5 SPATIAL INTERPOLATON METHODS

5.1 Introduction

Chapters 3 and 4 focused on capturing monitoring design tradeoffs using genetic algorithms. This chapter provides guidance to practitioners in choosing a plume interpolation method for any groundwater application by explicitly addressing opposing theoretical perspectives, ease-of-implementation, and effectiveness. More specific to the goals of this dissertation, Chapter 5 provides guidance in choosing the best estimation technique for evaluating the LTM performance criteria examined in Chapter 6.

5.1.1 Motivation & Scope

Plume interpolation consists of estimating contaminant concentrations at unsampled locations using the available contaminant data surrounding those locations. The goal of groundwater plume interpolation is to provide the best possible picture of the contaminant plume given the data limitations associated with sparse monitoring networks with irregular geometries. Beyond data limitations, contaminant plume interpolation has proven to be a difficult task because contaminant concentration fields are highly heterogeneous, anisotropic, and nonstationary phenomena (i.e., their mean and covariance vary widely across a given site). Contaminant concentrations commonly vary over several orders of magnitude within relatively short distances and in most cases are preferentially sampled in areas of high concentration. Preferential sampling and the highly skewed nature of groundwater contamination can combine to severely bias interpolation estimates and any subsequent decisions that must be made using these estimates (*Kitanidis* 1997, *Goovaerts* 1997, and *Chilès & Delfiner* 1999). Contaminant fate-and-transport simulation, inverse parameter estimation, site-characterization, and long-term monitoring (LTM) network design are a few examples of applications within which plume

interpolation plays a vital role (see *ASCE 1990b*, *Loaiciga* 1992, *Reed et al.* 2000 for more detailed reviews).

Historically, there have been two primary approaches to address the problem of contaminant plume interpolation. The first approach uses deterministic methods such as inverse distance weighted sums to estimate concentrations within the interpolation domain (*Gambolati & Volpi* 1979, *Gotway et al.* 1996). The second approach uses the stochastic framework of regionalized variable theory and geostatistics to obtain contaminant plume estimates *Matheron* (1965, 1971). *Chilès & Delfiner* (1999) define a *regionalized variable* "…to designate a numerical function [*c(x)*] depending on a continuous space index *x*, and combining high irregularity of detail with spatial correlation".

Of the two interpolation approaches, geostatistics has been the primary focus of previous studies within the water resources area (*Cooper & Istok* 1988a,b,c; *ASCE* 1990a,b). Additionally, there are several recent texts available that provide an array of theoretical perspectives on the application of geostatistics to both stationary and nonstationary phenomena (*Isaaks & Srivastava* 1989, *Kitanidis* 1997, *Goovaerts* 1997, and *Chilès & Delfiner* 1999). However, there have been very few studies that compare both deterministic and geostatistical interpolation methods in the water resources literature (*Gambolati & Volpi* 1979 and *Hughes & Lettenmaier* 1981). Given that spatial structure identification in geostatistics applications improves markedly as the number of data increases, and that many sites have limited data sets, it is important to consider both deterministic and geostatistical interpolation methods. Selecting an appropriate method can be a difficult task for the practitioner because of opposing theoretical perspectives given in the interpolation literature. For example, some geostatisticians recommend the use of more traditional ordinary kriging (OK) within local neighborhoods (*Journel & Rossi*

1989, *Goovaerts* 1997) when interpolating nonstationary phenomena (e.g. contaminant concentrations), while others recommend more advanced methods such as intrinsic kriging (IK) (*Chilès & Delfiner* 1999).

The attractive properties of geostatistical approaches, relative to deterministic methods, are their ability to: (1) incorporate knowledge of the underlying spatial structure of the phenomenon of interest into "unbiased" estimates and (2) provide estimation variances. Estimation variances are considered to be measures of the quality of the estimates (thus a measure of "local" uncertainty). As discussed in the next section, several studies have shown that estimating nonstationary phenomena such as contaminant concentrations with limited data is subject to bias for all forms of kriging. Moreover, as the level of available data decreases for a given site the estimation variances attained using geostatistical approaches significantly degrade in their ability to represent the local uncertainty in estimates. Identifying an optimal method for plume interpolation is a particularly difficult task given both conflicting perspectives in the literature and the limited availability of comprehensive performance comparisons of deterministic and geostatistical methods in the literature.

Gotway et al. (1996) and *Crawford & Hergert* (1997) discuss these difficulties in the context of mapping nonstationary soil properties for a regularly gridded data set. This series of studies presents an empirical comparison of OK within local neighborhoods, kriging with a trend (KT) [also termed *universal kriging*], median polish kriging (a transformation-based kriging approach for regularly gridded data), and inverse-distance weighting. Although these studies did not make definitive recommendations as to which interpolation approach was the best, they present a careful synopsis of their relative benefits and limitations to help guide the practitioner in her or his application. This chapter extends their comparison of interpolation methods to

scatter-point (i.e., non-gridded data) groundwater plume interpolation and provides a comparison of the following interpolation approaches:

- Ordinary Kriging
- Multigaussian Kriging (MGK)
- Intrinsic Kriging (ItK)
- Quantile Kriging (QK)
- Nonlinear Least Squares Inverse Distance Weighting (NLS)
- Inverse Distance Weighting of Power 2 (ID).

This chapter compares plume interpolation methods as a function of data availability. Additionally, it illustrates the consequences of each method's underlying assumptions. The methods were selected solely for addressing the problem of estimating contaminant concentrations at unknown spatial locations at a single snapshot in time and not for spatio-temporal interpolation.

5.1.2 Previous Work in the Estimation of Nonstationary Phenomena

The plume interpolation methods discussed in this chapter are based on the underlying premise that the values of contaminant concentrations at any two points become increasingly similar as the distance between the points decreases. *Gambolati & Volpi* (1979) presented a deterministic analysis of the geostatistical estimation method *kriging* and demonstrated its close relationship to inverse distance weighting and least squares methods because of this underlying premise. Studies comparing kriging methods and inverse distance-based interpolation schemes have generally found that the superiority of kriging is dependent on the number of available data, the level of anisotropy of the underlying phenomenon, and the presence of a trend (*Gambolati & Volpi* 1979, *Hughes & Lettenmaier* 1981, *Gotway et al.* 1996). These studies compared the

performance of inverse-distance methods with only one of the several possible forms of kriging and generally concluded that the primary strength of the geostatistical approach is that each estimate has a corresponding estimation variance (a measure of estimation error).

Kriging nonstationary data is a difficult task that has led to extensions of the original kriging estimator. However, there has been considerable debate about the relative merits of these extensions. The earliest forms of kriging, termed *simple kriging* (SK) and *ordinary kriging*, are based on an assumption of stationarity, first with respect to the mean and also with respect to the covariance or spatial increments (intrinsic hypothesis). The mean is assumed constant either throughout the entire domain or within local neighborhoods composing the domain, respectively. *Chilès & Delfiner* (1999) state that *Matheron* (1969) recognized the limitations of these assumptions and formulated the *kriging with a trend* (KT) approach for nonstationary phenomena.

The KT approach models the regionalized variable *c(x)* as the sum of a smoothly varying mean component *m(x)* termed a trend or drift and a second order stationary residual component *r(x)*. The KT estimator requires specification of a trend function to model *m(x)* and knowledge of the covariance structure of the residual component *r(x)*. Several studies have subsequently criticized the method primarily due to the "chicken-and-egg" (*Armstrong* 1984) conundrum practitioners face where the exact form of *m(x)* must be known to accurately model the spatial structure of *r(x)* while the exact form of *m(x)* cannot be determined without knowing the true spatial structure of *r(x)* (*Volpi & Gambolati* 1978, *Hughes & Lettenmaier* 1981, *Russo & Jury* 1987, *Journel & Rossi* 1989, *Crawford & Hergert* 1997). These studies show that the common practice of specifying *m(x)* as a second order polynomial can introduce significant bias in both

the kriged estimates and their corresponding estimation variances. For this reason, this method is not included in this chapter.

Recognizing the limitation of the KT approach, *Matheron* (1973) and *Delfiner* (1976) formulated the *intrinsic random function of order k* (IRF-k) theory for *intrinsic kriging* (see *Chilès & Delfiner* 1999, *Hughes & Lettenmaier* 1981). IRF-k theory was formulated to model nonstationary phenomena where the division of *c(x)* into *m(x)* and *r(x)* has no meaning. The ItK approach utilizes the concept of the *generalized increment of order k* in which a linear combination of sampled data is selected that will intrinsically filter a polynomial trend of order *k* without the need for explicitly specifying the trend as in the KT approach (*Chilès & Delfiner* 1999). Although IRF-k theory provides a theoretically elegant alternative to the KT approach, its application has been limited because determining the spatial structure of the underlying phenomena requires significantly more expertise (*Delfiner* 1976, *Kitanidis* 1983, *Marshall & Mardia* 1985, *Chilès & Delfiner* 1999).

Hughes & Lettenmaier (1981) compared the performance of ItK and traditional least squares approaches using hydrologic data and found that the relative superiority of ItK estimates is dependent on having a sample size greater than 50. Additionally, the study showed that for small data sets inference of spatial structure is subject to a bias that "…will generally lead the investigator to believe his [or her] estimates are more precise than they really are" (*Hughes & Lettenmaier* 1981). *Journel & Rossi* (1989) and *Goovaerts* (1997) recommend using OK within neighborhoods rather than KT and ItK when interpolating nonstationary natural phenomena. Note that these findings directly contradict the methodological recommendations given by *Chilès & Delfiner* (1999) who recommend the use of ItK in place of OK or KT for nonstationary phenomena.

In addition to nonstationarity, the distributional dependence of the estimation variance has led to additional kriging approaches. *Journel & Rossi* (1989) explicitly show that the estimation variance is a "model dependent ranking" of sampling configurations and is reflective of local accuracy only in cases where the studied phenomenon is multivariate Gaussian (i.e., normally distributed). The lognormal transform (i.e., $y = \ln[c(x)]$) is one of the most commonly applied transformations within water resources applications to try to normalize spatial phenomena such as transmissivities or contaminant concentrations (see *Cooper & Istok* 1988a and *ASCE* 1990a). Several issues require careful consideration when using the lognormal transform or log-kriging approach. The method implicitly assumes that the contaminant concentrations are multi-variate lognormal. *ASCE* (1990) state that this assumption cannot be verified using real-world data sets (i.e., data are often too sparse to even adequately evaluate the bivariate spatial distribution, let alone 3 or 4 point distributions). *Deutsch & Journel* (1998) warn that the back transform requires bias-correction to ensure y is an unbiased estimator of $c(x)$. Even with the bias-correction, the back-transform exponential function amplifies estimation errors (*Deutsch & Journel* 1998, ASCE 1990a). These drawbacks have led to a shift towards the transformations that are briefly described below.

Multigaussian kriging uses a normal score transformation of the sample data (i.e., mapping highly asymmetrical contaminant samples to a corresponding normal distribution) to improve its approximation of the multivariate Gaussian distribution with the intention of making estimation variances more reflective of the local accuracy of estimates (*Goovaerts* 1997). The assumptions and limitations of MGK are discussed in more detail in the methods section of this chapter. Other extensions of the OK system have used a variety of data transforms to formulate nonparametric (or distribution free) kriging systems that quantify local accuracy based on the

data's empirical cumulative distribution function (cdf) (*Journel* 1983, *Journel* 1988, *Chiueh et al.* 1997, *Journel & Deutsch* 1997, *Juang et al.* 2001). In addition to making the kriging approach distribution free, non-parametric kriging methods reduce estimation bias caused by highly variable phenomena with severely asymmetrical distributions such as contaminant concentrations. This chapter focuses on the rank-order transform (or quantile kriging) proposed by *Journel & Deutsch* (1998) and successively applied to a soil contamination application by *Juang et al.* (2001). As with MGK, the details of the QK approach are discussed in greater detail in the methods section of this chapter. Although transformation methods can render contaminant data to be more amenable to structure identification and kriging, *Goovaerts* (1997) warns that practitioners should carefully consider the effects of transformations before they apply them by paying close attention to any potential side effects (e.g. inflation of estimation error) incurred in such applications. The performance of each interpolation method considered in this chapter was explored using the three test cases described in the next section.

5.1.3 Test Cases

The test case data used in this research are drawn from a 50 million-node flow-and-transport simulation (for more details see *Maxwell et al.* 2000). The simulation provided realistic historical data for a steady-state hypothetical perchloroethylene (PCE) plume in a highly heterogeneous alluvial aquifer. The hydrogeology of the test case is based on an actual site located at the Lawrence Livermore National Laboratory in Livermore, California, currently being managed under the United States' Comprehensive Environmental Response, Compensation and Liability Act (CERCLA) program. One of the objectives of this chapter is to show how each of the interpolation approaches performs as a function of the amount of available data. To satisfy this objective, the test case data from *Maxwell et al*. (2000) were used to

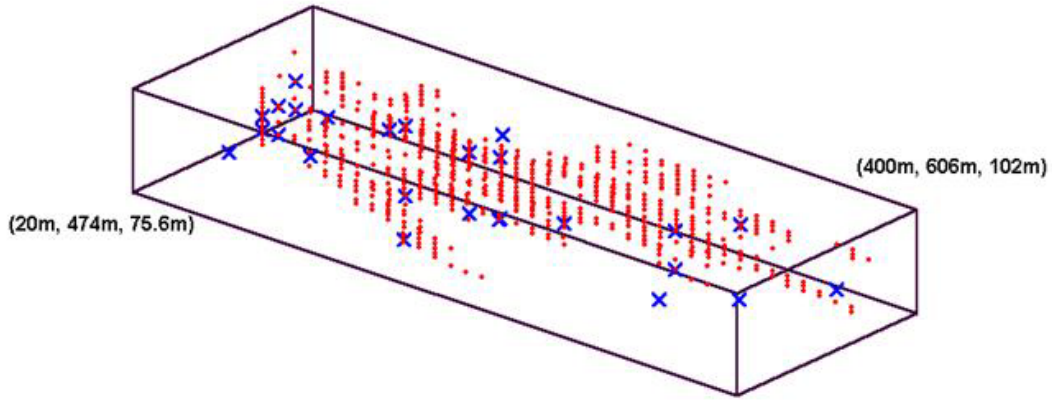
develop the three monitoring networks shown in Figure (5.1). The smallest network, shown in Figure (5.1a), consists of 20 monitoring wells that sample a total of 26 sampling locations within 1 million cubic meters of contaminated aquifer. The intermediate network, shown in Figure (5.1b), has 29 monitoring wells that sample 58 locations within 6 million cubic meters of contaminated aquifer. Finally, the largest network, shown in Figure (5.1c), has 59 wells that sample 124 sampling locations within 16 million cubic meters of contaminated aquifer.

	PCE Mean (mg/m^3)	**PCE Median (mg/m^3)**	**PCE Coefficient of Variation**
Large Test Case	29	1	11.4
Intermediate Test Case	164	11	3.6
Small Test Case	355	9	1.8

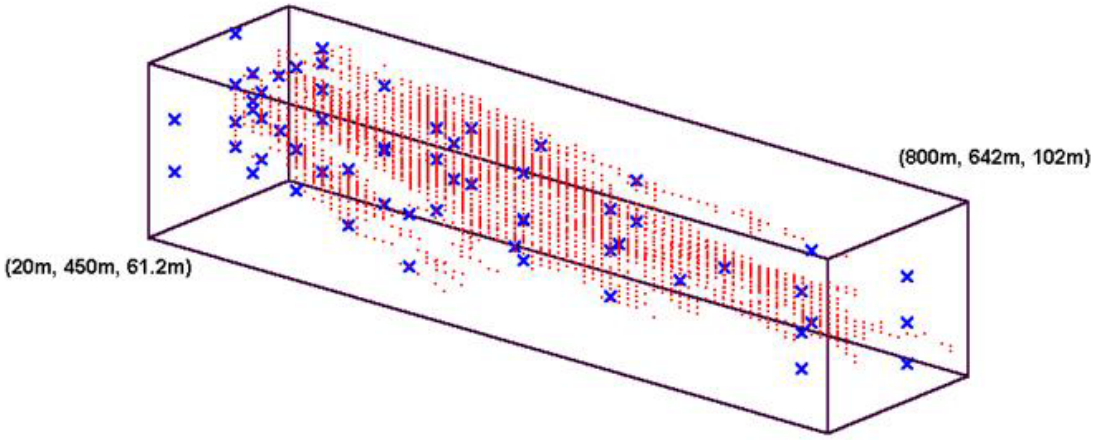
Table 5.1 Sample data statistics for each test case

This chapter makes the assumption that as the size and maturity of a contaminant plume increases, more monitoring points will be sampled for mapping contaminant concentrations. Table (5.1) presents the means, medians, and coefficients of variation computed using the PCE sample data attained from the test cases' monitoring networks. The statistics shown in Table (5.1) confirm that the sample data from these monitoring networks are both heavily skewed and highly variable. Moreover, Figure (5.1) shows that the scatter-point sampling locations are arbitrarily clustered in the body of plume, and preferentially sample the source area of the plumes. These test data sets are designed to be as representative as possible of real-world sites. Further, the use of the high resolution simulation data from *Maxwell et al.* (2000) enables comparison of the interpolation methods' performance at all estimated locations with the true values at these locations, while still using a realistic data set.

(a) Small Test Case



(b) Medium Test Case



(c) Large Test Case

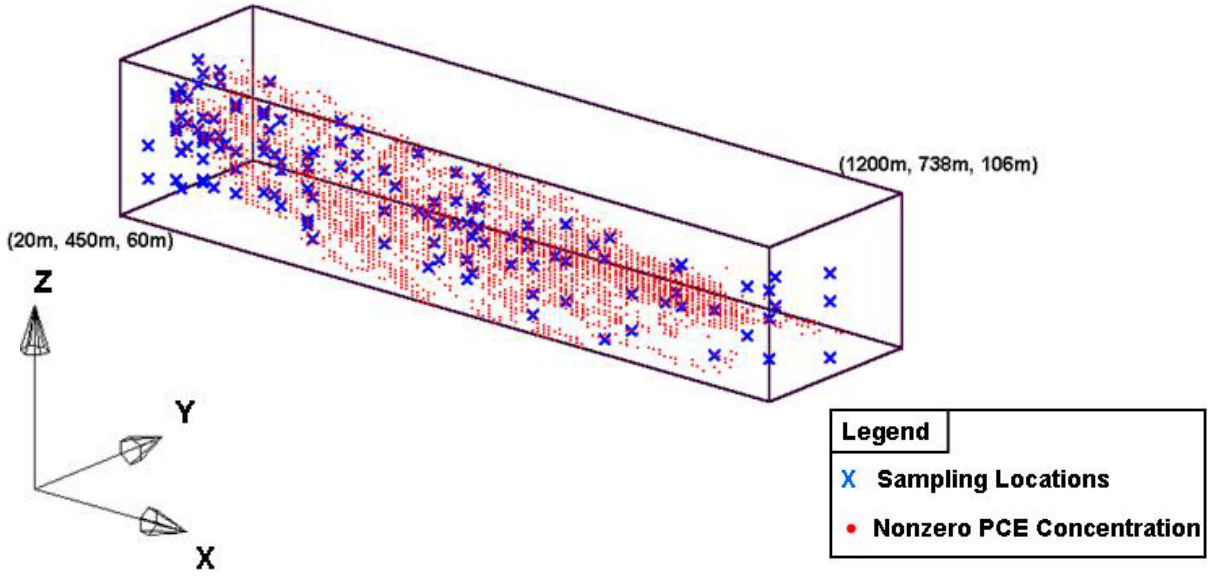


Figure 5.1 Test case monitoring networks

5.2 Methods

5.2.1 Geostatistical Methods

The geostatistical modeling approach requires both structural analysis and specification of a variogram model to represent the theoretical spatial structure. Structural analysis consists of analyzing available concentration data to determine the correlational structure. The analysis uses historical contaminant data to discretely approximate the *theoretical variogram* $\gamma(\bar{h})$, given in equation (5.1), where $\bar{h} = \bar{x}_0 - \bar{x}_1$ is a vector of separation distances in each spatial dimension between two locations $\bar{x}_0$ and $\bar{x}_1$.

$$\frac{1}{2} E\left[c(\bar{x}_0) - c(\bar{x}_1)\right]^2 = \gamma(\bar{h}) + \left[m(\bar{x}_0) - m(\bar{x}_1)\right]^2 \tag{5.1}$$

In equation (5.1), $c(\bar{x})$ is a regionalized variable representing concentration that can be decomposed into its mean, $m(\bar{x})$, and random, $r(\bar{x})$, components. Assuming that the random component $r(\bar{x})$ is stationary with a zero mean, half the expected value of the squared difference between concentrations at locations $\bar{x}_0$ and $\bar{x}_1$ (i.e., $1/2\, E\left[c(\bar{x}_0) - c(\bar{x}_1)\right]^2$) simplifies to yield the sum of the theoretical variogram $\gamma(\bar{h})$ and the squared difference between the mean concentrations at these two locations. The theoretical variogram $\gamma(\bar{h})$ represents the average dissimilarity between two concentrations separated by a distance $\bar{h}$. In the OK approach, the mean component $m(x)$ is assumed to be constant (or locally stationary), in which case the trend (or bias) term, $\left[m(\bar{x}_0) - m(\bar{x}_1)\right]^2$, in equation (5.1) is equal to zero and the theoretical variogram can be deduced directly from concentration samples. For nonstationary phenomena, the "chicken-and-egg" (*Armstrong* 1984) conundrum exists because the trend term is nonzero and a proper model for the theoretical variogram cannot be directly identified solely using

concentration data without making an assumption about the functional form of the trend. The next four sections of this chapter discuss different forms of the kriging system and describe the implicit assumptions that the practitioner must make when using each of the methods to interpolate nonstationary contaminant concentrations.

5.2.1.1 Ordinary Kriging within Local Neighborhoods

The OK estimator is formulated to provide the Best Linear Unbiased Estimate (or BLUE) for contaminant concentration at an unsampled location (for a detailed derivation see *Kitanidis* 1997 or *Goovaerts* 1997). The OK approach invokes the *intrinsic hypothesis*, which assumes that the mean or trend terms given in equation (5.1) are locally constant but unknown within neighborhoods surrounding the current unsampled location (i.e., the trend is modeled using a zeroth order polynomial). *Crawford & Hergert* (1997) state that, theoretically, locally defined neighborhoods of sample points should be used to discretely approximate the theoretical variogram [$\gamma(\bar{h})$ in equation (5.1)]. *Journel & Rossi* (1989) recommend using either data perpendicular to the general trend or local neighborhoods with radii defined to have lengths equal to 25 percent of the total length of the interpolated domain for approximating the variogram. The local approximation of the theoretical variogram is not done in practice because the limited number of data available in most applications would make discrete or experimental variograms unreliable (*Crawford & Hergert* 1997, *Goovaerts* 1997).

The OK approach assumes that the estimation error is normally distributed when computing a confidence interval for every estimate. Recall that the estimation variance is only a function of the theoretical model used to represent the variogram and the geometrical configuration of the sampling data (i.e., a function of the kriging weights only). *Journel & Rossi* (1989) showed that estimation variance is only directly reflective of true estimation error when

the underlying phenomenon is multivariate Gaussian. They also showed that asymmetric data distributions can significantly bias estimates of local uncertainty. The next two sections describe transformation-based kriging approaches that preserve the relative ranks and spatial structure of contaminant data while addressing estimation inaccuracies caused by highly asymmetric data distributions.

5.2.1.2 Multigaussian Kriging within Local Neighborhoods

The MGK approach to estimation utilizes the normal score transform of the sample data and models the contaminant concentrations as a multivariate Gaussian random function, $T(\bar{x})$ (for details on the transform see *Goovaerts* 1997). The normal score (NSCORE) transform maps the asymmetrical discrete sample distribution to a $N(0,1)$ normal distribution. *Deutsch & Journel* (1998) provide several programs that will perform all of the required steps for the transformation in the GSLIB software library. In the MGK approach, estimation is performed using OK on the transformed data, again assuming a zeroth order polynomial trend. The NSCORE estimates can then be back-transformed into concentration-space by using the same mapping discussed above.

The multivariate Gaussian assumption should be carefully validated using the methods described by *Goovaerts* (1997) because it has direct consequences on the estimation process. This transformation assumes extreme values have reduced correlation in space (relative to median values). In plume interpolation, often the extreme concentrations are pivotal in properly assessing risks and uncertainty for site management decisions. Assuming a reduced correlation structure for these extreme values can significantly affect risk and uncertainty calculations and lead to serious misrepresentations of the site conditions if the Gaussian assumption is inappropriate.

5.2.1.3 Quantile Kriging within Local Neighborhoods

The distributional dependence of OK and MGK, as well as their sensitivity to highly skewed data, led to the development of non-parametric or distribution-free kriging approaches using indicator (*Journel* 1983) or rank-order (*Journel & Deutsch* 1997) transformations of the data. Indicator kriging uses a binary transformation of sample data where the data value is set to either 1 or zero depending on whether it is greater than or equal to a user-specified threshold value. The indicator approach requires n_q threshold indicator transforms, each of which requires its own variogram and kriging system. *Goovaerts* (1997) states that at least five thresholds are required (i.e., $n_q = 5$) to quantify the local uncertainty of indicator kriging estimates. Correctly quantifying variograms for extreme concentration thresholds is generally not possible in plume interpolation applications due to data limitations (see *Chilès & Delfiner* 1999). Moreover, solving the OK system of equations for each of the thresholds for every point in the interpolation domain represents a severe computational limitation of the method.

These limitations motivated the development of quantile kriging (*Journel & Deutsch* 1997), which transforms concentrations into standardized ranks (or quantiles) using equation (5.2) and computes estimates in quantile space using OK (see *Journel & Deutsch* 1997 and *Juang et al.* 2001).

$$c(\bar{x}) = \frac{rank\ of\ sample\ i}{N+1} \tag{5.2}$$

The transform consists of ranking the concentration data in ascending order and dividing these ranks by one plus the total number of sample data N (for cases with multiple zero measures see the frequency analysis discussion in Appendix D). Figure (5.2) graphically illustrates quantile or standardized ranks transform. The i^{th} sample is assigned the probability that $c(\bar{x})$ is less than or equal to its concentration value [$F(c)$ in Figure (5.2)]. Note that quantiles [i.e., the probabilities

from the empirical cdf $F(c)$] are known to be uniformly distributed; this fact can be used in conjunction with the estimation variances to estimate the local uncertainty at unsampled locations in the interpolation domain. The empirical cdf of the sample data is also used to back transform estimates from quantile space to concentration space (for details see *Juang et al.* 2001).

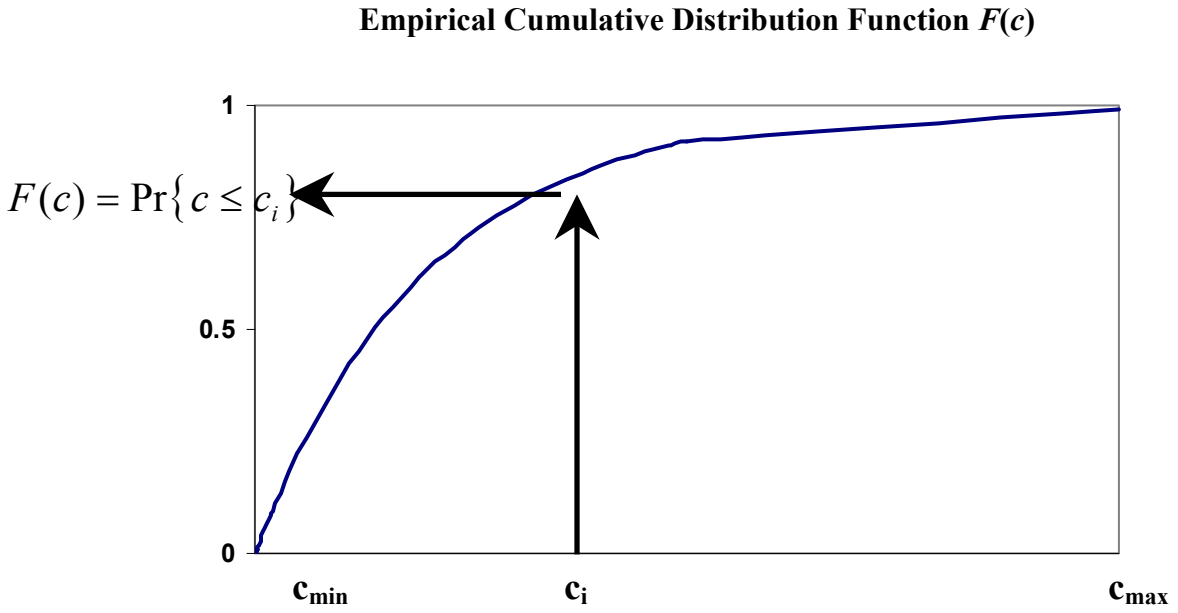


Figure 5.2 Illustration of rank-order transformation used in quantile kriging

This kriging approach has the following limitations: (1) it assumes that the empirical (or sample) cdf attained from available data is reflective of the true distribution of concentrations throughout the interpolation domain and (2) the approach probabilistically models a deterministic quantile space. As with any probabilistic statistic, the quality of the empirical cdf as a representation of the true concentration distribution in space varies as a function of the number of data (i.e., given infinite samples the empirical cdf should in theory approach the true concentration cdf). Note that regardless of its theoretical underpinnings, the primary criteria for evaluating geostatistical approaches such as quantile kriging is how well they work in practice.

This is in line with *Gambolati & Volpi* (1979), who show that the geostatistical approach can be viewed as a deterministic interpolation scheme.

Quantile kriging and MGK are both extensions of the OK approach, which assumes that the trend term in equation (5.1) is modeled as a zeroth order polynomial. The next section discusses both the KT and ItK approaches, which generalize the kriging system to consider higher order polynomial trends.

5.2.1.4 Intrinsic Kriging

The ItK approach was motivated by perceived theoretical shortcomings in both the OK and kriging with a trend (KT) systems. The KT and ItK approaches are closely linked, making it necessary to set the stage for the ItK description with a brief discussion of the KT method and its shortcomings. This is particularly relevant since KT is still actively applied in practice. The KT approach models concentrations as a regionalized variable, $c(\bar{x})$ that is composed of the sum of a smoothly varying mean (trend) component $m(\bar{x})$ and a second order stationary residual $r(\bar{x})$. The KT estimator for contaminant concentration requires the specification of a linear combination of known polynomial functions of spatial coordinates to model the trend component of the concentration [for a detailed derivation see *Goovaerts* 1997]. Given the assumed trend function, the residual of the concentration data and the trend are calculated and structural analysis of the residual is used to derive a residual covariance function $K_r(\bar{h})$. Inference of the residual covariance or variogram is difficult because the concentration data are not directly reflective of the residual function. *Goovaerts* (1997) recommends extending the recommendations of *Journel & Rossi* (1989) for the OK system to the KT approach by using either data perpendicular to the general trend or local neighborhoods with radii defined to have lengths equal to 25 percent of the total length of the interpolated domain. Again, this approach is

not feasible for most real-world contaminant data sets due to data limitations. The OK system is a special case of the KT system where the trend is assumed to be a zeroth order polynomial and is set equal to 1. *Goovaerts* (1997) shows that the only difference between the OK and KT systems results from the practitioner's arbitrary decision to explicitly model the local trend as either a constant or a higher order polynomial.

The KT and variations of OK discussed above all require the practitioner to accept the dichotomy of the regionalized variable representing contaminant concentrations into a smoothly varying mean component and a stochastic fluctuation. Several studies have shown that this assumption is often an arbitrary choice that can significantly bias both estimates and their respective estimation variances (*Volpi & Gambolati* 1978, *Hughes & Lettenmaier* 1981, *Russo & Jury* 1987, *Journel & Rossi* 1989, *Crawford & Hergert* 1997). Intrinsic random function of order k (IRF-k) theory avoids the above dichotomy by not requiring the explicit specification of a model for the mean trend. Instead, the ItK approach defines allowable linear combinations of the sample data, termed *generalized increments*, that are implicitly capable of filtering trends from the data, facilitating a more direct measure of the underlying spatial structure of contaminant concentrations. The approaches discussed in previous sections model spatial structure using pairs of sample data to deduce the variogram shown in equation (5.1). Alternatively, IRF-k theory defines *generalized increments of order k* to be $(k+1)$ point increments that are capable of filtering polynomial trends of degree less than or equal to k (*Chilès & Delfiner* 1999). Generalized increments, as the name implies, generalize the inference of spatial structure from a 2-point to a $(k+1)$ point measure of spatial correlation termed the *generalized covariance* $G(\bar{h})$.

In the ItK approach, concentration is modeled as an IRF-k where the generalized increments of $c(\bar{x})$ are assumed to be a stationary random function with a zero mean and a covariance equal to $G(\bar{h})$. It has been shown that the ItK system is identical to the KT system except for the substitution of the generalized covariance function $G(\bar{h})$ for the 2-point residual covariance function $K_r(\bar{h})$ (for details see *Chilès & Delfiner* 1999). The ItK method was formulated to avoid the estimation biases caused by the explicit specification of a functional trend model. Although the IRF-k approach elegantly unifies geostatistical theory, ItK requires significant expertise to implement relative to the more traditional approaches discussed above. Also, given the common occurrence of limited data sets with irregular sample network geometries for plume characterization, inference of a model for the theoretical generalized covariance function is a difficult process that again introduces bias into estimates and their corresponding estimation variances (see *Hughes & Lettenmaier* 1981 and *Goovaerts* 1997).

With scatter-point data, the traditional graphical approaches used to identify variograms cannot be used for ItK and instead identification of the appropriate generalized covariance model requires the use of "black-box" automated fitting procedures (see *Delfiner* 1976, *Kitanidis* 1983, *Marshall & Mardia* 1985, *Chilès & Delfiner* 1999). In this chapter, the *Delfiner* (1976) method was implemented using the Geovariances ISATIS version 3.4.0 advanced geostatistical software package. *Chilès & Delfiner* (1999) state that this approach "…has the advantage of being completely general, since it does not require the data points to be evenly distributed nor the underlying random function [i.e., contaminant concentrations] to be Gaussian". The method consists of the following tasks: (1) fitting a polynomial of degree less than or equal to *k* and (2) identifying the generalized covariance model with the best cross-validation scores. Cross-validation is performed by removing a "known" data point and estimating its values using the

remaining sample data. Residuals can then be computed as the difference between the data point's true value and its estimate (for more details see the discussion of equation (5.4) in the next section). Readers interested in the details of this fitting algorithm should see both *Delfiner* (1976) and *Chilès & Delfiner* (1999).

5.2.2 Deterministic Methods

Traditional inverse-distance weighting and least squares approaches have been compared to the kriging methods described above. The deterministic methods do not incorporate the underlying spatial structure of the phenomenon of interest and have generally been shown to be more sensitive to asymmetrical data distributions and bias-based estimation error (*Gambolati & Volpi* 1979, *Hughes & Lettenmaier* 1981, *Gotway et al.* 1996). These studies have also found that the relative superiority of kriging approaches is largely a function of the number of available data because, below a certain threshold, spatial structure (i.e., the spatial covariance) cannot be determined.

This chapter considers an array of plume interpolation methods that will enable the practitioner to define the most appropriate estimation technique given the data limitations of his or her site. Deterministic estimation techniques can be used for LTM design applications for smaller sites with limited data sets. *Reed et al.* (2000) also found that the inverse distance weighting techniques are excellent screening tools, which can be used to justify the added cost and analyses required for geostatistical approaches to LTM design. The next two sections present the deterministic interpolation methods considered in this chapter.

5.2.2.1 Nonlinear Least Squares Inverse Distance

The nonlinear least squares (NLS) method studied in this research is a variant of the scheme used by *Barry & Sposito* (1990) in their analysis of tracer plumes at the Borden site

located in Ontario, Canada. The *Barry & Sposito* (1990) interpolation method shown in equation (5.3) was selected because it requires minimal modeling assumptions and it has been successfully applied to three-dimensional historical contaminant data.

$$w(\bar{x}_j, \bar{x}_k) = 1/D(\bar{x}_j, \bar{x}_k)$$

$$D(\bar{x}_j, \bar{x}_k) = \left[\left(x_{j1} - x_{k1}\right)^2 / \alpha_1 + \left(x_{j2} - x_{k2}\right)^2 / \alpha_2 + \left(x_{j3} - x_{k3}\right)^2 / \alpha_3\right]^P$$

$$c_{est}(\bar{x}_j) = \sum_{\psi=1}^{N} w(\bar{x}_j, \bar{x}_k)\, c(\bar{x}_k) / w_t, \ \forall j \tag{5.3}$$

In equation (5.3), the j^{th} concentration estimate $c_{est}(\bar{x}_j)$ at location $\bar{x}_j$ is computed as a weighted sum of the available sample data. Each weight $w(\bar{x}_j, \bar{x}_k)$ is computed as an inverse function of the distance between the j^{th} unsampled location and k^{th} sample raised to the power P. The alpha parameters in equation (5.3) scale the relative importance of each spatial dimension in the distance calculations. Equation (5.4) presents the cross-validation based parameter estimation method that was used to fit the α parameters and distance exponent P shown in equation (5.3) to the contaminant data of each test case (*Barry & Sposito* 1990).

$$Minimize\ S = \sum_{k=1}^{N} \left\{ c(\bar{x}_k) - \sum_{i=1, i\neq k}^{N} w(\bar{x}_k, \bar{x}_i)\, c(\bar{x}_i) / \left(w_t - w(\bar{x}_k, \bar{x}_i)\right) \right\}^2 \tag{5.4}$$

The cross-validation method minimizes the sum of the squared residuals S between each of the actual PCE concentrations $c(\bar{x}_k)$ and their interpolated estimates based on the ($nsamp$ – 1) remaining data $c(\bar{x}_i)$. Equation (5.4) was solved using the Levenberg-Marquardt nonlinear least squares solution method (Moré 1977).

5.2.2.2 Inverse Distance Power-2

The primary strength of inverse distance-based interpolation is its numerical simplicity. Equation (5.3) describes the inverse distance power-2 approach (ID) when the alpha fitting

parameters are set equal to 1 and the distance exponent P is set equal to 2. Additionally, equation (5.3) shows that ID is simply a matrix multiplication of the inverse distance weights $w(\bar{x}_j, \bar{x}_k)$ by the vector of N sampled concentrations. Traditional inverse-distance estimation within local neighborhoods was tested in this chapter because it is commonly used in practice. Inclusion of this method enabled a comprehensive empirical performance analysis of interpolation methods ranging from advanced geostatistical estimators (i.e., the ItK approach) to one of the simplest interpolation methods used in practice.

5.2.3 Defining Estimation Neighborhoods

All of the methods considered in this chapter except for ItK and NLS require the specification of a neighborhood within which the sample data used to estimate concentration values are selected. Specification of an estimation neighborhood is particularly important for scatter-point contaminant data due to the ubiquity of preferential sampling at contaminated sites. Specifically, preferential sampling of areas of high concentration can lead to a bias that causes severe overestimation at unknown locations. In this chapter, each estimator used ellipsoid neighborhoods where, for each case, the axes were defined to be equal to one half of plume lengths along its transverse and longitudinal axes as recommended by *Cooper & Istok* (1988a) and *Goovaerts* (1997).

To reduce clustering effects, each of the ellipsoid neighborhoods were subdivided into octants within which a minimum of 1 and a maximum of 3 of the nearest sampled concentrations are selected to be used in estimation as recommended by *Chilès & Delfiner* (1999). This means that a maximum of 24 points were used to estimate concentrations at unknown locations. If there are no sample data within the estimation neighborhood surrounding an unknown location, then no estimate is computed for that location. The definition of the estimation neighborhood requires

that a balance be struck between the accuracy of estimates and the computational demands required to attain those estimates. For kriging systems, the computational complexity grows as a cubic function of the number of data used for estimation (e.g. doubling the number of data included in the neighborhood results in an 8-fold increase in the computational time required to solve the problem). Inverse distance approaches have the advantage that their computational complexity grows as a quadratic function of the number of data used in estimation. Each test case had a unique estimation neighborhood. To enable fair comparisons among the estimation methods, the same estimation neighborhood was used for all methods applied to a given test case.

5.2.4 Performance Measures

The performance of the interpolation methods was assessed using the following measures of their estimation errors: estimation error residual, standardized estimation error, confidence intervals, and root-mean-square error (RMSE). Each measure is summarized separately below.

5.2.4.1 Estimation Error Residual

The estimation error residual is shown in equation (5.5), which directly measures the difference between true and estimated concentration values. In this work estimation error residuals can be exactly calculated from the true simulation data. In practice, the evaluation of error residuals is approximated by using re-estimation scores, such as with cross-validation.

$$E_{rsd}(\bar{x}) = c_{true}(\bar{x}) - C_{est}(\bar{x}) \tag{5.5}$$

The kriging methods and NLS both seek to minimize these residuals to produce minimum error estimates of contaminant concentrations.

5.2.4.2 Standardized Estimation Error

In addition to computing minimum error estimates, kriging methods also produce estimation variances that can be used to measure the "local uncertainty" of the estimates. The term "local uncertainty" is taken from *Goovaerts* (1997) because uncertainty estimates are computed considering only a single point in the interpolation domain. Estimation variances are used to *standardize* error residuals by dividing the errors computed at each location $\bar{x}$ [shown in equation (5.5)] by the square root of its corresponding estimation variance, as shown in equation (5.6). Equation (5.6) was used to assess the cross-validation performance of each kriging method except QK by classifying estimates as being robust when their *standardized estimation error* falls within the interval [-2, 2], which is equivalent to a 95th percentile confidence interval for the normal distribution.

$$E_{Std}(\bar{x}) = \frac{c_{true}(\bar{x}) - c_{est}(\bar{x})}{\sigma\ (\bar{x})} \tag{5.6}$$

5.2.4.3 Confidence Intervals

A *robust* estimate can be defined as one in which its true value falls within the 95th percentile confidence interval surrounding its estimate, which can be computed using equation (5.7).

$$c_{est}(\bar{x}) \pm 2\sigma(\bar{x}) \tag{5.7}$$

This uncertainty calculation makes the following assumptions: (1) that the estimation errors are normally distributed at location $\bar{x}$ and (2) that $c_{est}(\bar{x})$ and $\sigma(\bar{x})$ represent the distribution mean and standard deviation and are independent. *Goovaerts* (1997) argues that these assumptions are "stringent" and rarely satisfied in practice due to highly asymmetrical sample distributions and spatially correlated estimation errors. Normality assumptions are avoided in the QK approach

because standardized ranks are known to be uniformly distributed allowing confidence intervals to be computed using equation (5.8) [see *Juang et al.* 2001].

$$c_{est}(\bar{x}) \pm \sqrt{3}\sigma(\bar{x}) \tag{5.8}$$

5.2.4.4 Root-Mean-Square Error

Finally, the root-mean-square error (RMSE), shown in equation (5.9), is the last performance measure that was used to rank the relative performance of the estimation methods.

$$RMSE = \sqrt{\frac{1}{n_{est}} \sum_{j=1}^{n_{est}} [c_{est}(\bar{x}) - c_{true}(\bar{x})]^2} \tag{5.9}$$

RMSE represents the average error computed for n_{est} estimates.

5.3 Results & Discussion

The next sections compare the relative performances of the interpolation schemes as the number of available sample data decreases. The geostatistical approaches required both structural analysis and cross-validation to specify models of the spatial "structure" or correlation of the contamination data. Each test case required the specification of 4 variogram models, one for each of the kriging approaches (yielding a total of 12 models among the 3 test cases), which were then cross-validated using analyses of the histograms of standard errors, plots of estimates versus their true values, and measures of estimate robustness (for detailed presentation of these results see Appendices B and C). Additionally, NLS minimizes cross-validation squared residuals to specify the fitting parameters of equation (5.3). Table (5.2) gives the NLS fitting parameters' values for each test case.

	(α_1, α_2, α_3)	**P**
Large Test Case	(1.8, 0.1, 0.4)	30
Intermediate Test Case	(306, 304, 305)	0.5
Small Test Case	(36, 60, 0.5)	1.2

Table 5.2 Nonlinear least squares fitting parameter values

In addition to cross-validation scores, this chapter evaluates the accuracy of estimated PCE concentrations relative to their *true* values using the simulation data from *Maxwell et al*. (2000). For each test case, the interpolation schemes are ranked based on their RMSE values as well as a detailed analysis of the spatial distribution of estimation errors using three-dimensional cross sections along the vertical and horizontal axes of the each plume. These cross-sectional views provide the most comprehensive means of visualizing the spatial distribution of estimation errors. Finally, the discussion of each test case concludes with an analysis of how well the *local uncertainty* estimates attained from the kriging approaches predict areas of high estimation error.

5.3.1 Large Test Case

The large test case represents a heavily sampled, mature PCE plume that extends over 1-km in the primary direction of groundwater flow within a highly heterogeneous alluvial aquifer. In this test case, 124 sample locations were available from 59 multi-level monitoring wells for the structural analysis required for each kriging system. Variogram models for the OK, MGK, and QK schemes were deduced using the standard trial-and-error approach of graphically fitting candidate models to experimental variograms followed by cross-validation to determine the models' performances.

The estimates from these kriging approaches were significantly improved when *geometric* anisotropies where incorporated into their variogram models (for detailed discussions of anisotropic variogram modeling see *Chilès & Delfiner* 1999 or *Goovaerts* 1997). These anisotropic models indicate that the PCE has shorter correlation ranges in the y- and z-directions (or in the transverse directions of the plume) relative to the x-direction (or along the longitudinal axis of the plume). Moreover, the vertical correlation range was shorter relative to both of the ranges specified in x- and y-directions as expected in an alluvial aquifer. For ItK, the second

order polynomial trend function ($1+z+z^2$) and an isotropic generalized covariance model were specified by black-box fitting (*Delfiner* 1976). Due to both computational limits and improved cross-validation performance, ItK estimates were computed using local neighborhoods. The next section overviews the cross-validation results for the kriging approaches and NLS.

5.3.1.1 Cross-Validation

Figure (5.3) shows the RMSE values attained from cross-validation for 120 of the 124 available sample data points. Four PCE concentrations, each of which exceeded 1000-mg/m^3 and were located in the source area of the plume are not included because these extremes severely impeded structural analysis for both OK and ItK. The calculation of the experimental variogram is very sensitive to extreme concentration values and in this case no spatial structure could be identified when these values where kept in the data set. The sensitivity of both variogram calculations and the kriging estimators to highly variable, strongly skewed concentration data is a motivating factor for the use of MGK and QK.

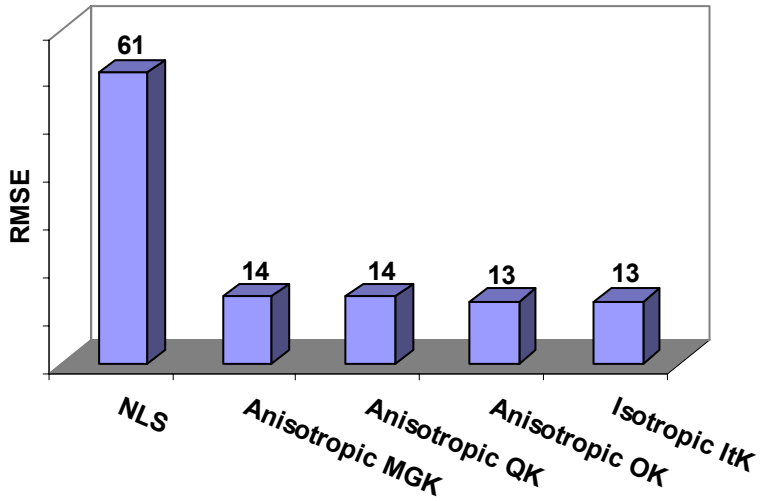


Figure 5.3 RMSE values from cross-validation for the large test case (mg/m^3)

The cross-validation results for MGK and QK presented in Figure (5.3) exclude the 4 largest samples to allow these methods to be compared to OK and ItK. However, these samples

were included in both the structural analysis and cross-validation used to specify their variogram models. Figure (5.3) shows that there is only a slight difference between the RMSE values for all of the kriging methods. These cross-validation scores must be considered cautiously. Although OK and ItK have slightly smaller RMSE values, these methods could not utilize the full data set in their cross-validation estimates, in particular the maximum concentration values that can be important in plume interpolation. Moreover, 80-percent of the monitoring wells in this test case sample multiple locations in space, a majority of which fall within 10-meters of each other. The close proximity of these sampling locations to one another greatly reduced the mean distance between unknown and known locations in the cross-validation calculations. The magnitudes of cross-validation RMSE scores are directly proportional to the mean distance between known and unknown locations and in this case are underestimated.

The RMSE values in Figure (5.3) clearly show that NLS has the highest expected error of the five methods discussed in this section. Again, caution must be exercised before drawing the conclusion that NLS is the worst of the methods. Kriging is a *low-pass filter* that by means of statistical expectation (or averaging) tends to "filter" extreme values from its estimates. Note that the distance exponent P shown in Table (5.2) for the large test case is equal to 30, which causes NLS to behave essentially as a nearest neighbor estimator. This yielded a high degree of variability in the NLS estimates. The actual quality of NLS estimates are a function of how well the 124 concentration samples represent the true site conditions, which will be captured by the estimation accuracy analysis below.

The four kriging approaches were also assessed in terms of the robustness of their estimates (i.e., the percentage of the estimates with standardized errors falling in the interval [-2, 2]). Conservatively, over ninety-eight percent of the MGK and QK estimates were robust. Only

90-percent of the OK and ItK estimates were classified as being robust due to their sensitivity to extremes, even without the 4 highest PCE samples.

Overall, cross-validation scores failed to identify which interpolation method is superior. This failure resulted due to clustered, multi-level monitoring points that preferentially sample zones of high concentration causing the cross-validation RMSE to be significantly underestimated, especially for the interpolation methods that are sensitive to extreme PCE concentrations.

5.3.1.2 Estimation Accuracy

The accuracy of interpolation estimates were tested relative to their *true* values for every point in the interpolation domains. Table (5.3) gives the grid dimensions of the interpolation domains.

	(xmin, ymin, zmin)	**(Δx, Δy, Δz)**	**(nx, ny, nz)**
Large Test Case	(20m, 450m, 60m)	(20m, 24m, 4.8m)	(59, 12, 9)
Intermediate Test Case	(20m, 474m, 61.2m)	(20m, 24m, 4.8m)	(39, 8, 8)
Small Test Case	(20m, 486m, 75.6m)	(10m, 12m, 4.8m)	(38, 11, 5)

Table 5.3 Interpolation grid specifications for each test case

Figure (5.4) gives the RMSE over the 6732 grid points in the large case domain for each interpolation method. The figure clearly shows the relative performances of the 6 interpolation methods considered in this chapter. The cross-validation RMSE values shown in Figure (5.3) predicted that the 4 kriging approaches should perform equally well, whereas Figure (5.4) shows that ItK and OK clearly fail to estimate PCE concentrations as accurately as QK and MGK. QK performed the best, with an absolute average error for PCE estimates of 8-mg/m^3.

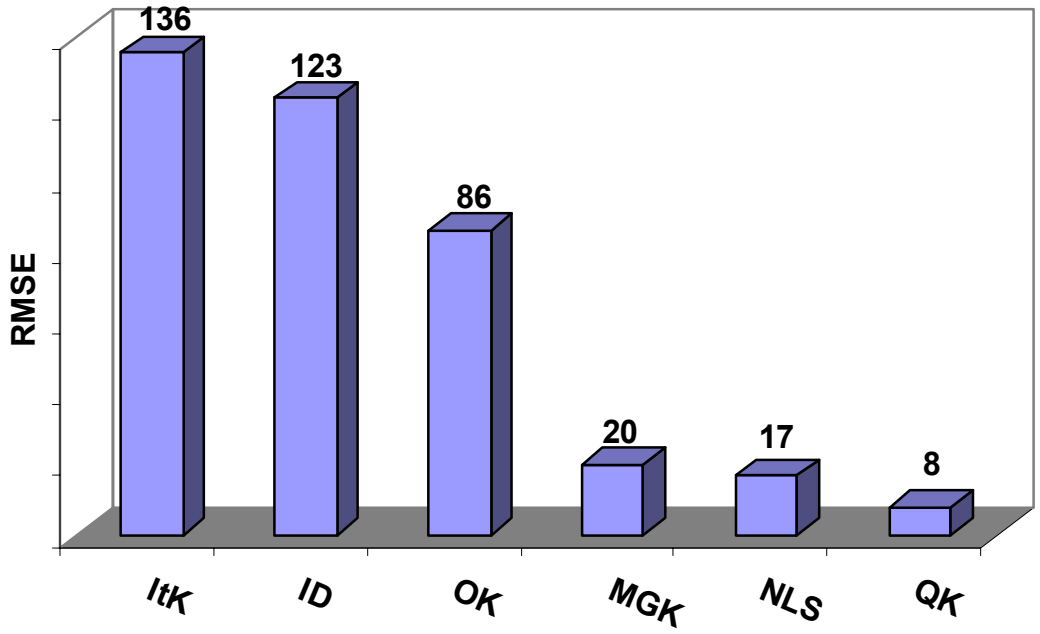


Figure 5.4 RMSE values from interpolation for the large test case (mg/m^3)

Figure (5.5) shows how the estimation errors are distributed throughout the interpolation domain for each of the interpolation methods shown in the order from the most to the least accurate approach. Figure (5.5a) shows that almost all of the QK estimates fell within ± 10-mg/m^3 of their true values. The figure shows that the zone of highest error occurs in the midsection of the plume centered at about 600-m in the x-direction. There are two primary factors contributing to the errors in this portion of the plume: (1) there is a reduced number of sample data available in this portion of the plume and (2) a group of monitoring wells clustered in this portion of the plume had PCE concentrations greater than 50-mg/m^3 but were not representative of the concentrations in this portion of the plume. Overall, the QK approach was fairly robust given the relatively poor spatial coverage of the monitoring wells in the midsection of the plume and preferential sampling of high concentrations throughout the plume.

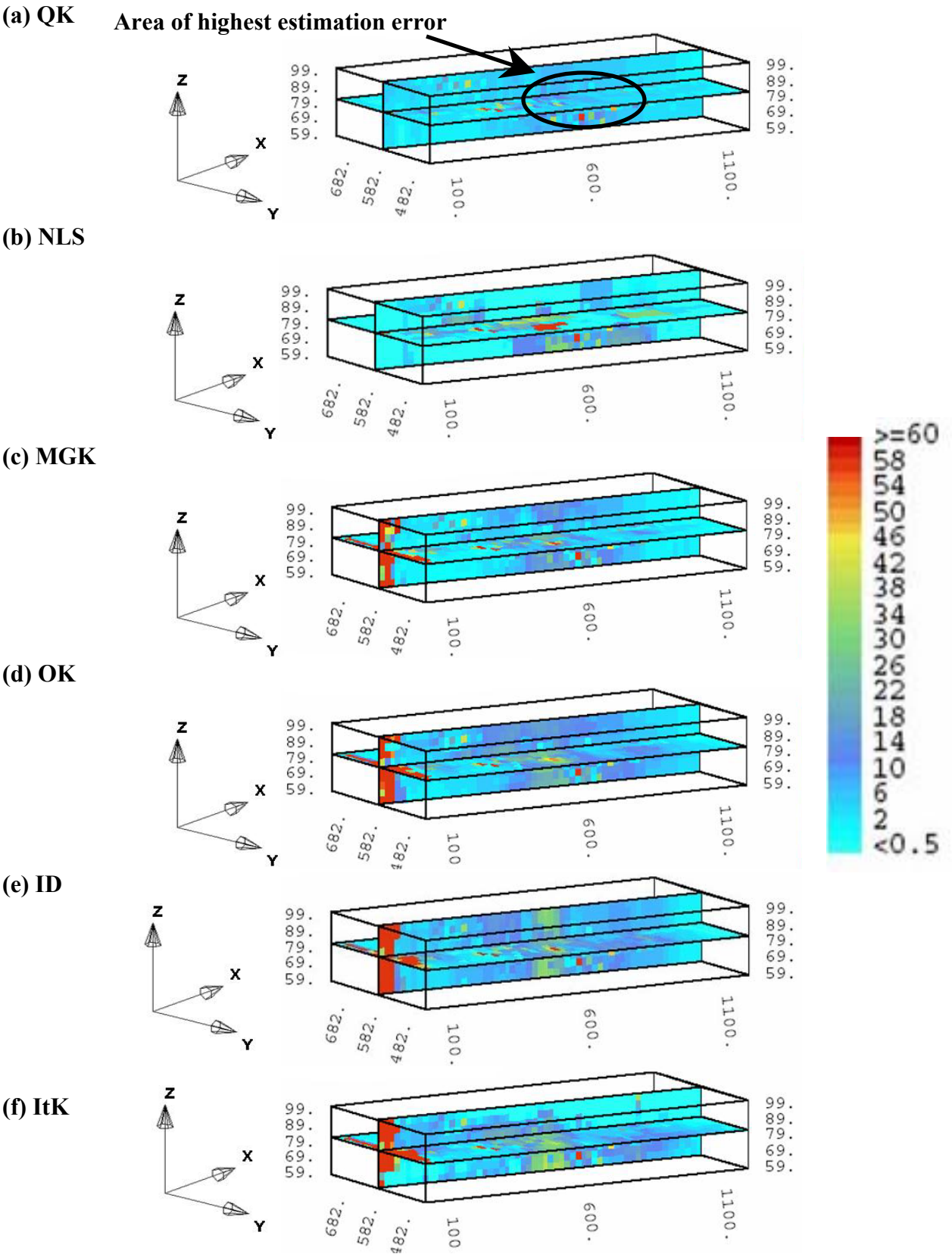


Figure 5.5 Spatial distributions of estimation errors for the large test case (mg/m3)

Figure (5.5b) shows that NLS actually outperforms QK in estimating the boundaries of the plume. The method's increased RMSE relative to QK resulted from an increased sensitivity to the poor spatial coverage and preferential sampling of high concentrations in the midsection of the plume. Figures (5.5a) and (5.5b) show NLS's sensitivity to these factors, clearly illustrating increased estimation errors in the midsection of the plume, many of which exceed $\pm$ 20-mg/m^3.

The spatial distribution of errors for MGK illustrated in Figure (5.5c) shows increased estimation errors in the midsection of the plume relative to both NLS and QK. Far more troubling, the figure shows that the MGK approach fails to produce accurate estimates in the source area of the plume. The extreme values within the plume's source area were considered during both structural analysis and cross-validation for MGK. However, the high estimation errors shown in Figure (5.5c) within the source area of plume occur because MGK implicitly assumes that (1) the data are symmetrically distributed and (2) they display a *destructurization effect*, which means that the extreme values are not highly correlated in space. The first assumption fails because of the tremendous variability of the PCE concentration data; even after transformation the data set remained skewed and asymmetrically distributed. The second assumption fails because both tails of the data cdf (i.e., the zero concentrations and the source area concentrations) are highly correlated in certain regions of the plume. Using the available software in the GSLIB library of *Deutsch and Journel* (1997), the PCE concentration data sets from all 3 test cases failed to match the theoretically expected behavior of a Multi-Gaussian phenomenon. These results are particularly important given the common use of normal and lognormal transforms within the groundwater community.

Figure (5.4) shows that the three remaining interpolation schemes had significantly higher RMSE values relative to QK, MGK, and NLS. The 4 extreme PCE samples located

within the source zone severely impeded the ability of OK, ID, and ItK to accurately estimate PCE concentrations within this zone of the plume. This is rather unsettling given that the RMSE values for OK and ItK from cross-validation do not accurately reflect the true performance of these methods. Figures (5.5d) – (5.5f) show the distributions of errors in the plume for each method. Figure (5.4) shows that ItK had the highest expected error of the six interpolation methods. Figure (5.5f) shows that the increased RMSE for ItK results from the method's failure to accurately estimate PCE concentrations in the source area and midsection of the plume. Although ItK is elegantly formulated for working with nonstationary phenomena, it was heavily biased in the present case by extreme concentration values; this result concurs with *Hughes & Lettenmaier* (1981). Both deterministic interpolation schemes (NLS and ID) perform nearly as well or better than OK. Given the ease with which these methods can be implemented, they have significant potential for use as screening tools (see *Reed et al*. 2000 for an example in monitoring design), although they have the drawback of not providing any measure of estimate uncertainty.

5.3.1.3 Local Uncertainty Analysis

Table (5.4) summarizes how well the kriging approaches were able to capture the true PCE concentrations within the 95th confidence intervals computed using the estimation variance for each estimate as shown in equations (5.7) and (5.8).

	QK	**MGK**	**OK**	**ItK**
Large Test Case	99	95	92	92
Intermediate Test Case	100	100	66	76
Small Test Case	100	100	100	89

Table 5.4 Percentage of true data falling in 95th confidence interval

For the 6372 points in the large test case's interpolation domain, all of the kriging methods were able to capture greater than 90-percent of the true data values within the estimated 95th percentile

confidence intervals. OK and ItK fell slightly below the expected value of 95 percent while QK exceeded this value by capturing 99-percent of the true data in its uncertainty estimates, suggesting that the uncertainty estimates for QK were conservative.

In addition to the results of Table (5.4), Figure (5.12a) shows the spatial distribution of predicted uncertainty computed for QK for the large test case. The figure plots the *interquantile ranges*, which are equal to the difference between the upper and lower bound estimates for the 95^{th} percentile confidence intervals at every point in the 3 interpolation domains considered in this chapter. The larger the distance between these values, the more uncertainty that is associated with an estimate. QK was used because the method was the least affected by extreme concentrations in both structural analysis and estimation. Moreover, QK uncertainty estimates have the advantage of being non-parametric. The purpose of the plots is to verify if zones of high predicted uncertainty correspond with zones of high estimation error. This comparison is important for applications such as site characterization or long-term monitoring where local uncertainty estimates are used to make regulatory or managerial decisions. Comparison of Figures (5.12a) and (5.5a) shows that the area of highest predicted uncertainty is centered at x = 600m [Figure (5.12a)] and corresponds well with the area of highest actual estimation errors [Figure (5.5a)]. The absolute estimation errors were highly correlated with the predicted estimation errors from kriging (i.e., the kriging estimation variances) for the 6372 grid points in the large case domain as evidenced by a rank correlation coefficient equal to 0.98 for the two quantities. The numerical values of kriging estimation variances did not predict the "actual" error values (see *Goovaerts* 1997, *Journel & Rossi* 1989), but they did successfully rank which areas of the plume have the highest expected error (i.e., areas with fewer sampling locations had both increased kriging estimation variances and actual estimation errors). In this test case, QK was

able to both successfully capture the true PCE concentrations and spatially predict the zone of highest estimation errors.

5.3.2 Medium Test Case

The medium test case shown in Figure (5.1b) consists of a total of 58 sample locations within a network of 29 multi-level monitoring wells. Reducing the data set size from 124 to 58 significantly impaired structural analysis. Anisotropy could no longer be identified for any of the kriging methods. Isotropic variogram models where specified for OK, MGK, and QK. The black-box fitting procedure used for ItK was unable to identify any correlation structure in the dataset and, as a result, was specified as a pure nugget variogram model. The fitting procedure failed to identify a model with structure due to an increase in the number of extreme PCE samples and a decrease in the overall number of data available for estimation.

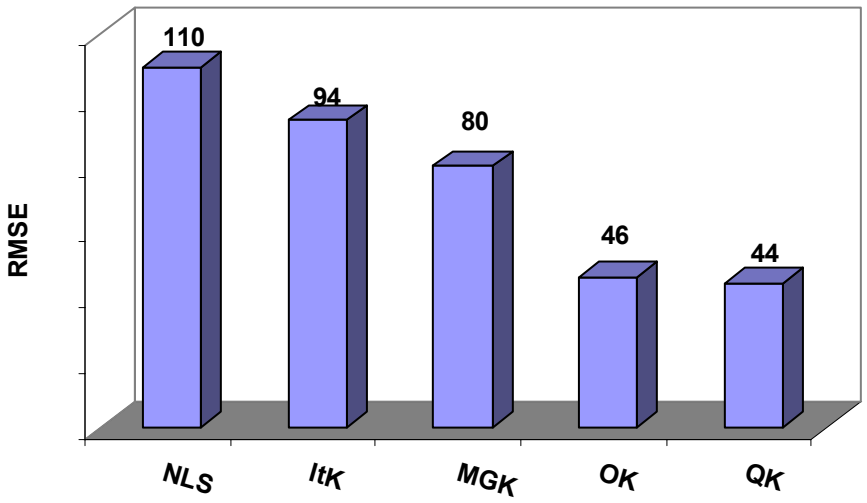


Figure 5.6 RMSE values from cross-validation for the medium test case (mg/m^3)

5.3.2.1 Cross-Validation

The cross-validation RMSE values shown in Figure (5.6) are significantly higher than those shown in Figure (5.3) for the large test case. Three factors contributed to the increased

cross-validation errors for this test case: (1) the reduction in the spatial coverage and total number of the sample data, (2) the sample data set is more heavily skewed toward high PCE concentrations, and (3) the proportion of monitoring wells that sample multiple locations along their vertical axis decreased from 80-percent for the large case to 66-percent for this case, which causes the average distance between known and unknown locations to increase. OK and ItK were again severely affected by high PCE concentrations and required the exclusion of 7 samples with values ranging from 500-mg/m^3 to 6000-mg/m^3. To facilitate performance comparisons with the 3 remaining interpolation methods, only the cross-validation residuals for 51 of the 58 available sample data points were used to compute the RMSE values shown in Figure (5.6). QK is clearly the superior interpolation scheme because the structural analysis performed for this approach did not require the exclusion of high PCE concentrations. Interestingly, the relative rankings of the 5 schemes changed significantly between large and medium test cases. Particularly, the RMSE scores for MGK and ItK increased significantly because both methods drastically overestimated concentrations for samples with no PCE that are in close proximity to high concentration samples in the source area of the plume. The next section shows that cross-validation again failed to properly rank the relative performances of the interpolation schemes.

5.3.2.2 Estimation Accuracy

Figure (5.7) presents the absolute errors for each of the six interpolation methods averaged over the 2496 points in the interpolation domain [see Table (5.3)]. Comparing the medium case RMSE values from cross-validation and interpolation, it is evident that cross-validation is again underestimating the expected errors for all of the interpolation methods. As was observed for the large test case, clustered sampling locations in areas of high concentration are biasing cross-validation residual calculations. This is particularly true for NLS, OK, and ItK,

all of which were heavily biased by extreme PCE concentrations. QK and MGK performed better than the remaining methods because their data transformations reduce the influence of the extreme PCE values on variogram identification and spatial overestimation.

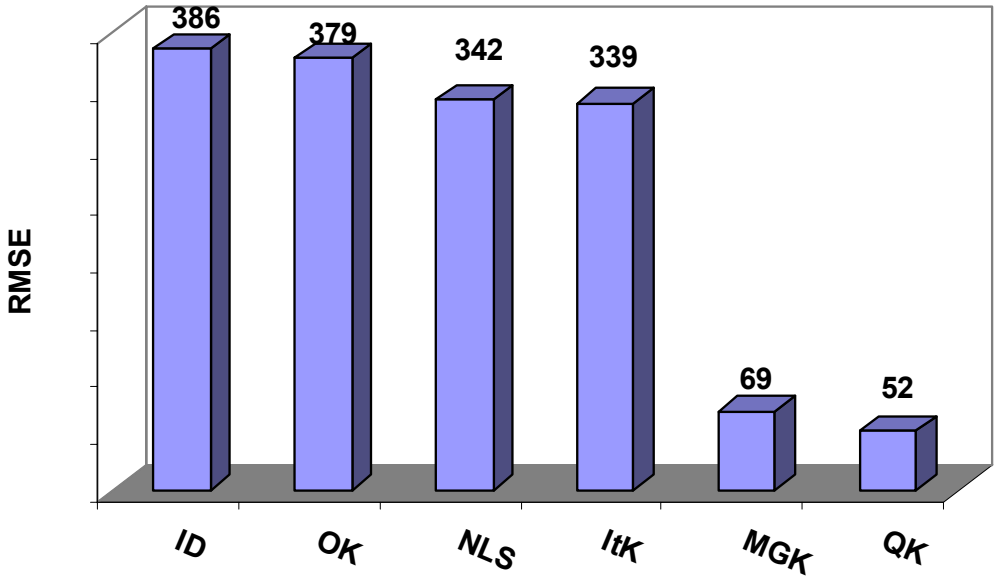


Figure 5.7 RMSE values from interpolation for the medium test case (mg/m^3)

Figures (5.8a)-(5.8f) show the spatial distributions of errors for each of the interpolation schemes. These figures clearly show that QK has the best distribution of error, with a majority of the errors falling below ± 25 -mg/m^3. MGK again fails to accurately estimate concentrations within the source area of plume due to the inappropriateness of the Multi-Gaussian assumption. Figures (5.8c) – (5.8f) show that the distributions of error for ItK, NLS, OK, and ID are all very similar. The methods all failed to produce accurate estimates in the source area of the plume and severely overestimate PCE concentrations in the area centered at x = 500m located near the bottom of the interpolation domain. This bias towards overestimation for ItK and OK also degraded the quality of their local uncertainty estimates as will be shown in the next section.

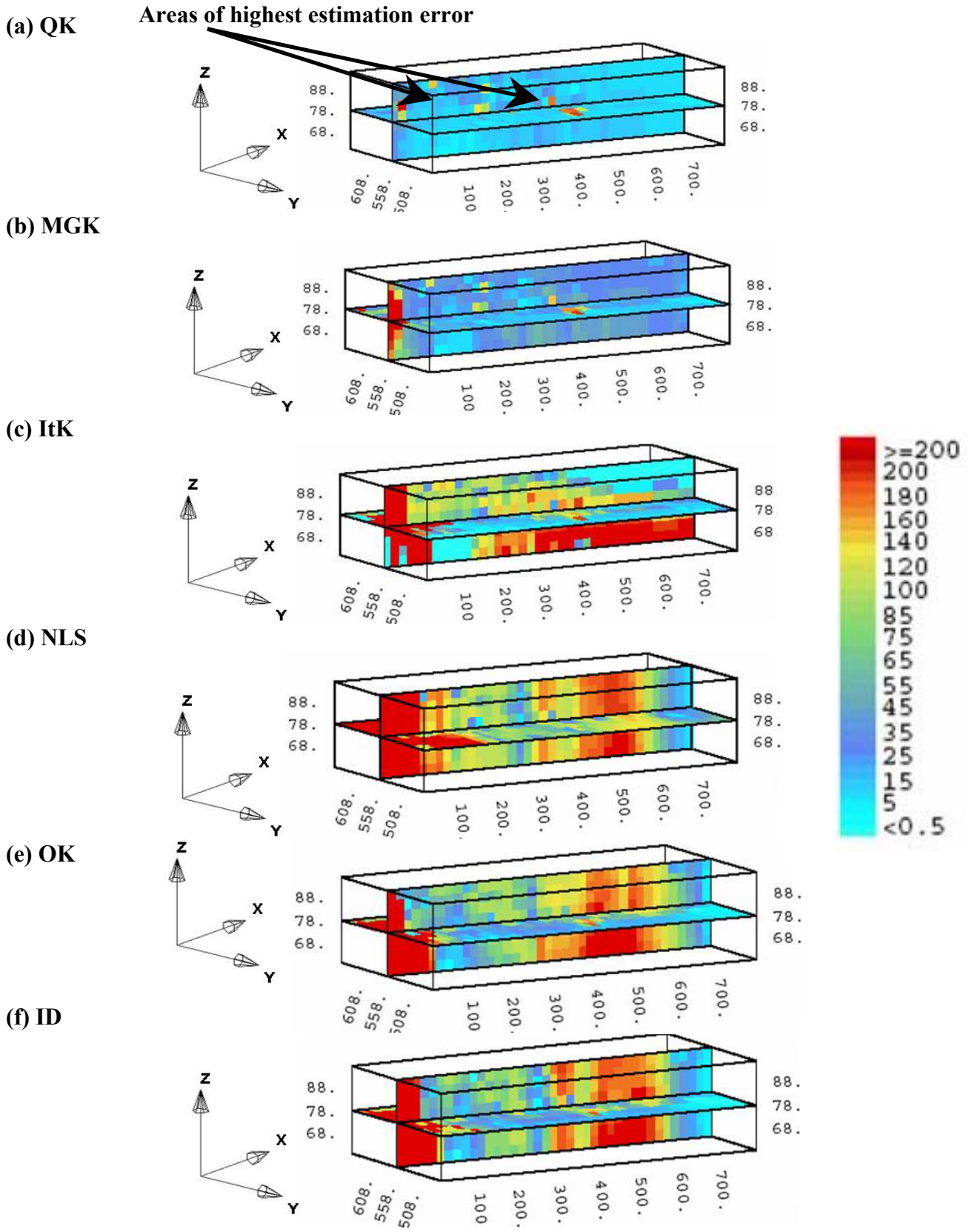


Figure 5.8 Spatial distribution of estimation errors for the medium test case (mg/m3)

5.3.2.3 Local Uncertainty Analysis

Table (5.4) shows that for this case, ItK and OK local uncertainty estimates were severely affected by estimation bias, capturing only 76- and 66-percent of the true data, respectively. Both methods underestimated the true extent of the uncertainty of their estimates. *Goovaerts* (1997) states that highly asymmetric data sets can significantly degrade local uncertainty estimates for kriging methods that require the assumption that estimation errors are normally distributed and spatially independent. The estimation errors for OK and ItK shown in Figures (5.8c) and (5.8e) were highly asymmetrical and correlated in space. Given that ItK and OK both failed to capture the true data with their local uncertainty estimates, these methods have no clear advantage over ID and NLS for this test case.

MGK and QK both produced conservative uncertainty estimates and captured all of the true data in their 95th percentile confidence intervals. Moreover, a comparison of Figures (5.12b) and (5.8a) again shows a close correspondence between areas projected to be highly uncertain and zones of high estimation error for QK with a rank correlation coefficient equal to 0.97.

5.3.3 Small Test Case

The small test case represents a sparsely sampled PCE plume extending approximately 400-m in the primary direction of groundwater flow, with a total of 20 multi-level monitoring wells sampling 26 locations. The high variability of the PCE concentrations and extremely limited size of the small test case data set combined to make identification of spatial structure impossible for OK, QK, and MGK. Pure nugget variogram models were specified for each of these methods. The black-box fitting procedure used for structural analysis in ItK was able to specify an isotropic general covariance model because this method did not use estimation neighborhoods (i.e., all 26 sample data were used in all of the fitting procedure's computations).

ItK could not use estimation neighborhoods for this test case because there were too few data in each neighborhood to fit a trend function. The limited size and heavily skewed nature of the small test case data set helped to identify which of the schemes were the most robust in facing these challenges.

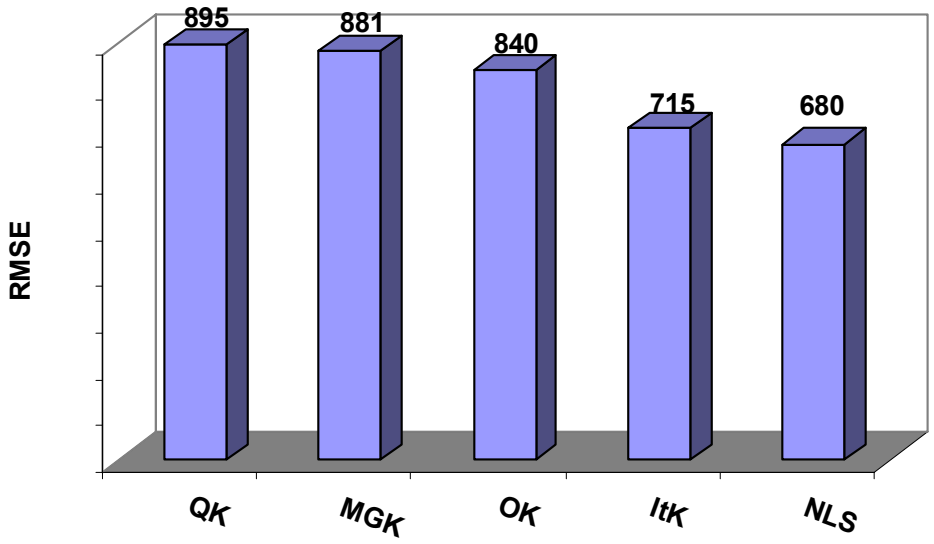


Figure 5.9 RMSE values from cross-validation for the small test case (mg/m^3)

5.3.3.1 Cross-Validation

Figure (5.9) shows that all of the interpolation methods had significantly higher cross-validation RMSE values for the small test case relative to the previous cases. The increased RMSE scores for this case resulted because the size and spatial coverage of this data set decreased while the proportion of extreme PCE samples increased relative to the large and medium test cases. Additionally, only 30-percent of monitoring wells for this test case sample multiple locations in space, which greatly increases the average distance between unknown and known locations in cross-validation calculations relative to the previous two test cases. Recall that as this average distance increases, RMSE increases.

The relative ranking of the interpolation schemes given in Figure (5.9) agrees with previous studies that conclude that kriging methods have no significant advantage relative to NLS and ID for data sets with less than 50 samples (*Hughes & Lettenmaier* 1981, *Gotway et al.* 1996). Interestingly, ItK produced more accurate cross-validation estimates, but the method's estimates were less robust (i.e., its local uncertainty estimates failed to capture the true PCE concentrations) relative to MGK and QK. Recall *Hughes & Lettenmaier* (1981) found that ItK suffers from a bias that causes estimation variances to be smaller then they should be for reduced data sets. However caution must be exercised when assigning performance rankings using cross-validation, because the next section again shows that the cross-validation scores for this test case do not reflect the true performances of the interpolation schemes.

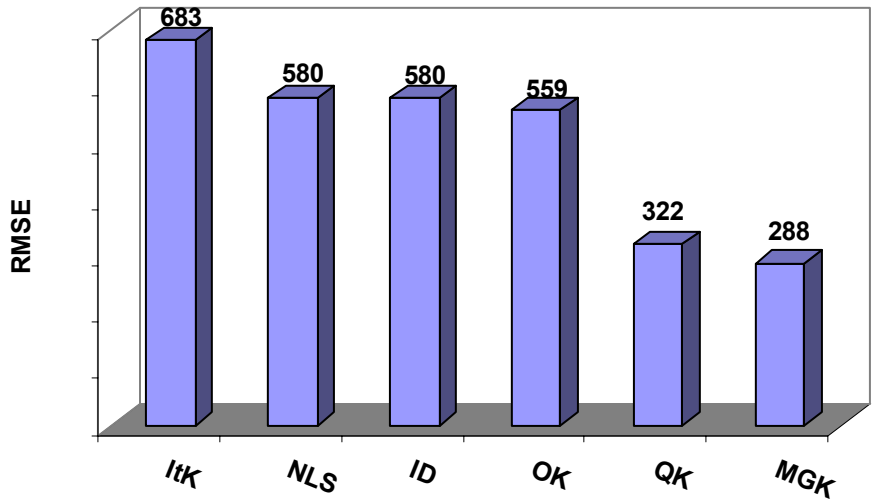


Figure 5.10 RMSE values from interpolation for the small test case (mg/m3)

5.3.3.2 Estimation Accuracy

Figure (5.10) presents the RMSE values computed relative to the true concentrations for the 2090 points in the small case interpolation domain. Comparison of the interpolation RMSE scores in Figure (5.10) with the cross-validation scores in Figure (5.9) shows that unlike the

previous cases, cross-validation actually overestimates the expected errors for all of the interpolation methods. The primary factor causing this shift from underestimation to overestimation lies in the reduction of the percentage of multi-level monitoring wells from 80-percent for the large case to 30-percent for the small case. Reducing the number of multi-level wells greatly increased the average distance between known and unknown locations in cross-validation computations, which resulted in a bias toward overestimating estimation error.

Figure (5.10) shows that MGK and QK were actually the best performing methods, in contrast to the worst performing methods as predicted in cross-validation [see Figure (5.9)]. Figure (5.11) provides a more detailed understanding of the relative performances of these methods, showing the spatial distributions of their estimation errors. Unlike the previous two test cases, QK had a higher RMSE value relative to MGK because the method overestimated zero concentrations near high concentrations in the source area of the plume as can be seen in Figure (5.11b). QK's performance was degraded because over 40-percent of the PCE sample data used to compute the empirical cdf had values exceeding 500-mg/m^3. Despite this fact, Figures (5.11a) and (5.11b) show that QK was less prone to overestimation than MGK in the leading edge of the plume. Finally, Figures (5.10) and (5.11) show there was a minimal difference in the performance of OK, ID, and NLS. Given ID's ease of implementation, it would be the preferred estimation method for practitioners who do not require uncertainty estimates.

5.3.3.3 Local Uncertainty Analysis

Table (5.4) shows that OK, QK, and MGK all captured 100-percent of the true data in the uncertainty estimates for the small test case. All of these methods produced very conservative uncertainty estimates because of their pure nugget variogram models (i.e., assumption of no spatial correlation), when spatial correlation is actually present in the spatial distribution of the

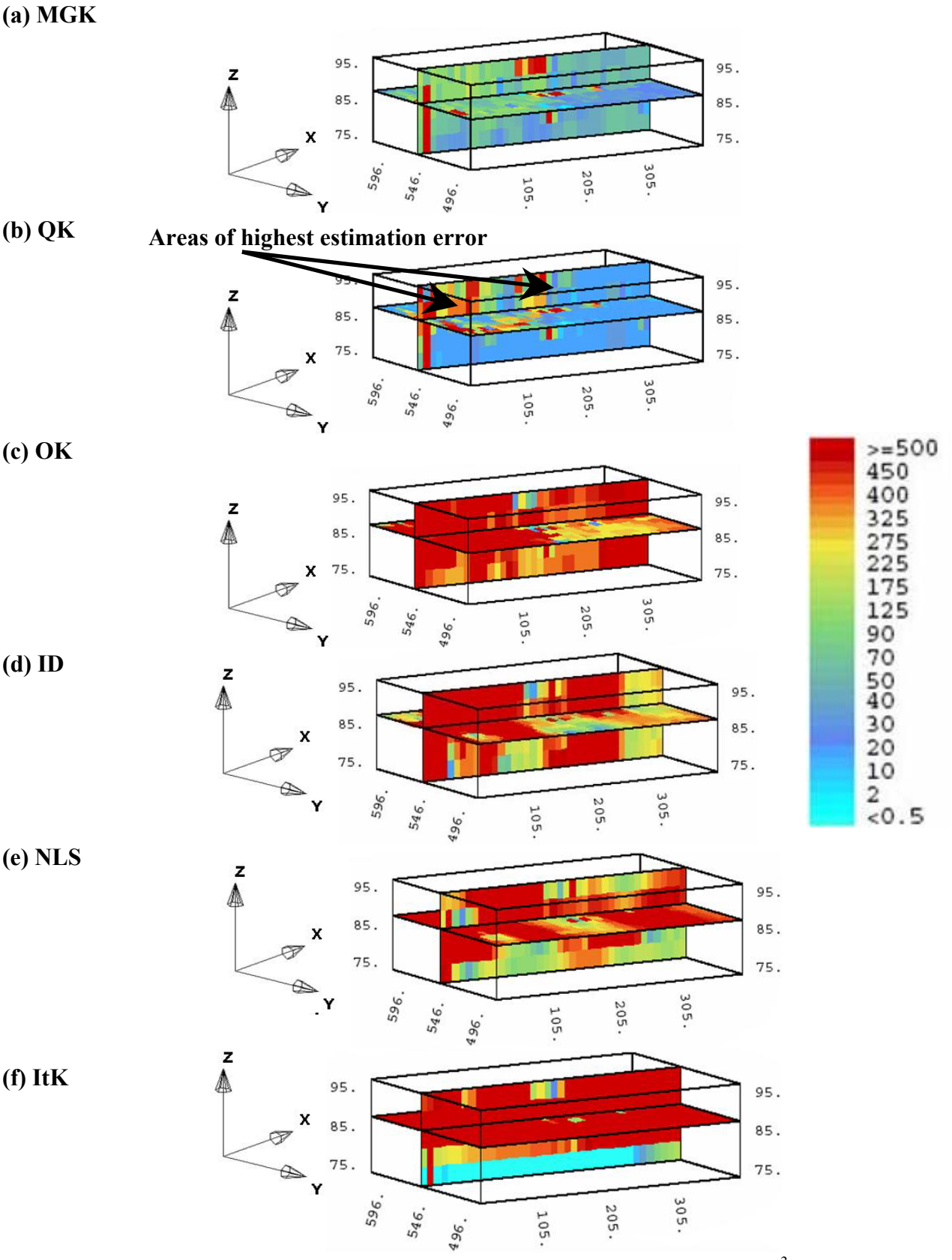
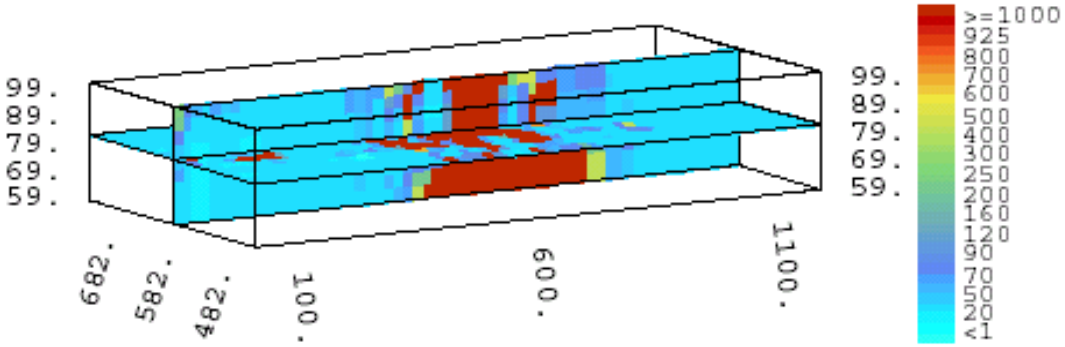
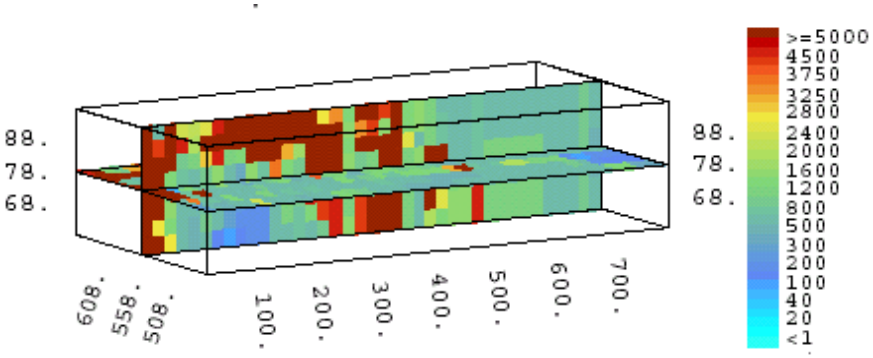


Figure 5.11 Spatial distribution of estimation errors for the small test case (mg/m^3)

(a) Large Case Uncertainty:



(b) Medium Case Uncertainty:



(c) Small Case Uncertainty:

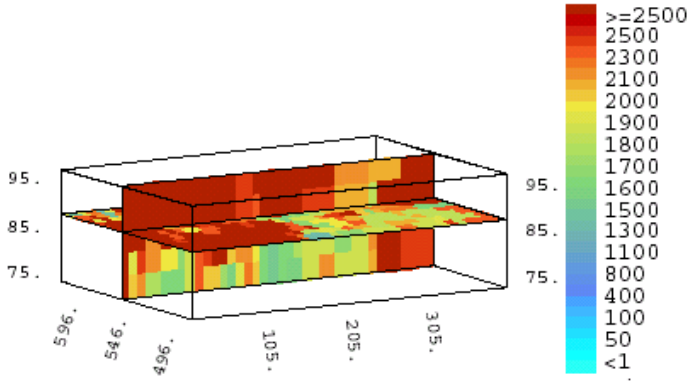


Figure 5.12 Spatial distributions of local uncertainty for QK (mg/m^3)

plume but is not quantifiable from the sample data. ItK was only able to capture 89-percent of the true data due to its susceptibility to under predicting estimates' uncertainties for small data sets. Figure (5.12c) is particularly interesting because even for a small data set where PCE concentrations are assumed spatially uncorrelated, the predicted zones of highest uncertainty are still correlated with the actual estimation errors shown in Figure (5.11b) with a rank correlation coefficient between the two quantities equal to 0.97.

5.4 Conclusions

This chapter provides a detailed description of the motivations and assumptions underlying each of 6 interpolation methods and an analysis of their performance for plume interpolation. Results from the 3 test cases show how robust the interpolation methods are as the number of sample data decreases and illustrate the difficulties practitioners face when interpolating contaminant concentrations at highly heterogeneous sites. The use of simulation data allowed this chapter to examine the spatial distributions of estimation accuracy and uncertainty predictions. An important result from this chapter is that performance ratings from cross-validation are misleading and should be used cautiously. Specifically, the chapter shows that clusters of monitoring wells used to preferentially sample zones of high concentration heavily bias cross-validation scores.

Although QK and MGK were the methods least affected by preferential sampling of high PCE concentrations, MGK's normality assumptions were inappropriate for all of the heavily skewed concentration data sets as evidenced by the method's failure to accurately estimate source area concentrations for all three test cases. The only significant advantage OK and ItK had over ID and NLS was their ability to produce uncertainty estimates. ItK had two primary drawbacks relative to the other kriging methods: (1) traditional graphical methods for structural

analysis could not be used for scatter-point concentration data and (2) it underestimated the uncertainty of its estimates in sparsely sampled areas of the contaminant plume. Both ItK and OK were heavily biased by preferential sampling and extreme concentrations. For all of the data sets, the reduced computational complexity and ease-of-implementation of NLS and ID make these methods useful as screening tools to determine if the expense of a geostatistical study is warranted for a data set. When selecting between these two methods, it is important to consider that NLS was only able to outperform ID when more than 50 sample data were available. Overall, the variability of PCE samples and preferential sampling controlled how well each of the interpolation schemes performed. QK was the most robust of the interpolation methods, showing the least bias from both of these factors and has the additional advantage of being non-parametric. Moreover, the method's non-parametric uncertainty estimates successfully predicted zones of high estimation error for each test case. Quantile kriging was the most robust of 6 groundwater plume interpolation methods, showing the least bias from both preferential sampling and the variability of contaminant data. These findings warrant further studies into the applicability of quantile kriging to data sets from other fields ranging from mining to life sciences, where nonstationary interpolation also plays a vital role.