

Wavelet denoising and cubic spline interpolation for observation data in groundwater pollution source identification problems

Ying Zhao, Qiang Fu, Wenxi Lu, Ji Yi and Haibo Chu

ABSTRACT

As the identified results of groundwater pollution source identification (GPSI) can influence the cost for the polluter in paying for remediating groundwater resources, it is important that the accuracy of the estimated result should be as high as possible. However, many factors can influence the result, such as noisy concentration data and incomplete concentration data. Thus, this paper is aimed at studying the difference between using the observation data before and after denoising and interpolating for solving GPSI problems. Four kinds of noise level and 20 groups of missing data were designed to test the performance of wavelet denoising and cubic spline interpolation, respectively. The results show that the denoising process can improve the estimated result for the GPSI problem, and the higher the noise level, the stronger this effect. In terms of interpolation, more accurate results can be made after interpolating if the missing data belong to the period after the source releases the pollutant. If the missing data are from when the pollution source is active, interpolation cannot help increase the estimated performance.

Key words | denoising, groundwater pollution, interpolation, source identification

Ying Zhao

Qiang Fu (corresponding author)

Ji Yi

School of Water Conservancy & Civil Engineering,
Northeast Agricultural University,
Harbin 150030,
China

E-mail: fuqiang@neau.edu.cn

Wenxi Lu

Key Laboratory of Groundwater Resources and
Environment Ministry of Education,

Jilin University,
Changchun 130021,
China

Haibo Chu

Department of Hydraulic Engineering,
Qinghua University,
Beijing 100084,
China

INTRODUCTION

Since groundwater pollution sources are buried in underground rock and soil media, the characteristics of these sources (including the number of pollution sources, their exact location, and release history) are difficult to make clear. However, information about groundwater pollution sources is connected with designs of remediation schemes, estimating risks, and conducting pollution liability assessments (Woodrow *et al.* 1986; Mirghani *et al.* 2012; Zhao *et al.* 2016). Therefore, studies of groundwater pollution source identification (GPSI) are especially important.

Comprehensive reviews on GPSI have been provided by several researchers (Atmadja & Bagtzoglou 2001; Michalak & Kitanidis 2004; Bagtzoglou & Atmadja 2005; Yeh 2015). Methods for GPSI can be divided into four categories: optimization approaches, probabilistic and geostatistical

simulation approaches, analytical solution and regression approaches, and direct approaches. No matter which approach is used, the effect of the error in the observation data is shown to have a strong sensitivity for the estimated result. Even very small measurement errors can make the estimated release history change dramatically (Skaggs & Kabala 1994; Singh *et al.* 2004). Thus, it is necessary to extract the actual concentration data from the noisy data first, then use the inversion method to identify the characteristics of the release sources. The technique for data denoising develops from simple nonparametric regression to spatially adaptive nonparametric methods (Percival & Walden 2000). The latter methods include variable-bandwidth kernel methods, variable-knot splines and wavelets (Hsu *et al.* 2005). This research chooses wavelet

doi: 10.2166/ws.2019.013

denoising over other methods because it handles abrupt changes better than the others (Donoho & Johnstone 1994).

In addition, identification accuracy is reduced if observation data are missing (Milnes & Perrochet 2007; Datta et al. 2011). The filling in of missing values, called data imputation, can be used to complete this incomplete data. The classic methods for dealing with incomplete data include Lagrange imputation, Newton imputation, Hermit imputation and cubic spline interpolation. The reason to choose cubic spline interpolation is its accuracy for the kind of missing data with a few missing values (Lakshminarayan et al. 1999).

Although the influence of both noisy observation data and missing data on the identified results is mentioned in many studies (Ayvaz 2010; Zhao et al. 2015), there has been no specific study on identifying the source characteristics of the observation data after denoising and interpolation has been performed. Thus, this study focuses on the differences of the identified source characteristics in the observation data before and after denoising and interpolating. Four kinds of noise level and 20 groups of missing data were designed to test the performance of wavelet denoising and cubic spline interpolation, respectively.

METHODOLOGY

A flow diagram of the proposed approach is presented in Figure 1.

A brief description of the approach steps is presented below.

Wavelet denoising

Noise is a phenomenon that affects all frequencies. In general, useful signals occupy the low-frequency domain. On the contrary, noise occupies the high-frequency domain. Wavelet analysis can separate the high-frequency components and low-frequency components effectively.

The first procedure of wavelet denoising applies a wavelet transform to a data object to obtain the vector of the wavelet coefficient. Second, we suppress or remove those elements above a certain threshold. Third, we reconstruct the denoised data using a cognate inverse wavelet

transform. The detail of how to denoise data with noise can be seen in Cao et al. (2003).

Cubic spline interpolation

The spline is a draftsman to draw a smooth curve slat. The spline function is obtained after the spline curve has been described by mathematical modeling. Its characteristics are piecewise polynomial and each piecewise polynomial has a certain connection, so the spline function is not only sufficiently smooth but also retains a certain discontinuity. Smoothness is to ensure the smooth graceful shape of the curve. Discontinuity can make it turn freely flexible when used.

Let $\Delta: a = x_0 < x_1 < \dots < x_m = b$ be a partition of $[a, b]$. Then s is said to be a cubic spline over Δ if $s \in C^2[a, b]$ and s restricted to $[x_{i-1}, x_i]$ is a cubic polynomial for $1 \leq i \leq m$. The space of all such cubic splines s is denoted by $SP(\Delta, 3)$. If, in addition, $s(x_i) = f(x_i)$ for $0 \leq i \leq m$, s is said to be an $SP(\Delta, 3)$ -interpolate of f (Lucass 1974).

The simulation–optimization method

The simulation model is the principal part of the simulation–optimization method. The governing partial differential equations for steady-state flow and transport in a two-dimensional aquifer system are shown in Equations (1) and (2), respectively (Pinder & Bredehoeft 1968; Singh & Datta 2006):

$$\frac{\partial}{\partial x_i} (T_{ij} \frac{\partial h}{\partial x_j}) = W \quad i, j = 1, 2 \quad (1)$$

$$\frac{\partial (bC)}{\partial t} = \frac{\partial}{\partial x_i} (bD_{ij} \frac{\partial C}{\partial x_j}) - \frac{\partial}{\partial x_i} (bCv_i) + \frac{C_s W}{\theta} \quad i, j = 1, 2 \quad (2)$$

Note that Equation (2) is related to Equation (1) through Darcy's Law, as follows:

$$v_i = -\frac{K_{ij}}{\theta} \frac{\partial h}{\partial x_j} \quad i, j = 1, 2 \quad (3)$$

where T_{ij} represents the transmissivity tensor [$L^2 T^{-1}$], h represents the hydraulic head [L], W represents the volumetric flux per unit area representing fluid inflow (negative sign)

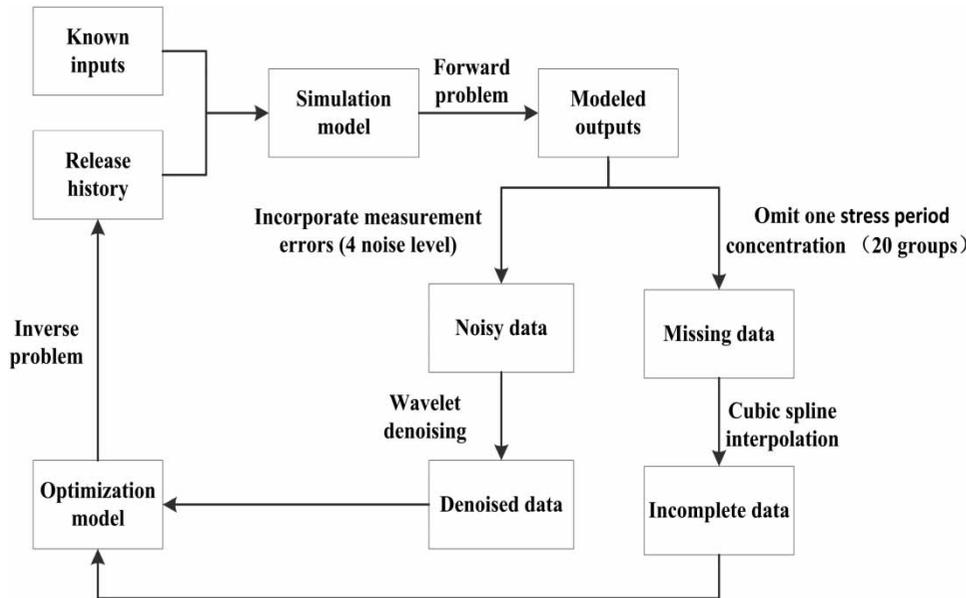


Figure 1 | Flow chart of the proposed approach.

and outflow (positive sign) [LT⁻¹], *b* represents the saturated thickness of the aquifer, *C* represents the dissolved concentration [ML⁻³], *D*_{*ij*} represents the hydrodynamic dispersion tensor [L²T⁻¹], *v*_{*i*} represents the average linear seepage velocity [LT⁻¹], *C*_{*S*} represents the dissolved concentration in a source or sink flux [ML⁻³], *θ* represents the effective porosity [-], and *K*_{*ij*} represents the hydraulic conductivity tensor [LT⁻¹].

The goal of the optimization model is to estimate the best source characteristics that minimize the error between the actual concentrations and model predictions. The optimization model is as follows:

$$\min \left[\sum_{t=1}^{N_t} \sum_{k=1}^{N_d} (C_k(t) - \tilde{C}_k(t))^2 \right] \tag{4}$$

Subject to:

$$\tilde{C} = f(\tilde{q}) \tag{5}$$

$$C^L \leq C \leq C^U \tag{6}$$

$$\tilde{q}^L \leq \tilde{q} \leq \tilde{q}^U \tag{7}$$

The meaning of the objective function shown in Equation (4) is to minimize the deviation between actual

observations and model predictions at each sampling location and each stress period. *N_t* stands for the number of stress periods, *N_d* stands for the number of sampling locations. *C_k(t)* represents the simulated pollution concentration at sampling location *k* and stress period *t* and $\hat{C}_k(t)$ means the measured pollution concentration at sampling location *k* and stress period *t*. In Equation (5), *f*(\tilde{q}) stands for the simulation model which transforms the source flux *q* into concentrations *c*. Equations (6) and (7) stand for the bounds of the concentrations and source fluxes, respectively.

Performance evaluation measurements

The identified results are evaluated in accordance with the normalized error (NE). The NE formula can be defined as follows (Mahar & Datta 2001):

$$NE = \frac{\sum_{t=1}^{N_t} \sum_{i=1}^n |\tilde{q}_i(t) - \hat{q}_i(t)|}{\sum_{t=1}^{N_t} \sum_{i=1}^n \hat{q}_i(t)} \times 100 \tag{8}$$

As the NE value goes to zero, the better the identified values are.

The parameter of wavelet denoising is evaluated by S_j . The formula for S_j can be defined as follows:

$$S_j = \frac{\max(w_{j,i})}{\sum_{i=1}^{N_j} |w_{j,i}|} \quad (9)$$

where $w_{j,i}$ stands for approximation coefficients and detailed coefficients. S_j reflects the content of useful and noisy information of the signal. If $S_j < 0.01$, it means the detailed coefficient only contains noisy information.

Incorporating measurement errors

The perturbed concentrations can be generated as follows:

$$\hat{C} = C + \varepsilon \cdot a \cdot C \quad (10)$$

where ε is a normally distributed error matrix, and a represents the level of noise. The noise level is low if $a < 0.10$. The noise level is moderate if $0.10 \leq a \leq 0.15$. The noise level is high if $a > 0.15$ (Singh & Datta 2006).

NUMERICAL APPLICATIONS

Basic introduction of the aquifer

The performance of both wavelet denoising and cubic spline interpolation is tested on a hypothetical case. The aquifer (Figure 2) is assumed to be uniform and the flow condition is steady-state. Boundary conditions of the aquifer model are linearly varying specified heads at the east and west sides, and no-flow at the north and south sides. There is one pollution source (S1) and six sampling wells (O1–O6). Observation pollution concentrations at each sampling location were gathered in each stress period by running the simulation model. The simulation time for this case is 60 months (5 years), which has been divided into 20 equal stress periods. Each stress period is 3 months. We assume that the released pollutant is typically conservative and the initial conditions, boundary conditions, and the length of the release periods are known. Moreover, the source is active in the first four stress periods (1 year). The detailed parameter values and the actual values of source fluxes are shown in Tables 1 and 2.

Note that the actual values of the source flux (Table 2) are not used in the calculation process. It only acts as the

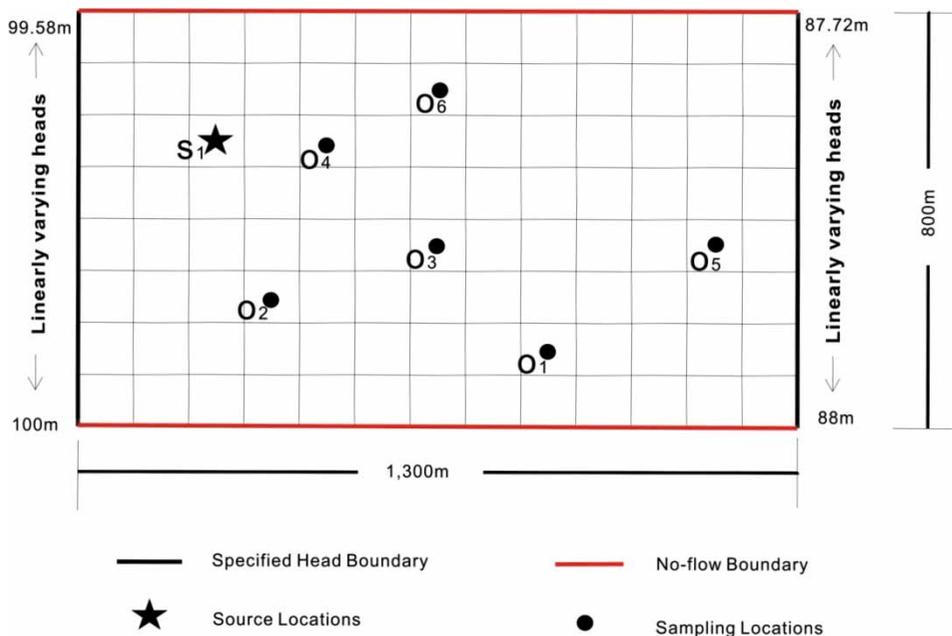


Figure 2 | Hypothetical aquifer model.

Table 1 | Aquifer parameters for the case

| Parameters | Values |
|--|----------------------|
| Hydraulic conductivity in x -direction, K_{xx} (m/s) | 0.0002 |
| Hydraulic conductivity in y -direction, K_{yy} (m/s) | 0.0002 |
| Effective porosity, θ | 0.25 |
| Longitudinal dispersivity, α_L (m) | 40 |
| Transverse dispersivity, α_T (m) | 9.6 |
| Saturated thickness, b (m) | 30.5 |
| Volume flux per unit area, W (m/s) | 1.0×10^{-9} |
| Initial concentration (mg/L) | 100 |

Table 2 | Actual values of the source fluxes in the first four stress periods (SP: stress period)

| Pollution sources | Source fluxes in each stress period (g/s) | | | |
|-------------------|---|-----|-----|-----|
| | SP1 | SP2 | SP3 | SP4 |
| S1 | 255 | 497 | 0 | 297 |

input of the simulation model to generate the observation concentrations at each sampling location, and the standard values evaluate the performance estimated results. This is also an advantage of the hypothetical case, as in the field study, the actual values are unknown. Therefore, the identified results could not be evaluated for the field study.

Performance of wavelet denoising

At first, the actual observation concentrations are perturbed by Equation (9). Then the observation concentrations with noise are denoised using wavelet denoising. Both code

programming and wavelet toolbox are used to train the parameter and filter. Wavedec functions and wrcoef functions are the most important functions when using Matlab. The parameters needing to be estimated for wavelet denoising are wavelet filter and denoising level. The Daubechies 4 filter, Coiflet 2 filter, and Symmlet 4 are the alternative wavelet filters due to their popularity for noised groundwater concentrations, and denoising level (N) is set from 1 to 4. The performance of the parameters is shown in Table 3.

If $S_j < 0.01$, it means the detailed coefficient only contains noisy information. It can be seen that only the combination of Daubechies 4 filter and denoising level 1 can reach this requirement. Thus, the denoising level is set to be 1, and the Daubechies 4 filter is chosen as the wavelet filter. After the parameters are estimated, the denoising results can be calculated. From the limitation of space, only one sample of denoising results is shown in Figure 3. To evaluate the accuracy of wavelet denoising, the Gaussian filter denoising method (Janecki 2011; Afshari et al. 2017) is employed. Figure 4 shows the comparison of residual results by wavelet and Gaussian filter under four different noise levels. It can be seen clearly that wavelet denoising performs more accurately than Gaussian filter denoising under all four noise levels. The average NE values for wavelet denoising and Gaussian filter denoising are 2.0% vs 3.2%, 3.8% vs 5.4%, 5.7% vs 7.6%, 7.5% vs 9.9% when noise level equals 0.05, 0.10, 0.15, 0.20, respectively.

The identified results for both the noisy observation concentrations and denoising observation concentrations are shown in Table 4. As the noise level increased, the accuracy

Table 3 | Performance of the estimated parameters for wavelet denoising

| Noise level | Criterion | Daubechies 4 | | | | Coiflet 2 | | | | Symmlet 4 | | | |
|-------------|-----------|--------------|-------|-------|-------|-----------|-------|-------|-------|-----------|-------|-------|-------|
| | | $N=1$ | $N=2$ | $N=3$ | $N=4$ | $N=1$ | $N=2$ | $N=3$ | $N=4$ | $N=1$ | $N=2$ | $N=3$ | $N=4$ |
| $a = 0.05$ | S_j | 0.002 | 0.011 | 0.025 | 0.033 | 0.010 | 0.011 | 0.033 | 0.025 | 0.012 | 0.015 | 0.029 | 0.017 |
| | NE | 1.99 | 4.87 | 10.39 | 14.16 | 1.98 | 4.50 | 10.53 | 14.08 | 2.00 | 4.16 | 10.69 | 14.32 |
| $a = 0.10$ | S_j | 0.003 | 0.02 | 0.02 | 0.03 | 0.011 | 0.013 | 0.034 | 0.026 | 0.012 | 0.015 | 0.030 | 0.017 |
| | NE | 3.81 | 5.76 | 10.87 | 14.62 | 3.69 | 5.36 | 11.02 | 14.46 | 3.70 | 4.97 | 11.11 | 14.66 |
| $a = 0.15$ | S_j | 0.004 | 0.018 | 0.025 | 0.036 | 0.010 | 0.014 | 0.035 | 0.027 | 0.012 | 0.015 | 0.031 | 0.017 |
| | NE | 5.65 | 4.68 | 11.36 | 15.09 | 5.42 | 6.25 | 11.52 | 14.85 | 5.43 | 5.82 | 11.54 | 14.99 |
| $a = 0.20$ | S_j | 0.004 | 0.020 | 0.026 | 0.037 | 0.010 | 0.015 | 0.036 | 0.027 | 0.012 | 0.014 | 0.032 | 0.017 |
| | NE | 7.52 | 7.62 | 11.85 | 15.56 | 7.23 | 7.16 | 12.03 | 15.25 | 7.24 | 6.69 | 12.01 | 15.32 |

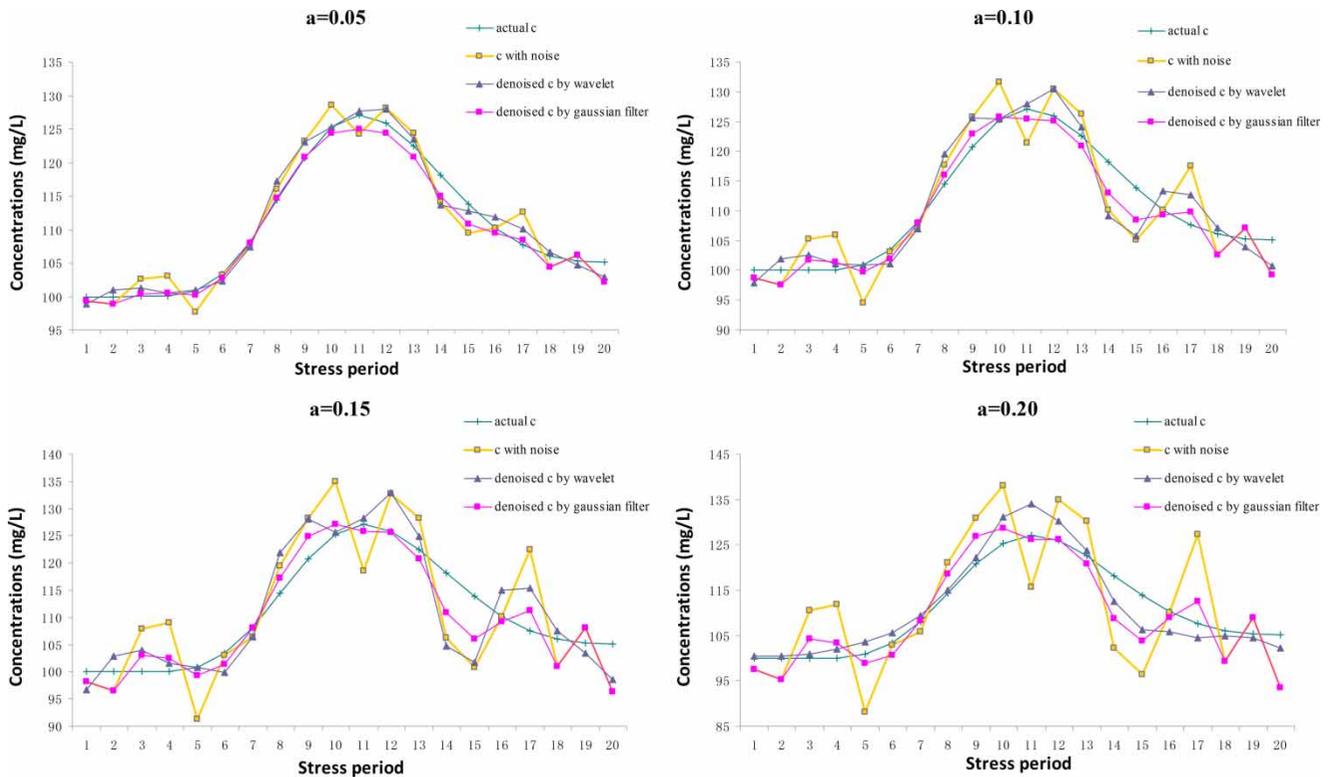


Figure 3 | Denoising results for different noise level concentrations (c is short for concentrations).

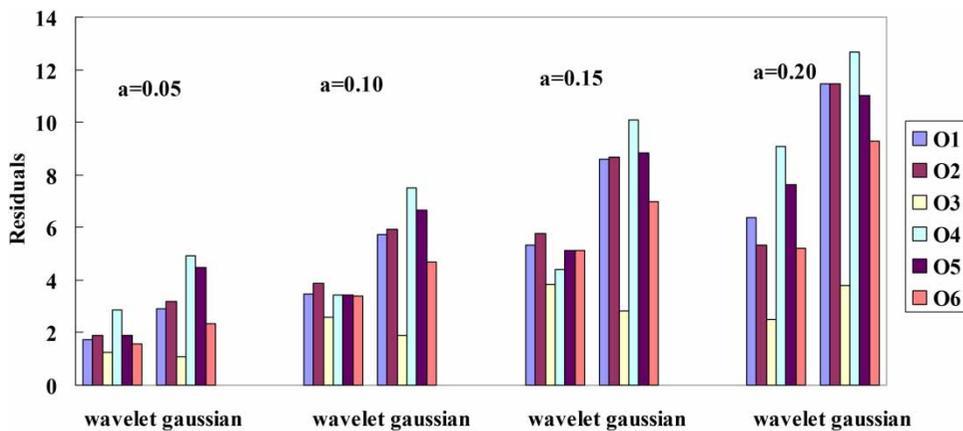


Figure 4 | Residual results by wavelet and Gaussian filter under four different noise levels.

of the identified results decreased for both the noisy (from 18.81% to 65.24%) and denoised data(19.30% to 51.39%). In contrast with the two groups of data, the identified results have a slight difference when noise level is low. However, when noise level increased from 0.10 to 0.20, the estimated result using denoised observation concentrations is much better than those using noisy observation concentrations for the same noise level.

Performance of cubic spline interpolation

To avoid any influence on the interpolation results, all the observation data are the output of the simulation model with no noise added. The performance of three different interpolation methods is shown in Figure 5.

The green line shows the interpolation performance by cubic spline interpolation. As the stress period goes up,

Table 4 | Estimated values of the source fluxes (SP: stress period)

| Pollution sources | Noise level <i>a</i> | Source fluxes at each stress period (g/s) | | | | NE of the identified results |
|-------------------|----------------------|---|-----|-----|-----|------------------------------|
| | | SP1 | SP2 | SP3 | SP4 | |
| Before denoising | <i>a</i> = 0.05 | 258 | 486 | 101 | 214 | 18.81 |
| | <i>a</i> = 0.10 | 253 | 492 | 181 | 138 | 33.00 |
| | <i>a</i> = 0.15 | 253 | 485 | 276 | 57 | 50.39 |
| | <i>a</i> = 0.20 | 247 | 492 | 357 | -19 | 65.24 |
| After denoising | <i>a</i> = 0.05 | 259 | 482 | 102 | 215 | 19.30 |
| | <i>a</i> = 0.10 | 247 | 517 | 139 | 163 | 28.65 |
| | <i>a</i> = 0.15 | 240 | 542 | 186 | 106 | 41.52 |
| | <i>a</i> = 0.20 | 238 | 559 | 224 | 60 | 51.39 |
| Actual values | <i>a</i> = 0 | 255 | 497 | 0 | 298 | |

the NE of the interpolation results decreases. The NE value after the seventh period is less than 1%. Moreover, except for the data in the first stress period, the other NE values are all below 10%. In fact, in estimating the first concentration data, it is not interpolation, but deduction of the back-up value (using the second to the 20th data to estimate the first value). For other interpolation results, the accuracy is not bad.

To verify the performance of cubic spline interpolation, linear interpolation and nearest interpolation methods are used for comparison. The Interp1 function in Matlab is the most important function used for interpolation. From Figure 5, it can be seen that cubic spline interpolation performs

obviously more accurately in all 20 stress periods. Thus, cubic spline interpolation is employed to interpolate incomplete concentrations. The performance of the identified results before and after interpolation is shown in Figure 6.

The blue line in Figure 6 illustrates the normalized errors using the incomplete observation concentrations, which missed the corresponding stress period concentration data. In the first four stress periods, the identified results get worse as we move from period 1 to 4. After the fourth stress period, the identified results do not have a significant trend. The reason why the first four stress periods are so important to the identified results is that the pollution source releases pollutant for those first four periods. We can infer that the pollutant concentrations in the period when the pollution source is released need to be gathered totally if high accuracy of the identified results is desired. The concentration data found in the period when the pollution source has finished releasing do not need to be collected totally because they have little influence on the estimated results.

The trend of the red line is consistent with the blue line. The results also indicate that the pollutant concentrations need to be gathered totally when the source releases the pollutant. Comparing the red line with the blue line, we can draw two conclusions. Firstly, for the first four stress periods, the NE of the identified results from the missing data is lower than that of the interpolated data. This means

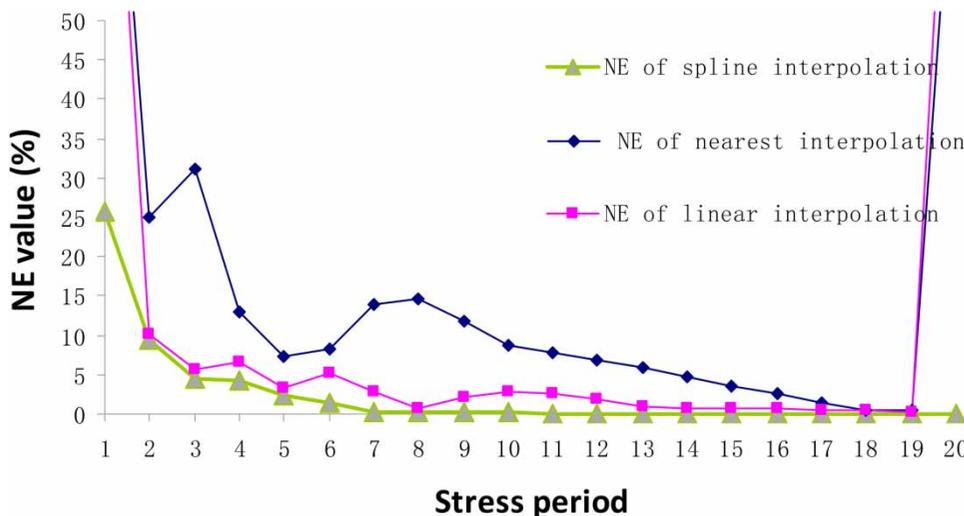


Figure 5 | Comparison of the performance of three different interpolation methods.

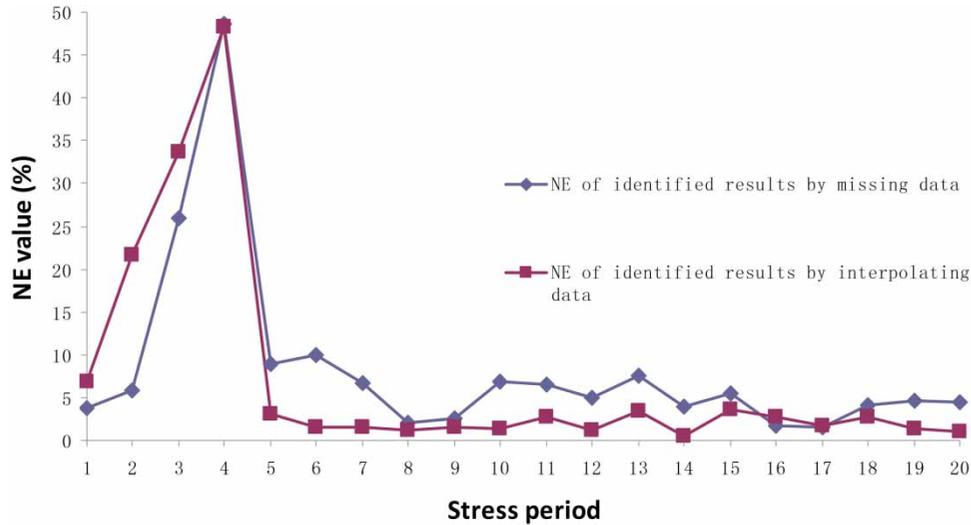


Figure 6 | Comparison of the identified results before and after cubic spline interpolation.

that when the source releases the pollutant, the actual observation data are much more reliable than the interpolated data. However, if the observation data are missing in the stress period after the end of the release time, the identified result from the interpolation data is more reliable than the identified result using just missing data.

CONCLUSIONS

This paper illustrates the difference between using observation data before and after denoising and interpolating for solving GPSI problems. The purpose of this research is to explore whether the identified results are more accurate after denoising and interpolating.

The results show that denoising has an effect on the estimated result for the GPSI problem, and the higher the noise level, the stronger this effect. In terms of interpolation, more accurate results can be made after interpolating if the missing data belong to the period after the source releases pollutant. If the missing data are from when the pollution source is active, interpolating does not increase the estimated performance.

Although denoising and interpolating can improve the accuracy of the identified results to a certain extent, the accurate actual concentration data are the most reliable

information to use. For real-world cases, the more accurate and complete the concentration data collected, the better the identified results could be.

ACKNOWLEDGEMENTS

This study is based on the work supported by the National Natural Science Foundation of China (41807196, 41672232), China's Post-doctoral Science Fund (2018M641793). The authors thank the editor and anonymous reviewers for their constructive comments and suggested revisions.

REFERENCES

- Afshari, H. H., Gadsden, S. A. & Habibi, S. 2017 [Gaussian filters for parameter and state estimation: a general review of theory and recent trends](#). *Signal Processing* **135**, 218–238.
- Atmadja, J. & Bagtzoglou, A. C. 2001 [State of the art report on mathematical methods for groundwater pollution source identification](#). *Environmental Forensics* **2**, 205–214.
- Ayvaz, M. T. 2010 [A linked simulation–optimization model for solving the unknown groundwater pollution source identification problems](#). *Journal of Contaminant Hydrology* **117** (1–4), 46–59.
- Bagtzoglou, A. C. & Atmadja, J. 2005 [Mathematical methods for hydrologic inversion: the case of pollution source identification](#). In: *Environmental Impact Assessment of*

- Recycled Wastes on Surface and Ground Waters: Vol. 3. The Handbook of Environmental Chemistry, Water Pollution Series (vol. 5, Part F)* (T. A. Kassim, series ed.). Springer-Verlag, Heidelberg, Germany, pp. 65–96.
- Cao, W. D., Chen, X. Y., Yang, X. R. & Wang, E. 2003 Discrete wavelets transform for signal denoising in capillary electrophoresis with electrochemiluminescence detection. *Electrophoresis* **24**, 3124–3130.
- Datta, B., Chakrabarty, D. & Dhar, A. 2011 Identification of unknown groundwater pollution sources using classical optimization with linked simulation. *Journal of Hydro-Environment Research* **5** (1), 25–36.
- Donoho, D. L. & Johnstone, I. M. 1994 Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81**, 425–455.
- Hsu, L., Self, S. G., Grove, D., Randolph, T., Wang, K., Delrow, J. J., Loo, L. & Porter, P. 2005 Denoising array-based comparative genomic hybridization data using wavelets. *Biostatistics* **6** (2), 211–226.
- Janecki, D. 2011 Gaussian filters with profile extrapolation. *Precision Engineering* **35** (4), 602–606.
- Lakshminarayan, K., Harp, S. A. & Samad, T. 1999 Imputation of missing data in industrial databases. *Applied Intelligence* **11**, 259–275.
- Lucass, T. R. 1974 Error bounds for interpolating cubic splines under various end conditions. *SIAM Journal of Numerical Analysis* **11** (3), 569–584.
- Mahar, P. S. & Datta, B. 2001 Optimal identification of groundwater pollution sources and parameter estimation. *Journal of Water Resources Planning and Management* **127** (1), 20–29.
- Michalak, A. M. & Kitanidis, P. K. 2004 Estimation of historical groundwater contaminant distribution using the adjoint state method applied to geostatistical inverse modeling. *Water Resources Research* **40** (8), W08302.
- Milnes, E. & Perrochet, P. 2007 Simultaneous identification of a single pollution point-source location and contamination time under known flow field conditions. *Advances in Water Resources* **30** (12), 2439–2446.
- Mirghani, B. Y., Zechman, E. M., Ranjithan, R. S. & Mahinthakumar, G. 2012 Enhanced simulation-optimization approach using surrogate modeling for solving inverse problems. *Environmental Forensics* **13** (4), 348–363.
- Percival, D. B. & Walden, A. T. 2000 *Wavelet Methods for Time Series Analysis*. Cambridge University Press, Cambridge, UK.
- Pinder, G. F. & Bredehoeft, J. D. 1968 Application of the digital computer for aquifer evaluations. *Water Resources Research* **4** (5), 1069–1093.
- Singh, R. M. & Datta, B. 2006 Identification of groundwater pollution sources using GA-based linked simulation optimization model. *Journal of Hydrologic Engineering* **11** (2), 101–109.
- Singh, R. M., Datta, B. & Jain, A. 2004 Identification of unknown groundwater pollution sources using artificial neural networks. *Journal of Water Resources Planning and Management* **130** (6), 506–514.
- Skaggs, T. H. & Kabala, Z. J. 1994 Recovering the release history of a groundwater contaminant. *Water Resources Research* **30** (1), 71–79.
- Woodrow Sterling, et al. versus Velsicol Chemical Corporation 1986 U.S. District Court for the Western District of Tennessee, 647 Fed. Suppl., 303.
- Yeh, W. G. 2015 Review: optimization methods for groundwater modeling and management. *Hydrogeology Journal* **23**, 1051–1065.
- Zhao, Y., Lu, W. X. & An, Y. K. 2015 Surrogate model-based simulation-optimization approach for groundwater source identification problems. *Environmental Forensics* **16** (3), 296–303.
- Zhao, Y., Lu, W. X. & Xiao, C. N. 2016 A Kriging surrogate model coupled in simulation-optimization approach for identifying release history of groundwater sources. *Journal of Contaminant Hydrology* **185–186**, 51–60.

First received 17 August 2018; accepted in revised form 4 January 2019. Available online 22 January 2019