



# Extension variance: an early geostatistical concept applied to assess nitrate pollution in groundwater

Maria-Theresia Schafmeister<sup>1</sup> · Michael Steffen<sup>2</sup> · Karl-Otto Zeissler<sup>2</sup> · Markus Zingelmann<sup>2</sup>

Received: 29 August 2022 / Accepted: 19 June 2023 / Published online: 6 July 2023  
© The Author(s) 2023

## Abstract

Monitoring and assessing groundwater quality according to European directives and national regulations is usually based on interpolation techniques, e.g. Kriging. However, contour maps of hydrochemical parameters often suggest a spurious local accuracy and can therefore lead to inappropriate action measures. Here, the early concept of extension variance combined with the Voronoi tessellation regionalization is proposed. The mosaic-like representation of pollutant concentrations in Voronoi polygons avoids misinterpretations caused by interpolation. The additional calculation of the extension variance, which is based on fundamental geostatistical assumptions, allows for estimating the probability that a given threshold is exceeded. This concept is further extended to hydraulically delimitable groundwater bodies, thus ensuring that hydraulic boundaries are considered. A method is here described for the assessment of groundwater quality with respect to nitrate concentration in the principal aquifer of the state Mecklenburg-Western Pomerania in Germany.

**Keywords** Geostatistics · Voronoi tessellation · Groundwater monitoring · Nitrate · EU Water Framework Directive

## Introduction

Nowadays the global concern is to maintain safe reserves of clean groundwater. The European Union (EU) Water Framework Directive 2000/60/EC aims to “achieve ‘good status’ for Europe’s rivers, lakes and groundwater by 2015” (EC 2000). One groundwater pollutant of major concern is nitrate as a consequence of overuse of fertilizer and intensive animal husbandry in agricultural regions. In fact, the status report on nitrate reports that 26.7% of the monitoring wells in agricultural regions in Germany exceed the threshold of 50 mg/L (BMUV 2020). Even though this percentage has been declining recently, effective measures still need to be taken in order to better monitor and improve the actual situation.

Accordingly, the State Geological Surveys of Germany are required to monitor groundwater quality in a suitable

piezometer network to facilitate good groundwater chemical status. In the course of this task, it was required that appropriate regionalization techniques be identified in order to unify the assessment procedure within Germany while disregarding state boundaries.

Among others, deterministic interpolation methods, e.g. inverse distance weighted- (IDW) or spline interpolations were suggested. However, these methods do not quantify the reliability of the spatial estimation result. Rather they suggest an accurate spatial regionalization where local concentrations can precisely be derived from contour plots, no matter how well the data are supported, i.e. whether there is an adequate number of sample locations (piezometers). It is the major advantage of geostatistical spatial estimators such as Kriging, that they incorporate the specific spatial variability structure of regionalized variables. Applied to nitrate pollution in the groundwater of Lower Saxony in Germany, Wriedt et al. (2019) discuss the capabilities of Simple Kriging (SK) and Ordinary Kriging (OK) in order to identify regions of high contamination load. Unfortunately, they did not make use of the Kriging standard deviation  $\sigma_K$ , which is also calculated contemporaneously in association with the actual Kriging estimate. The Kriging standard deviation represents the uncertainty that results from the data support, the sampling pattern and the specific spatial variability of the regionalized variable. Usually, it is displayed in addition to

---

Published in the special issue “Geostatistics and hydrogeology”.

✉ Maria-Theresia Schafmeister  
schaf@uni-greifswald.de

<sup>1</sup> Department of Applied Geology/Hydrogeology,  
University of Greifswald, Friedrich-Ludwig-Jahn-Str. 17a,  
17487 Greifswald, Germany

<sup>2</sup> Beak Consultants GmbH, Am St.-Niclas-Schacht 13,  
09599 Freiberg, Germany

or blended with the interpolated concentration map. Webster and Oliver (2014) criticize an imprudent use of Kriging techniques in geographic information systems (GISs) and provide helpful guidance in order to consider, among others, sample support, variogram and Kriging options, respectively, especially for environmental applications.

Advanced geostatistical techniques like SIMIK+ (Bárdossy et al. 2003) not only provide a measure of uncertainty as the Kriging error (i.e. Kriging standard deviation), they can incorporate supplementary information as well, e.g. type of geology and land use. Dokou et al. (2015) present a summary of advanced Kriging techniques, which especially contribute to the spatial interpolation of nonparametrically distributed groundwater quality data, including nitrate and trace metals; specifically, the Indicator approach is well suited for questions of risk assessment. Similarly, the spatial interpolation of groundwater quality parameters based on copula estimate local-distribution functions of the target parameters (Bárdossy and Li 2008) allows for the quantification of uncertainty. This approach is further enhanced by including secondary information from correlated data (Gnann et al. 2018). Knoll et al. (2020) suggest a machine-learning approach (Random Forest) which includes environmental factors influencing nitrate in groundwater, e.g. redox status of the aquifers.

Still, it is questionable whether an interpolated contour (or pixel) map adequately reflects the spatial distribution of groundwater pollutants. The adequacy strongly depends on the objectives that are to be achieved. Interpolated maps usually suggest a more or less smooth spatial structure, which is often not realistic (Bronowicka-Mielniczuk et al. 2019), but can provide a regional overview of the contaminant distribution. Small-scale variability is better reproduced by stochastic simulation techniques, e.g. sequential Gaussian or sequential indicator simulation (Gómez-Hernández and Srivastava 2021) or multiple point simulations (Strebelle 2002; Huysmans and Dassargues 2010). These regionalization techniques are well established as parametrization tools in groundwater flow and transport models (Renard 2007). The goal of a comparative groundwater quality assessment for delineable groundwater bodies is to determine the likelihood of thresholds being exceeded and to what extent, rather than to accurately determine the pollutant concentration at a particular point location. The latter may better be achieved by the aforementioned, however elaborate, regionalization techniques.

Nitrate in groundwater, which mostly results from human activities (BMUV 2020) and is enhanced by soil features, e.g. preferential flow paths or redox status, often exhibits an erratic spatial structure. It was therefore an early proposal to present the groundwater contamination more genuinely by plotting the measurements at their respective locations without interpolation in between (Schafmeister and Pekdeger 1994). As an alternative, the pollution may

be depicted spatially in Voronoi polygons, a natural neighbor technique; however, this implies an inherent error at any but the measurement location within the polygon.

Given an irregular sampling pattern, measurements can be attributed to areas of subregions around the sampling points, in which any point is closer to the sampling location than to its adjacent sampling locations. This method is named after the Russian-Ukrainian mathematician Georgy Voronoi (1868–1908) and is now known as Voronoi tessellation. Hydrologists know the same method as Thiessen polygons (after the American meteorologist Alfred Thiessen, 1872–1956) and apply it, e.g. in order to up-scale point-supported precipitation measurements from rain gauges to two-dimensional (2D) areal segments.

However, this natural-neighbor regionalization method implies a certain error at any point within the Voronoi polygon except the measurement location. This error can be expressed as the extension variance  $\sigma_c^2$  (Matheron 1971; David 1977). The extension variance quantifies the uncertainty that applies to an area (or volume) when a single-point measurement is attributed to this area (2D) or volume (3D). This approach, which combines the most plausible and simple regionalization technique with a simultaneously computed measure of uncertainty, is by no means a novel idea; however, it can easily be implemented, especially in questions of environmental risk assessment (Fuchs and Burger 2000).

The objective of this paper is to demonstrate the usefulness of the classical extension variance concept based on Voronoi-polygons as a natural-neighbor regionalization technique, in order to assess the chemical status of groundwater incorporating prediction uncertainties. This classical approach is effective and easily applicable in the course of administrative groundwater quality assessments. It was recently proposed as a standard method in Germany in the context of groundwater quality monitoring according to the goals of the EU Water Framework Directive (Zeissler and others, unpublished technical report for project 2020 0051, “Identifizierung eines Regionalisierungsverfahrens zur Bewertung des chemischen Zustands von Grundwasserkörpern nach Grundwasserverordnung und Anwendung des Verfahrens in M-V [Identification of a regionalisation procedure for assessing the chemical status of groundwater bodies according to the Groundwater Ordinance and application of the procedure in Mecklenburg-Western Pomerania]” 2021, unpublished report).

## Materials and methods

### Data

In order to demonstrate the concept of extension variance, a quality assessment of groundwater measurements of nitrate in the principal upper aquifer system of the state



**Fig. 1** Location of the study area. The boundary of Mecklenburg-Western Pomerania/Germany and 59 groundwater bodies are shown; colors indicate the water-table elevation

Mecklenburg-Western Pomerania (Germany) is used. Nitrate is introduced to groundwater via fertilizers and stock manure. As the oxidized species of nitrogen, nitrate is subject to redox processes and can be decomposed by lithotrophic or organotrophic means (oxidation of pyrite or organic matter, respectively). However, nitrate often reaches groundwater via preferential flow paths or if the buffering electron acceptors are missing. Under the drinking water ordinance, a concentration of 50 mg/L  $\text{NO}_3^-$  is considered admissible in Europe (BMUV 2010).

The quality assessment of groundwater is applied to smaller subregions, which can be considered delimitable with respect to the hydrogeological flow dynamics, e.g. water divides or surface waters. In Mecklenburg-Western Pomerania, which covers an area of 23,174 km<sup>2</sup>, 59 so-called groundwater bodies are identified, with an average size of 400 km<sup>2</sup> (Fig. 1). The principal upper Pleistocene aquifer system is built up by glacial sands and gravels with moderate to high hydraulic conductivities ( $1 \cdot 10^{-4} - 5 \cdot 10^{-4}$  m/s) alternating with low-permeability tills. Locally, it is connected to deeper, pre-Quaternary sedimentary layers. The average total thickness of the principal aquifer system is 50 m. It is mostly confined by Weichselian ground moraine slabs with several tens of meter thickness. Local shallow aquifers are not considered here. Beyond the northwest–southeast striking terminal moraines, sandy and gravelly outwash plains strike out. The average annual groundwater recharge is 100–150 mm. Mecklenburg-Western

Pomerania is a predominantly rural state, with more than 60% of its land used for agriculture (Stat-MV 2021).

Water samples were collected at 568 monitoring wells (Fig. 1) in Mecklenburg-Western Pomerania, which corresponds to a density of 1.2 wells per 50 km<sup>2</sup>. In accordance with BMUV (2010) the data set is compiled as the maximum concentration of relevant parameters (here nitrate) out of the annual average over a time span of 5 years (2015–2019) at each monitoring well.

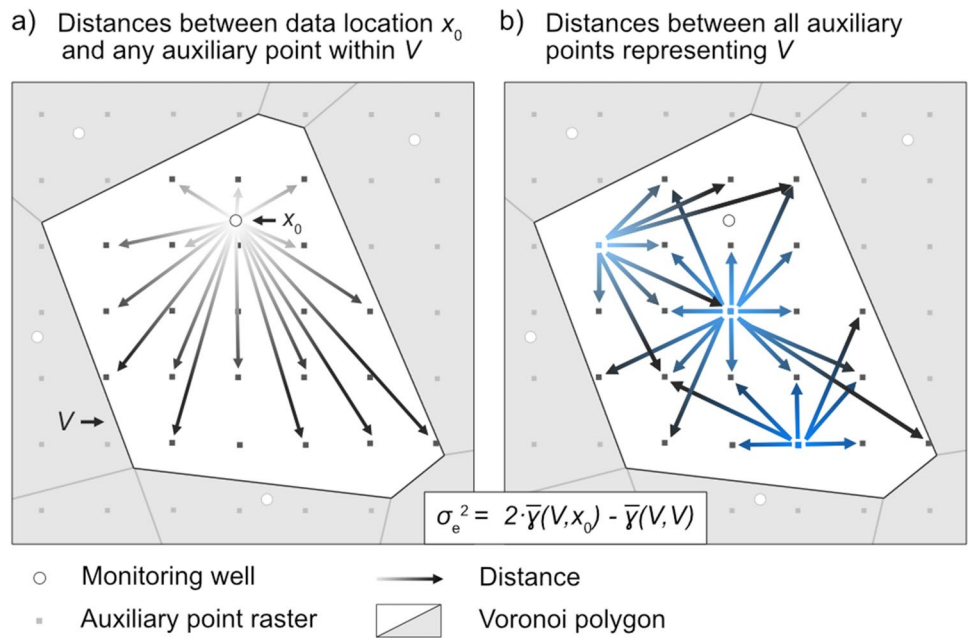
### Concept of extension variance

The early development of geostatistical regionalization methods for ore reserve estimations, e.g. Kriging, make use of this concept of extension variance, which expresses the error that occurs when a single measurement value is taken as an estimator for a bigger area or volume. The concept assumes that the regionalized variable  $Z(x)$  satisfies the intrinsic hypothesis, which assumes that the increments, i.e. the difference in variable values at distances  $h$ , are weakly stationary (Armstrong 1998). Here, weak stationarity of increments means that the expectation  $E$  (Eq. 1) and variance  $\text{Var}$  (Eq. 2) of  $[Z(x+h) - Z(x)]$  exist and are independent of location  $x$ .

$$E[Z(x+h) - Z(x)] = 0 \quad (1)$$

$$\text{Var}[Z(x+h) - Z(x)] = 2\gamma(h) \quad (2)$$

**Fig. 2** Close-up of a Voronoi polygon and auxiliary point raster, highlighting **a** distances between data location  $x_0$  and all auxiliary points representing  $V$ , and **b** distances between all auxiliary points. (Note: for clarity, not all possible distances are shown)



The extension variance is based on the knowledge of the spatial variability of regionalized variables, expressed as the (semi-)variogram  $\gamma(h)$ .

$$\sigma_e^2 = 2 \cdot \bar{\gamma}(V, x_0) - \bar{\gamma}(V, V) \tag{3}$$

where  $\sigma_e^2$  is the extension variance,  $\bar{\gamma}(V, x_0)$  is the average variogram value for all distances between all locations within the polygon  $V$  and the measurement location  $x_0$ , and  $\bar{\gamma}(V, V)$  is the average variogram value for all distances possible within  $V$ , which represents the estimation volume (3D) or area (2D). In the case of natural neighbor estimation on a Voronoi-polygon,  $V$  refers to the area of this polygon. Equation 3 clearly shows that the extension variance depends on the variogram and the polygon only; it is independent of the measurement value. The extension variance is influenced by (1) the regularity of the variable, (2) the geometry of the polygon  $V$  and (3) the location of the measurement point  $x_0$  relative to  $V$  (Fig. 2). Obviously, the error (extension variance) increases with increasing size of the polygon, i.e. the larger the distance between measurement points. Equally, the more marginally the measurement point  $x_0$  is located within a polygon, the less representative it may be, i.e. the error increases; however, an irregular, i.e. less continuous variable, also increases the error.

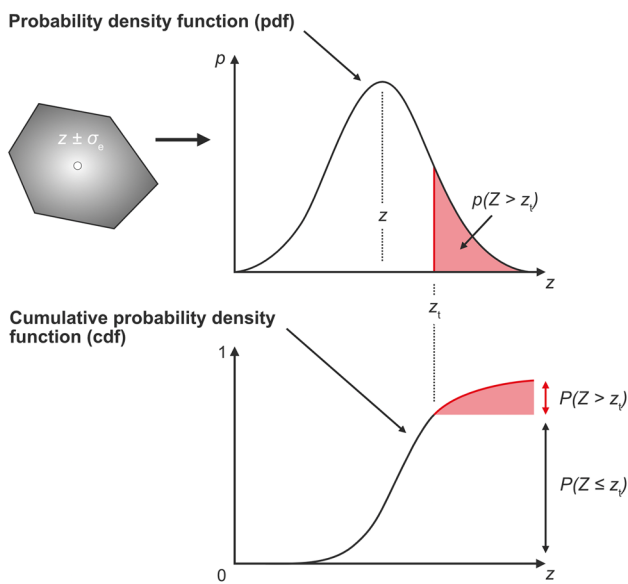
Thus, the concept of extension variance is an elegant tool to incorporate the sampling pattern into the quality assessment and can even be applied to optimize a monitoring network.

Since it is not possible to derive the average variogram  $\bar{\gamma}$  for all possible distances within a polygon, an auxiliary point raster needs to be defined that adequately represents the area of a Voronoi polygon, while still maintaining reasonable computation time. In this study, the entire investigation area was discretized into a regular  $500 \times 500 \text{ m}^2$  auxiliary point raster, i.e. a Voronoi-polygon with an average size of  $41 \text{ km}^2$  is represented by 164 auxiliary points allowing for more than 13,000 possible distances  $h$  within a polygon, from which the average variogram can be determined.

### Probability of exceedance

The extension variance  $\sigma_e^2$  can now be used to determine the probability that a threshold value is exceeded. Assuming a Gaussian probability distribution for  $Z(x), x \in V$ , it can be concluded that with a probability of 95%  $Z(V)$  varies between  $Z(x) \pm 2\sigma_e$ . The probability of exceedance can easily be derived from the cumulative probability density function (Fig. 3).

The final result will then reproduce the data value within a polygon  $V$  but, in addition, the uncertainty of this natural neighbor estimate, which results from the sampling pattern and the specific spatial variability structure, is visualized. All spatial analyses were run by ArcGIS Desktop 10.5.1 while the geostatistical computations are performed with R (R Core Team 2020; Ribeiro et al. 2020).



**Fig. 3** Voronoi polygon and data value  $z(x_0)$  (left). The extension standard deviation  $\sigma_z$  quantifies the possible data range within the polygon.  $P(Z > z_1)$  is the probability that the threshold  $z_1$  may be exceeded

## Results

### Creation and adaptation of polygons to groundwater bodies

The Voronoi tessellation is performed for the 568 monitoring locations. Due to the irregular sampling pattern, some

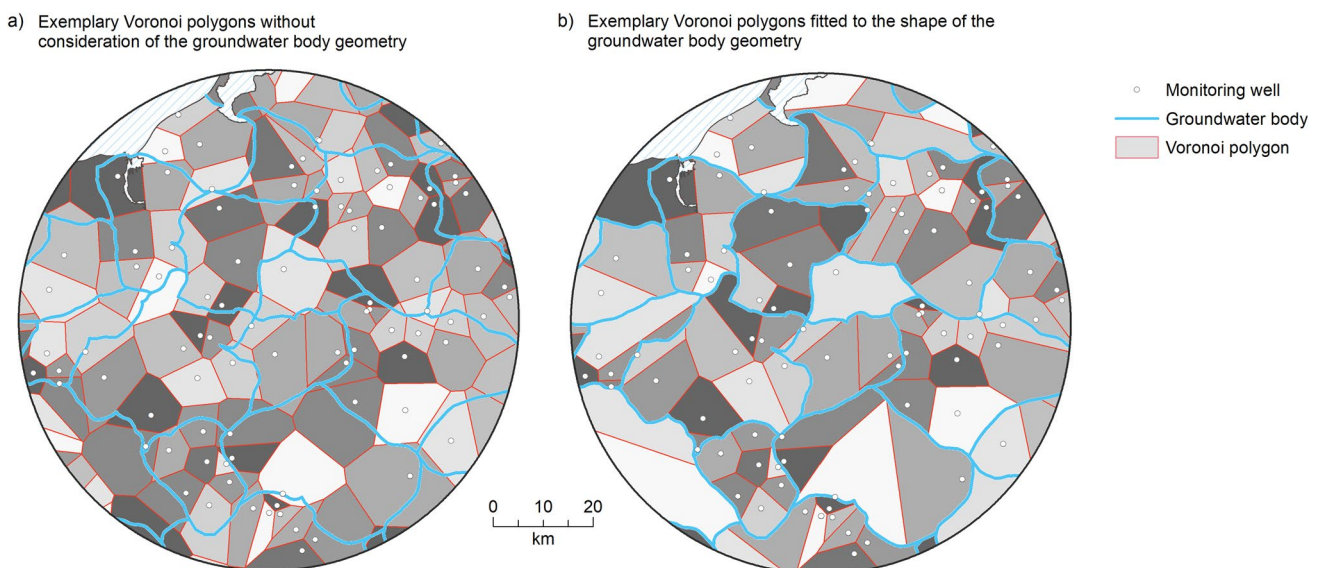
polygons are very small and others are as big as 250 km<sup>2</sup>. However, this first set of polygons needs to be adapted to the borders of the groundwater bodies. It makes sense that hydraulic boundaries which delimit the groundwater bodies should not be intersected by a polygon (Fig. 4). The final set of 568 Voronoi polygons is distributed to 59 groundwater bodies. The number of polygons which represent each individual groundwater body conforms to the number of its monitoring wells. Accordingly, some groundwater bodies can only be modelled by a single polygon, while others cover up to 16 polygons.

### Characterization of spatial variability of nitrate

The data set consists of 568 monitoring wells in Mecklenburg-Western Pomerania for which nitrate concentrations are reported. These values vary between 0.01 and 480 mg/L with a modal concentration of 0.3 mg/L. Thus, the nitrate concentration locally by far exceeds the maximum admissible value of 50 mg/L; in fact, this is the case at 109 monitoring locations, corresponding to 19%.

The frequency distribution of nitrate is strongly positively skewed; therefore, the geostatistical operations are applied to the ln-transformed data (Table 1). Thereafter, the results are back-transformed to the original concentration scale.

Figure 5 shows the experimental and modeled variogram and the histogram of ln-transformed data. The variogram exhibits a high small-scale variability, expressed by the nugget effect, which accounts for 40% of the variance as expressed by the total sill. This high degree of spatial irregularity of nitrate concentration reflects the high local variability of influencing factors, be it the lithological structure



**Fig. 4** **a** Original set-up of Voronoi polygons, and **b** Voronoi polygons adapted to the limits of the groundwater bodies

**Table 1** Frequency distribution and variogram parameters for nitrate concentrations (ln-transformed)

| Min   | Max  | Mean | Variance | Nugget-effect | Sill | Range    | Model     |
|-------|------|------|----------|---------------|------|----------|-----------|
| -4.60 | 6.17 | 0.74 | 6.7      | 3             | 4    | 10,000 m | Spherical |

of the soil and aquifer material, chemical zoning or the input function of nitrate.

### Voronoi regionalization, extension variance and derived probability of exceedance

Figure 6a shows the nitrate values attributed to 568 Voronoi polygons. Following the data values, 459 polygons (81%) fall below the threshold of 50 mg/L (green shadings). These regions correspond to the aquifers below ground moraine slabs. However, two more or less W–E striking areas show values beyond 50 mg/L, with two polygons in the SW and the center, respectively, even exceeding 300 mg/L nitrate (deep red). These zones approximately trace the outwash plains.

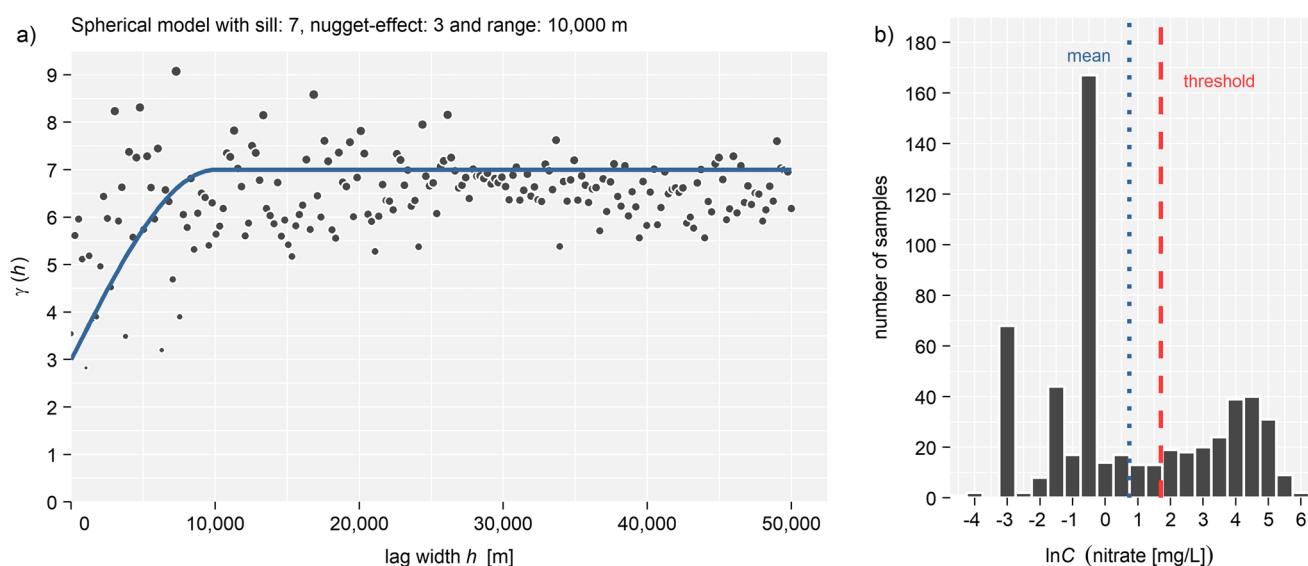
The extension variance as calculated in ln-scale is shown in Fig. 6b. The extension variance  $\sigma_e^2$  lies between 3 and 6.4, respectively, i.e. extension standard deviations  $\sigma_e$  between 1.7 and 2.5. Clearly, the uncertainty is higher (black polygons) in scarcely sampled areas, which are represented by big polygons. However, high uncertainties are also predicted for polygons represented by a monitoring well that is located far from the polygon center, i.e. closer to the border—see for example the two medium-sized, but oddly shaped polygons in the north (Rügen Island, A and B). The level of uncertainty is clearly independent of the data values, as was specified in Eq. (3).

Figure 7 shows the probability that the threshold of 50 mg/L is exceeded within a polygon. It is derived on the basis of the local cumulative probability function (cdf) in each Voronoi polygon. As can be expected, polygons with elevated or high nitrate values (yellow and deep red, respectively, in Fig. 6a) exhibit high exceedance probabilities, e.g. higher than 50% (light and deep red). This is observed along the W–E striking outwash plain regions. Still, there are polygons that have a high level of uncertainty. However, the exceedance probability remains below 50% (A in Fig. 6a) because the data value is low. Polygon B exhibits an elevated probability of exceedance.

### Quality classification of groundwater bodies

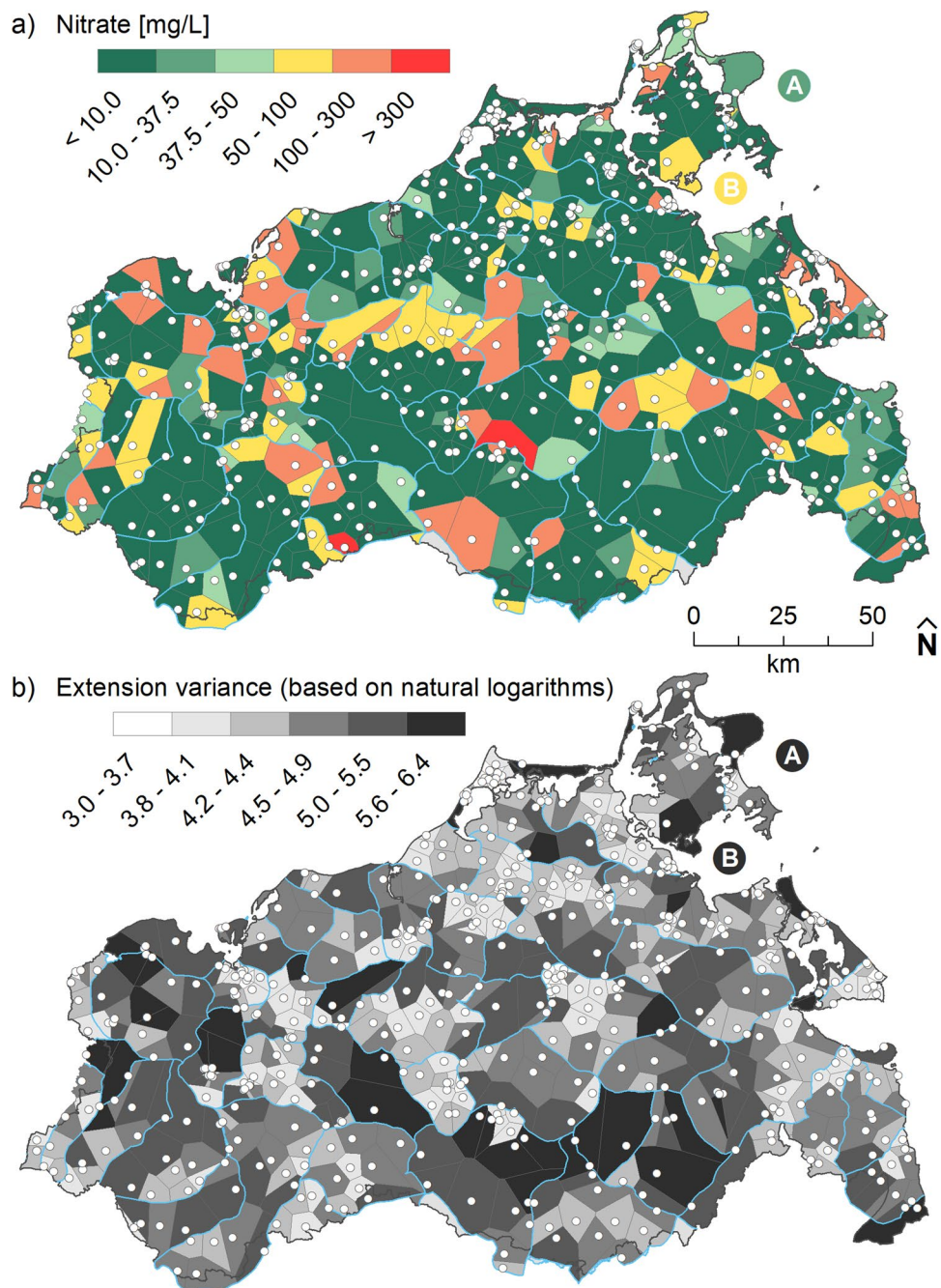
The ordinance on the protection of groundwater (BMUV 2010) requires that the assessment of the chemical status of groundwater applies to groundwater bodies. It is ruled that less than 20% surface area of a groundwater body exhibits a higher nitrate concentration than 50 mg/L. It is then classified as a ‘good status’.

The probabilities of exceedance that apply to the individual Voronoi polygons within a groundwater body are now aggregated in order to calculate the percentage ( $A_{exc}$ ) of the groundwater body surface area that might not fulfill the ‘good status’ criterion (Eq. 4)



**Fig. 5** a Variogram and b histogram of ln-transformed nitrate concentration data. The dashed red line indicates the threshold value of 50 mg/L

**Fig. 6** **a** Voronoi polygons shaded according to their respective data values  $z(x_0)$ , and **b** the extension variance  $\sigma_e^2$  (in ln-scale). *A* and *B* point at two polygons on the island of Rügen, which exhibit high uncertainty due to the marginal position of the well within the polygon

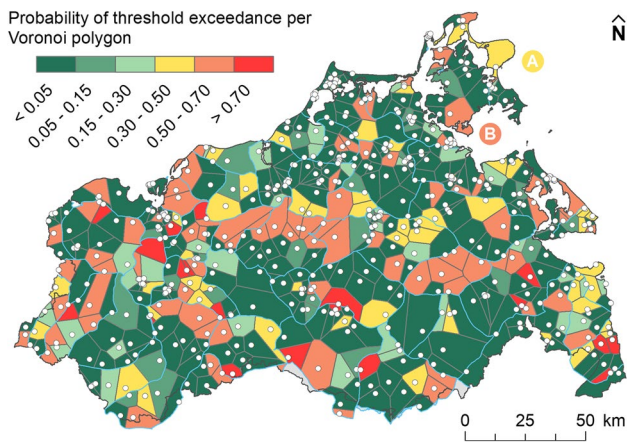


$$A_{exc} = \frac{\sum_{i=1}^m (a_i \cdot P_i)}{\sum a_i} \tag{4}$$

$A_{exc}$  is calculated as an area-weighted average, where  $a_i$  and  $P_i$  are the weights and probabilities, respectively, applying to  $m$  Voronoi polygons within the groundwater body and  $\sum a_i$  representing its total area.

Figure 8 depicts  $A_{exc}$  for all 59 groundwater bodies. According to this assessment, only ten groundwater bodies are clearly in ‘very good status’ (green), and another ten

are in ‘good status’ (yellow), which adds up to one-third of the groundwater bodies classifying as ‘good’. The remaining two-thirds of the groundwater bodies exhibit more than 20% potentially threatened by high nitrate concentrations (light and deep red). The less threatened groundwater bodies are located in the northeast and correspond to aquifers that are well protected by till sediments of the ground moraines; groundwater bodies that are highly uncertain with respect to a potential nitrate stress are identified for the less protected aquifers.

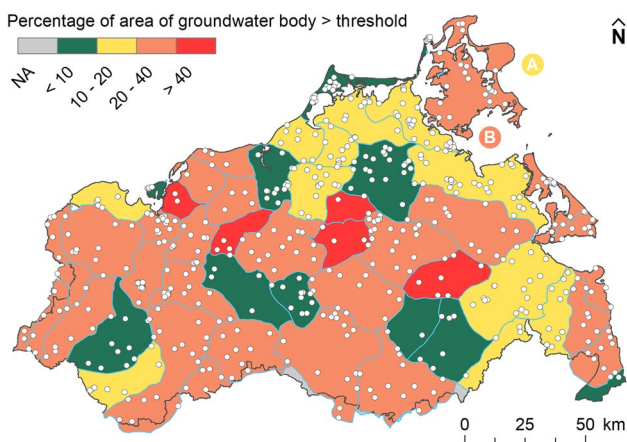


**Fig. 7** Map of probability of exceedance for 59 Voronoi polygons

It is important to note that  $A_{exc}$  does not localize high loaded spots within a groundwater body, rather it reflects the potential that the threshold concentration of 50 mg/L may be exceeded. This potential is derived by merging data values with an uncertainty resulting from the variable's structure and the spatial pattern of monitoring wells.

## Discussion and conclusion

The exercise described in the preceding demonstrates a simple and straightforward method, which helps to assess groundwater quality based on measured values and the uncertainty caused by the spatial monitoring pattern and the individual variability of the considered variable. This method seems specifically well suited for groundwater



**Fig. 8** Final map which indicates the qualitative status of groundwater bodies with respect to nitrate. Groundwater bodies in which less than 20% surface area might exceed the threshold of 50 mg/L nitrate (green-yellow) are considered as ‘good chemical status’

pollution parameters that often reveal an erratic spatial distribution structure. Compared to variables like water-table depth, the spatial distribution of pollutants depends on a multiple set of influencing factors, one of which is the human impact. Therefore, regionalization methods which smooth the spatial variability may not be suitable to assess the quality status of groundwater bodies—overestimations as well as underestimations may occur (Li and Heap 2011). Thus, the early concept of extension variance was revisited here. Originally (Matheron 1971; David 1977), this idea was applied in questions of ore reserve estimation, where an economic balance between sampling expense and assured reserve prediction was most important.

The discussed method offers some advantages over interpolation techniques. Interpolation algorithms are generally not suitable where data are scarce, i.e. wide distances need to be bridged, which is actually an extrapolation. However, they always produce descriptive maps that suggest reliable results. Only a few interpolation algorithms additionally quantify the estimation uncertainty: Kriging produces a measure of reliability ( $\sigma_K$ ), which is, in fact, a special case of  $\sigma_e$  for an orthogonal estimation grid, and thus offers quantitative information on the estimation confidence. In practice, however, this is often neglected.

The variogram inference presented here (Fig. 5; Table 1) shows a spatial continuity (range of influence) of 10 km. In other words, interpolating over larger data distances goes along with null estimation confidence. With an approximate data density of 1 well per 41 km<sup>2</sup> the overall estimation confidence in the investigation domain is generally poor.

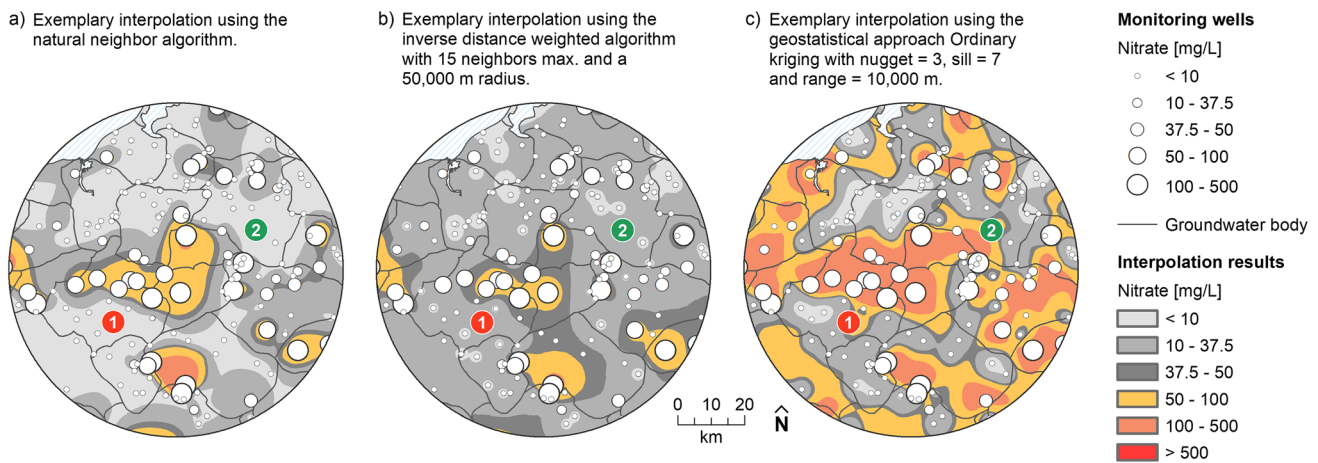
Furthermore, contour plots based on interpolation algorithms usually produce smooth representations of the variable in question. To an unexperienced user, these types of maps might suggest that pinpointing any location on the map will render a realistic value, i.e. the interpolated concentration ‘is true’. The mosaic-like visualization of data values in Voronoi polygons, however, prevents that kind of misunderstanding from the outset, since actually no interpolation is made.

A comparison of three popular interpolation algorithms applied to the same nitrate data set that was used in this study demonstrates the dependence of the assessment on the method chosen. In order to interpolate to a grid point  $x_0$ , interpolation functions calculate weights  $\lambda_i$  for the neighboring data values  $x_i$ :

$$\lambda_i = \frac{A(x_i)}{A(x_0)} \quad \text{Natural neighbor interpolation} \quad (5)$$

where  $A$  denotes the area of Voronoi polygons for data point  $x_i$ , and the intersection of a new polygon for grid point  $x_0$  and the former data polygon  $A(x_i)$ .





**Fig. 9** Comparison interpolated nitrate concentration, **a** natural neighbor interpolation, **b** inverse distance weighted (power 2), and **c** Ordinary Kriging (only the central part of Mecklenburg-Western Pomerania is shown)

$$\lambda_i = 1/d_i^\omega, \quad \text{here } \omega = 2 \text{ Inverse distance interpolation} \quad (6)$$

Usually distances  $d_i$  are squared in the IDW method. In Kriging, the weights  $\lambda_i$  depend on the variogram function and the individual sampling configuration:

$$\lambda_i = f[\gamma(x_0, x_i), \gamma(x_i, x_j)] \text{ Kriging} \quad (7)$$

with  $x_0$  referring to the estimation grid point, and  $x_{i,j}$  corresponding to the neighboring data locations.

The natural neighbor (NN) and IDW interpolation produce similar contour maps of nitrogen concentration. However, Ordinary Kriging provides extended areas where the threshold of 50 mg/L is exceeded. Actually, for the entire state Mecklenburg-Western Pomerania, the percentage of land surface in ‘good status’ is 77 and 80% according to NN and IDW, respectively. Kriging predicts only 40% of the total land surface to be in ‘good status’.

When looking closer at individual groundwater bodies (Fig. 9), e.g. 1 (center) and 2 (northeast), significant differences are revealed when the criterion that good status means less than 20% of the area exceeds 50 mg/L nitrate (BMUV 2010) is applied: For groundwater body 1, NN estimates 28% of the surface area where the threshold is exceeded, thus it is not in ‘good status’. However, according to IDW, groundwater body 1 must be considered in ‘good status (15%), and Kriging clearly certifies a poor quality (53%). For the same groundwater body, the combined Voronoi extension variance method predicts that 30% area of this groundwater body may exceed the threshold. Consequently, special attention is required here. Groundwater body 2 is clearly classified as being in ‘good status’, since apparently 50 mg/L is nowhere exceeded (NN 0.6, IDW 0.7%). In contrast, Kriging shows a high share of

the area as having elevated concentrations (38%). This is due to the fact that Kriging, as all interpolation methods applied here, does not account for hydraulic borders, which delimit the groundwater bodies. However, Kriging tends to preserve the spatial continuity. The interpolated concentrations in groundwater body 2 are strongly affected by some high-concentration wells in the adjacent western groundwater body, which may lead to an overestimation here. The assessment method proposed in this paper, however, predicts a ‘good status’ for groundwater body 2. The reason being is that this groundwater body is spatially relatively well represented by 16 piezometers.

Another advantage is that, provided the spatial variability structure (the variogram) of the variable in question can be inferred from measured data, the local probability density function for all polygons can be derived based on the extension standard deviation  $\sigma_e$  and data value  $z(x_0)$ . Once the local probability density function is calculated, prescribed thresholds can be modified at any time and the resulting exceedance probability determined. Furthermore, because  $\sigma_e$  is independent of the data value, it may even be calculated for additional piezometer locations and thus contribute to an optimization of the monitoring network (Schafmeister 1999; Tietze 1995).

The described method is fast and meets the groundwater protection ordinance’s requirements well; however, some aspects need to be considered. The approach, as discussed here, assumes that the variable in question follows a (1) parametric probability distribution, i.e. normal or ln-normal, and (2) is statistically homogeneous. Hydrochemical concentrations, especially for pollutants, often do not meet the former condition. If simple ln-transformation is not purposeful, a Gaussian anamorphosis (Wackernagel 2003) or the distribution independent indicator (Schafmeister 1999; Hoang et al. 2007) approach should be considered.

The condition of statistical homogeneity assumes that the distribution function of the considered variables holds throughout the entire investigation domain. Applied to groundwater quality data, this means that the physico-chemical conditions of the aquifers should be relatively constant. The principal aquifer system described previously (Mecklenburg-Western Pomerania) consists predominantly of unconsolidated porous material, which can be considered homogeneous with respect to hydrogeochemical interaction and hydrodynamic characteristics. Therefore, it seems justifiable to model the spatial variability of nitrate and other components, which were not discussed here, with a single variogram function. This variogram function is inferred from the whole investigation area, neglecting hydraulic boundaries; however, provided the intrinsic hypothesis can be assumed, it may well reflect the spatial covariance structure of the variable under consideration.

If applied to more heterogeneous geological and hydrogeochemical environments, the method needs to be constrained to statistically homogeneous regions. Investigations on the natural background values of hydrochemical compounds (Walter et al. 2012) suggest that this is especially relevant for complex fractured and karst aquifers, which dominate in the southern states of Germany.

The authorities in the state of Mecklenburg-Western Pomerania proposed the concept of extension variance (Zeissler and others, unpublished technical report for project 2020 0051, see previous details, 2021), as described in the preceding for the parameter nitrate, to be used as a standard routine method for a fast assessment of the groundwater quality according to the requirements in the groundwater protection ordinance (BMUV 2010), which attempts to assess the groundwater quality integrated for groundwater bodies. With respect to the nutrient nitrate, which is mainly applied as fertilizer, another ordinance needs to be obeyed—the Fertiliser Application Ordinance (BMEL 2017). This ordinance rules on how farmers may apply fertilizer, i.e. quantity, time and frequency; however, it requires that the actual nitrate concentration and its temporal development is known for specific agricultural land units, i.e. an accurate localization of elevated concentrations. The concept of merging extension variance with data values results in probabilities of exceedance that apply only to arbitrary Voronoi polygons whose spatial constellation is conditioned solely by the sampling pattern. The aggregation of probabilities of exceedance to groundwater bodies provides an integral evaluation of its quality rather than an exact localization of elevated concentrations. In this sense, the aforementioned method can only provide initial guesses, for which detailed information on the groundwater bodies is lacking and additional monitoring, i.e. a denser piezometer pattern and more frequent sampling, is required.

The assessment method discussed here is based on fundamental geostatistical assumptions and one of the earliest geostatistical measures of uncertainty (or estimation confidence), i.e. the extension variance. Although during seven decades of geostatistical development, many advanced approaches and tools have been developed, this classical concept is still useful, especially in questions of environmental risk assessment, when hard data are still a precious commodity.

**Acknowledgements** The authors acknowledge the great support of the State Agency for the Environment, Nature Conservation and Geology, Mecklenburg-Western Pomerania (LUNG-MV), who not only provided the data but also explained and interpreted the legal aspects of the ordinance. Valuable comments from Hamid Kardan and two anonymous reviewers contributed significantly to the improvement of the manuscript.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

## Declarations

**Conflict of interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Armstrong M (1998) Basic linear geostatistics. Springer, Heidelberg, p 155. <https://doi.org/10.1007/978-3-642-58727-6>
- Bárdossy A, Li J (2008) Geostatistical interpolation using copulas. *Water Resour Res* 44:W07412. <https://doi.org/10.1029/2007W R006115>
- Bárdossy A, Giese H, Grimm-Strele J, Barufke KP (2003) SIMIK+ – GIS-implemientierte Interpolation von Grundwasserparametern mit Hilfe von Landnutzungs- und Geologiedaten [SIMIK+ - GIS-implemented interpolation of groundwater parameters using land use and geological data]. *Hydrol Wasserbewirtsch* 47(1):13–20
- BMEL (2017) Düngeverordnung - Verordnung über die Anwendung von Düngemitteln, Bodenhilfsstoffen, Kultursubstraten und Pflanzenhilfsmitteln nach den Grundsätzen der guten fachlichen Praxis beim Düngen 2 [Fertilizer ordinance - ordinance on the application of fertilizers, soil additives, cultivation substrates and plant auxiliaries in accordance with the Principles of Good Fertilizing Practice 2], last modified 10.08.2021. [http://www.gesetze-im-internet.de/d\\_v\\_2017/index.html#BJNR130510017BJNE000200000](http://www.gesetze-im-internet.de/d_v_2017/index.html#BJNR130510017BJNE000200000). Accessed 26 Aug 2022
- BMUV (2010) Verordnung zum Schutz des Grundwassers [Ordinance on the protection of groundwater], last modified 2017. <https://>

- [www.bmu.de/gesetz/erste-verordnung-zur-aenderung-der-grundwasserverordnung](http://www.bmu.de/gesetz/erste-verordnung-zur-aenderung-der-grundwasserverordnung). Accessed 11 Aug 2022
- BMUV (2020) Nitratbericht 2020 - Gemeinsamer Bericht der Bundesministerien für Umwelt, Naturschutz und nukleare Sicherheit sowie für Ernährung und Landwirtschaft [Nitrate report 2020 - joint report of the Federal Ministries of the Environment, Nature Conservation and Nuclear Safety and Food and Agriculture]. <https://www.bmu.de/download/nitratberichte/>. Accessed 9 Aug 2022
- Bronowicka-Mielniczuk U, Mielniczuk J, Obroślak R, Przystupa W (2019) A comparison of some interpolation techniques for determining spatial distribution of nitrogen compounds in groundwater. *Int J Environ Res* 13:679–687. <https://doi.org/10.1007/s41742-019-00208-6>
- David M (1977) Geostatistical ore reserve estimation, vol 2. Elsevier, Amsterdam, pp 1–364
- Dokou Z, Kourgiyalas NN, Karatzas GP (2015) Assessing groundwater quality in Greece based on spatial and temporal analysis. *Environ Monit Assess* 187:774. <https://doi.org/10.1007/s10661-015-4998-0>
- EC (2000) Directive 2000/60/EC of the European Parliament and of the Council establishing a framework for Community action in the field of water policy. OJ L327, 22.12.2000. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32000L0060:en:NOT>. Accessed 9 Aug 2022
- Fuchs M, Burger H (2000) Geostatistik und die Polygonmethode [Geostatistics and the method of polygons – Materials of the Federal Environmental Agency]. UBA-Texte 49/00, Umweltbundesamt, Berlin
- Gnann SJ, Allmendinger MC, Haslauer CP, Bárdossy A (2018) Improving copula-based spatial interpolation with secondary data. *Spat Stat* 28, pp 105–127. <https://doi.org/10.1016/j.spasta.2018.07.001>
- Gómez-Hernández JJ, Srivastava RM (2021) One step at a time: the origins of sequential simulation and beyond. *Math Geosci* 53:193–209. <https://doi.org/10.1007/s11004-021-09926-0>
- Hoang DN, Schafmeister MT, Bui H (2007) Assessing the risk of contamination of the coastal Quaternary aquifers in Namdinh area (Vietnam): a case study based on indicator kriging. In: Candela L, Vadillo P, Bedbur E, Trevisan M, Vanclooster M, Viotti P, López-Geta JA (eds) Water pollution in natural porous media at different scales: assessment of fate, impact and indicators. WAPO2, Barcelona, Spain, pp 187–193
- Huysmans M, Dassargues A (2010) Application of multiple-point geostatistics on modelling groundwater flow and transport in a cross-bedded aquifer. [https://doi.org/10.1007/978-90-481-2322-3\\_13](https://doi.org/10.1007/978-90-481-2322-3_13)
- Knoll L, Breuer L, Bach M (2020) Nation-wide estimation of groundwater redox conditions and nitrate concentrations through machine learning. *Environ Res Lett* 15. <https://doi.org/10.1088/1748-9326/ab7d5c>
- Li J, Heap AD (2011) A review of comparative studies of spatial interpolation methods in environmental sciences: performance and impact factors. *Ecol Inform* 6(3–4):228–241. <https://doi.org/10.1016/j.ecoinf.2010.12.003>
- Matheron G (1971) The theory of regionalized variables (English translation). Les Cahier due Centre de Morphologie Mathématique, ENSMP, Paris, p 212
- R Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, <https://www.R-project.org/> Accessed June 2023
- Renard P (2007) Stochastic hydrogeology: what professionals really need? *Groundwater* 45:531–541. <https://doi.org/10.1111/j.1745-6584.2007.00340.x>
- Ribeiro PJ, Diggle P J, Schlather M, Bivand R, Ripley B (2020) geoR: analysis of geostatistical data. R package version 1.8-1. <https://CRAN.R-project.org/package=geoR>
- Schafmeister MT (1999) Geostatistik für die hydrogeologische Praxis [Geostatistics for hydrogeological practice]. Springer, Heidelberg, 172 pp
- Schafmeister MT, Pekdeger A (1994) Der Einsatz geostatistischer Verfahren zur Regionalisierung hydrogeologischer Prozesse und Parameter [The use of geostatistical methods for the regionalisation of hydrogeological processes and parameters]. In: Matschullat J, Müller G (eds) Geowissenschaften und Umwelt. Springer, Berlin, Heidelberg, pp 145–150. [https://doi.org/10.1007/978-3-642-79021-8\\_18](https://doi.org/10.1007/978-3-642-79021-8_18)
- Stat-MV (2021) Statistische Berichte, Bodenfläche nach Art der tatsächlichen Nutzung in Mecklenburg-Vorpommern 2020 [Statistical reports, land area by type of actual use in Mecklenburg-Western Pomerania 2020]. <https://www.laiv-mv.de/Statistik/Ver%C3%B6ffentlichungen/Statistische-Berichte/C/>. Accessed 12 Aug 2022
- Strebelle S (2002) Conditional simulation of complex geological structures using multiple-point statistics. *Math Geol* 34:1–22
- Tietze J (1995) Geostatistische Verfahren zur optimalen Erkundung und modellhaften Beschreibung des Untergrunds von Deponien [Geostatistical methods for optimal exploration and model description of the subsurface of landfills]. PhD Thesis, D Vol. 9, Berliner Geowissenschaftliche Abh., Berlin
- Wackernagel H (2003) Gaussian anamorphosis with Hermite polynomials. In: Multivariate geostatistics. Springer, Heidelberg. [https://doi.org/10.1007/978-3-662-05294-5\\_33](https://doi.org/10.1007/978-3-662-05294-5_33)
- Walter T, Beer A, Brose D, Budziak D, Clos P, Dreher T, Fritsche HG, Hübschmann M, Silke Marczinek S, Peters A, Poeser H, Schuster HJ, Wagner B, Wagner F, Wirsing G, Wolter R (2012) Determining natural background values with probability plots. In: Malina G (ed) Groundwater quality sustainability. IAH Selected Papers on Hydrogeology, vol 17. CRC Press, Boca Raton, pp 331–342. <https://doi.org/10.1201/b12715-32>
- Webster R, Oliver MA (2014) A tutorial guide to geostatistics: computing and modelling variograms and kriging. *Catena* 113, Elsevier, Amsterdam, pp 56–69
- Wriedt G, de Vries D, Eden T, Federolf C (2019) Regionalisierte Darstellung der Nitratbelastung im Grundwasser Niedersachsens [Regionalized representation of nitrate pollution in the groundwater of Lower Saxony]. *Grundwasser* 24:27–42. <https://doi.org/10.1007/s00767-019-00415-0>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.