Science of the Total Environment 652 (2019) 1339–1346

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Statistical modelling of groundwater contamination monitoring data: A comparison of spatial and spatiotemporal methods

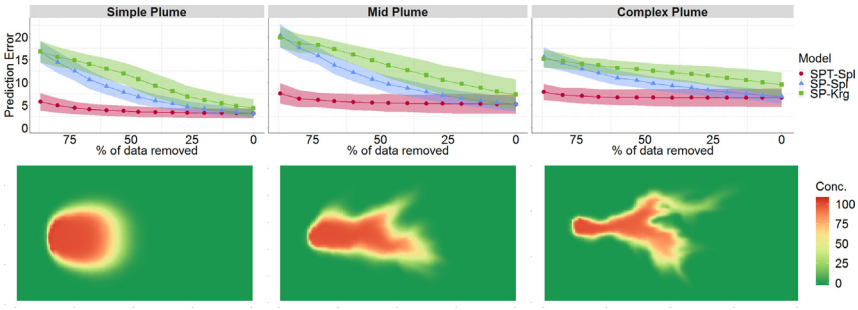
M.I. McLean[a,*], L. Evers[a], A.W. Bowman[a], M. Bonte[b], W.R. Jones[c]

[a] University of Glasgow, School of Mathematics and Statistics, University Place, Glasgow G12 8QQ, United Kingdom of Great Britain and Northern Ireland
[b] Shell Global Solutions International B.V., Kessler Park 1, Rijswijk 2288GS, the Netherlands
[c] Shell Global Solutions, 40 Bank Street, Canary Wharf, London E14 5NR, United Kingdom of Great Britain and Northern Ireland

HIGHLIGHTS

- Target current industrial practice of using repeated spatial analyses
- Spatiotemporal modelling methods result in clearer more accurate plume estimations.
- Kriging and spline models are compared on real and hypothetical plumes.
- Significantly smaller sample sizes can be used while retaining estimation accuracy.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 3 August 2018
Received in revised form 28 September 2018
Accepted 16 October 2018
Available online 22 October 2018

Editor: José Virgílio Cruz

Keywords:
Groundwater monitoring
Spatiotemporal
Kriging
Penalized splines
Statistical modelling
Groundwater contamination

ABSTRACT

Field monitoring of groundwater contamination plumes is an important component of managing risks for downgradient receptors and remedial strategies that rely on monitored natural attenuation. Collection of groundwater quality data can however take a considerable effort and be associated with high cost. Here, we investigated the relative merits of analyzing groundwater quality data using spatial compared to spatiotemporal statistical modelling and assessed the accuracy of both methods and implications for data collection requirements. The aim of this was to determine whether the quantity of data collected can be reduced, while retaining the same level of estimation accuracy, by analyzing groundwater contamination data using a spatiotemporal model which "borrows strength" across time, rather than a spatial model for individual sampling events. To capture the variability encountered under field conditions, we used three hypothetical groundwater contamination plumes with increasing complexity, and site data for a large groundwater gasoline additive plume. The results show that spatiotemporal methods can increase efficiency markedly so that, in comparison with repeated spatial analysis, spatiotemporal methods can achieve the same level of performance but with smaller sample sizes.

1. Introduction

Groundwater quality monitoring is routinely carried out at sites where contamination of groundwater resources has occurred, for example due to the release of hazardous substances, or at sites where the potential consequences of contamination are high, for example in aquifers where groundwater is extracted for human consumption (Fouillac et al., 2009). In both cases, the groundwater monitoring strategy involves assessing the spatial pattern of certain environmental parameters (i.e. the contamination plume) as well as the trend through time (Loaiciga et al., 1992) . Collecting groundwater quality data requires installation of groundwater wells, periodic sampling and laboratory analyses. All of which come at considerable

* Corresponding author.
E-mail address: m.mclean.1@research.gla.ac.uk (M. McLean).

https://doi.org/10.1016/j.scitotenv.2018.10.231

cost and effort, as well as the risk of water resource deterioration due to perforation of aquitards when installing monitoring wells (Bonte et al., 2015) or short circuiting of different aquifers for monitoring wells with long filters (Elci et al., 2001) . Following data collection, the groundwater quality data set is reviewed and analyzed to define the appropriate actions. In most cases, the spatial characteristics of an environmental parameter or plume of groundwater contamination are assessed separately from temporal features. The most common practice for spatial analyses is to apply a statistical model to separate monitoring events (e.g. Ricker, 2008) or apply a single spatial model to a data set consisting of multiple time steps within a certain time period (binning) (e.g. Aziz et al., 2003). Spatial modelling techniques can be as simple as manual interpolations of a plume extent or fitting a concentration trend surface using Kriging or inverse distance weighting (Nas and Berktay, 2010; Reed et al., 2004; Wu et al., 2005) . Temporal analyses are often performed on data collected from individual wells using methods such as the (seasonal) Mann-Kendall test for trends (Donohue et al., 2001; Hirsch and Slack, 1984) or a Student's t-test when testing the significance of the change in water quality between different periods (Harris et al., 1987) . Although several authors have proposed methods to assess the regional consistency between trends observed for different sampling points (e.g. Helsel and Frans, 2006, van Belle and Hughes, 1984), truly integrated spatiotemporal (ST-) modelling of groundwater quality data is rarely done. This contrasts with other environmental disciplines that apply ST-modelling for monitoring acid deposition, forecasting precipitation or stream flow (Kyriakidis and Journel, 1999) and estimating nutrient concentrations in surface water (Knotters and Brus, 2010) . Only quite recently, integrated spatiotemporal statistical modelling tools for groundwater quality data have become available, for example the open-source software GWSDAT (Jones et al., 2015) . The joint modelling of both space and time in a single ST-modelling framework leads to a more coherent interpretation of site groundwater characteristics (Evers et al., 2015) , which in turn, leads to a more accurate interpolation of subsurface groundwater concentrations between sampling points. The increased accuracy of ST-modelling compared to spatial (S-)modelling could be used to optimize groundwater monitoring networks or improve analyses of existing data sets which could ultimately result in a reduced need for intrusive investigations. Although the arguments for ST-modelling are strongly appealing, the actual benefit of ST- over S-modelling in accurately interpolating a groundwater plume has, to our knowledge, not yet been quantified. In this paper, we aim to fill this gap through a systematic comparison of ST- and S-modelling using penalized splines (p-splines) and Kriging. This comparison is performed for three contamination plumes with increasing complexity generated by a numerical groundwater and solute transport model. A case study of a large contaminated site is also discussed. The results of this work will be relevant to researchers seeking to analyse groundwater quality monitoring networks, as well as to industrial regulatory staff who need to define monitoring requirements for contaminated sites.

2. Methods

2.1. Spatial Kriging

Spatial Kriging is one of the most popular spatial modelling techniques. It was first proposed as a method for prediction from a spatial Gaussian process. Given a set of current observations, $\mathbf{Z} = z(\mathbf{s}_1), \ldots, z(\mathbf{s}_n)$, at spatial locations s_i, $i \in 1, \ldots, n$, the Kriging predictor for a set of new spatial locations, $\mathbf{Z}_0$, is obtained by minimising the mean square prediction error (MSPE) and subsequently using the conditional distribution property of a multivariate Gaussian Distribution.

Application of the aforementioned property allows the optimal predictor in a MSPE sense to be derived for new observation locations, $\mathbf{Z}_0$ given current observation locations, $\mathbf{Z} = (z(\mathbf{s}_1), \ldots, z(\mathbf{s}_n))^\top$, as:

$$\mathbf{E}[\mathbf{Z}_0|\mathbf{Z}] = \mu_Z + \mathbf{C}_0^\top \mathbf{K}^{-1}(\mathbf{Z} - \mu_Z \mathbf{1}) \tag{1}$$

and;

$$\mathrm{var}(\mathbf{Z}_0|\mathbf{Z}) = \mathbf{K}_0 - \mathbf{C}_0^\top \mathbf{K}^{-1}\mathbf{C}_0 \tag{2}$$

where μ_Z is the estimated mean of the current observations, $\mathbf{C}_0$ is the matrix of covariances between the prediction locations and the observed locations, $\mathbf{K}$ is the covariance matrix for the observed locations and $\mathbf{K}_0$ is the covariance matrix for the predicted locations.

There are several functions that can be used to describe the covariance structure of these observations, such as the Exponential and Matérn. Classically the parameters in these covariance functions are estimated using maximum likelihood methods. Alternatively a Bayesian approach can be used, (Diggle and Ribeiro, 2007).

2.2. Spatial P-splines

Splines are a common nonparametric regression approach which allows the relationships between a response and covariates to be described in a flexible manner. For the simple one dimensional case with response y_i and covariates x_i; $i \in \{1, \cdots, n\}$, the model is:

$$y_i = f(x_i) + \epsilon_i \tag{3}$$

where $\epsilon_i \sim N(0, \sigma^2)$ describes the random variation and $f(x)$ is a non-parametric regression function of the covariates. In the splines approach, $f(x)$ can be represented by a linear combination of spline basis functions, $B_j(x)$, and corresponding basis coefficients α_j i.e. $f(x) = \sum_{j=1}^{m} \alpha_j B_j(x)$. Estimating the basis coefficients can be achieved by a least squares approach. A common choice of spline basis are b-splines due to their efficient construction from polynomial pieces of a chosen degree (Fahrmeir et al., 2013).

To extend this further to data indexed over space (x_{1i}, x_{2i}) the non-parametric function $f(x)$ can be expressed as:

$$f(x_1, x_2) = \sum_j \sum_k \alpha_{jk} B_j(x_1) B_k(x_2) \tag{4}$$

which is analogous to taking tensor products of the marginal b-spline bases, this can be computed efficiently through row-wise Kronecker products (Lee and Durban, 2011). For computational simplicity the basis functions are constructed over equally spaced knots across the spatial domain.

Choosing the number of basis functions to control the level of smoothness is subjective and involves a bias-variance trade off; increasing the number of basis functions achieves a more flexible function with lower bias but higher variance, while decreasing the number of basis functions reduces the variance but increases the bias. To overcome this difficulty, p-splines were proposed by Eilers & Marxs (1996) who suggested using a relatively high number of basis functions ~ 20–40 per dimension and adding a penalty term on the basis coefficients to control the smoothness of the function. The basis coefficients are estimated as the values of $\boldsymbol{\alpha}$ which minimise the penalised least squares expression:

$$PLS(\boldsymbol{\alpha}) = ||\mathbf{y} - \mathbf{B}\boldsymbol{\alpha}||^2 + \lambda \boldsymbol{\alpha}^\top \mathbf{D}_d^\top \mathbf{D}_d \boldsymbol{\alpha} \tag{5}$$

This leads to the estimated basis coefficients:

$$\hat{\boldsymbol{\alpha}} = (\mathbf{B}^\top \mathbf{B} + \lambda \mathbf{D}_d^\top \mathbf{D}_d)^{-1} \mathbf{B}^\top \mathbf{y} \tag{6}$$

where $\mathbf{D}_d$ is a *dth* order difference matrix and λ, the *smoothing parameter*, is a non-negative value which penalises the overall smoothness of the function. As λ increases the function is pulled towards a linear fit.

A good choice of smoothing parameter is important and this can be tackled through an optimality criteria such as Cross-Validation (CV) and Akaike's Information Criterion (AIC) (Wood, 2006). Alternatively a Bayesian approach can be adopted, (Evers et al., 2015) , where λ is chosen as the value that maximises the log posterior density of λ, known as the MAP (maximum a posteriori) estimate.

2.3. Spatiotemporal P-splines

The spatial splines method can be extended to space-time data with relative ease. The addition of a further dimension significantly increases the number of basis functions used to build these models and, as would be expected, increases the computation time. For spatiotemporal data the non-parametric function $f(x)$ can be expressed as

$$f(x_1, x_2, t) = \sum_j \sum_k \sum_l \alpha_{jkl} B_j(x_1) B_k(x_2) B_l(t) \tag{7}$$

To extend a p-splines model into a third dimension an extra set of basis functions for the temporal component is added. Again, row-wise Kronecker products can be taken to construct the desired tensor product structure.

The penalty term for spatiotemporal data can take two forms, either with one global smoothing parameter, smoothing over space and time equally or, two separate smoothing parameters, one for space and one for time. Eq. (8) defines the least squares estimates of the basis coefficients for a spatiotemporal model with separate space and time smoothing parameters.

$$\hat{\boldsymbol{\alpha}} = \left(\mathbf{B}^\top \mathbf{B} + \lambda \left(\mathbf{D}_s^\top \mathbf{D}_s + \lambda_{rel} \mathbf{D}_t^\top \mathbf{D}_t\right)\right)^{-1} \mathbf{B}^\top \mathbf{y} \tag{8}$$

where λ is the overall smoothing parameter, λ_{rel} is an adjustment of λ for the temporal penalty and $\mathbf{D}_s$ and $\mathbf{D}_t$ are difference matrices for space and time respectively. There are grounds to argue for a separate smoothing parameter for each spatial dimension as well as for time; however, tuning 3 parameters would be very time consuming. To by-pass this, the number of basis functions in the spatial components can be scaled by the vertical and horizontal ranges of the study region.

2.4. Model parameters and specifications

There are two natural approaches to selecting the degree of smoothness, namely Generalised Cross Validation (GCV) and the Bayesian MAP criterion (Evers et al., 2015) . Simulations suggest that GCV is more effective for spatial data (where MAP oversmooths) and MAP is more effective for spatiotemporal data (where GCV undersmooths). These choices therefore allow the different spline models to perform to the best advantage.

In this paper, for the spatial p-splines model, 15 basis functions were used for the easting component, these were then scaled by the study region dimensions for the northing component. For the spatiotemporal splines model, 15 basis functions were also used for the easting and time components and again the northing component was scaled by the spatial dimensions. From other work, this number of basis functions appeared to work well. For the Kriging model a Matérn covariance function was used with $\kappa = 1.5$, the model parameters were optimised using maximum likelihood.

2.5. Hypothetical and real-life contamination plumes

In order to compare effectiveness of the different statistical modelling techniques, three hypothetical contaminant plumes were generated with increasing complexity (simple, mid, complex), and site data was used for a large gasoline additive plume. This case study was also used in earlier work, (Evers et al., 2015) , to assess the criterion for determining the optimal smoothing parameter in a spatiotemporal p-splines model.

The hypothetical plumes were developed using a groundwater flow and contaminant transport model based on MODFLOW (Harbaugh et al., 2000) for groundwater flow, and MT3D (Zheng, 1990) for solute transport. The model considered solute transport by advection and dispersion. The hypothetical 'mid' and 'complex' plumes were simulated using a heterogeneous hydraulic conductivity field which was generated with the procedure, (Frenzel, 1995) . The 'simple' plume assumed homogeneous conductivity with a fixed K = 15m d^{-1}, considered representative of a sand aquifer (Pickens and Grisak, 1981) . For the 'mid' and 'complex' plumes the generated fields had a mean K = 15 m d^{-1} and correlation length = 0.1. The standard deviations of these scenarios were $\log_{10}(\text{sd(K)}) = 0.4$ and 0.9 respectively. These parameter values were chosen based on a summary of field data by Gelhar (Gelhar et al., 1992) who show that at the majority of sites the standard deviation is between 0.4 and 1. Longitudinal dispersivity of the contaminant plumes was fixed at 6 m for each of the three scenarios, where 6 m is 1/10th of the plume width (60 m) (Gelhar et al., 1992). Data were generated at 6-month(182 days) intervals over a period of 10 years. An overview of all model parameters are presented in Table 1, with Fig. 1 showing the final state of each plume.

Table 1
A summary of the model parameters and specifications used for MODFLOW/MT3D for the hypothetical plume study.

Model aspect	Model value
Model discretisation	
Model domain	1 × 1 km^2
Vertical discretisation	1 layer, 10 m thick, confined
Flow parameters (MODFLOW)	
Horizontal conductivity (K)	**Simple plume:** fixed K=15 m d^{-1} (sand). **Mid plume:** mean K = 15 m d^{-1}, standard deviation $\log_{10}$ (sd(K)) = 0.4 **Complex plume:** mean K = 15 m d^{-1}, standard deviation $\log_{10}$(sd(K)) = 0.9
Porosity (n)	0.25
Boundary conditions	Constant head cells at east (0 m) and west (−1 m) boundaries resulting in a hydraulic gradient of 0.001 m/m. Active cells elsewhere Groundwater recharge rate of 1 mm d^{-1}
Time discretisation	Total simulation time: 2 years
Transport parameters (MT3D)	
Longitudinal dispersivity	6 m
Horizontal and vertical transverse dispersivity	0.1 m
Diffusion	Ignored
Advection	Method of characteristics (MOC) scheme (time discretisation based on a Courant number of 0.75)
Boundary conditions	Fixed concentration cells on the west boundary (0 mg l^{-1}) representing clean groundwater inflow, active concentration cells elsewhere The contamination source area (100 m^2) is represented by assigning a concentration groundwater recharge of 100 mg l^{-1} to 36 model cells. Outside the contaminated area, the concentration of recharge is 0 mg l^{-1}.
Initial conditions	Assigned concentration of 0 mg l^{-1}

Observation measurements were obtained by interpolating the MT3D simulated concentrations at a set of predefined times and locations. 15% measurement error was added on the log scale to represent multiplicative error. This error value represents noise from sampling and analytical variations and is based on a comparison of blind duplicate samples in a large unpublished groundwater quality dataset for a Shell site (personal communication J. Smith, Shell Global Solutions). Simulations of each hypothetical plume can be viewed at: https://marnie-svst.shinyapps.io/gw-app/

2.6. Assessing model performance

2.6.1. Hypothetical plume - Mean square prediction error (MSPE)

For the hypothetical plumes MSPEs were used to assess the models' predictive performance. 200 random well networks consisting of 45 wells were simulated for each plume i.e. 900 observations per simulation. For each simulation two types of data removal were considered. The first, referred to as 'observation removal', consisted of removing three individual observations at each sampling time, which is around 5% of the data. The second consisted of removing all of the observations from three wells at each step of data removal, this is referred to as ' well removal'. At each round of data removals each comparison model was built at each sampling time using data record up until this time and MSPEs were calculated using the 'true' concentrations simulated for the three plume types. The MSPE at time t is defined as:

$$MSPE_t = \frac{1}{N}\sum_j \left(y_{\mathbf{s}_j t} - \hat{y}_{\mathbf{s}_j t}\right)^2 \tag{9}$$

where $y_{\mathbf{s}_j t}$ and $\hat{y}_{\mathbf{s}_j t}$ are the true and statistically fitted values, respectively, at spatial prediction location $\mathbf{s}_j = (x_{1j}, x_{2j})$ and prediction time t, and N is the total number of gridded prediction locations at time t.

These MSPEs were then summed over all time points and averaged over the simulations to give mean total MSPEs (MTMSPE) for each stage of data removal.

$$MTMSPE = \frac{1}{m}\sum_m \sum_t MSPE_t \tag{10}$$

where m is the number of simulations and t are the number of time points for which predictions were made.

2.6.2. Case study - k-fold cross-validation

For the case study, the 'true' underlying surface is unknown and all that is available to test predictive performance is observed values at individual locations. This is done with cross-validation (CV), which involves leaving out each observation in turn, building a model with the remaining data then using this model to predict the value for the missing data. Due to the sparsity of the data, each of the chosen times where predictions were made (382, 759, 1144 and 1508 days) only had ~10 observations, many of which were in a very close vicinity. This is a very small number of observations from which to make reliable predictions and thus for spatial splines, observations from within a 1-month time window were also included to improve the model's stability. The CV score for SPT and SP models are computed as:

$$CV = \frac{1}{N}\sum_i \left(y_i - \hat{y}_i^-\right)^2 \tag{11}$$

where $\hat{y}_i^-$ is the predicted value at the $\mathbf{s}_i^{th}$ spatial location at time t using a model that was built without the observation at location $\mathbf{s}_i$

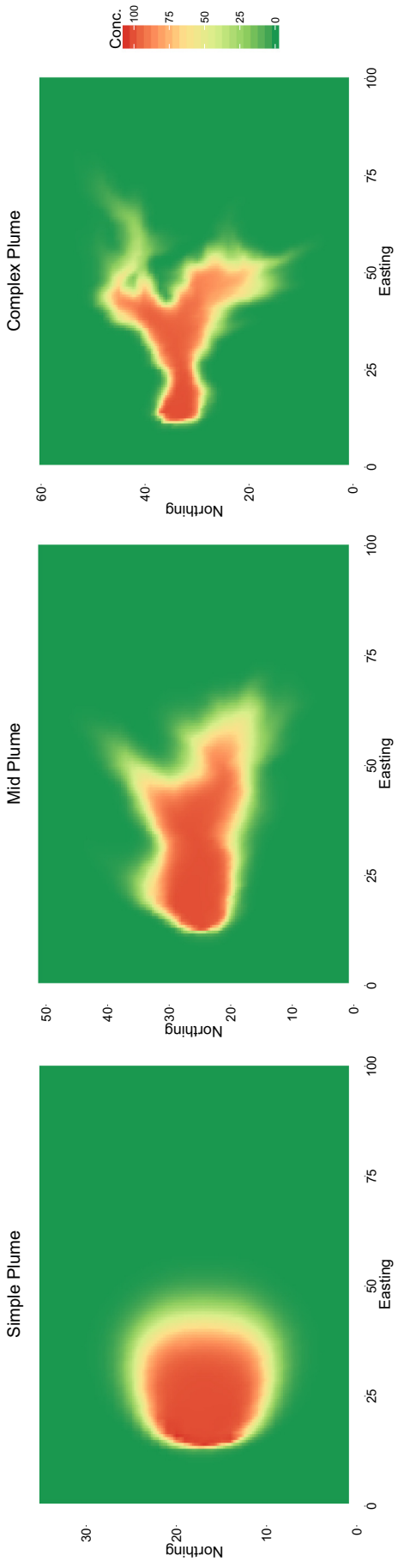


Fig. 1. Final state of each of the three contaminant plumes used for the simulation study.

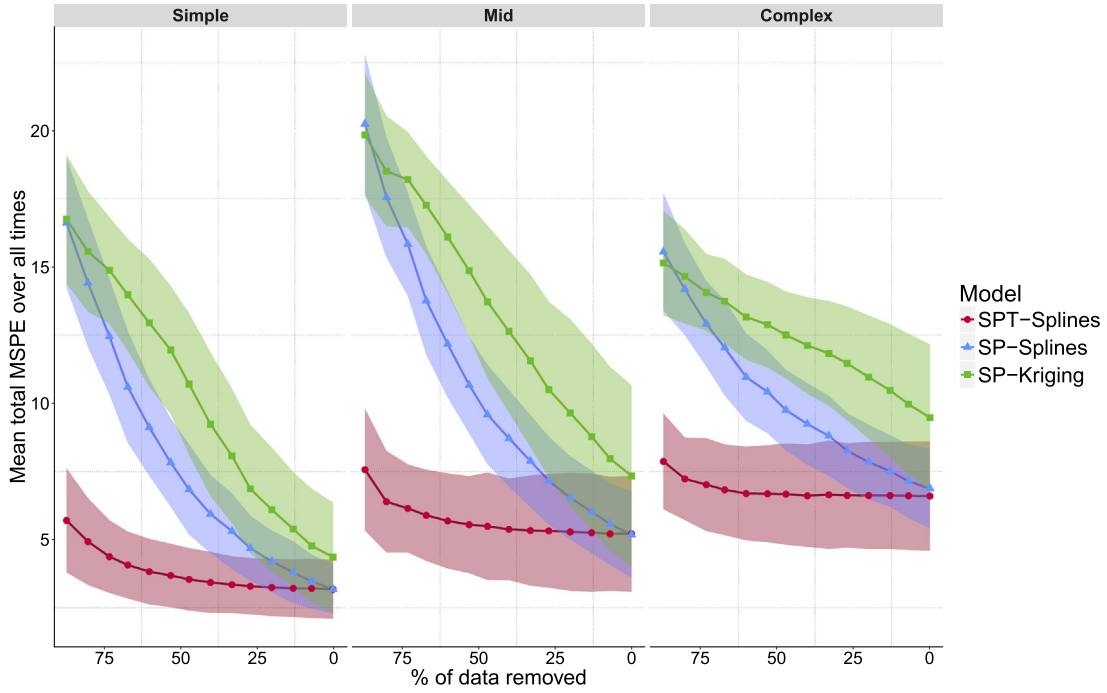


Fig. 2. Mean total Mean square prediction error over all time points at each stage of data removals for 200 random well network simulations. Bands indicate plus and minus one standard deviation.

at time t; y_i is the actual observed value at this location and N is the number of observations.

However, in this form, computing the CV score is very time consuming as it requires N models to be built. To improve computational time and effort, the observations can be divided into k groups (folds) and each group of observations is left out in turn, the CV score is then computed in the same way but is averaged over k rather than N. A low value of CV indicates the model has predicted well. For this study 10 folds were used.

3. Results

3.1. Hypothetical contaminant plumes

Fig. 2 (and the supplementary information) shows that when all the data are used the spline models have a slightly superior performance in comparison with Kriging. As data are removed, the performances of the spatial methods deteriorate markedly in contrast with a much gentler rate of decline with the spatiotemporal

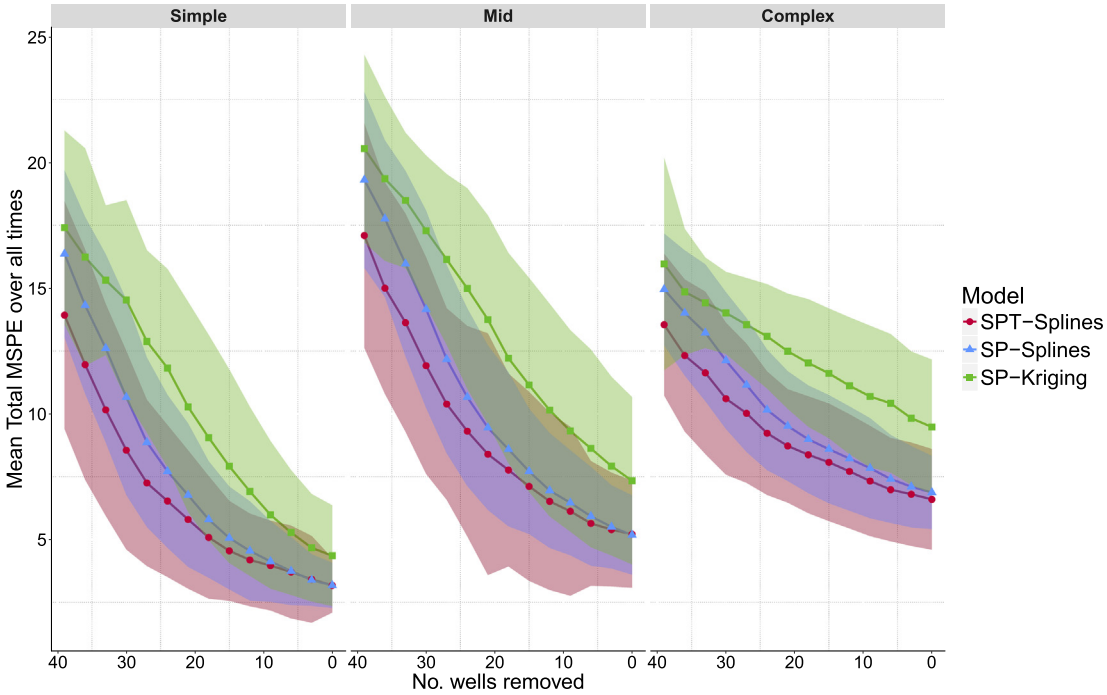


Fig. 3. MTMSPE over all time points at each stage of well removals for 200 random well network simulations. Bands indicate plus and minus one standard deviation.

model. This demonstrates the benefit of 'borrowing strength' across space and time. This particular example involves a substantial number of wells and observations, to allow the relative performances of the spatial and spatiotemporal models to be demonstrated effectively. Clearly, the effectiveness of any model will be considerably reduced when the data become very sparse. Fig. 2 also shows, as expected, a slight decrease in performance of all methods as the complexity of the plume increases.

With the large amount of data in this example, there is little difference in the performance of spatial and spatiotemporal splines when all the data are used. One way of quantifying the relative performances is to observe that the accuracy of estimation of spatial splines when 25% of the data are removed is comparable to the accuracy of estimation of spatiotemporal splines when approximately 75% of the data are removed. This applies across all three plume complexities and therefore suggests scope for reducing sampling frequency when a spatiotemporal model is used for analysis.

Predicted surfaces from one simulation of the 'observation removals' can be viewed in an R Shiny application (Chang et al., 2017) at: https://marnie-svst.shinyapps.io/gw-app/.

Fig. 3 shows results from the scenario where wells are removed rather than observations. This is a more challenging situation as the ability to 'borrow strength' is significantly reduced. That is reflected in the MSPE results where the spatiotemporal model still delivers the best performance but the rate of deterioration with well removal, while slower, is more similar to those of the spatial methods.

3.2. Case study

The shape and direction of the estimated contaminant plume are consistent with the south-east/north-west gradient in the groundwater flow for the tree statistical models (Fig. 4). Looking at the first time point, the spatiotemporal splines model is able to capture the release of the contaminant in the south east corner, while this is not the case for the spatial models due to there being a lack of samples from this region. Following the location of the leak being identified, both models are able to track the depletion of the plume, but the, spatiotemporal model provides a more sharply defined plume shape in comparison with the spatial model. The spatial models estimated surface is particularly poorly compared with the spatiotemporal model at time 1144 days, which is likely to be due to the models over-smoothing.

Due to the sparse nature of the case study data and the large number of sampling events ~500, the effects of data being removed on the estimation accuracy could only be fairly assessed via 'well removal'. The estimation accuracy was also only computed at the times of interest i.e. 382, 759, 1144 and 1508 days rather than at all of the sampling times as in the hypothetical plumes study. The results of removing wells iteratively for the real life plume are similar to those from the well removal simulations for the hypothetical plumes for some of the prediction times (Fig. 5). Removing wells does not seem to have as much of a negative effect on the CV scores for the spatiotemporal models compared with the spatial models when predicting at 382 days and 759 days. This is particularly evident after 40% of the wells have been removed at time 759 days. However, this is not the trend when predicting at times 1144 days and 1508 days. Here the spatiotemporal model performs equally as well as the spatial models as wells are removed, with the spatiotemporal model predictions deteriorating as more than 40% of the wells are removed.

However, that the CV score only measures what is happening at the well locations specifically and not the whole region of interest. The predicted surfaces produced by the spatiotemporal model are consistent with what is known about the leak, and, are able to capture the underlying state of the groundwater in more detail. The spatiotemporal model is able to interpolate the plume more accurately to show the location of the release, whereas the spatial models interpolate a constant surface, specifically at times 1144 days and 1508 days. Thus, although the spatial model performs better in terms of CV score as wells are removed, the predictions achieved by the spatiotemporal model are more meaningful.

4. Conclusions & discussion

4.1. Limitations and modelling considerations

The simulation study presented demonstrates that a spatiotemporal model results in a smoother, clearer and more accurate prediction through time compared to spatial modelling of individual time steps. To achieve the equivalent accuracy with a spatial model, the network needs to be sampled much more extensively.

It is worth noting however, that using a spatiotemporal model is not always straightforward and there are several complications that can be experienced. One is ballooning where unusually high or low predictions arise in data poor regions. This is evident from the narrowing of the standard deviation bands in Fig. 2 with increasing data removal. Ballooning occurs when closely neighbouring observations exhibit a steep gradient followed by a region with no data; this causes the model to continue predicting the steep gradient into the region with no data, (Evers et al., 2015) . As data is removed, the distance between neighbouring observations increases resulting in a more shallow gradient and hence a reduction in the likelihood of ballooning.

Ballooning can be prevented by either increasing the number of basis functions, with increased computational cost, or by increasing the smoothing parameter λ. Also the well network has an influence on ballooning with a gridded network likely to perform better. In reality however, positioning of monitoring wells will generally be dependent on the conceptual site model, location of source zones and receptors, and practical restrictions.

The reduction in sampling frequency that can be achieved by a spatiotemporal model very much depends on how data are removed, with eliminating portions of observations being significantly more effective than eliminating entire wells. There is potential to reduce the number of samples further by considering the sampling design along with using a spatiotemporal model.

Kriging can be extended to the spatiotemporal setting. We have also seen evidence of 'ballooning' in Kriging models when a Matérn covariance structure is used. Splines out-performed Kriging in the spatial setting and given spatiotemporal Kriging requires a significant computational effort compared with spatiotemporal splines it was left out of this study. One benefit to modelling with splines model is that there is no assumption of stationarity and isotropy.

4.2. Environmental and technological implications

The comparison of spatial versus spatiotemporal modelling showed that a similar or higher amount of information can be obtained with fewer observations when using a spatiotemporal modelling approach. Leveraging the data obtained in this way can result in a reduced need for drilling and sampling of monitoring wells, while better characterizing and managing environmental risks. Freely available tools such as GWSDAT http://www.api.org/GWSDAT can facilitate environmental professionals to use spatiotemporal models. The next logical steps to further improve groundwater monitoring may include optimization of the monitoring network design using spatiotemporal modelling and directly integrating continuous field monitoring using telemetry systems at a smaller number of monitoring wells feeding into a spatiotemporal model.

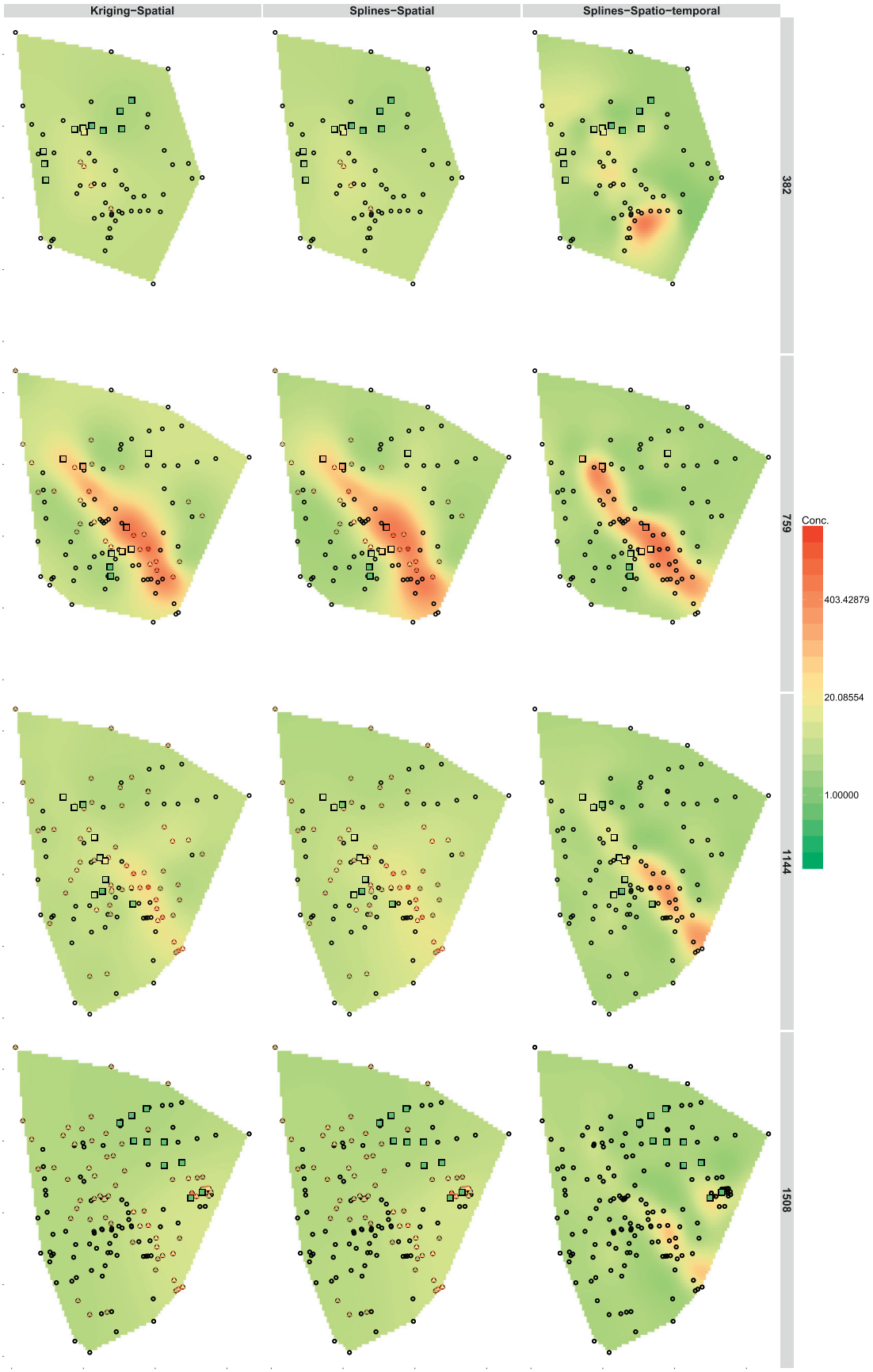


Fig. 4. Predicted contaminant plumes at 4 time points (382, 759, 1144 & 1508 days) for spatial and spatiotemporal models. Black squares indicate wells sampled at the time of interest and are filled by their observed concentrations; red triangles indicate observations that fell within the 1 month time window and that were used for the spatial models and black circles indicate the location of wells that had been sampled at some time before the interpolation time.

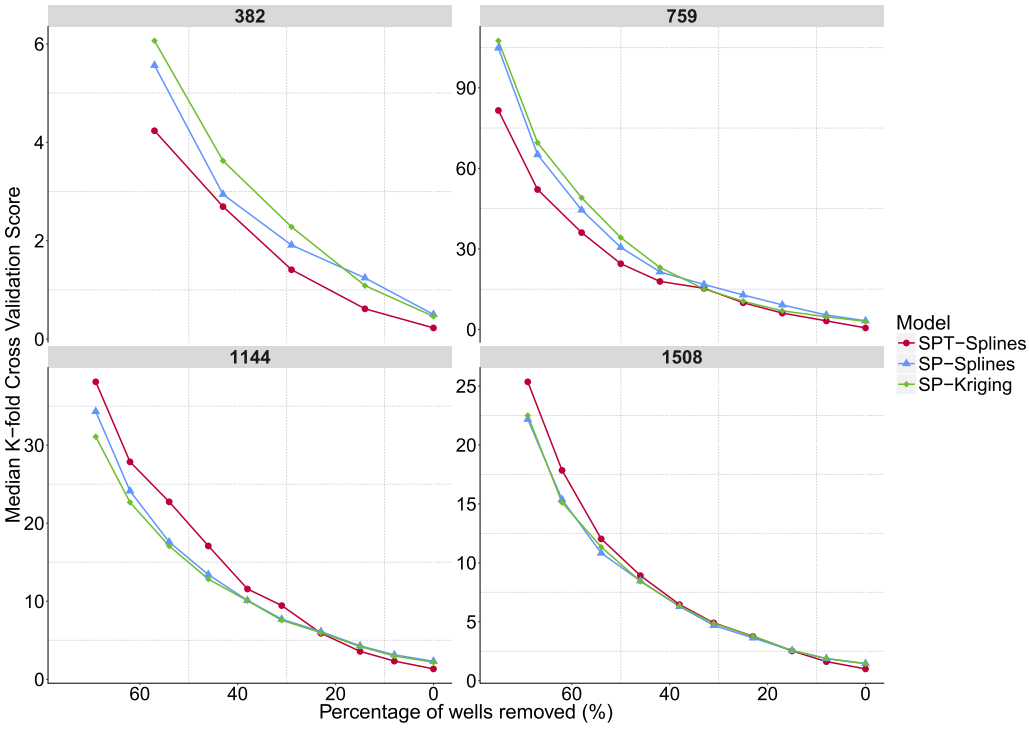


Fig. 5. K-fold cross validation score for each prediction time (382, 759, 1144 & 1508 days) and each model with increasing proportions of wells being removed.

Acknowledgements

This work was part funded by Shell Global Solutions (UK) Ltd. The authors acknowledge contributions from numerous colleagues who contributed to this article: Dr Matthew Lahvis, Dr George Devaull, Dr Michael Spence, Dr Philip Jonathan, Mr Chet Clarke and Professor Jonathan Smith of Shell Projects & Technology. The views expressed are those of the authors and may not reflect the policy or position of Royal Dutch Shell plc.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2018.10.231.

References

Aziz, J.J., Ling, M., Rifai, H.S., Newell, C.J., Gonzales, J.R., 2003. Maros: a decision support system for optimizing monitoring plans. Ground Water 41 (3), 355–367.
Bonte, M., Zaadnoordijk, W.J., Maas, K., 2015. A simple analytical formula for the leakage flux through a perforated aquitard. Groundwater 53 (4), 638–644.
Chang, W., Cheng, J., Allaire, J., Xie, Y., McPherson, J., 2017. shiny: Web Application Framework for R. R package version 1.0.5. https://CRAN.R-project.org/package=shiny.
Diggle, P., Ribeiro, P., 2007. Model-based Geostatistics.3. Springer Series in StatisticsSpringer.
Donohue, R., Davidson, W.A., Peters, N.E., Nelson, S., Jakowyna, B., 2001. Trends in total phosphorus and total nitrogen concentrations of tributaries to the swan - canning estuary, 1987 to 1998. Hydrol. Process. 15 (13), 2411–2434.
Eilers, P.H.C., Marx, B.D., 1996. Flexible smoothing with b-splines and penalties. Stat. Sci. 11 (2), 89–102.
Elci, A., Molz, F.J., Waldrop, W.R., 2001. Implications of observed and simulated ambient flow in monitoring wells. Ground Water 39 (6), 853–862.
Evers, L., Molinari, D.A., Bowman, A.W., Jones, W.R., Spence, M.J., 2015. Efficient and automatic methods for flexible regression on spatiotemporal data, with applications to groundwater monitoring. Environmetrics 26 (6), 431–441.
Fahrmeir, L., Kneib, T., Lang, S., Marx, B., 2013. Regression: Models, Methods and Applications. Springer Berlin Heidelberg.
Fouillac, A.M., Grath, J., Ward, R., 2009. Groundwater Monitoring. Water Quality Measurement SeriesWiley.
Frenzel, H., 1995. A Field Generator Based on Mejia's Algorithm. Institut fr Umwelt-physik, University of Heidelberg.
Gelhar, L.W., Welty, C., Rehfeldt, K.R., 1992. A critical review of data on field-scale dispersion in aquifiers. Water Resour. Res. 28 (7), 1955–1974.
Harbaugh, A.W., Banta, E.R., Hill, M.C., McDonald, M.G., 2000. Modflow-2000, the U.S. Geological Survey Modular Ground-water Model-user Guide to Modularization Concepts and the Ground-water Flow Process. Open-file Report, U. S. Geological Survey.
Harris, J., Loftis, J.C., Montgomery, R.H., 1987. Statistical methods for characterizing ground-water quality. Ground Water 25 (2), 185–193.
Helsel, D.R., Frans, L.M., 2006. Regional Kendall test for trend. Environ. Sci. Technol. 40 (13), 4066–4073.
Hirsch, R.M., Slack, J.R., 1984. A nonparametric trend test for seasonal data with serial dependence. Water Resour. Res. 20 (6), 727–732.
Jones, W.R., Spence, M.J., Bonte, M., 2015. Analyzing groundwater quality data and contamination plumes with GWSDAT. Groundwater 53 (4), 513–514.
Knotters, M., Brus, D.J., 2010. Estimating space-time mean concentrations of nutrients in surface waters of variable depth.W08502–W08502. Water Resour. Res. 46,
Kyriakidis, P.C., Journel, A.G., 1999. Geostatistical space-time models: a review. Math. Geol. 31 (6), 651–684.
Lee, D.J., Durban, M., 2011. P-spline ANOVA-type interaction models for spatio-temporal smoothing. Stat. Model. 11 (1), 49–69.
Loaiciga, H.A., Charbeneau, R.J., Everett, L.G., Fogg, G.E., Hobbs, B.F., Rouhani, S., 1992. Review of groundwater quality monitoring network design. J. Hydraul. Eng. 118 (1), 11–37.
Nas, B., Berktay, A., 2010. Groundwater quality mapping in urban groundwater using GIS. Environ. Monit. Assess. 160 (1–4), 215–227.
Pickens, J.F., Grisak, G.E., 1981. Scale-dependent dispersion in a stratified granular aquifer. Water Resour. Res. 17 (4), 1191–1211.
Reed, P.M., Ellsworth, T.R., Minsker, B.S., 2004. Spatial interpolation methods for nonstationary plume data. Ground Water 42 (2), 190–202.
Ricker, J.A., 2008. A practical method to evaluate ground water contaminant plume stability. Groundwater Monit. Remediat. 28 (4), 85–94.
van Belle, G., Hughes, J.P., 1984. Nonparametric tests for trend in water quality. Water Resour. Res. 20 (1), 127–136.
Wood, S.N., 2006. Generalized Additive Models: An Introduction with R. Chapman & Hall/CRC Texts in Statistical ScienceTaylor & Francis.
Wu, J., Zheng, C., Chien, C.C., 2005. Cost-effective sampling network design for contaminant plume monitoring under general hydrogeological conditions. J. Contam. Hydrol. 77 (1), 41–65.
Zheng, C., 1990. MT3D. A Modular Three-dimensional Transport Model for Simulation of Advection, Dispersion and Chemical Reactions of Contaminants in Groundwater Systems, Report to the US Environmental Protection Agency. 170.