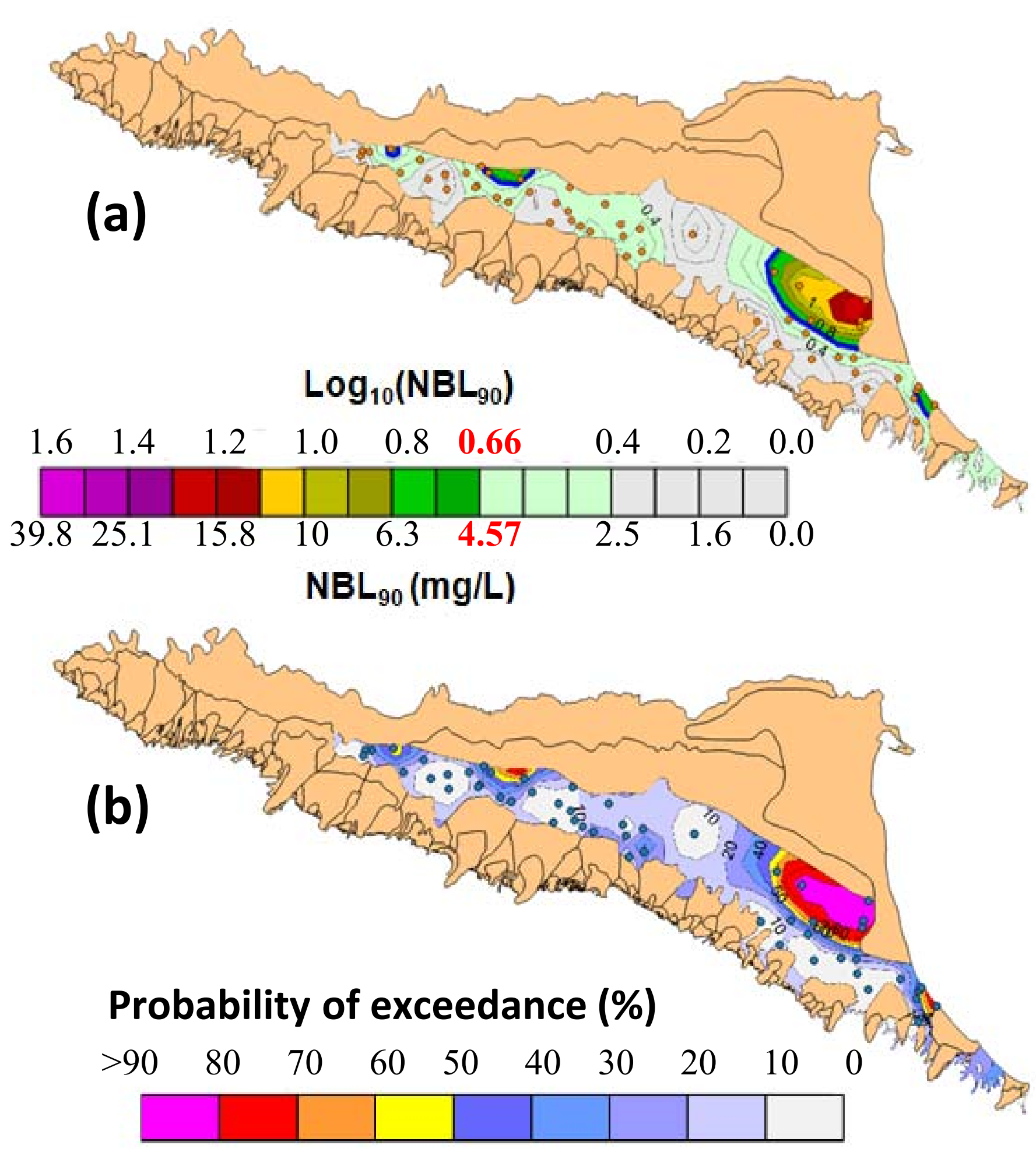
Table 7

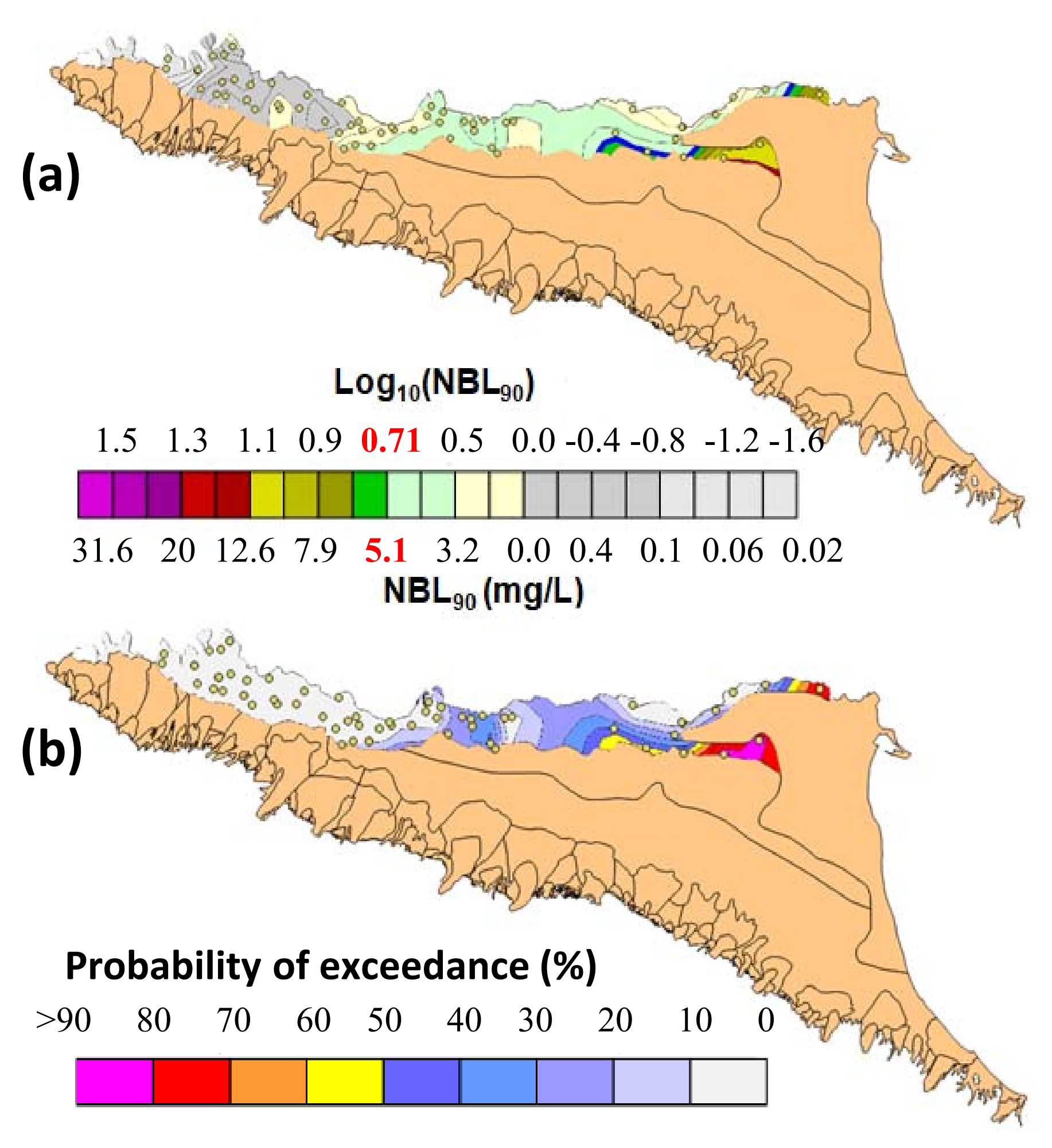
Estimated parameters of variogram models of $AsLog_{10}$ NBL_{90} values and associated posterior probability (*p*) based on *KIC* (4). Here, PN = Pure Nugget, Sph = Spherical model, Exp = Exponential model; n = nugget, a = range (practical range for Exponential model), c = sill; (*) NBL_{90} values are given in (μg/L).

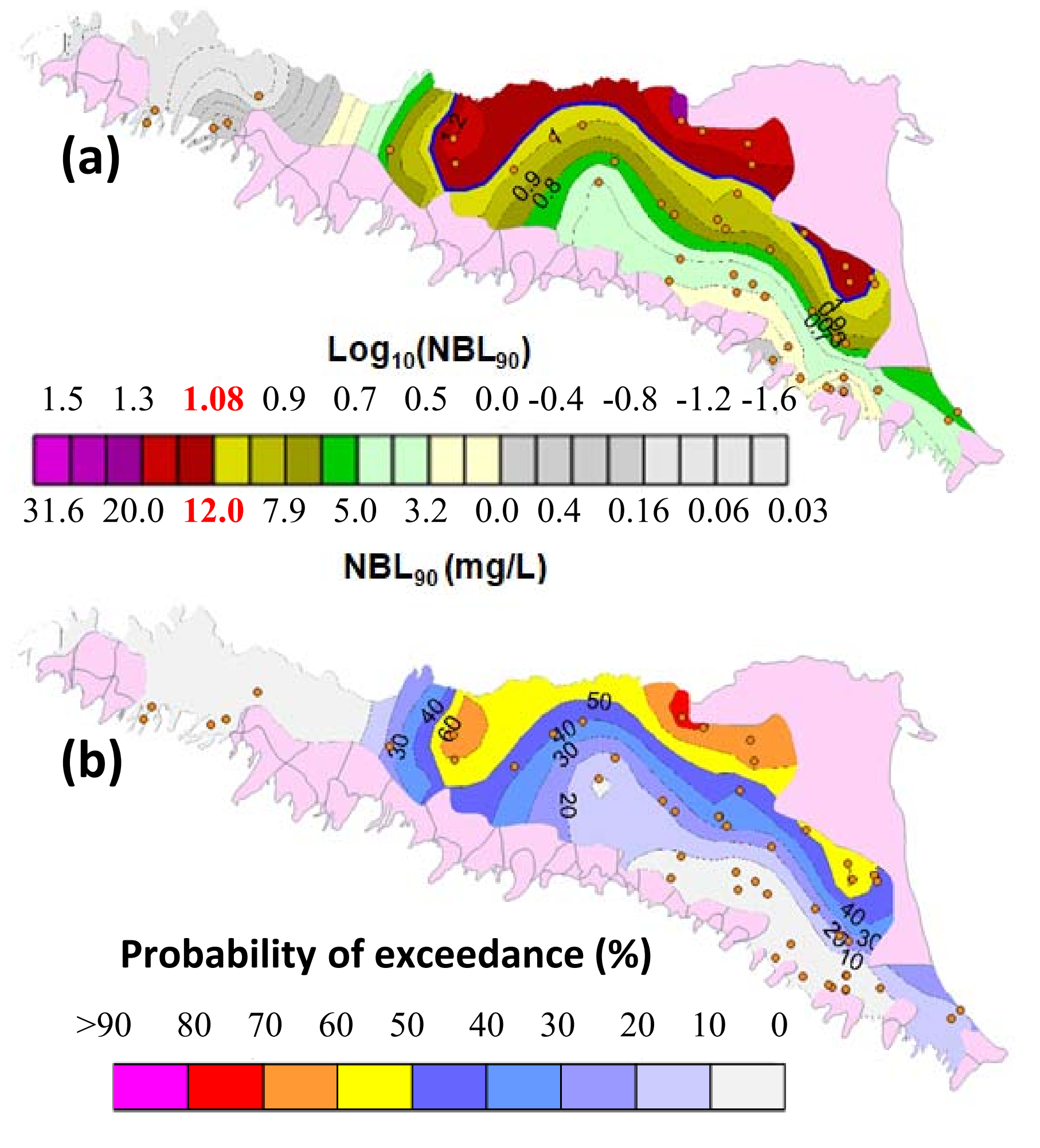
	Groundwater body 0610				Groundwater body 0630				Groundwater body 2700			
Model	n (*)	a (km)	c (*)	*p* (%)	n (*)	a (km)	c (*)	*p* (%)	n (*)	a (km)	c (*)	*p* (%)
PN	0.33	-	-	100	0.252	-	-	0.002	0.238	-	-	6.7
Sph	-	-	-	-	-	12.78	0.256	0.016	-	11.57	0.243	31.6
Exp	-	-	-	-	-	14.57	0.257	0.030	-	11.57	0.243	31.6
Sph	-	-	-	-	0.132	19.80	0.125	99.77	0.07	14.42	0.172	15.0
Exp	-	-	-	-	10^{-5}	14.58	0.257	0.177	0.07	14.42	0.172	15.0

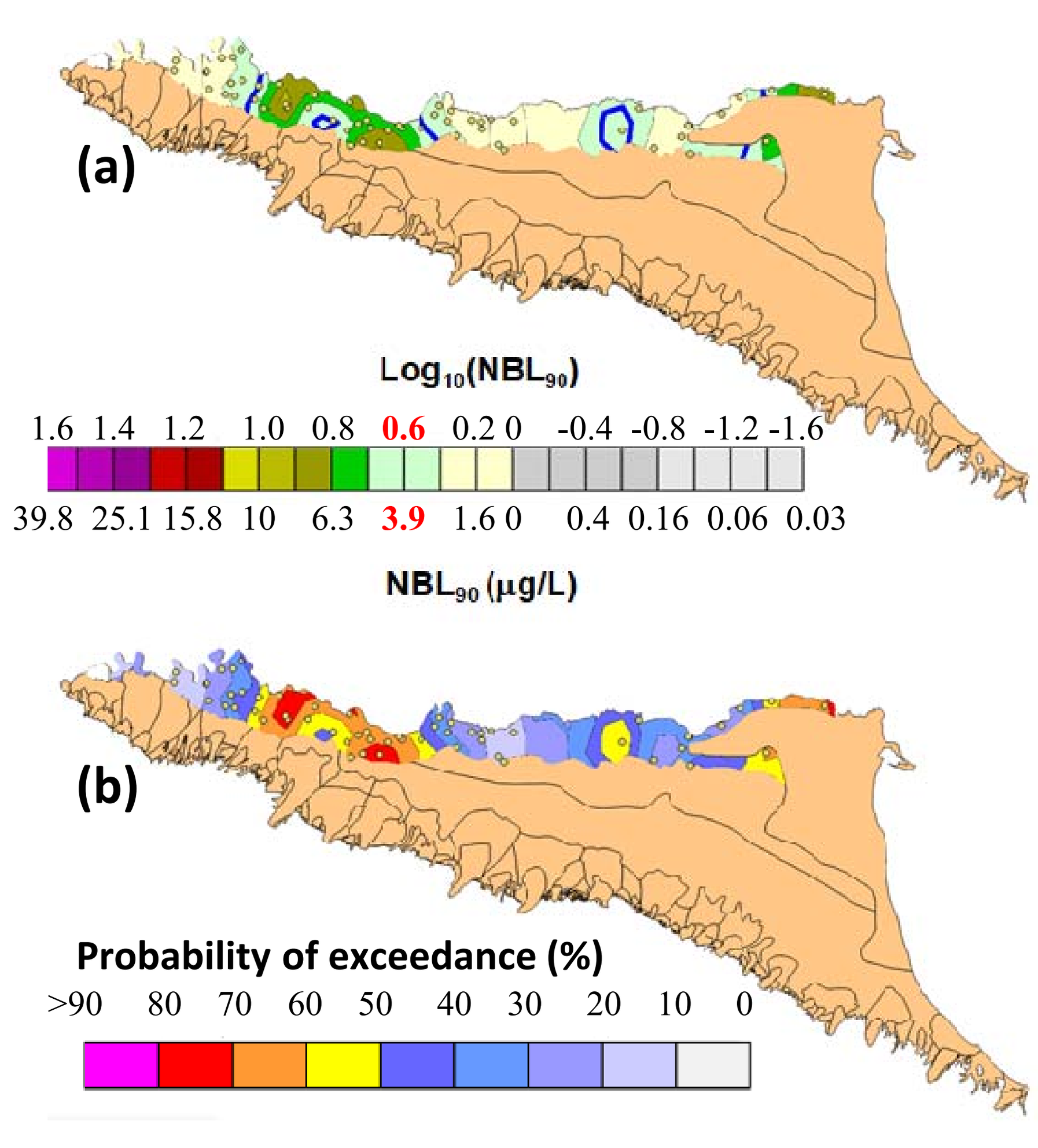

(a)
0630
0610

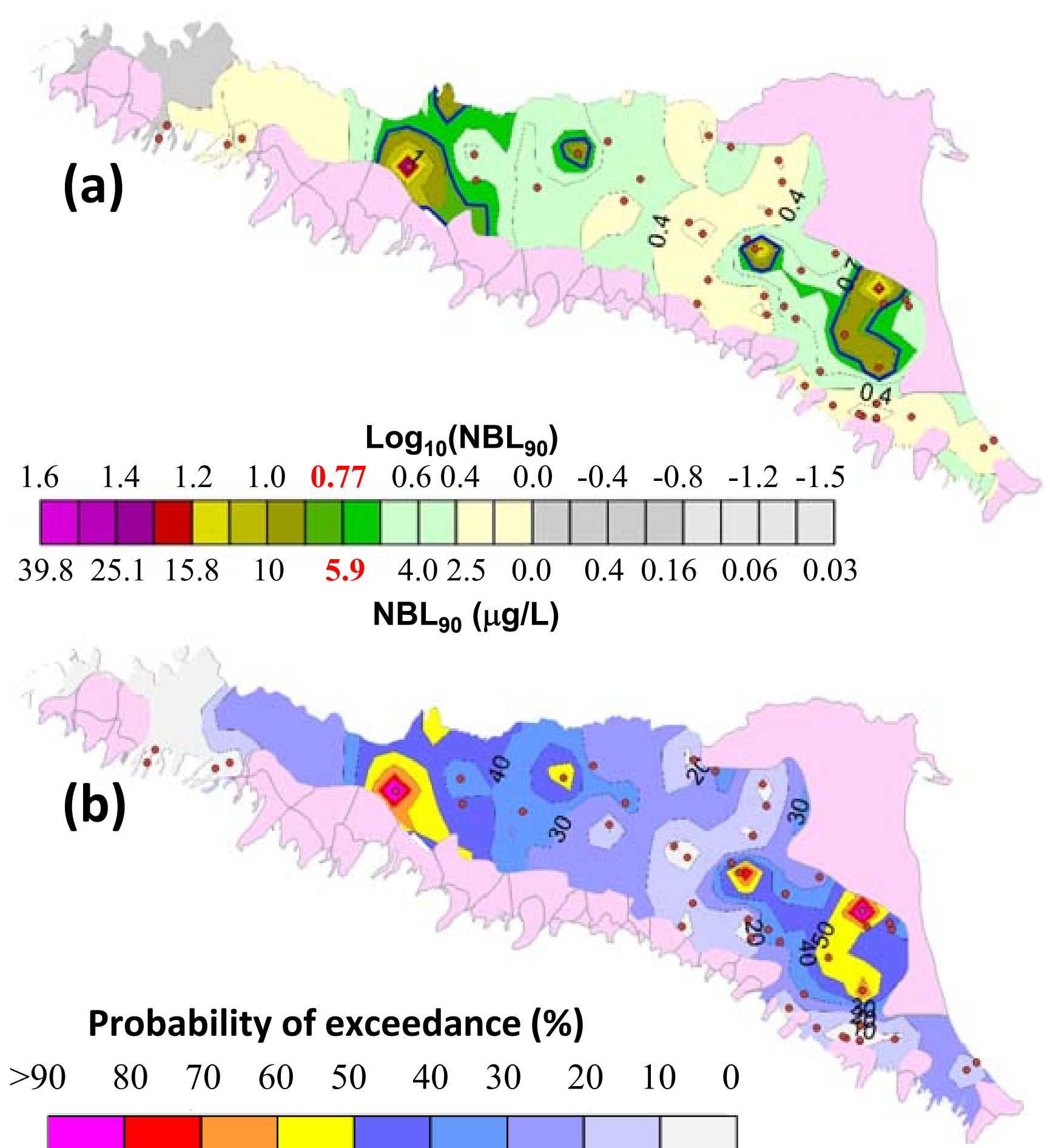

(b)
2700


(a)
Log10(NBL90)
1.6 1.4 1.2 1.0 0.8 0.66 0.4 0.2 0.0
39.8 25.1 15.8 10 6.3 4.57 2.5 1.6 0.0
NBL90 (mg/L)
(b)
Probability of exceedance (%)
>90 80 70 60 50 40 30 20 10 0


(a)
Log10(NBL90)
1.5 1.3 1.1 0.9 0.71 0.5 0.0 -0.4 -0.8 -1.2 -1.6
31.6 20 12.6 7.9 5.1 3.2 0.0 0.4 0.1 0.06 0.02
NBL90 (mg/L)
(b)
Probability of exceedance (%)
>90 80 70 60 50 40 30 20 10 0


(a)
Log10(NBL90)
1.5 1.3 1.08 0.9 0.7 0.5 0.0 -0.4 -0.8 -1.2 -1.6
31.6 20.0 12.0 7.9 5.0 3.2 0.0 0.4 0.16 0.06 0.03
NBL90 (mg/L)
(b)
Probability of exceedance (%)
>90 80 70 60 50 40 30 20 10 0


(a)
Log10(NBL90)
1.6 1.4 1.2 1.0 0.8 0.6 0.2 0 -0.4 -0.8 -1.2 -1.6
39.8 25.1 15.8 10 6.3 3.9 1.6 0 0.4 0.16 0.06 0.03
NBL90 (µg/L)
(b)
Probability of exceedance (%)
>90 80 70 60 50 40 30 20 10 0

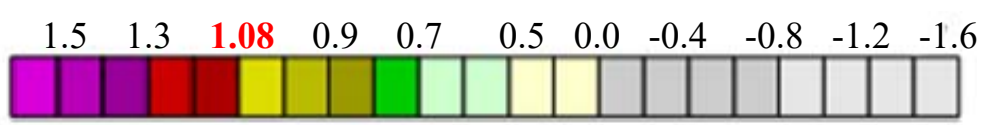

(a)
0.4
0.4
0.4
Log10(NBL90)
1.6 1.4 1.2 1.0 0.77 0.6 0.4 0.0 -0.4 -0.8 -1.2 -1.5
39.8 25.1 15.8 10 5.9 4.0 2.5 0.0 0.4 0.16 0.06 0.03
NBL90 (µg/L)
(b)
40
30
20
30
20
40
Probability of exceedance (%)
>90 80 70 60 50 40 30 20 10 0

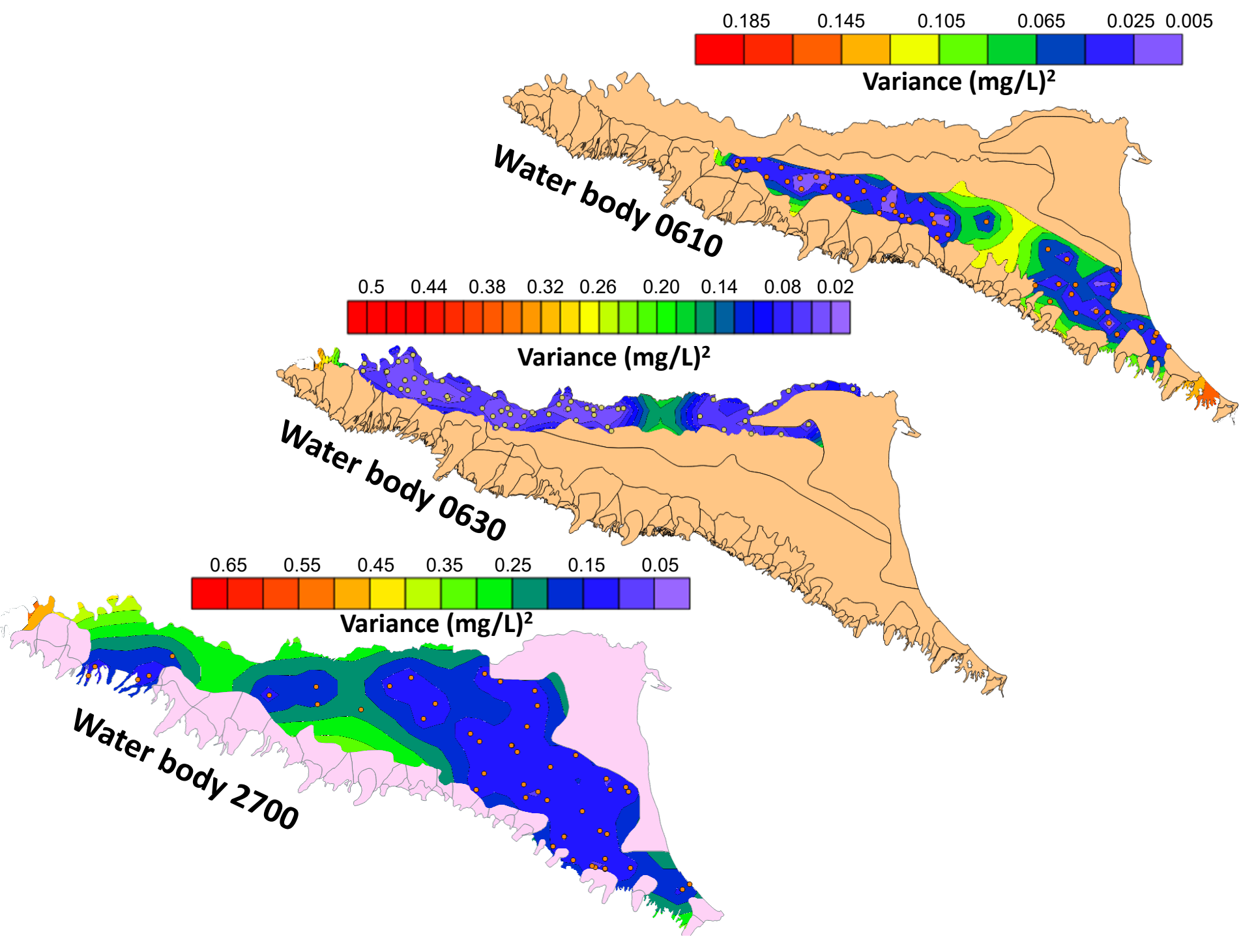
a) Spherical model without nugget

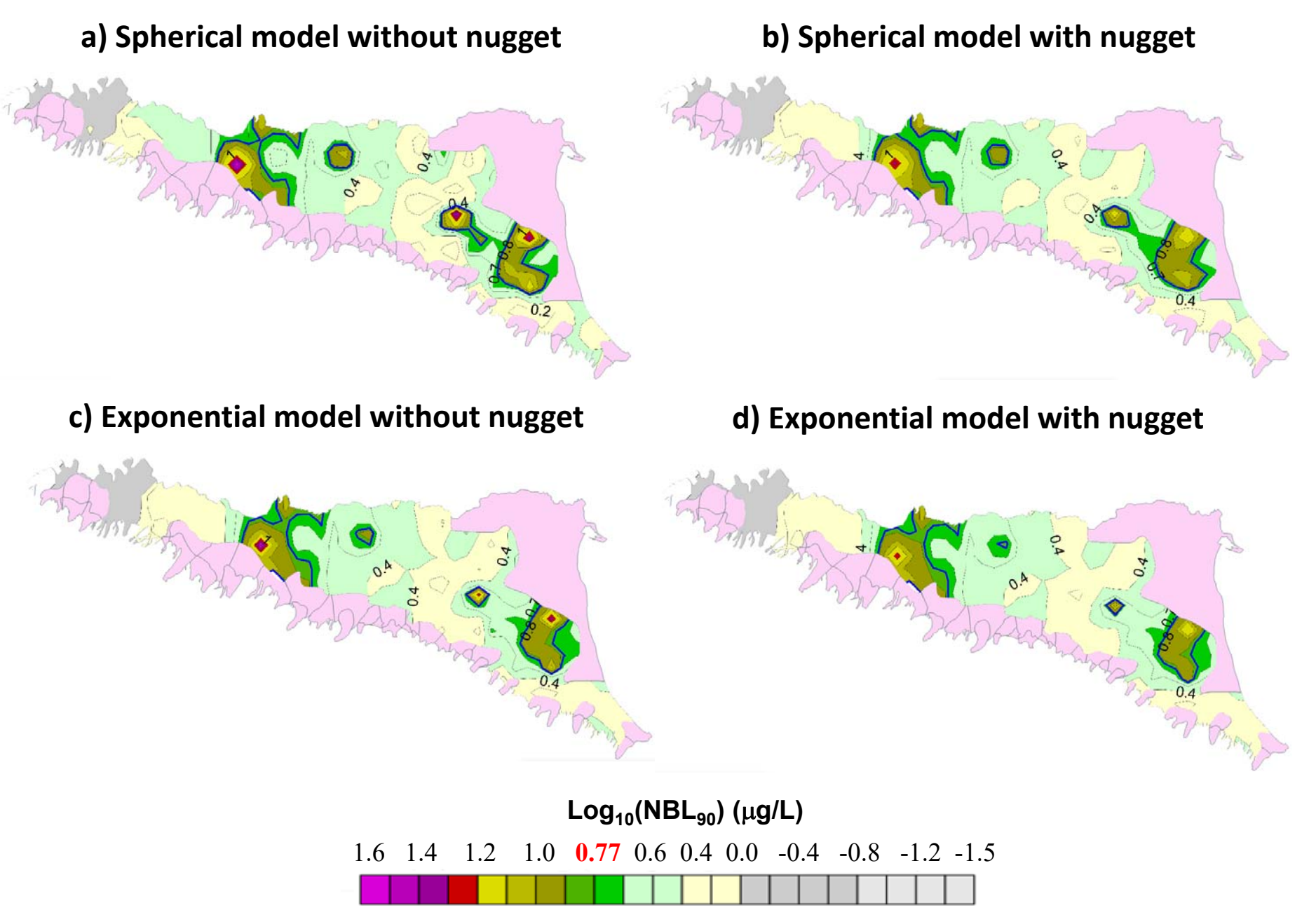
b) Spherical model with nugget

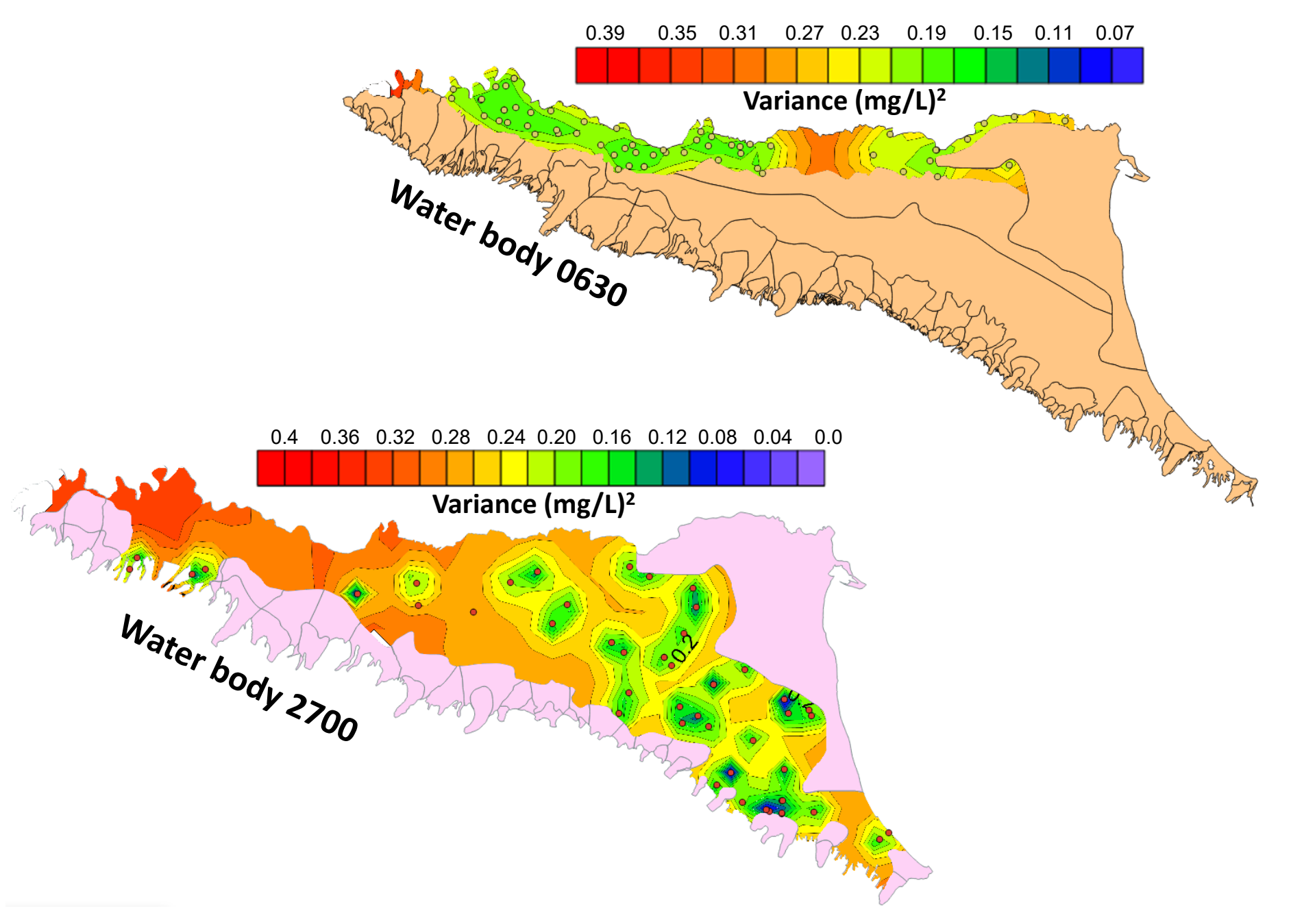
c) Exponential model without nugget

d) Exponential model with nugget




0.185 0.145 0.105 0.065 0.025 0.005
Variance (mg/L)²
Water body 0610
0.5 0.44 0.38 0.32 0.26 0.20 0.14 0.08 0.02
Variance (mg/L)²
Water body 0630
0.65 0.55 0.45 0.35 0.25 0.15 0.05
Variance (mg/L)²
Water body 2700


a) Spherical model without nugget
b) Spherical model with nugget
c) Exponential model without nugget
d) Exponential model with nugget
Log10(NBL90) (µg/L)
1.6 1.4 1.2 1.0 0.77 0.6 0.4 0.0 -0.4 -0.8 -1.2 -1.5


0.39 0.35 0.31 0.27 0.23 0.19 0.15 0.11 0.07
Variance (mg/L)²
Water body 0630
0.4 0.36 0.32 0.28 0.24 0.20 0.16 0.12 0.08 0.04 0.0
Variance (mg/L)²
Water body 2700
0.2
0.4

Electronic Supplementary Material (for online publication only)

Click here to download Electronic Supplementary Material (for online publication only): Supplemental material.docx