Environmental Science Processes & Impacts

Interpolation of extensive routine water pollution monitoring datasets: methodology and discussion of implications for aquifer management

Journal:	*Environmental Science: Processes & Impacts*
Manuscript ID:	EM-ART-03-2014-000190.R1
Article Type:	Paper
Date Submitted by the Author:	13-May-2014

Complete List of Authors:	--, Yuval; Technion IIT, Civil and Environmental Engineering Rimon, Yaara; Technion IIT, Civil and Environmental Engineering Graber, Ellen; The Volcani Center, Agricultural Research Organisation, Institute of Soil, Water and Environmental Sciences Furman, Alex; Technion IIT, Civil and Environmental Engineering

The methodologies and guidelines provided by the paper enable better use of existing groundwater monitoring data and thus a more efficient utilisation of available aquifer water. Given the increasing demand for fresh water in the world, the understanding that the paper provides is a step forward in water conservation efforts.

Interpolation of extensive routine water pollution monitoring datasets: methodology and discussion of implications for aquifer management

Yuval,*[a], Yaara Rimon[a], Ellen R. Graber[b] and Alex Furman[a]

Received Xth XXXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX
First published on the web Xth XXXXXXXXXX 200X
DOI: 10.1039/b000000x

A large fraction of the fresh water available for human use is stored in groundwater aquifers. Since human activities such as mining, agriculture, industry and urbanisation often result in incursion of various pollutants to groundwater, routine monitoring of water quality is an indispensable component of judicious aquifer management. Unfortunately, groundwater pollution monitoring is expensive and usually cannot cover an aquifer with the spatial resolution necessary for making adequate management decisions. Interpolation of monitoring data is thus an important tool for supplementing monitoring observations. However, interpolating routine groundwater pollution data poses a special problem due to the nature of the observations. The data from a producing aquifer usually includes many zero pollution concentration values from the clean parts of the aquifer but may span a wide range of values (up to a few orders of magnitude) in the polluted areas. This manuscript presents a methodology that can cope with such datasets and use them to produce maps that present the pollution plumes but also delineates the clean areas that are fit for production. A method for assessing the quality of mapping in a way which is suitable to the data's dynamic range of values is also presented. Local variant of inverse distance weighting is employed to interpolate the data. Inclusion zones around the interpolation points ensure that only relevant observations contribute to each interpolated concentration. Using inclusion zones improves the accuracy of the mapping but results in interpolation grid points which are not assigned a value. That inherent trade-off between the interpolation accuracy and coverage is demonstrated using both circular and elliptical inclusion zones. A leave-one-out cross testing is used to assess and compare the performance of the interpolations. The methodology is demonstrated using groundwater pollution monitoring data from the coastal aquifer along the Israeli shoreline. The implications for aquifer management are discussed.

1 Introduction

Groundwater is an essential resource, especially in arid and semi-arid countries. In many cases, the groundwater aquifers are vulnerable to contamination from natural and anthropogenic sources. The vulnerability of the groundwater can be assessed using various means[1–4] however, the large uncertainty in these assessments and the operational requirements of water production call for more accurate estimations of the spatial distribution of pollutants. Probably the most important mechanism for detection of pollutants and protection of the groundwater is a pollution monitoring network[3]. With the increasing demand for fresh water and the dwindling resources, the prevalence of comprehensive aquifer monitoring is rising. However, due to the cost of the monitoring operations, the networks cannot usually cover the aquifers at an adequate spatial resolution. Interpolation of monitoring data is thus used extensively to enhance the information that the raw monitoring data can provide[5–7]. Beside enhancing the spatial coverage, the interpolation of pollution monitoring observations provides many additional benefits. It facilitates a visualisation of the spatial distribution of pollutants, avoiding bias due to variability in the spatial density of the monitoring. Interpolated pollution maps on a regular gird also enable quantification of the fraction of the aquifer area unfit for water production. In addition, using interpolated maps, the spatial distributions of different pollutants can be examined and compared at the same set of locations for the purpose of pollution source allocation and can be used as an input for statistical methods like Principle Component Analysis[8–10] which require data on a regular grid.

Two types of interpolation cases are usually considered in the groundwater literature. One involves plumes of high pollution concentration, usually associated with a known point source[6,11,12] and the other involves monitoring of major ions[13–15]. In both these cases the observed pollution concentration data follow well- behaved statistical distributions, and the geostatistical tools developed over the last decades provide good interpolation solutions. However, many anthropogenic pollutants are very rare in the natural environment and routine observations of their concentrations are expected to be, and in many cases are indeed, of zero concentration[16,17]. In polluted

locations though, the concentrations of anthropogenic pollutants can be very high, in some cases many orders of magnitude beyond the limits set by the standards. Interpolation of such data over a large water producing aquifer is problematic. It is very desirable to delineate the areas from which water production should be permitted, but to do so while indicating at the same time the extent of the polluted areas. Clearly, interpolation between high localised pollution values and zero, or very low values, results in an unrealistic smearing of the pollution beyond the true polluted area and leads to inefficient use of groundwater resources. This paper examines the potential to achieve that goal using simple interpolation schemes which use for the estimation of the interpolated value at each grid point only selected observations within inclusion zones around it. In sparsely monitored areas, and most aquifer areas can be considered as such, this results in grid points at which interpolation is not carried out. The inherent trade-off between the interpolation coverage extent and the quality of the interpolations is discussed and demonstrated. The interpolation testing and comparisons are carried out using a comprehensive cross-testing scheme which is based on a quality measure suitable for the monitoring data's dynamic range of values and for water management purposes.

2 Methods

2.1 Study area

Our work is demonstrated using data from Israel's coastal plain aquifer. The aquifer covers about 1800 km^2 along the coast between the Mt. Carmel ridge in the north to the Sinai peninsula in the south. Its longitudinal extent is 12-20 km, from the shoreline in the west to limestone structures in the east. The aquifer is mostly composed of sand, sandstone and calcareous sandstone (Kurkar) layers on top of a clay base, which is just a few m below the surface in the east to a maximal depth of about 200 m close to the shoreline. The aquifer is dissected by loam and clay lenses, pinching out 2-5 km from the shoreline, which are thought to effectively divide its western part into hydrologically separate sections[16,17]. The total water storage capacity is 20 billion m^3 with a natural average recharge of about 250 Mm^3 (million cubic meters) a year[16,18]. Irrigation excess and artificial recharging contribute another 100-150 Mm^3 a year of water. The water head is between 40-50 m above sea level in the east to a few m above sea level in the west, dictating a general east to west flow (Fig. 1). The flow rate is 3-5 m a year through the aquifer bulk, with faster local flows. Being an unconfined aquifer throughout its entire area and situated below a densely populated, industrialised and agricultural region, the coastal aquifer is under constant threat of pollution from urban sewage drainage, industrial spills and leaching of nitrates, herbicides, pesticides and excess irrigation water (significant proportion of which is recycled effluent water with relatively high salinity levels). In addition, historically the aquifer has suffered significant seawater intrusion in locations of intensive production.

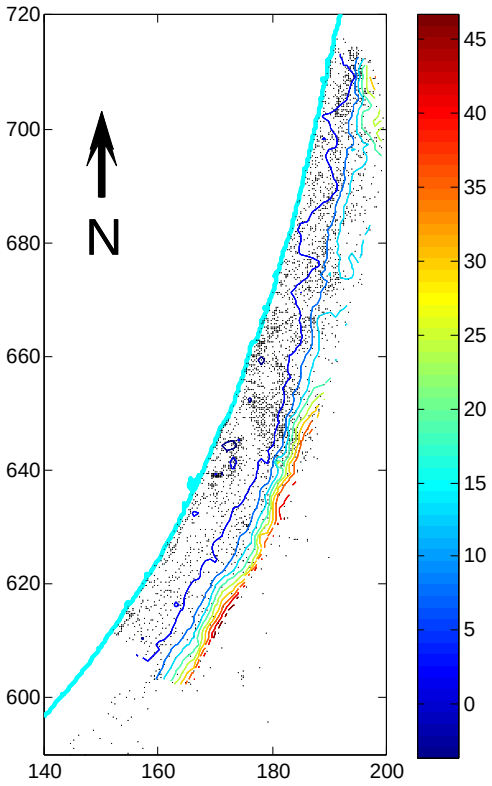

Fig. 1 A map showing the locations of the wells in the aquifer (black dots), and the water head levels (coloured thin continuous lines). The colourbar on the right provides the head levels colour-key in m. The thick cyan line marks the Mediterranean coast. The coordinates are in the New Israeli Grid in km.

2.2 Data

The monitoring network in the coastal aquifer includes about 1050 active production wells and 750 monitoring wells operated by various agencies (Fig 1). A total of 49 pollutants are monitored. This study deals with 31 of the most commonly occurring determinands, which are monitored sufficiently frequently (Table 1). The monitoring frequency per pollutant in a well is between one to six years, depending on the last observed concentration. This adds up to about 150-250 observations per pollutant per year. The monitoring records of 2000-2011 were examined for this study, but due to lower quality of

the data from the early years, records mainly from the period 2006-2011 were used. Over the whole period, the concentration of the least monitored pollutant was observed in 423 wells while the concentration of the most monitored one was observed in 1882 wells (see details in Table 1).

The water samples were analysed by several different laboratories over the years as part of the legal monitoring requirements. All the data are archived in a database of the Israel Water Authority. As part of this study, data quality assurance was carried out. The data were processed such that the different laboratory reporting limits for detection and quantification will not affect subsequent investigations using the data. This entailed setting observations below the quantification threshold to zero. Cases where the quantification threshold record was larger than 20% of the standard environmental threshold were examined individually and processed according to instructions by experienced Israel Water Authority personnel. When in doubt, data were removed from the set. Descriptive statistics of the records of each pollutant were calculated and presented visually. Aberrant values were examined individually and a decision regarding their fate was taken following consultation with the relevant laboratory and Water Authority personnel.

Water head gradients in the aquifer were used to construct the elliptical inclusion zones of the Inverse Distance Weighting (IDW) interpolation (see section 2.3). Water level is routinely measured in the wells. For each well, the median of its water level observations during 2009-2010 was considered. These median values were projected to the regular grid used for the pollution data interpolation using linear Delaunay triangulation method implemented by the Matlab® griddata function[19]. The water head gradient direction and magnitude were then calculated for each point on that regular grid using a finite differences scheme.

2.3 Interpolation methodology

The IDW estimator at a desired interpolation point (x,y) is given by the weighted sum of the data points z_i observed at locations (x_i,y_i)

$$z(x,y) = \sum_{i=1}^{N} \frac{w_i(x_i,y_i)z_i(x_i,y_i)}{\sum_{j=1}^{N} w_j(x_i,y_i)}. \quad (1)$$

The weights $w_i(x_i,y_i)$ are given by

$$w_i(x_i,y_i) = \frac{1}{{d_i}^p}, \quad (2)$$

where d_i is the Euclidean distance between (x,y) and (x_i,y_i) and p is a positive real number. The optimal value of p can be determined by cross-validation using a line search of a range of values. For simplicity this study assumes $p = 2$, which is a very common practice[20]. In a large aquifer, especially whose flow pattern is interrupted by impermeable barriers, the N observations used for the interpolation at a given point should be a small subset of the complete dataset[20]. This subset should include only observations at locations which given the advection and dispersion time scales may be associated with the concentration at the interpolation grid point. These locations can be defined as those residing in an inclusion zone around it.

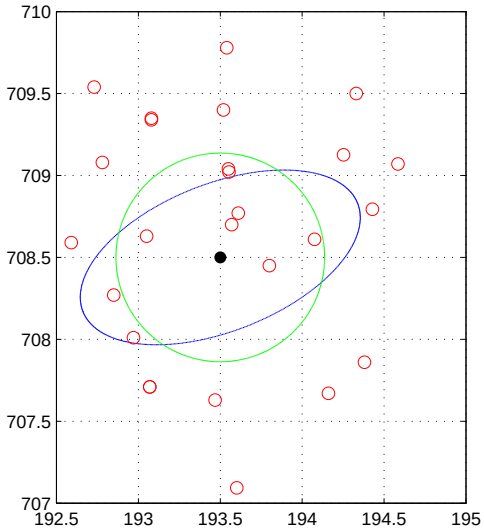

Fig. 2 A diagram showing the two types of inclusion zones (large circle and ellipse) and the well locations (small circles) in the vicinity of grid point (193.5E, 708.5N). The coordinates are in the New Israel Grid in km. The head gradient is in direction 68.5 degrees from the north. The head magnitude is 0.0016 m/m and R was set at 4, resulting in ellipse axes ratio of 2.0023 (see text and Fig. 3). The minor ellipse semi-axis was set at $a = 450$ m. To have a circular zone with area equal to that of the ellipse, the circle's radius is thus $\rho = (a*a*2.0023)^{0.5} = 636.76$ m.

We considered the two types of inclusion zones shown in Fig. 2. The first is a circle with radius ρ, including only data points at a distance shorter than ρ from the interpolation point. The second type of inclusion zone is an ellipse with minor and major semi-axes a and b, respectively. The major axis of the ellipse should be directed along the water advection direction, dictated by the local water head gradient. The ratio between the ellipse axes should manifest the advection and transverse dispersion ratio and was set as a value reciprocally proportional to the logarithm of the magnitude of the head gradient at the interpolation point. This ratio is in the range $[1\ R]$, where R is the maximal ratio between the axes, encountered at the interpolation location where the head gradient is the largest.

As R approaches unity, at locations where the head gradient is the smallest, the ellipse approaches a circle with radius a. Figure 3a shows a graph of the ratio between the ellipse axes as a function of the head gradient magnitude. The value of R is set a-priory based on the transverse dispersion expected in the aquifer. For example, in the coastal aquifer the transverse dispersion is assumed usually to be around 2. Thus in the example in Fig 3a, $R = 4$ for points with the largest (and very rare, see Fig. 3b) head gradient but $R = 2$ for points where the head gradient is 0.001 m/m, close to the statistical mode of the gradient magnitude values.

The choices of ρ and a have an important impact on the interpolations. Small ρ and a increase the chances that only relevant observations will be included in the computation of an interpolated value. However, the smaller the number of observations used to produce an interpolation value, the larger is its statistical uncertainty. When just very few values are used, an aberrant error in any of them will have a large impact on the result. In the limit where ρ or a are so small that no observations are included, the uncertainty is infinite i.e., we know nothing on the value there. The trade-off in the choice of ρ and a is demonstrated in section 3. For comparison we applied also an IDW interpolation where all observations in the aquifer are used in the calculation of each interpolated point. We will refer to this interpolation as infinite inclusion zone interpolation. In the comparison between the circular and elliptical inclusion zone interpolations we always made sure that the total area covered by the inclusion zones around the interpolation points is equal i.e., following the example demonstrated in Fig. 2 at all the interpolation points.

2.4 Assessment of the quality of the interpolation

Any interpolation scheme must be tested in an objective way to assess its performance at producing unobserved values. Given a reasonably large and representative data set, the most comprehensive approach for such an assessment, an approach that was adopted by this study, is the use of a leave-one-out cross-testing scheme[21]. This scheme entails carrying out M interpolations for a data set of M observations. Each of these interpolations uses $M-1$ data points to create the interpolation model, and predict the value of the one left out. The process results in a set of M independently predicted values that can be compared to the M observations.

The assessment of the performance of the interpolation can be carried out using a proper statistical measure. The performance of predicting values with reasonably behaving empirical probability density function is usually tested using the root mean square error (RMSE) or some form of correlation measure. However, our water pollution data exhibit very peculiar density functions, including many zero or very small values, and a very long and thin tail due to a few very large values spanning many orders of concentration magnitude. Figure 4 demonstrates this phenomenon for the pollutant nickel. For many other pollutants the phenomenon is even more extreme and is difficult to demonstrate using intuitive visual means. Using RMSE and correlation as performance measures may be very misleading in such cases as errors in predicting a few of the very large data values might bias the measure in a very disproportional way. Data values transformations may be used to alleviate this problem but then the performance measures will depend on the specifics of the transformation and will thus be hard to assess. Moreover, the success in predicting zero concentration exactly, an important trait for management decisions, may not have a large impact on the traditional performance measures. A more reasonable approach is considering the success at correctly assigning the cross-tested interpolation values into categorical concentrations levels. We use for this purpose the Success Rate (SR,[22]) at predicting pollutant concentrations. The SR is given by

$$SR = \frac{L}{M}, \tag{3}$$

where L is the number of times that the leave-one-out cross-tested interpolated values of the pollutant fall within the same pre-defined concentration brackets as the corresponding observations. Isolated observation points that do not have other observations within their inclusion zone do not have interpolation counterparts. An issue to consider is whether M should be taken as the total number of observations or the number of observations for which interpolations were produced. (Interpolations are not produced for points which have no data observations in their inclusion zone.) We consider and present SR calculations using the latter, thus providing a performance measure of the interpolations that did take place. The issue of locations for which interpolations are not carried out due to lack of observations in their vicinity will be discussed in the context of the areal coverage that the interpolation can provide.

The computational work required for implementing the processes described in this section was carried out using computer codes written in-house by the authors in the Matlab® programming language.[19].

3 Results

Figure 5 demonstrates the mapping properties of three types of interpolation schemes for the sample of maximal observed chromium values in each well during the five year period 2006-2010. The three maps were prepared using the same input data and the same 500x500 m interpolation grid, extending over the parts of the aquifer area covered by wells. The choice of chromium is due to the fact that its maps provide good visual demonstration of the points that will discussed

below. Figure 5a was produced using all the data for each interpolation point (infinite inclusion zone). Figure 5b was produced using circular inclusion zones with a radius such that the total area which they cover around the interpolation grid points equals the area covered by the corresponding ellipses with minor semi-axes of 1000 m used to produce Fig. 5c. The map in Fig. 5a shows a smooth pattern, with no zero values at all, that covers the full interpolation grid. It is very different from the patterns in Figs. 5b-5c which are composed of clusters of very high and very low (including zero) concentration values and which do not cover the full grid area. The differences between 5b and 5c are hardly noticeable. Figure 6 shows the fractions of the aquifer area covered by chromium concentrations within percentile brackets of the Israeli chromium concentration drinking water standard of 50 ppb. Using an infinite inclusion zone results in full coverage of the interpolation grid. However, none of the grid points was assigned zero concentration. The fractions of the area which were assigned concentrations in between 0-25%, 25-50%, 50-75% and 75-100% of the standard brackets are 0.52, 0.23, 0.08 and 0.05, respectively. Most important of all, in 0.11 of the interpolation grid points, the concentrations exceed the standard. The fractions of interpolation points yielded by the two interpolations using inclusion zones are very similar to each other. The fractions of the grid points that were not assigned an interpolation value in the circular and elliptical inclusion zones maps were 0.38 and 0.36, respectively. In both cases, fractions of 0.12, 0.46, 0.15 and 0.006 of the grid points have values in the 0-25%, 25-50%, 50-75% and 75-100% ranges of the standard brackets, respectively. The fraction of grid points where the drinking water standard is exceeded is 0.021.

The process and analysis described above for chromium was carried out for all 31 pollutants using three inclusion zone sizes (ellipses with minor semi-axes of 500, 1000 and 1500 m, and the corresponding circular zones). It should be noted that a 1000 m distance scale is quite conservative and that most pollution plumes in the coastal aquifer are not expected to have advanced much farther than 1 km from their source. Figure 7 summarises some interesting results noted in the maps of pollutants 11-31 (the non-major ions). The figure provides the fractions of interpolation points for which the two types of inclusion zone interpolations could not provide a result due to lack of observations in the inclusion zones (Figs. 7a-7c). It also provides the fractions of the interpolation points which were assigned concentrations exceeding the health standards (Figs. 7d-7f). The most notable point that Fig. 7 shows is that as the inclusion zones increase, the fraction of interpolation points not assigned a concentration decreases, but the fraction of points assigned concentrations exceeding the health standard increases. An additional interesting point is that the fractions of the interpolation points not assigned a concentration is smaller using the elliptical inclusion zones. This is hardly noticeable in Fig. 7a (500 m minor ellipse semi-axes and the corresponding circle radii) but is clear in the other two plots with the differences between them at the $\alpha = 0.078$ and $\alpha = 0.0025$ statistical significance levels for the 1000 and 1500 m minor ellipse semi-axes, respectively (Mann-Whitney test[23]). The differences in the fractions of points exceeding the standard between the two methods are very small and inconsistent. It must be emphasised that by design the total area around the interpolation points covered by the two methods is exactly equal. Thus, the elliptical inclusion zones enabled delineating similar portions of the aquifer within certain brackets of the health standard as the circular zones, but did so covering more (almost 20% in a few cases) of the total aquifer area. This fact has small but practically important significance for assessing the state of the aquifer. It may demonstrate an advantage achieved by incorporating the additional information of the water head level by using the elliptical inclusion zones.

The performance of the interpolation was assessed using cross-testing for the data of all the 31 pollutants, and the quantitative results are given in Fig. 8. The testing was of interpolations using the maximum concentrations observed in 2006-2010 in each well. Three types of interpolation were tested: using infinite inclusion zone, elliptical inclusion zone, and circular zones such that the total area covered by the circles around the wells observing the pollutant was equal to that covered by the ellipses (see Fig. 2). The SR was calculated using four concentration brackets. Three brackets had their upper limits at zero concentration, the standard for the pollutant and three times the standard. The fourth bracket is of concentrations above three times the health standard. Figures 8a, 8b and 8c show the results for minor ellipse semi-axis inclusion zones (and their corresponding circle radii) of 500, 1000 and 1500 m, respectively. For all the major ions, with the exception of potassium, the SR achieved by the three interpolations are very similar and are above 0.8. No clear differences can be noticed between the results of different inclusion zone sizes. For potassium and most of the non-major ion pollutants, the infinite inclusion zones interpolation are clearly inferior, achieving SR which in most cases in between 0 and 0.3 while the two interpolations using limited inclusion zones achieve SRs between 0.52-1.00. (The perfect SR achieved by all three interpolation for pollutant 14, aldicarb, is due to the fact that only zero concentrations were observed for this pollutant in 2006-2010, see Table 1). The differences between the cross-testing results of the interpolations using circular and elliptical inclusion zones are very small, usually in the order of 1% or less. The most important point that Fig. 8 shows is the trade-off of decreasing SR values (for potassium and the non-major ions) as the inclusion zones increase.

4 Discussion and conclusion

Interpolation of groundwater pollution data over a large aquifer is a difficult task given the extreme pollution data distributions and the fact that in most cases there are many pollution sources of various types scattered over the aquifer area. The applicability of using local inclusion zones around the interpolation points was studied in this paper. The quality of the results was assessed using the Success Rate (SR), defined in in section , which is a simple measure in the range 0 to 1 suitable to both the dynamic range of the observations and to management decisions regarding the aquifer. The interpolations using limited inclusion zones proved clearly superior compared to using all the observations to produce each interpolation point (infinite inclusion zone), with SRs in the range 0.5-10. No significant differences were found between using using circular and elliptical inclusion zones. However, maps produced using elliptical inclusion zones managed to provide similar information regarding the pollution's spatial extent at a smaller price in terms of compromising their total aerial coverage.

Aquifer management must consider a wide range of contaminants, derived from both point and diffuse sources. Long term monitoring of a large aquifer may not be able to keep track of the sources and in many cases they are unknown. The approach taken in this work is to provide a solution for point sources in the aquifer, and simultaneously capture also the impact of diffuse sources through the use of judiciously chosen interpolation zones. As the Coastal aquifer is comprised predominantly of sand, the differences in transport of the various contaminants due to various adsorption/exchange interactions are minimal, thus, a single inclusion zone was used for all the species. The choice of the inclusion zone scale was mainly driven by the known advection and transverse dispersion in the aquifer. The flow can be considered homogeneous at the small spatial scale of transport dictated by the advection and transverse dispersion magnitudes. An application of our methodology to other aquifers may require using different inclusion zones for the different contaminants.

In risk assessment of contaminated lands the source-pathway-receptor linkages paradigm is conventionally applied. This paradigm is mainly relevant for point sources, but nevertheless, does form the underlying rationale for the approach used in this study. In our case, the groundwater at the sampling point is simultaneously the source (the linkage between original on-surface sources and the current location of the contaminant/s in the groundwater is not known), and the pathway (the flowing groundwater is the sole means of transport and dispersion of the contaminant). Adjacent areas of the aquifer that are not sampled are the potential receptors. We examined how the chosen interpolation method can influence which part of the aquifer is considered to be a receptor.

This study emphasises the clear trade-off that practitioners must keep in mind when producing interpolation maps using limited inclusion zones. The SRs indeed improve as the zones become smaller. However, that advantage is achieved at the expense of the total fraction of covered area. Using limited inclusion zones leaves a fraction of the aquifer uncovered i.e. with complete uncertainty regarding the pollutant concentration there. We believe this is a realistic approach which enables safe water production where the pollutants can be mapped reliably but one that sets clear limits on it where the data are scarce and interpolation cannot be carried out with sufficient confidence.

The fact that no difference was noted between the SRs achieved by the two inclusion zone type interpolations is intriguing. A possible explanation is that the SRs were calculated by cross-tested interpolations using data from all the wells, the majority of which are located in areas of small head gradient (see Fig. 1), where the ellipses approach circles. The possible advantage of the use of elliptical zones in areas of strong water gradients cannot have much weight in the SR calculation given the small number of wells in these parts of the study aquifer. However, in other aquifer systems, there may be a more significant differences between the two limited zone type approaches.

This study was born out of the need of the Israeli Water Authority to improved its management of the Coastal aquifer. We believe that the methodologies and understandings developed in this paper enable more efficient management of that water resource both in terms of water production and pollution remediation. Identifying areas where interpolation cannot be carried out with sufficient confidence helps in directing future monitoring programmes. Delineating polluted areas with high level of accuracy focuses the remediation efforts and enables production in areas of the aquifer shown to be clean. Pollution maps serve as an excellent intuitive visual aids but can also serve as a basis for more complex geostatistical analyses and as an input for a decision support system (DSS[24]). Decision making is thus made more informed process and the population benefits from from more ample but safer water sources. It must be noted though that three dimensional effects are not accounted for and our method cannot be used for accurate estimations of volumetric properties. Nor did we use any additional information of the sub–surface and the terrain, which could contribute for improved estimations. These issues are left for future research.

5 Acknowledgements

The authors would like to thank the Israel Water Authority for providing the data for this study and the funds to carry it out.

References

1 M. Soutter and A. Must, *Journal of Contaminant Hydrology*, 1998, **32**, 25–39.
2 A. Al-Hanbali and A. Kondoh, *Hydrology Journal*, 2008, **16**, 499–510.
3 H. Baalousha, *Agricultural Water management*, 2010, **97**, 240–246.
4 S. Javadi, N. Kavehkar, M. Mousavizadeh and K. Mohammadi, *Agricultural Science and Technology*, 2011, **13**, 239–249.
5 P. Reed, T. Ellsworth and B. Minsker, *Groundwater*, 2012, **42**, 190–202.
6 T. Kistemann, J. Hundhausen, S. Herbst, T. Claßen and H. Färber, *International Journal of Hygiene and Environmental Health*, 2008, **211**, 308–317.
7 M. Morio, M. Finkel and E. Martac, *Environmental Modelling & Software*, 2010, **25**, 1769–1780.
8 S. Carroll and A. Goonetilleke, *Environmetrics*, 2005, **16**, 257–274.
9 C. Güler, M. Kurt, M. Alpaslan and C. Akbulut, *Journal of Hydrology*, 2012, **414-415**, 435–451.
10 A. Melo, E. Pinto, A. Aguiar, C. Mansilha, O. Pinho and I. Ferreira, *Environmental Monitoring and Assessment*, 2012, **184**, 4539–4551.
11 M. Varljen, M. Barcelona and H. Wehrmann, *Environmental Monitoring and Assessment*, 1999, **59**, 31–46.
12 P. Adhikary, C. Dash, R. Bej and H. Chandrasekharan, *Environmental monitoring and assessment*, 2011, **176**, 663–676.
13 G. Liu, W. Wu and J. Zhang, *Agriculture, Ecosystems & Environment*, 2005, **107**, 211–220.
14 Adhikary, H. P.P, Chandrasekharan, D. Chakraborty and K. Kamble, *Environmental monitoring and Assessment*, 2010, **167**, 599–615.
15 B. Tutmez and Z. Hatipoglu, *Ecological Informatics*, 2010, **5**, 311–315.
16 A. Furman and H. Abbo, *Water Policy in Israel*, Springer, Netherlands, 2013, ch. 8, pp. 125–136.
17 *Hydrological Service of Israel: Utilization and state of the water resources of Israel to fall 2009*, The Hydrological Service, Jerusalem, 2011.
18 H. Gvirtzman, *Israel Water Resources; Chapters in Hydrology and Environmental Sciences*, Yad Ben-Zvi Press, Jerusalem, 2002.
19 Matlab, *Matlab version 7.9.0.529 (R2009b)*, The MathWorks Inc., Natick, Massachusetts, 2010.
20 E. Isaaks and R. Srivastava, *An Introduction to Applied Geostatistics*, Oxford University Press, New York, 1989.
21 A. Ko, P. Cavalin, R. Sabourin, de Souza Britto and A., *IEEE Transections on Pattern Analysis and Machine Intelligencejhjjj*, 2009, **31**, 2168–2178.
22 Yuval, D. Broday and P. Alpert, *Environmental Pollution*, 2012, **166**, 65–74.
23 E. Lehmann and H. D'Arbera, *Nonparametrics: Statistical Methods Based on Ranks*, Springer, New York, 2006.
24 K. Westphal, R. Vogel, P. Kirshen and S. Chapra, *Journal of Water Resources Planning and Management*, 2003, **129**, 165–177.

[a] *Department of Civil and Environmental Engineering, Technion, Israel Institute of Technology, Haifa, Israel*
[b] *Institute of Soil, Water and Environmental Sciences, The Volcani Center, Agricultural Research Organisation, Israel*
[*] *Technion IIT, Haifa 32000, Haifa, Israel. Fax: +972 3 8292676; Tel: +972 4 8228898; E-mail: lavuy@tx.technion.ac.il*

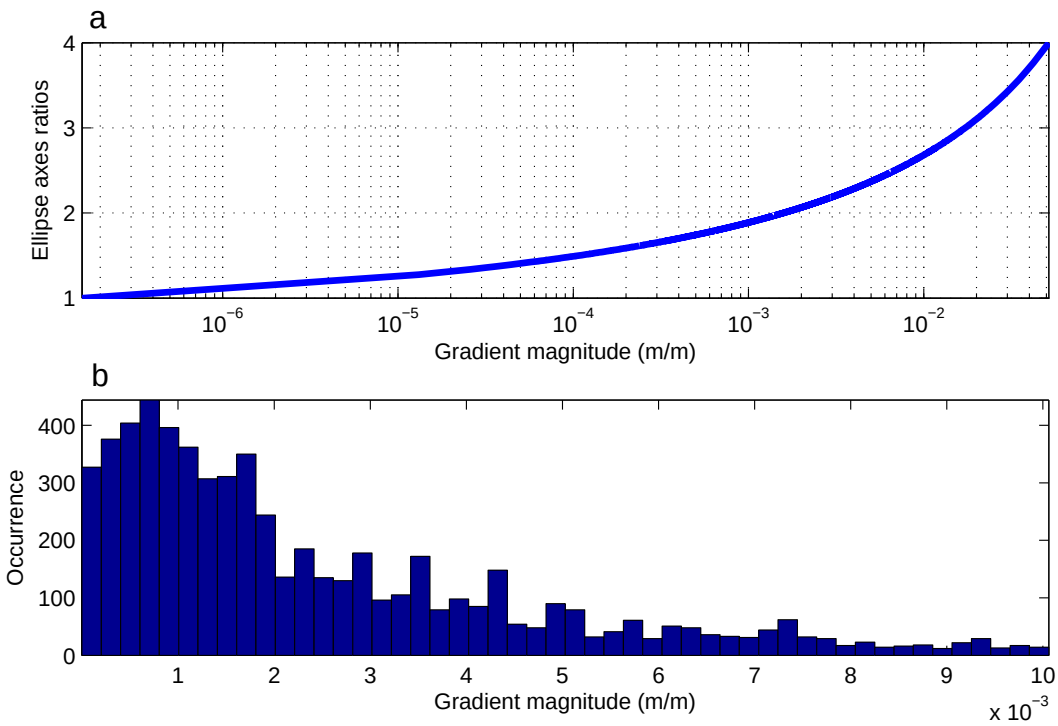

Fig. 3 (a) The ratio between the axes of an elliptical inclusion zone as a function of the head gradient magnitude. (b) A histogram of the head gradient magnitudes at the well locations in the coastal aquifer.

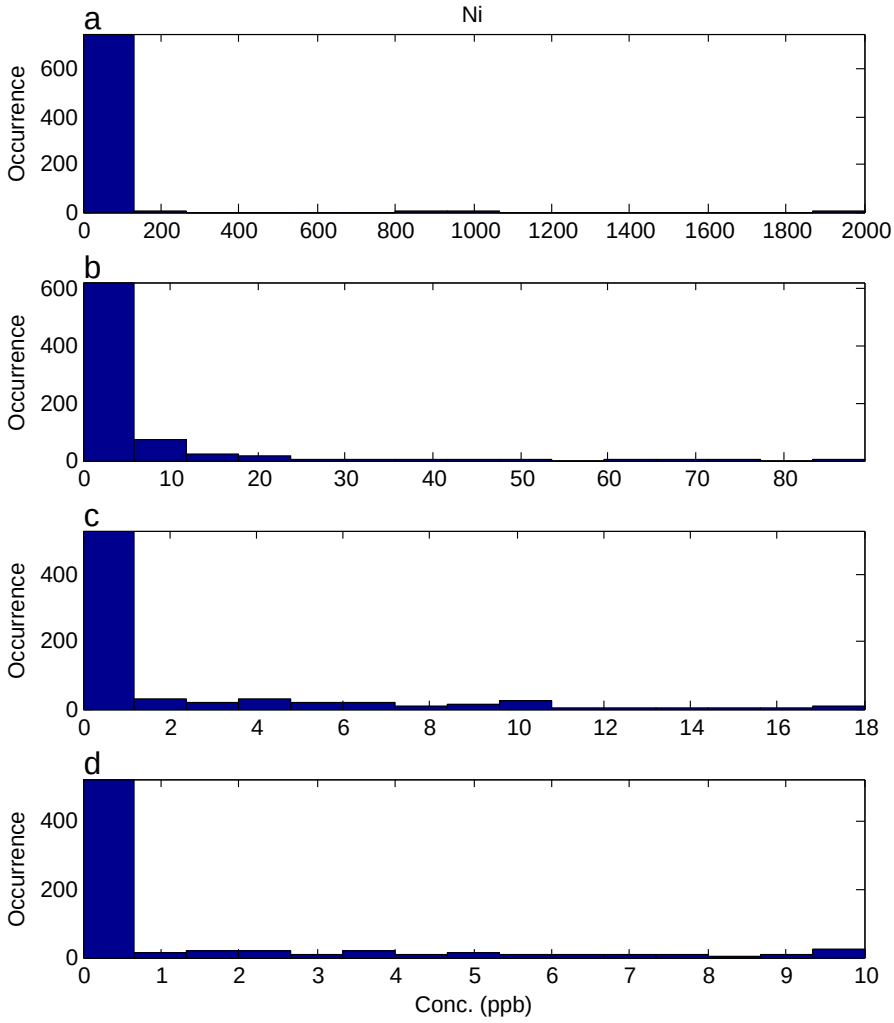

Fig. 4 Histograms of the monitoring nickel concentrations. The maximal value observed in each well in 2000-2010 was considered. At any data value range presented in the figure the great majority of the data are very small with a few very large ones. (a) Using all the values. (b) Using only values below the 99% percentile. (c) Using only values below the 95% percentile. (d) Using only values below the 90% percentile.

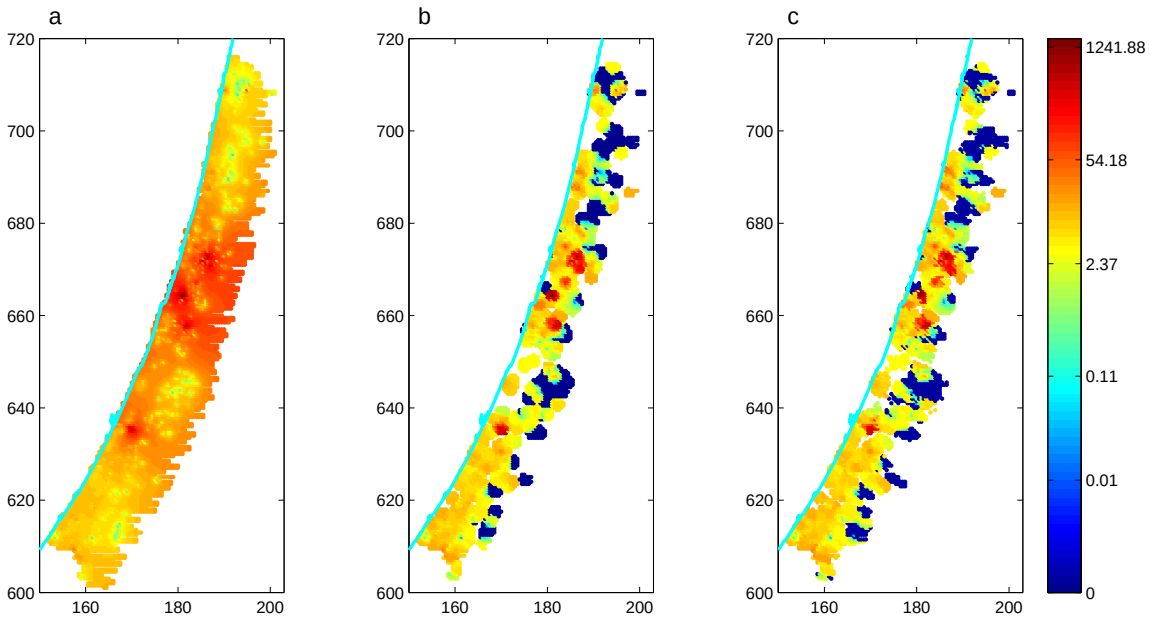

Fig. 5 Interpolation maps of chromium concentrations. The continuous cyan line marks the coast line. The coordinates are of the New Israeli Grid in km. The concentrations are colour coded. Identical colour coding was used for the three maps and is given in the colour bar in ppb units. (a) Using an infinite inclusion zone (all data points are used to produce each grid value). (b) Using a circular inclusion zone. (c) Using an elliptical inclusion zone. The radius of the circular inclusion zones was set such that the total area covered by the circular zones is equal to the area covered by ellipses with minor semi-axis of 1000 m. Note the gaps in the areal coverage of maps (b) and (c) compared to that of map (a).

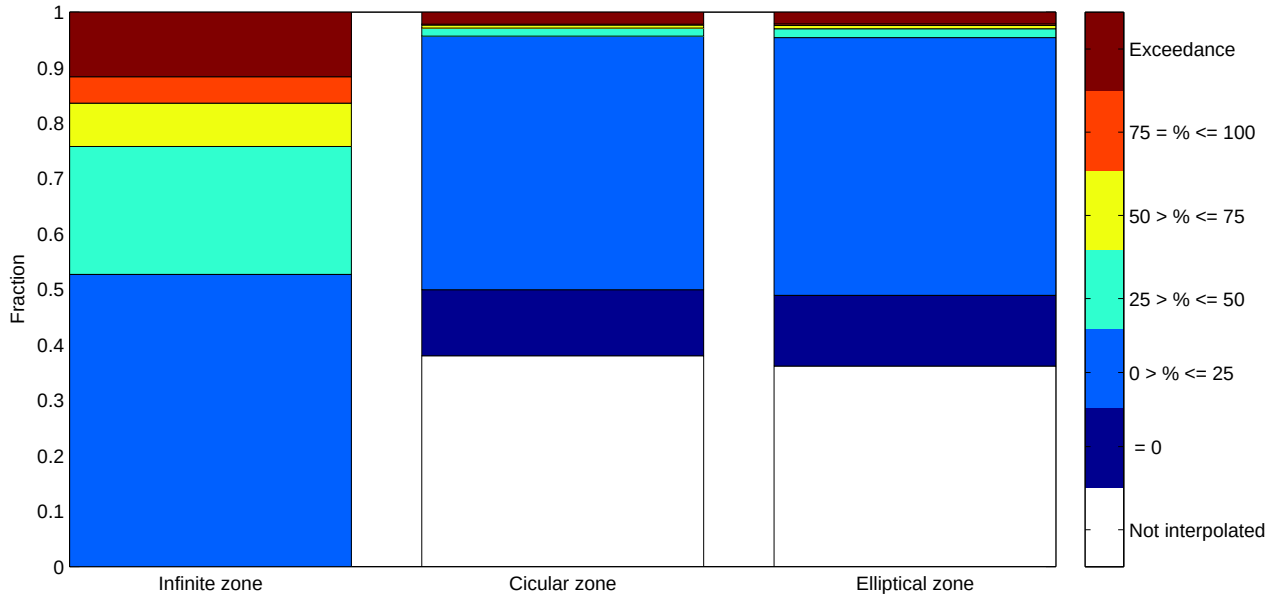

Fig. 6 The fractions of the aquifer area covered by chromium concentration at different brackets. The brackets are between percentages of 50 ppb, the Israeli Standard for chromium.

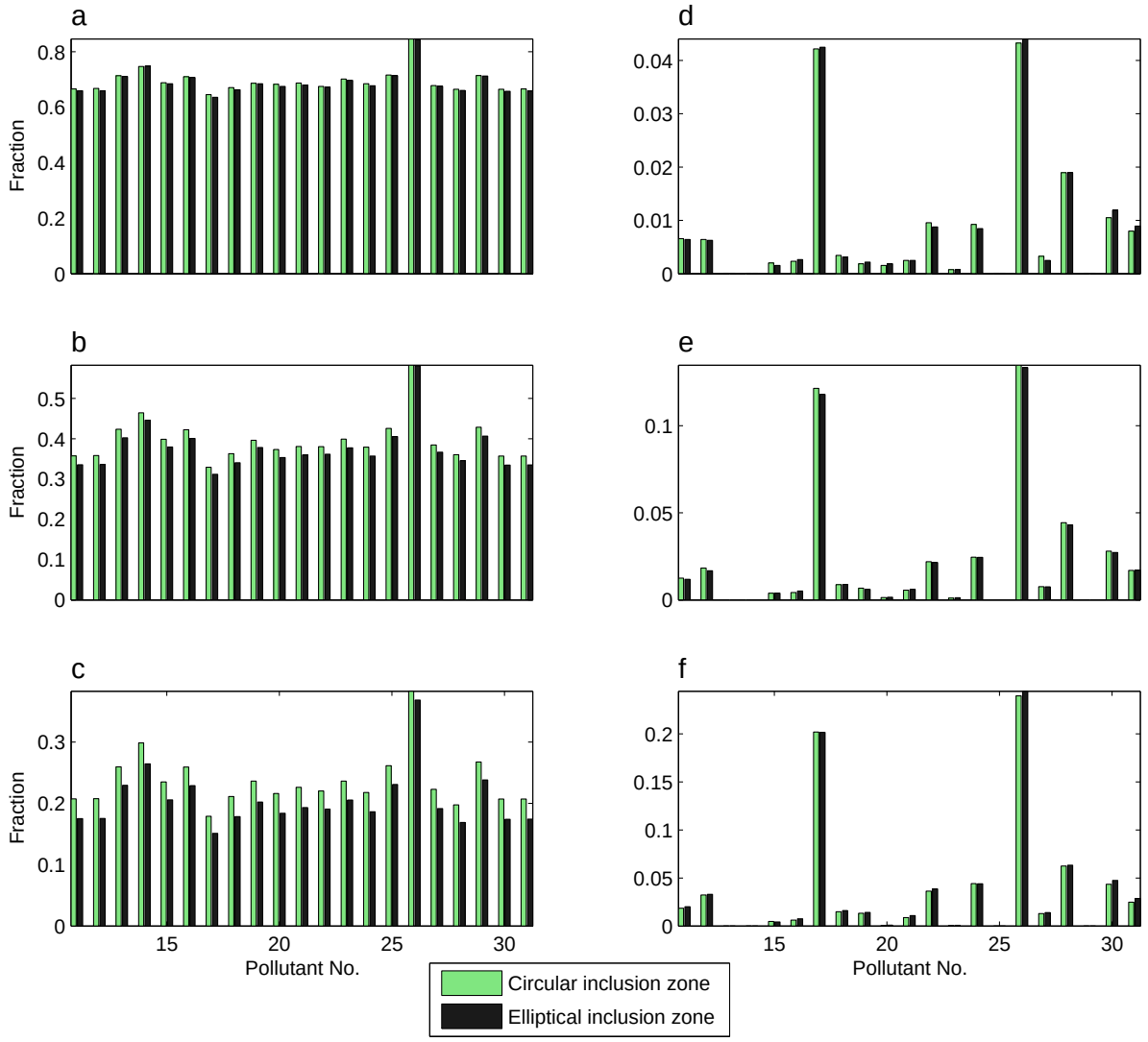

Fig. 7 (a) The fractions of the interpolation points that were not assigned a concentration value for pollutants 11-31 (non-major ions, see Table 1) using elliptical inclusion zones with minor semi-axes of 500 m and the corresponding circular zones. (b) Like (a) but for minor ellipse semi-axis of 1000 m. (c) Like (a) but for minor ellipse semi-axis of 1500 m. (d) The fractions of the interpolation points that were assigned a concentration value exceeding the health standard of pollutants 11-31 (non-major ions, see Table 1) using interpolation elliptical inclusion zones with minor semi-axes of 500 m and the corresponding circular zones. (e) Like (d) but for minor ellipse semi-axis of 1000 m. (f) Like (d) but for minor ellipse semi-axis of 1500 m.

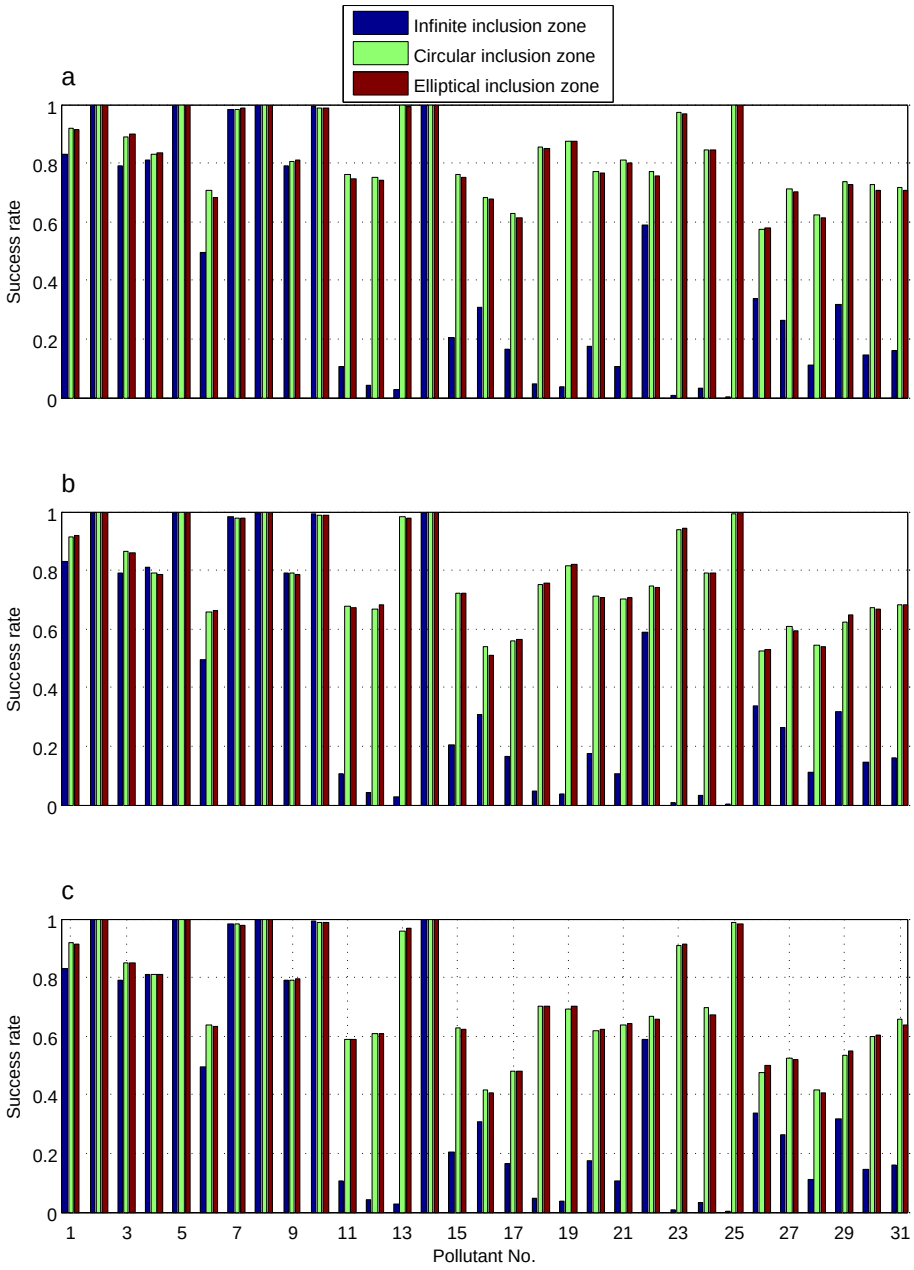

Fig. 8 The cross-tested success rate values achieved for each pollutant by interpolations using elliptical, circular and infinite inclusion zone. The interpolation was carried out for the whole aquifer area. (a) The radius of the circular inclusion zones was set such that the total area covered by the circular zones around the interpolation points is equal to the area covered by ellipses with minor semi-axis of 500 m. (b) The same as (a) but using ellipse minor semi-axes of 1000 m. (c) The same as (a) but using ellipse minor semi-axes of 1500 m.

Table 1 A list of the pollutants monitored in the coastal aquifer whose data were used in this study. The table provides for each pollutant its group affiliation, the number of wells at which it was monitored during the study period and the number of wells at which all the records are of zero concentration. The pollutant groups are: 1-major ions, 2-volatile organic compounds, 3-heavy metals, 4-fuels, 5- pesticides/herbicides

No.	Name	Symbol	Pollutant group	No. of wells	No. of zeros
1	Boron	B	1	1089	10
2	Calcium	Ca	1	1068	0
3	Chloride	Cl	1	1882	0
4	Fluoride	F	1	464	52
5	Bicarbonate	HCO_3	1	948	0
6	Potassium	K	1	1067	0
7	Magnesium	Mg	1	989	0
8	Sodium	Na	1	1060	0
9	Nitrate	NO_3	1	1780	13
10	Sulphate	SO_4	1	1029	2
11	1,1-Dichloroethene	1,1-DCE	2	862	733
12	1,2-Dichloroethane	1,2-DCA	2	865	732
13	Alachlor	Alchalor	5	618	600
14	Aldicarb	Aldicarb	5	423	423
15	Arsenic	As	3	777	457
16	Atrazine	Atrazine	5	641	403
17	Benzene	BNZ	2, 4	1097	784
18	Carbon Tetrachloride	CCl_4	2	845	758
19	Cadmium	Cd	3	784	672
20	Chloroform	$CHCl_3$	2	816	672
21	cis-1,2-Dichloroethene	c-1,1-DCE	2	838	735
22	Chromium	Cr	3	827	139
13	1,2-Dibromo-3-chloropropane	DBCP	3, 4	705	691
24	Ethylene dibromide	EDB	2, 3, 4	814	729
25	Lindane	Lindane	5	602	563
26	Methyl tert-butyl ether	MTBE	2, 4	566	303
27	Nickel	Ni	3	803	514
28	Lead	Pb	3	977	599
29	Simazine	Simazine	5	600	410
30	1,1,1-Trichloroethene	TCE	2	883	616
31	Tetrachloroethene	PCE	2	876	681