Assessing Interpolation Methods for Accuracy of Design Groundwater Levels for Civil Projects

Tara Zirakbasha\*, Ryan Admiraalb, Anastasia Boroninac, Martin Andad, Parisa A. Bahrie

a Ph.D. Candidate, College of Science, Health, Engineering and Education, Murdoch University, 90 South Street, Perth, 6150, Western Australia

b Doctor, Senior Lecturer, School of Mathematics and Statistics, Victoria University of Wellington, Kelburn, Wellington 6012, New Zealand, Ryan.Admiraal@vuw.ac.nz

c Doctor, Hydrogeologist, Neva Groundwater Consulting, 16 Newton Street, Bayswater 6053, Western Australia, aboronina@hotmail.com

d Doctor, Academic Chair Environmental Engineering, College of Science, Health, Engineering and Education, Murdoch University, 90 South Street, Perth, 6150, Western Australia, M.Anda@murdoch.edu.au

e Professor, Discipline of Engineering and Energy, College of Science, Health, Engineering and Education, Murdoch University, 90 South Street, Perth, 6150, Western Australia, E-mail: [P.Bahri@murdoch.edu.au](mailto:P.Bahri@murdoch.edu.au)

\*Corresponding author: E-mail: T.Zirakbash@murdoch.edu.au, Tel. +61 8 9360 2172

Keywords: interpolation methods, spatial distribution, prediction accuracy

# Abstract

This research provides a comparison of natural neighbour interpolation with other interpolation methods commonly implemented in ArcGIS. It also evaluates the relative performance of interpolation methods for various spatial data distributions, including line transects. Further, it characterizes locations which are associated with large prediction errors.

To assess the relative performance of interpolation methods, a validation procedure was used consisting of 75% training data and 25% test data. Statistical error measures were used to measure the predictive performance of the interpolation methods, and the spatial distribution of errors was used to characterize areas where interpolation methods performed poorly.

Results showed that Topo to Raster, natural neighbour, ordinary kriging, and empirical Bayesian kriging methods consistently outperformed other interpolation methods for a variety of spatial distributions of the data. However, natural neighbour interpolation was unsuitable for linear transects. In general, the accuracy of most of the interpolation methods increased with narrow spatial data distributions. Also, spatial distribution of large prediction errors was predominantly similar, regardless of the interpolation method used, and was related to changes in physical characteristics.

# Introduction

Urban development can impact groundwater level, and, in turn, groundwater level variations may present major risks to infrastructure and environment and cause economic losses. For example, variation in groundwater levels caused by urbanization can lead to subsidence, inundation of road surfaces, or drying surface water bodies as well as create favourable conditions for insect breeding. Groundwater level planning and design in urban development depend on maximum groundwater levels and no standardized method currently exists for accurately determining maximum groundwater levels (Boronina et al. 2015; Fürst et al. 2015). Maximum groundwater levels are maximum levels expected to occur at a specified period of time in an area. Evaluation of the spatial variability of groundwater levels for any urban planning and design is usually achieved by interpolation of groundwater level point data.

Interpolation of point measurement values is used across many different disciplines and applications, including groundwater level estimation, for understanding the spatial variation of a specific variable. A number of studies have been carried out to understand the most appropriate interpolation methods (Adhikary and Dash 2017; Arun 2013; Azpurua and Ramos 2010; da Silva et al. 2018; de Amorim Borges et al. 2016; Elumalai et al. 2017; Hsu et al. 2017; Journel and Huijbregts 1978; Kamińska and Grzywna 2014; Kumar 2007; Mair and Fares 2010; Nikroo et al. 2010; Ohmer et al. 2017; Salekin et al. 2018; Varouchakis and Hristopulos 2013b; Xie et al. 2011). There have also been several broad reviews conducted on interpolation methods which emphasize the importance of the selection of an appropriate interpolation method (Burrough et al. 2015; Caruso and Quarta 1998; Lam 1983; Li and Heap 2008; Li and Heap 2011, 2014; Ly et al. 2013). Previous studies highlight that various factors affect the performance of interpolation methods, including sample density, data variation, sample size, sampling design, distribution of the data, data quality, and the resolution used (Bater and Coops 2009; Burrough et al. 2015; Li and Heap 2011, 2014; Ly et al. 2013; Simpson and Wu 2014). Different studies have found different interpolation methods to be most suitable, even for assessing the same type of variable (e.g., groundwater levels) (Adhikary and Dash 2017; Arslan 2014; Kamińska and Grzywna 2014; Kumar 2007; Mirzaei and Sakizadeh 2016; Rusu and Rusu 2006; Sun et al. 2009; Yao et al. 2014), and results vary significantly according to interpolation method (Elumalai et al. 2017; Ohmer et al. 2017). This highlights the importance of the careful selection of interpolation method based on the purpose of the study and nature of the available data in order to improve the accuracy of interpolation, leading to better management strategies (Jie et al. 2013; Li and Heap 2014; Ohmer et al. 2017).

There are a number of software packages that provide functionality for interpolating spatial data (Li and Heap 2014). However, geographic information systems (GIS) provides a powerful tool that enables different interpolation techniques to be applied. Hence, ArcGIS is commonly used for interpolation in various studies and is widely used by groundwater specialists (Arun 2013; Bhunia et al. 2018; Elumalai et al. 2017; Mirzaei and Sakizadeh 2016; Nikroo et al. 2010; Ohmer et al. 2017; Simpson and Wu 2014; Szypuła 2016; Wijeratne and Manawadu 2016).

One method developed by Boronina et al. (2015) for estimating maximum groundwater level for a road design project in Western Australia (WA) is called design groundwater levels (DGWL). The DGWL method requires a number of decisions to estimate maximum groundwater levels at different measurement points. This method incorporates various types of data, such as groundwater levels in open drains or surface water bodies. A time period of groundwater elevation data is considered and a year in the selection period that corresponds to a reasonable maximum is selected, with consideration of changes in groundwater levels owing to the construction of drainage. At all locations, maximum groundwater level for the same year is estimated. Using these estimates of maximum groundwater levels at the measurement points, interpolation of groundwater elevation can be carried out using NaN interpolation method.

The DGWL method was found to be more efficient than two other methods in WA for estimating maximum groundwater levels (Zirakbash et al. 2018). For interpolating maximum groundwater levels using the DGWL method, Boronina et al. (2015) recommended natural neighbour interpolation (NaN), as it preserves the defined data value for the cell where the monitoring boreholes are located. However, this technique has not been frequently used in environmental and water resource assessments, Boronina et al. (2015) being a rare exception in their application of NaN with the DGWL method, and NaN is generally not suggested for interpolation of groundwater level elevation (GISResources 2014; Li and Heap 2011; Ohmer et al. 2017). The advantages and disadvantages of the NaN interpolation method have been discussed in the literature in which the advantages are related to its exact and local features as well as creation of a continuous surface everywhere except at the reference points (Dumitru et al. 2013; Sambridge et al. 1995). Also, the studies highlighted that the NaN method is able to handle highly irregular distributions of data points with various scale-length, however its function is limited to the area bounded by the set of data points (Dumitru et al. 2013; Sambridge et al. 1995).

Groundwater levels are usually monitored in a short distance on both sides of structures for road or infrastructure designs, so the distribution of the groundwater level monitoring network is spatially narrow and is distributed along the road or infrastructure. NaN interpolation is claimed to allow for the development of accurate surface models from data sets that are linear in spatial distribution (GISResources 2014). However, there has been minimal effort to assess the impact of the spatial distribution of data on the performance of NaN interpolation. Although in many cases the spatial distribution of the groundwater monitoring network is heterogeneous in the study area (Izady et al. 2017; Nikroo et al. 2010; Njeban 2018; Ohmer et al. 2017; Sun et al. 2009), few studies have assessed the impact of heterogeneity in the spatial data distribution on interpolation accuracy. One such study by Nikroo et al. (2010) examined the effects of heterogeneity on interpolation accuracy, specifically considering kriging interpolation methods for groundwater data following an elliptical ring distribution. They found that residual kriging using a second-order function to eliminate groundwater elevation trends produced the lowest root mean squared error (RMSE) in prediction. To the authors' knowledge, an assessment of relative performance of different interpolation methods, including deterministic and geostatistical methods, for a narrow spatial distribution of data has not been previously considered.

Building on the results of Boronina et al. (2015) the overarching objective of the research presented in this paper is to assess the accuracy of NaN in comparison with other interpolation methods integrated into ArcGIS for groundwater level analysis. Additionally, the relative performance of the interpolation methods is assessed considering various spatial distributions of data including three narrow spatial patterns following a road structure. Also, where interpolation methods perform poorly, key features of these locations that might explain the reason for poor estimates of groundwater levels are characterized.

# Study Area and Data Description

## 2.1. *Study Area*

The study site is related to Gateway WA, the largest transport infrastructure project ever undertaken by Main Roads WA (Fig. 1 [right]). The main interchange required a three-level interchange, and a portion of the interchange had to be constructed below ground level. The area is located on a flat, low-lying, often swampy plain inundated with high seasonal groundwater levels with no subsoil drains allowed in the area. The cost of the project was highly sensitive to maximum groundwater levels. Consequently, it was crucial to accurately determine maximum groundwater levels in the study area, and DGWL (using NaN interpolation) was used to estimate the spatial changes of groundwater levels in the study area (Boronina et al. 2015).

The study area is located on the Swan Coastal Plain in the Perth Basin, which contains a multilayered unconfined aquifer and covers an area of 185 km2 (Fig. 1). The superficial aquifer is comprised of “Superficial formations”, an informal name used within WA for late Tertiary to Quaternary sediments. The dominant geological units contained within the study area are composed of Guildford Clay, Bassendean Sand, and the Tamala Limestone (Fig. 1. [left]), in order of deposition. Within the study area, Bassendean Sand, which is largely comprised of quartz sand, overlies the Guildford Clay, which consists of low permeability material and has perched water table (Davidson 1995).

The site is physically bounded by the Darling Scarp to the east, Helena River to the northeast, Swan River to the north and west, and Canning River to the south. The regional groundwater flow direction is from the Darling Scarp in the east towards the Swan River in the west.

Rainfall data collected from the Perth Airport station inside the study area shows a downward trend over the last few decades from 900mm in 1945 to below 700mm in 2016.

## 2.2. *Data* *Description*

For the DGWL method, all available data types, consisting of water levels in wetlands and permanent natural surface water bodies, open channel drains, compensating basins, boreholes with short-term and long-term (>30 years) records as well as ground elevation data and aerial photographs, were used to produce estimates of maximum groundwater levels at a number of different locations of the previously described study site (Boronina et al. 2015). The maximum groundwater level elevation point data calculated by the DGWL method for the Gateway project was provided by Main Roads WA for this study, and frequencies of data points by data type are as presented in Table 1.

# Methodology

## 3.1. *Interpolation Methods*

In this study, commonly used interpolation methods integrated into ArcGIS Spatial Analyst and Geostatistical Analyst tools are compared using the default ArcGIS input parameters for a given interpolation method.

Given that Boronina et al. (2015) recommended NaN over other interpolation methods for maximum groundwater level assessment, this study particularly focus on the performance of NaN relative to other interpolation methods. A brief description of the interpolation methods used in this study is provided in Table 2.

Estimation based on nearly all spatial interpolation methods can be represented as a weighted average of sampled data (Li and Heap 2008). The interpolation methods can be divided into two categories—deterministic (or non-geostatistical) and geostatistical. Deterministic interpolation methods are based on the values of sample points and proximity to the point of interest (e.g., IDW, NaN) or on mathematical formulas (e.g., RBF, LPI) that determine the shape of the predicted surface (Johnston et al. 2001). Geostatistical methods (e.g., kriging), on the other hand, are based on statistical models that can incorporate autocorrelation/spatial dependence between measurements, allowing them to not only produce a continuous prediction surface but also provide an accompanying measure of variability associated with the prediction (Johnston et al. 2001; Theodossiou and Latinopoulos 2006).

## 3.2. *Validation and Measurement Error*

To assess the relative performance of interpolation methods, it is important that predictions are compared against actual values for some set of measurements not used in producing the prediction surface. This process of validation provides a measure of the level of error in predictions associated with each interpolation method and involves partitioning the dataset into a training set for developing the prediction surface and test set for comparing true values against predicted values in order to assess the model performance. For the purpose of this research, this validation process also helps to assess the location and magnitude of error for the predicted points. A number of measures of error, listed in Table 3, can be used to assess predictive performance to inform which model provides the most accurate prediction. Most commonly, root mean squared error or mean absolute error are used, but results are presented based on all presented forms of measurement error for reference.

To assess the relative performance of NaN against other interpolation methods for various measures of prediction error, a Friedman rank-sum test was used to test for a global difference across all interpolation methods (Friedman 1937, 1939). (A repeated measures analysis of variance was found to be inappropriate, even after exploring transformations of the data.) Where this test was statistically significant (p-value < 0.05), post-hoc Wilcoxon signed-rank tests with Bonferroni-corrected p-values were carried out for all comparisons of NaN with other interpolation methods (Dunn 1961; Wilcoxon 1945).

## 3.3. *Interpolation Workflow for total data points*

Maximum groundwater level data are used to compare different existing interpolation methods in ArcGIS. In order to minimise the effect of input parameters on the quality of interpolation results, a common processing extent and cell size of 30m were used for all interpolation methods. The following workflow, illustrated in Fig. 2, was used for the validation routine for estimating measurement error for the interpolation methods:

* Maximum groundwater level data were randomly split into disjoint subsets of 75 percent training and 25 percent test data.
* Training data were used for each interpolation method in Table 2 to predict maximum groundwater levels for the test data.
* The predicted maximum groundwater levels produced by each interpolation method were compared with observed maximum groundwater levels for the test dataset to produce an error measurement.
* The error measurements (presented in Table 3) were used to evaluate the relative performance of the various interpolation methods.
* Ten replicates of the validation routine were carried out to produce ten estimates of measurement error corresponding to each interpolation method.

## 3.4. *Narrow spatial data distribution analysis*

In order to assess how the performance of interpolation methods changes based on the narrow spatial distribution of the data, three narrow spatial patterns of data were considered where only groundwater level measurement points along a road were kept. This resulted in limiting the range of input values and the extent of groundwater level change. Using this reduced dataset, data were further divided and analysed for three different spatial narrow patterns to assess whether or not different patterns affect the accuracy of predictions produced by the various interpolation methods and locations of large error in prediction. The three patterns are depicted in Fig. 3 and include:

* Pattern A: all data points along the road,
* Pattern B: up-gradient points, and
* Pattern C: two-flow directions with a common turning point.

The validation routine outlined in Fig. 2 was again followed, with the same training and test sets used for the three spatial patterns to compare the performance of the interpolation methods for the different spatial patterns.

# Results and Discussion

## 4.1. *Comparison of interpolation methods using all data points*

Validation results of interpolation methods across the 10 replicates for the full dataset are depicted in Fig. 4 (a-d), and a summary of the results is provided in Table 4. Friedman’s test applied to any of the measures of prediction error presented in Table 2 produces a highly statistically significant result (minimum test statistic of χ2 = 73.4 on 9 degrees of freedom, p-value < 5e-12), and post-hoc Wilcoxon signed-rank tests with a Bonferroni correction to *p*-values suggest that T2R, NaN, OK, and EBK perform similarly and outperform IDW, RBF, LPI, and UK, which perform similarly and in turn outperform SI, which outperforms TSA. This order of model performance was highly consistent across the ten replicates of the validation routine with only minor changes in the order of models for a handful of replicates.

As this research is particularly aimed at comparing the performance of NaN relative to other methods, a comparison of percentage changes in MAE and RMSE is provided for the various interpolation methods with respect to NaN in Table 5 and Fig. 5. Mean percentage changes in MAE and RMSE for T2R, OK, and EBK relative to NaN are not statistically significant (Bonferroni-corrected Wilcoxon signed-rank test *p*-values of 1.00) with only T2R producing slightly better (again, not statistically significant) empirical results with an estimated 2.2% reduction in MAE and 1.3% reduction in RMSE relative to NaN. Mean percentage changes in MAE and RMSE for all other methods relative to NAN are significantly greater than 0 (*p*-values < 0.05). The TSA method, a global and inexact method, is the worst performing interpolation method compared with other methods. Global methods use all available data for estimation, whereas local methods operate within a small area around the point of interest (Burrough et al. 2015). Also, exact methods generate a same value as the observed value at a sampled point and all other methods are inexact (Burrough et al. 2015).

Regarding the NaN interpolation method, due to the nature of how weights for neighbours are calculated based on triangulation (Theissen polygons), predictions cannot be expanded beyond the spatial distribution of the input data sets and additionally cannot be produced at the boundary of the input dataset. The predicted surface by NaN interpolation method compared with another method(T2R) is presented in Fig 6. (c and d) as examples.

Predictions of maximum groundwater level should be within a certain accepted tolerance of the truth with the appropriate tolerance varying based on the application. In this case, it may additionally be of interest to know what percentage of predictions fall within this acceptable tolerance. For this groundwater level data, a tolerance of 0.5m is considered. The percentage of predictions for the test set that fall within this tolerance is calculated for each repetition of the validation routine and each interpolation method. These results are shown in Table 5 and Fig. 5(c), again with Wilcoxon signed-rank tests comparing the various interpolation methods to NaN. Again, TSA performs miserably according to this measure of predictive of performance, but the difference between the other interpolation methods is less pronounced, and none of them is significantly different than NaN (even if empirical results for T2R, NaN, OK, and EBK are slightly better than the rest). Interestingly, SI performs much more credibly according to this criterion than against previous measures of error (MAE, RMSE, R2). This is due to predictions within tolerance being closer to the 0.5m tolerance and predictions outside of the tolerance of 0.5m being significantly worse, relative to other interpolation methods. The error histograms of the T2R and NaN models in one of the replicates are presented as the examples (Fig. 6 [a,b]).

For T2R, NaN, OK, and EBK range of errors resulting from all ten replicates is small which shows their proper accuracy despite the various spatial data distribution. However, although the mean of MAE for these methods is less than 0.5m (Table 4), in some replicates, the magnitude of the difference between predicted and observed values is more than 3.5m in several locations (Table 8), which are particularly undesirable for designing urban structures and bring about safety risks and economic loss.

A notable outcome is that, although the spatial distribution of the training datasets differs in each replicate of the validation routine, generally the location of large prediction errors is the same in all replicates and for all interpolation methods. This would suggest that the location of large errors is not due to the type of interpolation method and deficiencies of certain interpolation methods in certain situations but rather some artefact of the data or study area, such as key physical characteristics of the area. For instance, large prediction errors occur near the physical boundary of the study area and input data points as well as where there are large fluctuations in groundwater levels within a short distance for the input data. This is consistent with the findings of Xie (2011), who found that the uncertainty in contamination prediction by interpolation methods located in local maxima and local minima regions as well as the boundary of the contaminated area. Large prediction errors predominantly occur as a result of changes in physical properties of locations. Topography, changes in lithology with subsequent variation of hydraulic properties in addition to changes caused by human interventions such as urban structures like major roads, bridges and especially drains are among physical properties that can cause large prediction errors. This is consistent with what is seen in the upper right region of the right panel of Fig. 1, where changes in topography and lithology cause a large difference in groundwater elevation in a short distance, and this region tends to have larger prediction errors (Fig. 7-9).

## 4.2. *Comparison of interpolation methods with narrow spatial data distribution*

To compare the performance of the various interpolation methods for spatial data following clearly defined systematic trends, MAE, RMSE and R2 were again used to assess the predictive performance of the interpolation methods. Table 6 shows MAE, MSE, and R2 for a single run of the validation routine for each of the interpolation methods for various narrow spatial patterns. Empirical results are similar to those for the full data set with T2R, NaN, OK and EBK resulting in smaller MAE and RMSE relative to other interpolation methods and TSA producing the highest prediction error.

As noted previously, one weakness of the NaN method is its lack of ability to produce predictions on the boundaries and beyond the extent of input data points. This weakness is especially pronounced where input data follow a narrow spatial distribution, as predictions cannot be produced for many locations, meaning that estimates for MAE, RMSE, and R2 are based on fewer predictions and, consequently, are subject to higher levels of variability. This would likely explain why rankings of T2R, NaN, OK, and EBK based on MAE and RMSE would suggest that NaN performs the worst of the four (other than for Pattern B, for which it is second worst) and roughly on par with SI. By contrast, rankings based on empirical results using the full dataset saw NaN being outperformed only by T2R. This deficiency of NaN relative to OK, T2R, and EBK is particularly obvious in Pattern C where empirical RMSE for NaN is more than 30% higher than that of OK, T2R, and EBK (Table 7). This would suggest that, although NaN is still among the best performing interpolation methods in terms of prediction accuracy, it is less suitable when the spatial distribution of data is narrow.

Again, additionally, the frequency of groundwater level predictions within a 0.5m tolerance of the true groundwater level was considered. Empirical results show T2R, NaN, and SI producing the greatest proportion of predictions within this tolerance except for Pattern C, where NaN and SI perform worse. In spite of this worse relative performance, predictions within tolerance increase for NaN and SI for Pattern C relative to Patterns A and B. This is consistent with what is observed in terms of MAE and RMSE for Pattern C relative to Patterns A and B, where MAE and RMSE decrease for this spatial distribution. This improvement in predictive performance for Pattern C relative to Patterns A and B is noticeable and much more pronounced for most other interpolation methods, including T2R, OK, EBK, IDW, LPI, UK, and TSA. The range of errors in these methods is smaller for Pattern C as well (Fig. 11), and the improved predictive performance could be due to more gradual changes in physical properties of the area in this pattern compared to changes in physical properties for Patterns A and B. Examining the locations of large prediction errors for data distributed according to Patterns A, B, and C, these are largely consistent with those based on the full dataset. This could be a consequence of variation in physical properties which cannot be accounted for by the interpolation methods.

## 4.3. *Relative performance of interpolation methods using all available data and data following specific spatial distribution patterns*

Regardless of whether all available data or only data following a specific spatial distribution are used, T2R, NaN, OK, and EBK seem to perform best in terms of reducing MAE and RMSE and also perform similarly in terms of producing predictions within a specified tolerance. Results based on the full dataset suggest no significant difference between these four interpolation methods but provide evidence of them performing better than other interpolation methods. For data following specific spatial distribution patterns, it appears that the accuracy of the NaN method is reduced, possibly due to the inability of the method to produce predictions on the boundaries or outside of geographic space of the input data.

Comparing the empirical results for RMSE for the three spatial distribution patterns with the distribution of RMSE based on the full dataset, empirical RMSE was no worse or decreased for data following the three spatial distributions as compared to the distribution of RMSE for the full dataset for all interpolation methods other than IDW and RBF (Fig. 11). Also, the range of error and similarly the value of maximum and minimum errors of all methods is declined in narrow spatial data distributions. It could result from more data density relative to the area of investigation and a smaller range of input value in assessing narrow spatial data distribution. For MAE, on the other hand, Patterns A and B frequently resulted in a higher empirical MAE relative to the distribution of MAE for the full dataset for not only IDW and RBF but also other interpolation methods (T2R, NaN, OK, UK). As is observed for RMSE, Pattern C seems to result in a significant drop in MAE almost regardless of interpolation method.

The proportion of predictions within a 0.5m tolerance for the best performing interpolation methods (T2R, NaN, OK, and EBK) have been assessed for all spatial distribution patterns and for the full dataset (Table 5 and Table 7). Better prediction accuracy appears to be related to a lack of data points in regions with different hydraulic characteristics where large prediction errors were observed to occur. Existence of large errors is predominantly related to the nature of initial data where a large difference in groundwater level happens within a short horizontal distance, typically the resulted of either significant natural physical changes in the landscape or by human interventions (right panel of Fig. 1).

For example, there are large prediction errors in the southeast (Fig. 10), and the lack of points from this region for spatial Pattern C is a likely reason why prediction error is lower for this spatial distribution as compared to the other two spatial patterns which contain data points from this region.

# Conclusion

The accuracy of ten interpolation methods for a variety of spatial data distributions, including three narrow patterns, were compared using groundwater elevation data from a road project in Western Australia. It suggested higher prediction accuracy for T2R, NaN, EBK and OK interpolation methods (all of which are exact and local methods) compared with other methods. The worst performing interpolation method was TSA, which is a global and inexact method.

The NaN method performed well relative to other interpolation methods, regardless of spatial distribution, but its inability to predict values on the boundary or beyond the extent of input data meant that its performance relative to other interpolation methods decreased for data distributed according to specific narrow patterns. That being said, there does not appear to be a significant difference in the relative performance of interpolation methods based on whether the full dataset or only data distributed according to specific spatial patterns are used. In general, for most interpolation methods prediction accuracy seems to improve when applying interpolation methods to data following specific spatial patterns. This improvement in predictions could be a result of greater sampling density and smoother change in groundwater level and less heterogeneity as a result of reducing the extent of the area.

One of the objectives of this study was to characterize locations with large prediction error. These places are predominantly in regions:

• where large differences in groundwater levels occur in a short horizontal distance;

• near the boundary of the study area;

• where there are key changes in in physical characteristics such as lithology, hydraulic gradient and topography;

• or where human interventions cause major changes in natural hydraulic properties, such as drainage and artificial recharge points.

The accuracy of prediction is largely controlled by the physical properties of aquifer and their variability as well as hydrodynamic boundaries. These factors need to be addressed by a sufficient density of groundwater monitoring and visual assessments of groundwater level contours in the aforementioned types of locations. This will be important especially for major engineering projects, that failure in proper prediction might result in risks for human and the environment or economic losses by conservatism in design or time overrun of projects.

# Data Availability

Some or all data, models, or code used during the study were provided by a third party. Direct requests for these materials may be made to the provider as indicated in the Acknowledgements.

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

# Acknowledgement

The Authors acknowledge the Government of Western Australia, Department of Water and Environmental Regulation and Main Roads WA for providing data. T. Zirakbash gratefully acknowledges Australian Government for Australian Commonwealth Government Research Training Program (RTP) Scholarship.

# Notation List

n= number of observations or samples

Pi = Predicted value

Oi = Observed(measured) value

Pave = mean of predicted value

Oave = mean of observed value

χ2 = chi-square

References

Adhikary, P. P., and Dash, C. J. (2017). "Comparison of deterministic and stochastic methods to predict spatial variation of groundwater depth." *Applied Water Science*, 7(1), 339-348.

Arslan, H. (2014). "Estimation of spatial distrubition of groundwater level and risky areas of seawater intrusion on the coastal region in Çarşamba Plain, Turkey, using different interpolation methods." *Environmental monitoring and assessment*, 186(8), 5123-5134.

Arun, P. V. (2013). "A comparative analysis of different DEM interpolation methods." *The Egyptian Journal of Remote Sensing and Space Science*, 16(2), 133-139.

Azpurua, M. A., and Ramos, K. D. (2010). "A comparison of spatial interpolation methods for estimation of average electromagnetic field magnitude." *Progress in electromagnetics research*, 14, 135-145.

Bater, C. W., and Coops, N. C. (2009). "Evaluating error associated with lidar-derived DEM interpolation." *Computers & Geosciences*, 35(2), 289-300.

Bhunia, G. S., Shit, P. K., and Maiti, R. (2018). "Comparison of GIS-based interpolation methods for spatial distribution of soil organic carbon (SOC)." *Journal of the Saudi Society of Agricultural Sciences*, 17(2), 114-126.

Boronina, A., Clayton, R., and Gwynne, C. (2015). "Recommended methodology for determination of design groundwater levels." *Aust. Geomech. J*, 50(3), 11-21.

Burrough, P. A., McDonnell, R., McDonnell, R. A., and Lloyd, C. D. (2015). *Principles of geographical information systems*, Oxford University Press.

Caruso, C., and Quarta, F. (1998). "Interpolation methods comparison." *Computers & Mathematics with Applications*, 35(12), 109-126.

Chen, S., Cowan, C. F., & Grant, P. M. (1991). "Orthogonal least squares learning algorithm for radial basis function networks." *IEEE Transactions on neural networks*, 2(2), 302-309.

da Silva, A. S. A., Stosic, B., Menezes, R. S. C., and Singh, V. P. (2018). "Comparison of Interpolation Methods for Spatial Distribution of Monthly Precipitation in the State of Pernambuco, Brazil." *Journal of Hydrologic Engineering*, 24(3), 04018068.

Davidson, W. A. (1995). *Hydrogeology and groundwater resources of the Perth region, Western Australia*, Geological Survey of WA.

de Amorim Borges, P., Franke, J., da Anunciação, Y. M. T., Weiss, H., and Bernhofer, C. (2016). "Comparison of spatial interpolation methods for the estimation of precipitation distribution in Distrito Federal, Brazil." *Theoretical and applied climatology*, 123(1-2), 335-348.

Dumitru, P. D., Plopeanu, M., and Badea, D. (2013). "Comparative study regarding the methods of interpolation." *Recent advances in geodesy and Geomatics engineering*, 1, 45.

Dunn, O. J. (1961). "Multiple comparisons among means." *Journal of the American statistical association*, 56(293), 52-64.

Elumalai, V., Brindha, K., Sithole, B., and Lakshmanan, E. (2017). "Spatial interpolation methods and geostatistics for mapping groundwater contamination in a coastal area." *Environmental Science and Pollution Research*, 24(12), 11601-11617.

ESRI "What is Empirical Bayesian Kriging?", <http://desktop.arcgis.com/en/arcmap/10.3/guide-books/extensions/geostatistical-analyst/what-is-empirical-bayesian-kriging-.htm>. (19,04,2019).

Friedman, M. (1937). "The use of ranks to avoid the assumption of normality implicit in the analysis of variance." *Journal of the american statistical association*, 32(200), 675-701.

Friedman, M. (1939). "A correction: The use of ranks to avoid the assumption of normality implicit in the analysis of variance." *Journal of the american statistical association*, 34(205), 109.

Fürst, J., Bichler, A., and Konecny, F. (2015). "Regional frequency analysis of extreme groundwater levels." *Groundwater*, 53(3), 414-423.

GISResources (2014). "Types of interpolation methods." <http://www.gisresources.com/types-of-interpolation-methods\_2/>. (19,04,2019).

Hsu, S., Mavrogianni, A., and Hamilton, I. (2017). "Comparing spatial interpolation techniques of local urban temperature for heat-related health risk estimation in a subtropical city." *Procedia Engineering*, 198, 354-365.

Hutchinson, M. (1988). "Calculation of hydrologically sound digital elevation models." *Proc., Proceedings of the Third International Symposium on Spatial Data Handling*, International Geographical Union Sydney.

Hutchinson, M. (1989). "A new procedure for gridding elevation and stream line data with automatic removal of spurious pits." *Journal of hydrology*, 106(3-4), 211-232.

Hutchinson, M. (1996). "A locally adaptive approach to the interpolation of digital elevation models." *Proc., Proceedings, Third International Conference/Workshop on Integrating GIS and Environmental Modeling*, National Center for Geographic Information and Analysis Santa Barbara, CA, 21-26.

Hutchinson, M. (2000). "Optimising the degree of data smoothing for locally adaptive finite element bivariate smoothing splines." *ANZIAM Journal*, 42, 774-796.

Hutchinson, M. (2011). "ANUDEM Version 5.3. Fenner School of Environment and Society, Australian National University, Canberra."

Izady, A., Abdalla, O., Ahmadi, T., and Chen, M. (2017). "An efficient methodology to design optimal groundwater level monitoring network in Al-Buraimi region, Oman." *Arabian Journal of Geosciences*, 10(2), 26.

Jie, C., Hanting, Z., Hui, Q., Jianhua, W., and Xuedi, Z. (2013). "Selecting proper method for groundwater interpolation based on spatial correlation." *Proc., 2013 Fourth International Conference on Digital Manufacturing & Automation (ICDMA)*, IEEE, 1192-1195.

Johnston, k., Ver Hoef, J. M., Krivoruchko, K., and Lucas, N. (2001). "Using ArcGIS Geostatistical Analyst." *GIS by ESRI*.

Journel, A. G., and Huijbregts, C. J. (1978). *Mining geostatistics*, Academic press London.

Kamińska, A., and Grzywna, A. (2014). "Comparison of deteministic interpolation methods for the estimation of groundwater level." *Journal of Ecological Engineering*, 15(4), 55--60.

Kis, I. M. (2016). "Comparison of Ordinary and Universal Kriging interpolation techniques on a depth variable (a case of linear spatial trend), case study of the Sandrovac Field." *Rudarsko-Geolosko-Naftni Zbornik*, 31(2), 41.

Krivoruchko, K. (2012). "Empirical bayesian kriging." *ArcUser Fall*, 2012, 6-10.

Kumar, V. (2007). "Optimal contour mapping of groundwater levels using universal kriging—a case study." *Hydrological Sciences Journal*, 52(5), 1038-1050.

Lam, N. S.-N. (1983). "Spatial interpolation methods: a review." *The American Cartographer*, 10(2), 129-150.

Ledoux, H., and Gold, C. (2005). "An efficient natural neighbour interpolation algorithm for geoscientific modelling." *Developments in spatial data handling*, Springer, 97-108.

Li, J., and Heap, A. (2008). "A review of spatial interpolation methods for environmental scientists geoscience Australia record 2008/23, p 154."

Li, J., and Heap, A. D. (2011). "A review of comparative studies of spatial interpolation methods in environmental sciences: performance and impact factors." *Ecological Informatics*, 6(3-4), 228-241.

Li, J., and Heap, A. D. (2014). "Spatial interpolation methods applied in the environmental sciences: A review." *Environmental Modelling & Software*, 53, 173-189.

Liang, L., and Hale, D. (2010). "A stable and fast implementation of natural neighbor interpolation." *CWP-657. Ctr. for Wave Phenomena, Colorado School of Mines, Golden*.

Ly, S., Charles, C., and Degré, A. (2013). "Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale: a review." *Biotechnologie, Agronomie, Société et Environnement= Biotechnology, Agronomy, Society and Environment [= BASE]*, 17(2), 392-406.

Mair, A., and Fares, A. (2010). "Comparison of rainfall interpolation methods in a mountainous region of a tropical island." *Journal of Hydrologic Engineering*, 16(4), 371-383.

Mirzaei, R., and Sakizadeh, M. (2016). "Comparison of interpolation methods for the estimation of groundwater contamination in Andimeshk-Shush Plain, Southwest of Iran." *Environmental Science and Pollution Research*, 23(3), 2758-2769.

Nikroo, L., Kompani-Zare, M., Sepaskhah, A. R., and Shamsi, S. R. F. (2010). "Groundwater depth and elevation interpolation by kriging methods in Mohr Basin of Fars province in Iran." *Environmental monitoring and assessment*, 166(1-4), 387-407.

Njeban, H. S. (2018). "Comparison and Evaluation of GIS-Based Spatial Interpolation Methods for Estimation Groundwater Level in AL-Salman District—Southwest Iraq." *Journal of Geographic Information System*, 10(04), 362.

Ohmer, M., Liesch, T., Goeppert, N., and Goldscheider, N. (2017). "On the optimal selection of interpolation methods for groundwater contouring: An example of propagation of uncertainty regarding inter-aquifer exchange." *Advances in Water Resources*, 109, 121-132.

Peralvo, M., and Maidment, D. (2004). "Influence of DEM interpolation methods in drainage analysis." *Gis Hydro*, 4.

Pilz, J., and Spöck, G. (2008). "Why do we need and how should we implement Bayesian kriging methods." *Stochastic Environmental Research and Risk Assessment*, 22(5), 621-632.

Rusu, C., and Rusu, V. "Radial basis functions versus geostatistics in spatial interpolations." *Proc., IFIP International Conference on Artificial Intelligence in Theory and Practice*, Springer, 119-128.

Salekin, S., Burgess, J., Morgenroth, J., Mason, E., and Meason, D. (2018). "A Comparative Study of Three Non-Geostatistical Methods for Optimising Digital Elevation Model Interpolation." *ISPRS International Journal of Geo-Information*, 7(8), 300.

Sambridge, M., Braun, J., and McQueen, H. (1995). "Geophysical parametrization and interpolation of irregular data using natural neighbours." *Geophysical Journal International*, 122(3), 837-857.

Šiljeg, A., Lozić, S., and Radoš, D. (2015). "The effect of interpolation methods on the quality of a digital terrain model for geomorphometric analyses." *Tehnicki vjesnik/Technical Gazette*, 22(5).

Simpson, G., and Wu, Y. H. (2014). "Accuracy and effort of interpolation and sampling: can GIS help lower field costs?" *ISPRS International Journal of Geo-Information*, 3(4), 1317-1333.

Sun, Y., Kang, S., Li, F., and Zhang, L. (2009). "Comparison of interpolation methods for depth to groundwater and its temporal and spatial variations in the Minqin oasis of northwest China." *Environmental Modelling & Software*, 24(10), 1163-1170.

Szypuła, B. (2016). "Geomorphometric comparison of DEMs built by different interpolation methods." *Landform Analysis*, 32, 45-58.

Theodossiou, N., and Latinopoulos, P. (2006). "Evaluation and optimisation of groundwater observation networks using the Kriging methodology." *Environmental Modelling & Software*, 21(7), 991-1000.

Varouchakis, E., Hristopulos, D., and Karatzas, G. (2012). "Improving kriging of groundwater level data using nonlinear normalizing transformations—a field application." *Hydrological sciences journal*, 57(7), 1404-1419.

Varouchakis, E. A., and Hristopulos, D. T. (2013a). "Improvement of groundwater level prediction in sparsely gauged basins using physical laws and local geographic features as auxiliary variables." *Advances in water resources*, 52, 34-49.

Varouchakis, Ε., and Hristopulos, D. (2013b). "Comparison of stochastic and deterministic methods for mapping groundwater level spatial variability in sparsely monitored basins." *Environmental monitoring and assessment*, 185(1), 1-19.

Wijeratne, V., and Manawadu, L. (2016). "VALIDATION OF SPATIAL INTERPOLATION TECHNIQUES IN GIS."

Wilcoxon, F. (1945). "Individual Comparisons by Ranking Methods." *Biometrics Bulletin*, 1(6), 80-83.

Xie, Y., Chen, T.-b., Lei, M., Yang, J., Guo, Q.-j., Song, B., and Zhou, X.-y. (2011). "Spatial distribution of soil heavy metal pollution estimated by different interpolation methods: Accuracy and uncertainty analysis." *Chemosphere*, 82(3), 468-476.

Yao, L., Huo, Z., Feng, S., Mao, X., Kang, S., Chen, J., Xu, J., and Steenhuis, T. S. (2014). "Evaluation of spatial interpolation methods for groundwater level in an arid inland oasis, northwest China." *Environmental earth sciences*, 71(4), 1911-1924.

Zirakbash, T., Boronina, A., Anda, M., and Bahri, P. A. (2018). "Maximum groundwater level for urban development: Evaluation of different calculation methods in Western Australia." *Computer Aided Chemical Engineering*, Elsevier, 2533-2538.

Zou, L., Kent, J., Lam, N. S.-N., Cai, H., Qiang, Y., and Li, K. (2016). "Evaluating land subsidence rates and their implications for land loss in the Lower Mississippi River Basin." *Water*, 8(1), 10.