THE M-SCALE MODEL: A MULTI-SCALE MODEL FOR DECISION SUPPORT OF ON-SITE REMEDIATION

by

Meng-Ying Li

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Environmental Engineering)
in The University of Michigan
2008

Doctoral Committee:

Professor Peter Adriaens, Co-Chair

Professor Anna M. Michalak, Co-Chair

Professor Daniel G. Brown

Professor Steven J. Wright

Noémi Barabás, Limno-Tech Inc.

To Hui-Ju and Jen-Yun

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

Contaminated sediment is defined by the U. S. Environmental Protection Agency (U. S. EPA) as "soils, sand and organic matter or minerals that accumulate on the bottom of a water body and contain toxic or hazardous materials that may adversely affect human health or the environment." (U. S. EPA, 1998). The sediment environment, particularly surface water sediments, is characterized by three major factors that impact the effectiveness and economics of remedial decision-making:

- Diffuse contamination: Contaminant sources in sediments are derived from, among others, direct industrial discharge and combined sewage overflows from urban areas (point sources), contaminant transport via the groundwater-surface water interface, agricultural run-off, and exchanges at the air water interface (diffuse contamination). Once released in the sediments, the contaminant is distributed over wide areas as a result of the hydrodynamics of the system, including tidal effects, storm events, and shipping. These events further exacerbate the problem because point sources become diluted, contaminants are

exported over large areas, and the various point and diffuse sources mixed in time and place.

- Large areas of impact: According to an estimate by the U. S. EPA, approximately 10% or 1.2 billion cubic yards of the sediment underlying the country's surface water is sufficiently contaminated with toxic pollutants to pose potential risks to fish and to humans and wildlife that eat fish (U. S. EPA, 1998). This represents the upper five centimeters of sediment where many bottom dwelling organisms live, and where the primary exchange processes between the sediment and overlying surface water occur. For example, an assessment of 1,543 surficial sediment samples collected during 1991 through 1997 in 25 estuaries and marine bays representing a total area of approximately 7300 km2 indicated toxicity in approximately 6% of the combined area (Long, 2000). Adverse ecological effects from contaminated sediments include fin rot, increased tumor frequency, and reproductive toxicity in fish. In addition, contaminated sediments can also pose a threat to human health when pollutants in sediments accumulate in edible aquatic organisms (U. S. EPA, 1998, and reference therein). Sediment contaminants include polycyclic aromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs), and various metals and metalloids (e.g. arsenic). These contaminants are the most frequently reported contaminants in sediments (U. S. EPA, 1998), and often dominate the ecological and human health risk associated with contaminated sediments.

- Ecological risk via the foodchain: Ecological and human risk associated with the sediments ecosystem is imparted by the uptake from, and propagation of contaminants through the trophic foodchain. In dynamic river systems, effective and sustainable risk management of sediments, contaminants and their sources must be carried out on a river basin scale. The primary concern here

is associated with hydrophobic contaminants and metals or metalloids because of either their acute toxicity or propensity to bioaccumulate from benthic organisms to higher fish species. Comprehensive ecological risk assessments are generally used to: i) identify and characterize the current and potential threats to the environment from a hazardous substance release, ii) evaluate the ecological impacts of alternative remediation strategies, and iii) establish cleanup levels in the selected remedy that will protect those natural resources at risk.

The consequence is that typical site management questions attempt to address the following issues:

- What are the major pathways contributing to ongoing exposure?
- What is the estimated time to achieve a target level of exposure under various remedial approaches?
- What is the expected stability of the system under the 10 year wind/flood event? 100 year event?

Investigation of contamination in the sediment features more complexity than soil or terrestrial investigations, not only due to the sampling limitation by the simple fact that it lies under the water, but also that the contaminant is already an integration of different sources within a watershed or coastal region, which is difficult to distinguish and model separately according to the site-specific characteristics (Apitz et al., 2005). Consequently, although it is widely understood that sediment contamination can pose a substantial level of risk to an aquatic ecosystem, the remediation of the contamination in the sediment inevitably becomes an even more complex challenge because in addition to the investigation of the contamination, the remediation also involve a variety of different treatment strategies which reduce different amount of

contamination, with different cost-effectiveness under different time span to apply the treatment. Furthermore, the clean-up efficiency of the treatments are difficult to evaluate and require further use of a variety of validation and risk assessment tools, while the validation is necessary to relieve the uncertainty for the stakeholders in mind that the residual contamination is attenuated below a certain detrimental level (Apitz et al., 2005).

It is customary then that to address this questions, it is necessary to start with a conceptual site model (CSM). The CSM represents an abstraction of the interactive processes within the system under investigation, and at least includes the buried sediment, the mixed layer of sediment that exposes the foodchain to the contaminants, the tropic foodchain, and the water phase. The CSM is then queried using a range of investigative and analytical tools (see Fig. 1.1, adopted from Dekker et al., 2007) including geostatistical tools to allow for data interpolation between measured locations. This dissertation will focus on the latter.

1.1 Current State of Remediation Strategies

The current practice for remediation of contaminated sediments is primarily limited to three strategies: dredging/excavation, capping, and monitored natural recovery (MNR). Dredging and disposal is the most practiced approach for remedial actions, with capping and natural recovery receiving increasing consideration as in situ remedial options (Cushing, 1999; Adriaens et al., 2006). Dredging is applied for high-concentration contamination, while the application of dredging requires a good understanding of the spread of contamination. In addition, the practice of dredging may also result in a substantial amount of resuspension and at times result in more ecological problem or ending up with little improvement with all the expenses

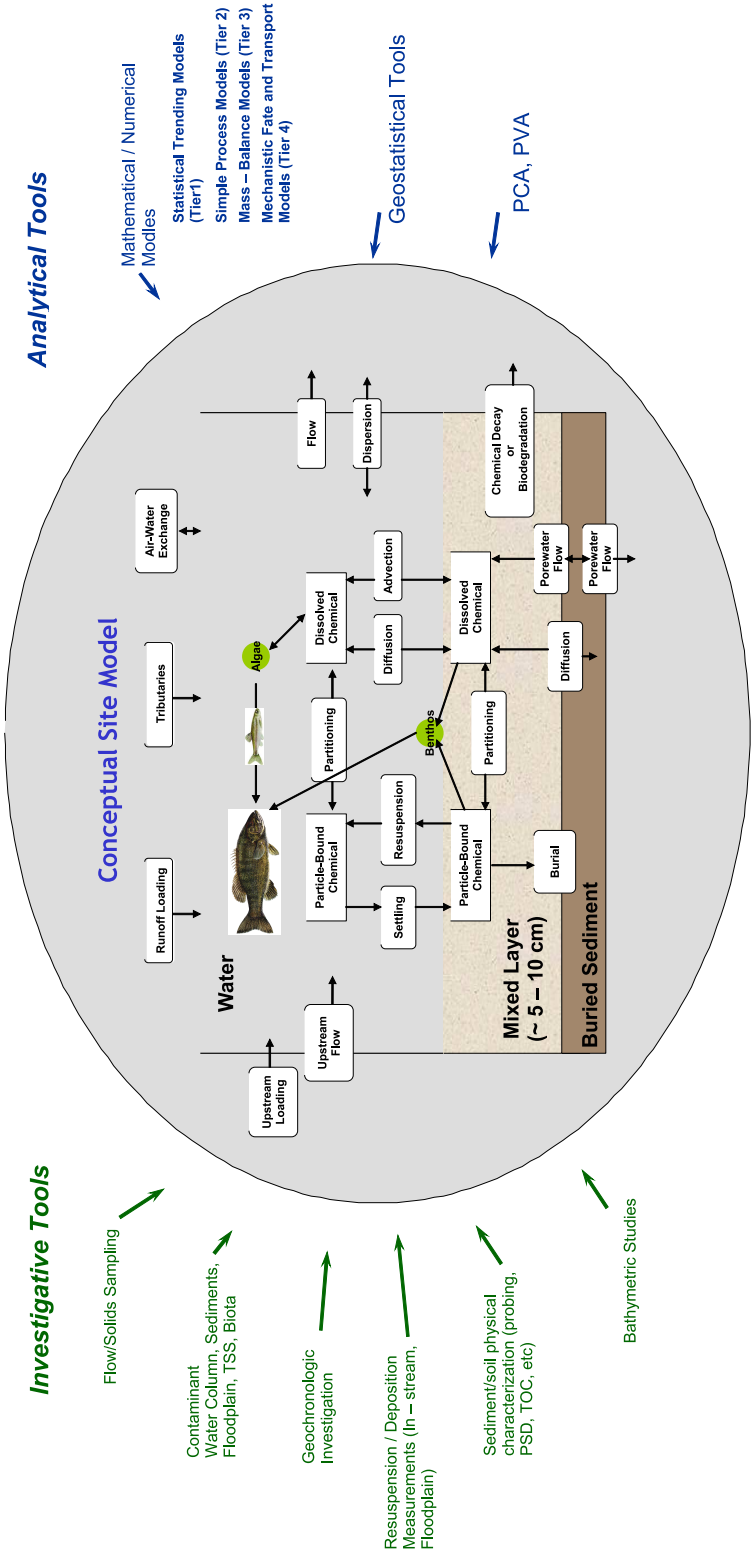

Figure 1.1: Overview of conceptual site model integration with field and modeling tools to assess risk and exposure for remedial decision-making (Dekker et al., 2007).

(Thibodeaux and Duckworth, 2001). In other words, a monitoring approach for the cleanup effectiveness is essential before and after the cleanup decision for the dredging activity.

Among the three remediation strategies, capping and MNR, which are also known as in-situ remedies, are considered for lower level of contaminant. The advantage of in-situ sediment treatment is that it has the potential for overall protectiveness and permanence, while satisfying the regulatory preference for reduction of toxicity, mobility, or volume through treatment. There is also great potential for reduction in cost, relative to dredging and disposal, by eliminating the need for sediment removal as well as ex-situ sediment water removal, treatment, and solids disposal. The application of in-situ remediation, however, also requires a good knowledge of contamination distribution over the study area because of the potential of high risks not observed a priori, which may exist under a certain level of uncertainty.

Similar approaches which allow for the quantification of the impact of a given technology (or technology suite) on the baseline (time trends, risk levels, etc...) should be used to consider implementation of remediation strategies as a function of the prevailing hydrogeochemical regime of the river or floodplain. Hence, the design of a technology with ultimate practical translation needs to be informed by process characteristics governing contaminant transport in sediments, which needs to rely on data-driven models based on field and laboratory measurements.

Spatial analysis for all three treatment strategies are needed, consequently, to fully understand and quantify the potential of high-risk area with its corresponding uncertainty using probabilistic and statistical techniques. In practice, if dredge-and-dispose is decided as the treatment, the area to dredge needs to be evaluated using spatial analysis tools, and further evaluation is required for the risk of resuspension of residual contaminants and eventual bioaccumulation of contaminants in the food

chain. If in-situ approaches are used, on the other hand, it is also important to have a spatial monitoring tool for both the assessment of capping strategies and the assessment of the potential of a possible MNR.

1.2 Current State of Spatial Analysis for Sediment Attributes

The level of toxicity or hazard is usually represented by measurable values scientifically studied to indicate their impact on human health or environment, and understood by spatial analysts as *attributes*, or *attribute values* (p.7 of Goovaerts 1997b). With the quantification of the level of toxicity or hazard, consequently, the first step in the contaminated sediment management, i.e., defining the contamination problem, consists of different perspectives of the problem by government agencies or stakeholders (p.5 of Barabás 2003), which sums up into a *regulatory threshold* for the attributes of the study area (see for example, Iannuzzi et al. 1995). Locations where attribute values violate this regulatory threshold will be considered contaminated and pose detrimental effect on human or ecology (Apitz et al., 2005).

For locations where measurements are not available, however, no attribute value can be compared to the regulatory threshold to decide whether a cleanup decision is needed, thus an *estimation* should be inferred on the basis of measurements from other locations by a scientifically developed model in order to decide the action to take for that certain location (p.4 of Chilès and Delfiner 1999). The *estimation uncertainty* (p. 489 of Isaaks and Srivastava 1989) of this inferred value by the certain spatial analysis model usually characterizes the conditional cumulative distribution function (p.262 of Goovaerts 1997b), to indicate the possible true attribute values with their correspondent likelihood of occurrence. The management of contaminated site is

illustrated in Figure 1.3., which poses two questions: (i) what is the impact of the choice of spatial interpolation technique, and (ii) how does this influence the area classified as contaminated per a given regulatory threshold.

With the estimation and estimation uncertainty evaluated, the rule that decides which remediation action to take is called *classification* of the contamination, by which a true/false statement is assigned as outcomes of the judgment by comparing the regulatory threshold to the inferred value Ramsey et al. (2002), which may also involve a certain level of uncertainties in correspond to the estimation uncertainty. The classification may take categorical values such as "no further action," "immediate treatment," or, if uncertainty exceed the expectation of decision-makers, "require future sampling and analysis." (Apitz et al., 2005)

1.3 Challenges and Research Needs in Contaminated Sediment Management

The decision making process of remediation of contaminated sediment covers not only the remediation strategies, but also how the decision makers are assisted by scientists to make the decision. The complete process is generically called the *management* of the contaminated sediment, and defined by Apitz and Power (2002) as "the process of making decisions and taking actions on sediments, taking into consideration a wide range of factors." This process depends on a complex procedure in which the characterization of contamination and the risk-assessment of the contaminant has to be integrated with the negotiation process of decision making (Barabás, 2003). A simplified flowchart for the sediment contaminant management is shown in Fig. 1.2.

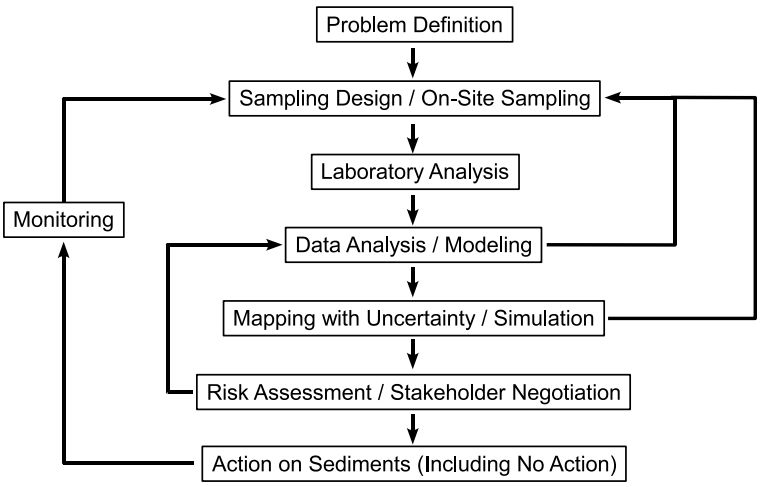

Figure 1.2: Flow chart of contaminated sediment management.

Although it is widely understood by modelers of environmental systems that no sampling approach provides complete understanding of the complex sediment systems (Chapman et al., 2002), there is a challenge to communicate to the general public how uncertainty impacts remedial action. It is also summarized by Apitz et al. (2005) that it is important for scientists to carefully balance a possible over-conservative cleanup design with the potential costs of a risky design on the basis of a sparse sample set, and to understand and communicate their conclusions to the general public along with the explanation of the risk and cost for potential uncertainties that come along with the outcomes.

It is crucial to provide a spatial description of the possible extent of the contaminated area as the basis for communication, otherwise the risk-assessment or negotiation among stakeholders may not proceed. A spatial estimation/uncertainty map acts as an important communication tool that helps the decision making, on a condition that the outcomes meet the goals of the assessment objectives. Considering that the contaminant distribution is usually heterogeneous and uncertain in space (Levin, 1992), the estimation/uncertainty maps to meet the assessment objectives, consequently, is by itself a complicated process. This is due to the consideration that

the maps should be able to address the concerns and priorities of all stakeholders who will be part of the decision process, including the general public, risk assessors, and engineers designing different remedy options (Apitz et al., 2005).

The main purpose of this research, observing all the challenges described above, is to specifically discuss the impact of spatial estimation on describing the extent of sediment contamination, and address the uncertainties associated with spatial estimation for remedial actions.

1.4 Dissertation Outline

This dissertation, aiming at providing a decision-support tool that facilitates communication among stakeholders of a contaminated sediment management project, follows the following structure:

The current chapter (Chapter 1) describes the challenges and research needs of contaminated sediment management, introducing the components to be included in order to fill the communication gaps among stakeholders.

In Chapter 2, the background literature on contaminated sediment management is presented, which includes an explanation of the complexity of the contaminated sediment. Followed by the literature review are concepts of site characterization using scattered samples, and a further review of spatial estimation/uncertainty evaluation approaches applied to facilitate site characterization and remedial decision support. The knowledge gap for spatial analysis is described along with the statement of research hypothesis for this dissertation at the end of the chapter.

In Chapter 3, the knowledge gap is proposed to be resolved by the M-Scale model,

a spatial statistical model developed in this dissertation. To facilitate the understanding of the M-Scale model, the theoretical development of conventional ordinary kriging (OK) and constrained kriging (CK) approaches are demonstrated preceding the statistical derivations of the M-Scale model. The development of the M-Scale model will be detailed following the description of OK and CK, with application procedure of the model listed at the end of the chapter.

In Chapter 4, the analytic results for artificial data using OK, CK and the M-Scale models will be demonstrated to assess the performance of the M-Scale model comparing to other models. Subsequently, the applicability of the M-Scale model is explored using the Passaic River datasets (Chapter 5), which also provides the user with diagnostic tools to choose a spatial statistical model among alternatives for the specific goal of remedial decision.

Chapter 6 describes the importance of knowledge of artificially induced error for on-site remedial decisions, and explores the applicability of cross-validation for a sparse sample set. An example using the Anacostia River datasets will be demonstrated with analysis of the benefit to decision making when knowledge of artificially induced error is obtained.

The final chapter (Chapter 7) addresses conclusion for the results observed, with future research proposed for the development of a more sophisticated estimation model.

CHAPTER 2

Background

Dredging and disposal of contaminated sediments is the management option employed in the great majority of remedial actions, with capping and natural recovery receiving increasing consideration as in situ remedial options (Cushing, 1999). Although caps may prove durable enough to prevent exposures if properly designed, and natural recovery may be sufficiently rapid and irreversible to be protective at some sites, statutory criteria and the precautionary principle have sustained a long-standing preference for dredging in remedial decision-making. Superfund criteria in particular have been difficult standards for in situ remedies to satisfy. In addition to complying with applicable rules and regulations, Superfund remedial actions must be protective of human health and the environment and meet additional "balancing" criteria for remedial selection, which include long-term effectiveness and permanence; and reduction of toxicity, mobility, or volume through treatment. In situ remedies are considered for low level wastes, but that close scrutiny should be applied to consideration of these remedies in cases presenting high potential risk and uncertainty.

The advantage of in situ sediment treatment is that it has the potential for overall

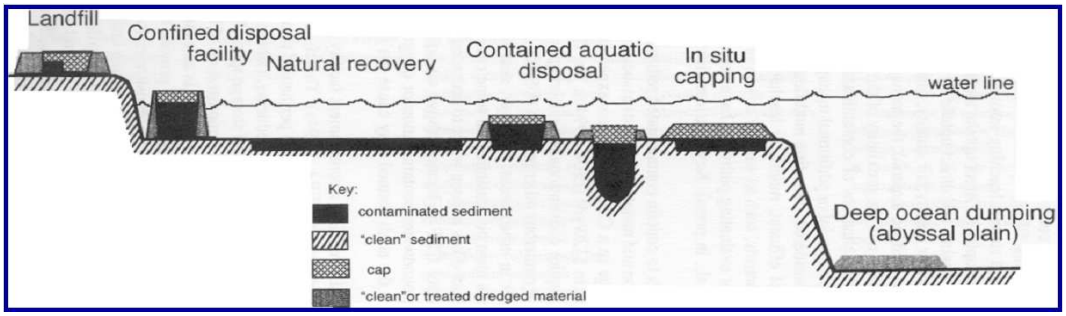

Figure 2.1: In-place sediment remediation strategies.

protectiveness and permanence, while satisfying the regulatory preference for reduction of toxicity, mobility, or volume through treatment. There is also great potential for reduction in cost, relative to dredging and disposal, by eliminating the need for sediment removal as well as ex situ sediment dewatering, treatment, and solids disposal. The U. S. Navy's guidance for addressing contaminated sediments (SPAWAR Systems Center (SSC) San Diego and Battelle Science & Technology International, 2003), the Remediation Technology Development Forum (RTDF) Weight of Evidence Framework for Evaluation of MNR (Davis et al., 2004), and EPA's Contaminated Sediment Remediation Guidance for Hazardous Waste Sites (U. S. EPA, 2005) are example frameworks that guide the sequence of decisions and also lay out opportunities for decision efficiencies. These frameworks address the sequence of management decisions from establishment of the site to closure, with a site conceptual model supporting each of those decisions. An important input into these decisions is the forecasted rate of natural recovery as a baseline for comparison with active remedies and as a component of MNR. (Fig. 2.1). Among the remediation strategies, dredging and disposal have become less favorable not only for political and social reasons, but also due to the problems associated with the uncertainties of defining highly contaminated areas, the tradeoffs of sediment disturbance and redeposition, and long term cost of managing confined disposal facilities (CDFs). In-situ contained aquatic disposal (CAD) is less attractive because the approach still requires dredging prior

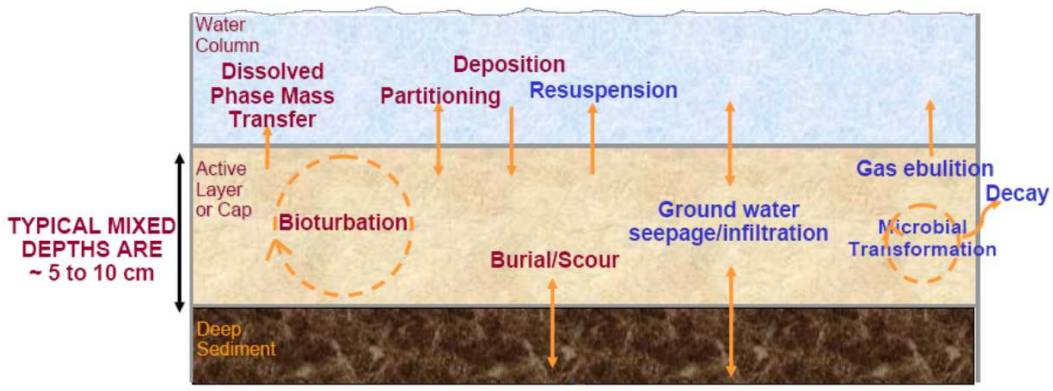

Figure 2.2: Components of contaminant transport/transformation.

to disposal. Alternative approaches such as passive (mitigation of exposure pathways only) and active (exposure pathway mitigation and toxicity reduction) capping have received increasing attention as contaminated sediment management options. However, their application has also been limited due to the uncertainties about long-term stability and exposure pathways under various environmental conditions.

2.1 Complexity of Contaminant Transport/Transformation in the Sediment Environment

The critical processes affecting contaminants in sediments are bioturbation, erosion/resuspension, diffusion, advection, biogeochemical interactions, and biotic or abiotic transformation reactions (National Research Council, 1997), as illustrated in Fig. 2.2. In addition to the mixture of different processes in the sediment with knowledge gaps in their physical behavior, conceptual models built for each processes require also site-specific informatiosn in order to be applicable for the remediation project. The following sections will provide a brief overview of some of the challenges faced in assessing the processes that impact in-place sediment remediation, describing

the complexity in ground water advection and sediment biogeochemistry along with how the assessment becomes even more challenging when these attribute values are sampled at different spatial scales.

2.1.1 Partitioning

Many chemicals of concern in contaminated sediment sites are hydrophobic; therefore, they have a propensity to partition to particulate matter. The degree to which contaminants can desorb from particles determines the mass available for biota exposure, as well as the rate of transport out of the system from other processes. Research shows that the partitioning behavior of the contaminant present can be influenced by a variety of factors, including chemical composition (Means et al., 1980; Karickhoff et al., 1979), sediment size and composition (Rutherford et al., 1992; Huang et al., 2003), hydraulics and hydrodynamics (Wu and Gschwend, 1986), and water chemistry (Elzerman and Coates, 1987). Often, a 3-phase partitioning approach is used, where the fraction in the dissolved phase is disaggregated between sorbate bound to dissolved organic matter and truly dissolved sorbate. The simplest relationships assume linear partitioning (Karickhoff, 1984), and thus the partitioning coefficient is the ratio of the concentrations in the sorbed phase to the dissolved phase:

$$K_p = \frac{C_s}{C_w}$$

where K_p = partitioning coefficient [L^3/M]; C_s = concentration in sorbed phase [M/M]; C_w = concentration in dissolved phase [M/L^3] (enclosed in the brackets are measurement dimensions. [M]: mass unit, [L]: length unit). Often, concentrations in the above equation are normalized to total suspended solids (TSS), or particulate organic carbon (POC). The 3-phase approach has advantages in relating partitioning to

bioavailability, but dissolved organic carbon concentrations are often very small and difficult to measure. Most fate and transport models assume adsorption and desorption kinetics to be in instantaneous equilibrium. This assumption may be adequate when exposure times are long and the hydraulics of the system is relatively stable (Wu and Gschwend, 1986). However, considerable research has shown that desorption kinetics in natural systems are often quite slow (e.g., on the order of weeks to years to reach equilibrium) and significantly differ from theoretical predictions under the equilibrium assumption (Gong and Depinto, 1998; Pignatello et al., 1993). Therefore, the equilibrium assumption may not always be valid, particularly in cases of high solute turnover (i.e., storm events, etc; see Wu and Gschwend 1986) and complex biotransformation processes (Bertelsen et al., 1998). In fact, in-situ concentrations have been found that differ by 1 to 5 orders of magnitude of those predicted by the equilibrium partitioning models (Zhou et al., 1999; Readman et al., 1987; Cullen, 2002).

2.1.2 Groundwater Advection in Sediments

The interaction of groundwater with streams and lakes has been extensively studied, yet little is known about the constraints such interaction may impose upon in-situ contaminated sediment remediation. This phenomenon can affect sediments physically by promoting or suppressing resuspension, depending on the direction of water exchange and grain-size. The extent to which groundwater advection affects the chemical and physical stability of contaminants in sediment and under caps is an important factor for sediment cap design (Myers et al., 1991; Winter, 2002). The effect of seepage rates must be validated in the laboratory as well as in the field, because groundwater fluxes may have effects on biogeochemical processes and redox conditions over time (Apitz and Power, 2002).

Erosion rates of fine-grained cohesive sediments are often modeled as a linear function of the applied shear stress above a certain threshold (Kandiah, 1974). The coefficient of the linear function varies with a number of site-specific parameters, including the environment in which the sediment was deposited (Krone, 1999; Lau, 2000), the pH of the sediment pore water (Dennett et al., 1995), salinity (Parchure and Mehta, 1985; Gularte et al., 1995) as well as bulk density of the sediment (Parchure and Mehta, 1985; Krone, 1999).

In general, the strength of these sediments is controlled by the physical-chemical bonds between the surface-charged particles. A vertically upward flow of pore fluid, associated with groundwater discharge will induce a mechanical stress that could result in resuspension of fine material (Reddi and Govindaraju, 1995). Another important mechanism in the marine environment is that a fresh groundwater discharge could reduce the ionic concentration of the pore fluids, consequently reducing the strength of the inter-particle forces. Pore water seepage could also reduce the consolidation rates in the sediment, decreasing the bulk density as well as the erosion resistance.

Particle Deposition and Erosion

Considerable research has been conducted in this area, but much of the theory has been developed using non-cohesive sediments (Ibad-zade and S. P. Ghosh (translator), 1992; Cheng, 1997b), which are of less relevance than cohesive sediments for understanding contaminant fate and transport. According to Stokes Law (Henderson, 1966), particle settling is dictated by particle diameter and density, but important factors causing non-ideal settling include particle shape and concentration (Cheng, 1997b), flow velocity and turbulence (Cheng, 1997a), and flocculation (Lick et al.,

1993; Droppo and Ongley, 1994). Flocs formed by fluid shear and differential settling differ in time to form, character, and settling rates: differential settling is slower and forms larger flocs with lower settling velocities (Lick et al., 1993). Deposition onto and attachment to the sediment bed have been described as probabilistic processes that are affected by turbulence at the sediment-water interface and by cohesiveness of the solid material (Krone, 1962). Sediment scour depends on hydraulic shear stress rising above a critical level, sufficient to dislodge particles (Kandiah, 1974). Cohesive sediment scour has been observed to depend on sediment bulk density (Jepsen et al., 2000), surface and porewater chemistry (Lee and Mehta, 1994), algal colonization (Ravens and Gschwend, 1999), and gas formation (Jepsen et al., 2000), in addition to bottom shear velocity.

Groundwater Seepage

Contaminant transport through the groundwater-surface water interface (GSI) is governed by a combination of complex hydraulics in and around the sediment bed, and a transport environment in the sediment bed that frequently exhibits sharp gradients in temperature, salinity, redox chemistry, biological population, and physical disruption. Mechanisms of groundwater flow and exchange with surface water can vary significantly from river systems (McBride and Pfannkuch, 1975; Winter et al., 1998; Conant Jr, 2004; Conant et al., 2004) to coastal environments (Robinson et al., 1998), and directionality of exchange can vary across reaches or even at a scale of meters (See for example, Cable et al. 1996, 1997; Conant Jr 2004; Conant et al. 2004). Where surface water concentrations of contaminants are significantly lower than porewater concentrations, the bulk exchange coefficient is essentially equal to the Darcy velocity. The porewater concentration may be less than expected based on the solid-phase concentration, where transport through the sediment bed is too rapid

to allow equilibrium to be reached.

2.1.3 Sediment Biogeochemistry and Ebullition

Microbial activities can impact contaminant mobility in both positive and negative ways (Apitz et al., 2005). The impacts may include the influence on chemical adsorption/desorption, generating ebullition of gases (Fendinger et al., 1992), and the chemical transformation on the contaminants themselves (Adriaens et al., 1999, 2003). The chemical transformations can lead to less or more toxic compounds, depending on the contaminant and on the transformation pathways.

Ebullition

Ebullition of gases resulting from microbial activities (e.g. methane, hydrogen sulfide, dinitrogen gas, and carbon dioxide generation from methanogenesis, sulfate reduction, and denitrification, respectively) can cause destabilizion of the sediment, and result in the desorption of organic contaminants from the sediment particles into the gas bubbles, thus facilitating the convective and diffusive transport of the contaminant (Palermo et al., 2002).

Because of the high sulfate concentrations in saline environments, sulfate-reduction is often a dominant process in marine sediments (Capone and Kiene, 1988), although fermentation (Ollivier et al., 1994), denitrification (Bonin et al., 1994; Nowicki, 1994), iron reduction (Lovley et al., 1991), and methanogenesis (Ollivier et al., 1994) have also been demonstrated. These conditions may increase the reduction reactions such as dechlorination of PCBs and dioxins (Adriaens et al., 1999), and aromatic ring destabilization of PAHs (Chang et al., 2003). Natural dechlorination rates for these

compounds range from seven to ten years for one chlorine to be removed, depending on the rates of carbon mixed into the sediments (Murphy and Schramke, 1998). Lesser chlorinated compounds have been assumed to move up to upper sediment layers, or out of the sediments into the water column (Gevao et al., 1997; Lohmann et al., 2000; Fu et al., 2001).

The formation, movement, emission and effect of the microbial gases in marine sediments has been well documented (Claypool and Kvenvolden, 1983; van Weering et al., 1997; Casper et al., 2000). The process of bubble formation, and its potential impact on sediment stability and contaminant mobility is shown in Fig. 2.3. This figure shows that the pathways from gas formation to sediment destabilization can be conceived as a complex sequence of events comprised of the following steps:

1. Metabolic generation of gas resulting from respiratory processes
2. Gas bubble formation, growth, and formation of channels in sediments
3. Gas dissolution and diffusion in the sediment
4. Contaminant mobilization in the gas phase via desorption and diffusion processes
5. Entrainment of sediment fines in particle wake during upward migration
6. Reduction of sediment bulk density and critical shear stress resulting in resuspension

It is generally observed that the generation of biochemical gases depends on the amount of organic matter in sediment and on the thickness of sediment cover (Ginsburg and Soloviev, 1997). However, limited direct measurements of gas diffusion in porous matrices have been reported, and the available information is generally

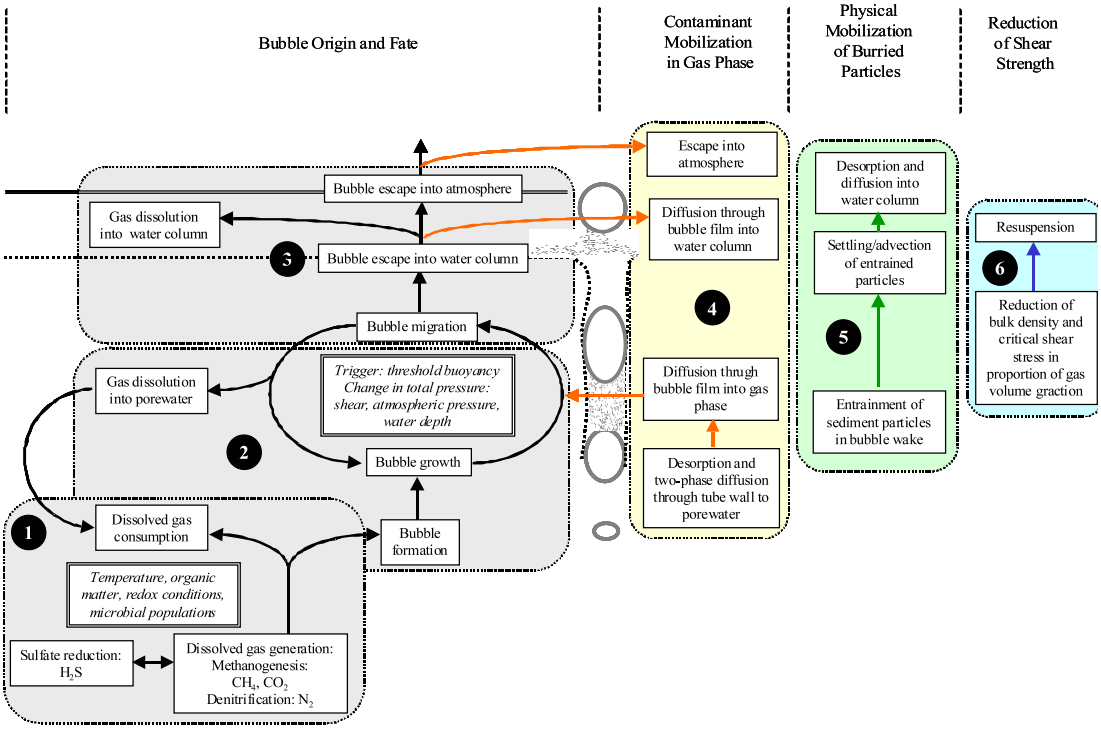

Figure 2.3: Gas bubble generation processes and the impact of gas bubbles on sediment and contaminant stability.

restricted to methane, carbon dioxide, and oxygen (Elberling, 1996; Rothfuss and Conrad, 1998). Fig. 2.4 shows the rates of ebullition observed in freshwater, tidal, estuarine, and marine environments. This table indicates that ebullition rates varies in tens of orders of magnitude from site to site, with uncertainties that are also site-specific, as shown in the minimum and maximum number observed in the different researches.

The significant uncertainties restricted the development of models that account for sediment/cap stability as a function of gas ebullition (Palermo et al., 2002). The impact of biogenic gases on contaminant fluxes in uncapped or capped sediments has rarely been incorporated into the modeling and/or remediation decision making at contaminated sites. This is in part due to the assumption that these gases would have a negligible impact on the system, relative to other fate processes (Reible et al., 2003). While ebullition may be insignificant at some sites, many sites show evidence

that biogenic gases can have an impact on both physical and chemical stability of sediments, and also capped sediments. For example, gas bubbles rising through the sediment column have been shown to strip contaminants from the porewater, carrying them through the sediment, to the water column (Martens and Klump, 1980) and into the atmosphere (de Angelis and Scranton, 1993). The presence of gas bubbles decreases bulk density, rendering the sediment more susceptible to erosion, while gas bubbles rising through the sediment can directly mix and transport buried sediments to the surface by entraining sediment particles in the wake of the bubble. Accordingly, ebullition may in part explain observed excess contaminant fluxes due to non-resuspension processes (e.g. Hartman and Hammond (1984); Erickson et al. (2005)). Thus, ebullition can modify natural recovery rates, and in the case of a cap, ebullition can increase the effective thickness of a cap, as well as dictate the type of suitable capping material (e.g. organic content, grainsize). However, probably a more important reason that gas ebullition is neglected is the dearth of information concerning the various mechanisms through which these gases can affect contaminant stability/fate.

The potential importance of gas formation has been observed at the following sites. Cap failure due to methane ebullition in freshwater environments has been experienced at the EPA/Manistique and the Oxbow, WI sites, where successive geostextile layers were lifted and exposed by methane formed in the contaminated sediment (Palermo et al., 2002). Methane outgassing has been observed at the Simpson-Tacoma site as well, where it did not cause physical disruption of the cap, but concerns of chemical transport by the gases prompted additional sampling of the gases (Stivers and Sullivan, 1994). The outcome of this follow-up sampling is currently unpublished. Cap design that accounted for ebullition was demonstrated at the Stryker Bay Superfund site in Duluth, Minnesota, on the St. Louis River just upstream of Lake Superior. At this site, ebullition actively transported both PAHs and NAPL

Salinity	Ebullition Rates in cm/d ("-" means no data reported) Min	Max	Avg	Stdev	N(data)	% CH_4			
fresh	0.03	0.03	-	-	7	44	88	induced by rapidly falling barometric pressure	Casper et al. (2000)
fresh	-	0.26	-	-	-	-	-	cited from Ward and Frea, 1979	Adams et al. (1990)
fresh	-	0.02	-	-	-	-	-	cited from Chau et al., 1977	Adams et al. (1990)
fresh	-	0.26	-	-	-	-	-	cited from Howard et al., 1971	Adams et al. (1990)
fresh	-	0.10	-	-	-	-	-		Adams et al. (1990)
fresh	-	-	3.22	0.56	3	-	-	flowthrough microcosm in laboratory, with SO4, numbers estimated from graph	
fresh	0.02	0.08	0.05	0.02	33	73	100	June-October, recalculated from graph	Strayer and Tiedje. (1978)
fresh	-	0.92	0.08	0.16	143	-	-	June and July measurements, recalc from graphs	Joyce and Jewell (2003)
fresh	6.72	44.80		-	-	-	-	Outbursts during pressure release	Richardson (1998)
fresh	-	-	10.54	-	-	-	-	daily, cited from Cicerone and Shetter, 1981	Chanton and Martens (1988)
fresh	0.32	47.45	-	-	-	-	-	daily, range from many environments	Chanton and Martens (1988)
tidal fresh	0.01	0.01	0.01	0.00	-	-	-	annual rate	Chanton and Martens (1988)
estuarine	-	10.07	-	-	10	85.5	85.5		Martens and Klump (1980)
marine	-	-	0.06	-	10	82	90	June-October	Martens and Klump (1980)
marine	-	3.12	-	-	10	82	90	same as above, converted to mmol	Martens and Klump (1980)
-	-	-	-	-	-	46	95	cited from Kuznetsov, 1968; Howard et al., 1971; Chen et al., 1972, Ward and Frea, 1979; Chanton et al., 1988	

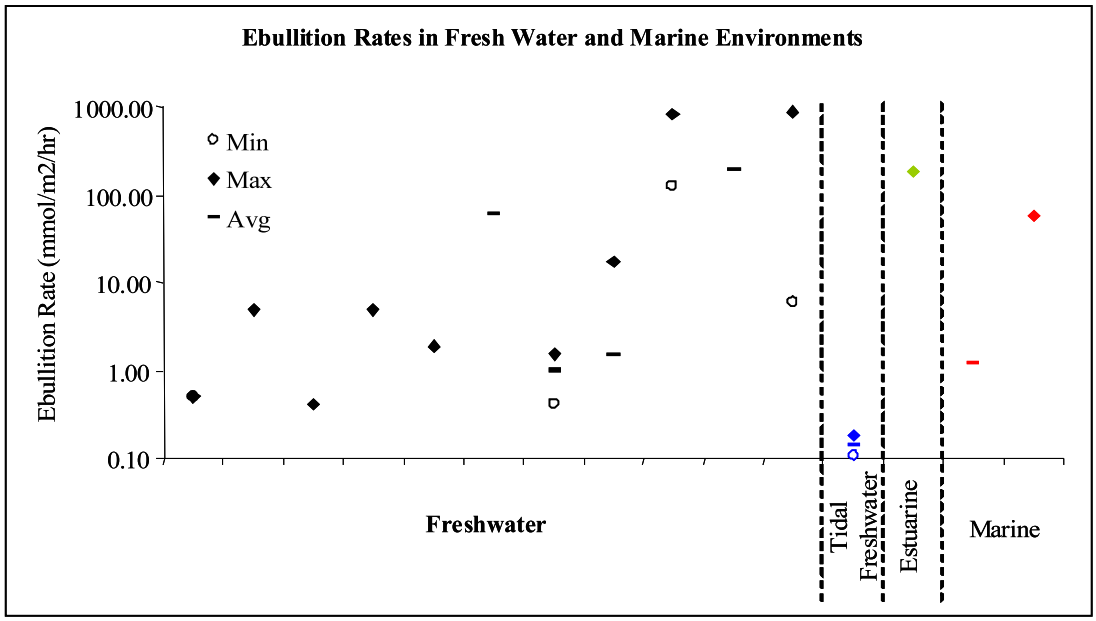

Figure 2.4: Gas bubble generation processes and the impact of gas bubbles on sediment and contaminant stability.

to the water surface. A transient model of post-capping sediment consolidation and associated porewater movement, groundwater advection and gas ebullition predicted that a 3-ft thick cap would be necessary to suppress methanogenesis to levels at which groundwater advection of dissolved gases and sediment strength would prevent the formation of free-phase gases (Huls et al., 2003).

Diffusive Mass Transfer and Bioturbation

Diffusive mass transport of porewater contaminants across the sediment-water interface is restricted by the thickness of the benthic boundary layer, which is very difficult to either measure or to relate to system properties (Thibodeaux, 1996). Bioturbation, which encompasses a diverse set of mixing processes mediated by benthic organisms, is generally thought to be the most important mechanism for reworking sediments and releasing porewater contaminants in sediments (Reible et al., 1991). Bioturbation increases flux by one to two orders of magnitude over molecular diffusion alone (van der Loeff et al., 1984). Bioturbation is controlled by a variety of biotic (organism size and seasonal life cycles, population density, deposition of organic matter, and species diversity) (Widdows et al., 1998; Berelson et al., 1999; Robbins et al., 1977; Wheatcroft et al., 1994; Petersen et al., 1998) and abiotic (temperature, sedimentation and erosion conditions and sediment chemistry) factors (Rhoads and Boyer, 1982; Schaffner et al., 1987). The importance of these multiple factors coupled with the spatial and temporal heterogeneity of benthic communities (Gerino et al., 1998; Berelson et al., 1999) has made it difficult to determine which factors are most important in driving biological mixing.

Biochemical Transformation

Biochemical transformation processes can occur due to chemical and biological processes. Biodegradation can occur due to growth metabolism or catabolism (Mills et al., 1985). For environmental transformations, redox conditions are particularly important because of their determining role in the microbial ecology and energetics (Capone and Kiene, 1988; Ollivier et al., 1994; Bonin et al., 1994; Nowicki, 1994; Lovley et al., 1991; de Angelis and Scranton, 1993). Biogenic gas production may also affect contaminant partitioning and sediment stability. For PCBs and other persistent sediment contaminants, rates of degradation are generally very slow (Murphy and Schramke, 1998), so that biodegradation is not generally a quantitatively important remediation process. However, biotransformations may be important for converting chemicals to more labile, mobile forms (Rhee et al., 1993; Adriaens et al., 1999), and may also decrease or exacerbate toxicity, altering risk without significantly changing total concentration (Apitz et al., 2002).

2.1.4 Uncertainty Associated with Sediment Conceptual Model Attributes

The following is a summary of uncertainties in the components of the complex contaminant transport/transformation.

Partitioning

Accurate modeling of the fate and transport of hydrophobic organic compounds (HOCs) depends greatly on appropriate characterization of the site conditions as well as the physiochemical properties of contaminants of concern. Site-specific data are

often limited, requiring assumptions regarding the partitioning coefficients. The use of non-site-specific partitioning data can contribute significant uncertainty to fate and transport modeling, and the extent of this uncertainty is rarely evaluated. Laboratory analysis of samples collected from both the water column and sediment bed of the contaminated site is required for accurate representation of the in-situ partitioning, but even when these are available there is uncertainty associated with translation from the laboratory to behavior at field scale. The influence of partitioning uncertainty on the uncertainty in predictions needs to be quantified.

Particle Deposition and Erosion

The main uncertainties in deposition processes are associated with particle size distributions and shapes, the degree of particle aggregation/disaggregation as a function of shear stress and particle properties, and the effects of fluid shear and bottom roughness on deposition. The main uncertainties for erosion (resuspension) process include the effects of depth and associated consolidation on critical shear stress, and the true resuspension rates of cohesive sediments with a range of compositions, ranging from virtually all clay and fine silt with high organic content to a significant fraction of sand but still enough clay/silt to impart cohesive properties. The literature includes a very wide range of estimates for these parameters, reflecting potential measurement artifacts and the generally unsettled state of measurement technologies. Additional sources of uncertainty include armoring processes and the extent to which erosion rates change over time and amount of material eroded, and quantifying the effect of sediment porosity on resuspension, including the impact of gas bubble formation as it affects sediment column stability.

Groundwater Seepage (Advection)

The high degree of spatial heterogeneity and variability of seepage fluxes and porewater concentrations implies a high degree of uncertainty in local contaminant fluxes, although this uncertainty is reduced at a more integrated spatial scale. An analysis of observed ranges of seepage fluxes and distribution coefficients for PCBs indicates that contaminant fluxes would be significantly reduced by a low permeability cap, and would be further reduced if the cap contained adsorptive materials, and that these conclusions hold across the range of site-specific parameter values.

Ebullition and Non-Resuspension Mass Transfer

Large knowledge gaps exist about the ebullition process, particularly with respect to the mechanistic/theoretical aspect of processes and empirical measurements of rates and their dependence on environmental factors. The greatest uncertainties surround the process of bubble formation and growth, and the physical transport of contaminants. Bubble sizes and residence times need to be better understood in order to properly estimate the extent of contaminant partitioning into the gas phase. The rate and extent to which migrating bubbles mix sediments is also an important uncertainty. Because gas generation rates are the driving force behind ebullition, it is important to better define the microbial, chemical and physical factors that affect it on all spatial and temporal scales. Diurnal, seasonal and weather related variabilities, as well as spatial variabilities, all contribute to predictive uncertainty. The interaction between groundwater seepage and ebullition is also not well enough understood. Reducing these uncertainties in process understanding and quantitative effects would greatly facilitate the incorporation of ebullition into existing modeling frameworks.

Diffusive Mass Transfer and Bioturbation

Chemical transport within the upper layers of bed sediments is a very complex process that will continue to challenge the efforts of environmental chemists, benthic biologists, and engineers. Aside from radionuclide tracer data, the laboratory and field data needed to verify mechanistic models for a specific site are usually very limited. While molecular diffusivity can be predicted with reasonable accuracy based on chemical characteristics and sediment porosity, biodiffusion is much more difficult to predict without extensive knowledge of local benthic populations and processes. Biodiffusion releases to the water column at rates excluding molecular diffusion must therefore be considered unless ruled out by site-specific benthic studies.

Biotransformation

The leading models contain simplifications and assumptions of site-specific parameters to facilitate application with limited data, generally represented as 1st-order decay rates. Given the wide range of degradation rates provided in the literature, and the hazards of translating laboratory rates to the field, there is considerable uncertainty in predicting biochemical transformation fluxes at any given site.

2.1.5 Overall Significance and Relative Uncertainty

To gain the most benefit from improvement of process representations one should focus on those processes to which the surface sediment response to alternative in situ technologies is most sensitive (i.e. where the process plays a significant role in governing the rate of change in surface sediment concentrations over time) and for which there is high degree of uncertainty/variability. For example, it does not pay to

reduce process uncertainty for a process that does not significantly affect the change in exposure from surface sediments over time.

The relative significance of processes in a system-level context can best be assessed by comparing their rates on an equivalent basis. To do that we can initially compare the estimated half-time for natural attenuation of a chemical in a surface sediment layer if the process of concern is the only one leading to that attenuation (i.e., a simple washout half-time). This is not a definitive definition of significance because the relative significance of processes and their half-time for exposure change over time may vary as a function of the particular in situ technology being applied. Nevertheless, a screening assessment of significance can be obtained by comparing the attenuation half-times with no remediation action. For this comparison, we assume the following common parameter values (minimum, median, maximum): bulk density: 1.0×10^5, 2.25×10^5, 5×10^5 g/m^3; surface sediment mixed layer depth: 5, 10, 15 cm; particle density 2.0, 2.25, 2.5 g/m^3; porosity 0.8, 0.9, 0.95; equilibrium partition coefficient 105,106,107 cm^3/g. Estimates of process parameters determining mass transfer in sediment-surface water systems can vary over as many as three orders of magnitude, and for some processes, measurement issues and heterogeneities make it difficult to reduce this uncertainty, even with site-specific data. For this reason, a probabilistic approach is needed to quantify the uncertainty in any process and its impact on prediction. Process prediction uncertainty can be difficult to evaluate on a generic basis, but can be estimated by developing probability distributions from the rates reported in the literature. In doing so, it should be recognized that the range of reported rates include both artificially induced error as well as the influence of all of the factors leading to stochastic variability in the environment. A simple Monte Carlo analysis is used to develop a half-time distribution for the processes of interest using estimated distributions of process-governing parameters based on the review of parameter variability and uncertainty. The characteristics of the resulting half-time

Table 2.1: Representative rates and attenuation half-times.

Process	Range of Observed Rates	Median Washout Half-time	Ratio (75%/25%) Washout Half-times
Net Sedimentation	-2 to 5 cm/y	0.5 to 15 yrs[a]	N/A[b]
Gas Ebullition			
Gas Phase Transport (Stripping)[c]	0 to 47 cm/d (0 to 17000 cm/y)	20,000,000 yrs	22
Particle Entrainment	Unknown	Unknown	Unknown
Groundwater Seepage	0 to 125 cm/d (0 to 46000 cm/y)	3,700 yrs	25
Bioturbation	0.001 to 30 cm^2/y	500 yrs	20
Molecular Diffusion in Porous Media	0.3 to 30 cm^2/y	1,100,000 yrs	9
Biotransformation	10^{-4} to 10^{-6}/d	55 yrs	4

[a]Applies only when net depositional. Not a median due to unknown distribution shape.
[b]Unknown distribution shape for the sedimentation/erosion rates
[c]By partitioning to bubble phase (does not account for particle entrainment and diffusion enhancement)

distributions are presented in Table 2.1.

An overall assessment of the relative magnitude of predictive uncertainty for the transport and transformation processes of interest can be made by combining the knowledge gained from the significance and uncertainty/variability analysis presented above with an evaluation of the other factors leading to prediction uncertainty (theoretical understanding, model representation and process parameterization, and site-specific information), as shown in Table 2.2.

2.2 Uncertainty Modeling and Remedial Decision Making

To assess the dependence between site-specific attributes, model estimations are usually required, to interpret the information derived from the limited number of

Table 2.2: Qualitative assessment of uncertainty.

Process	Theoretical Understanding (Mathematical Formulation)	Model Representation (Process Parameterization)	Site-specific Information (Process Variability & Availability of Data)	Overall Uncertainty
Partitioning	+++	+++	++	+++
Net Sedimentation	++	+++	++	++
Gas Ebullition	+	+	+	+
Groundwater Seepage	+++	++	+	++
Diffusive Mass Transfer and Bioturbation	+++	++	+	++
Biotransformation	+++	++	+	++

++++ (Low uncertainty) $\longrightarrow$ + (High uncertainty)

samples collected, and to aid in future sampling guidance. The choice of methods or models to be used, to this end, is the first step in estimation. Unfortunately, the choice is often made on a largely subjective and intuitive basis (Maidment, 1992). While some subjectivity may always be involved, the choice of an estimation method usually depends on consideration of deterministic or stochastic approaches, and whether a particular method and its parameter values are suited to the application.

2.2.1 Deterministic and Stochastic Models for Sediment Remediation

Deterministic models, derived based on physics laws that are generally accepted, are applied for the simplicity of their formulation or for the ease of explanation to a broader audience. Stochastic models on the other hand, are often used to measure the uncertainty accompanied with restriction of sample size due to the cost of taking samples. In addition to using the stochastic model as an uncertainty assessment tool,

the applications may also include the forecasting or estimation of a certain attribute based on currently available measurements.

A mixed deterministic and stochastic approach may provide an alternative that evaluates the parameters of a deterministic model from the specific site measurements and leaves out the residuals for stochastic analysis. The mixed approach, however, is usually categorized as a stochastic approach based on the fact that the models are taking statistic components into consideration. For example, a kriging with a trend model incorporates a varying local trend within a certain neighborhood expressed by a deterministic function of the coordinates (Goovaerts, 1997b). The deterministic components of the model require inter-disciplinary knowledge from experts in flow modeling and lab experiments, who should be involved in the stochastic estimation in the cases where physical models are available for attribute quantification.

Risk assessment is an integral component of decision-making for contaminated sediment management, and an important driving force is the presence of uncertainty. Uncertainties exist at all stages of site assessment and modeling. Even when preventive measures are based on sound planning, sampling (Batley et al., 2002) and analytical methods (U. S. EPA, 2000) aimed at eliminating and minimizing some sources of error and bias, uncertainties always remain. Nevertheless, the element of uncertainty in the management of contaminated sediments is often treated inadequately (Suter, 1999), or not addressed at all. For example, unquantified uncertainties can force extremely conservative estimates (Linkov et al., 2002), which can lead to high inefficiencies in the allocation of resources for remediation. Most frequently, uncertainty is expressed in the form of confidence intervals, or by incorporation of safety factors based on best "professional judgment." But these measures and strategies are rarely tested and vary along with sampling and measurement methods, models, and site variability and other site characteristics.

The applications of geostatistical techniques to the spatial analysis of sediments are recent and few. Barabás et al. (2001) modeled spatial uncertainty in dioxin contamination in three dimensions using indicator kriging in Passaic River sediments; Murray et al. (2002) mapped the extent of DDE contaminated sediment thickness on the Palos Verdes shelf using sequential indicator simulation, while Fioole et al. (1998) developed the SURFIS computer program, to integrate geostatistical methods in order to account for random error in the optimization of dredging of contaminated sediments with digital terrain models (DTMs). Ouyang et al. (2002) studied DDT in river sediment, and used kriging to "characterize" its spatial distribution. Mear et al. (2006) characterized geostatistical spatial structure for the fine-grained content of the superficial sediments in eastern Bay of the Seine, France, for the indication of heterogeneous patches in the sediment. Mear et al. also performed conditional simulation of the same attribute, while locally calculated mean value of the 80 conditional realizations to generate the estimation map. In the research by Ouyang et al. (2003), heavy metal was estimated in river sediment.

When systems are better understood and mechanistic models available, the environmental attributes and parameters of the mechanistic models can be treated as random variables. Uncertainties of these inputs (parameters/coefficients and variables) can be expressed in the form of value distributions. These uncertainties are then propagated to the model output via mathematical analysis for simple models and via stochastic simulation (Kleijnen, 1997). The uncertainty evaluation of input parameters for the deterministic models must often rely on professional judgment, in particular when values of parameters are scattered in the literature and were derived in different experiments (or sites), by different investigators, under different experimental (or site) conditions (Linkov et al., 2002; Steinberg et al., 1997). Once input distributions are quantified, values for each input variable and model parameter are randomly drawn from their probability distributions using, for example, a Monte

Carlo approach, preserving correlations among the inputs if applicable. The resulting combination of input values is used in the model (e.g. sediment fate and transport, contaminant flux) to generate a realization (Rossi et al., 1993; Clarke and McFarland, 2000; Papadopoulos and Yeung, 2001). The process is repeated, while the nature of the sampling ensures that the realizations are equally probable, allowing a statistical summary of the model outputs.

2.3 Principles of Spatial Estimation and Contaminant Delineation

In observation of the complexity of contaminant assessment in the sediment environment, spatial estimation approaches have been applied to provide information for decision making (Englund and Heravi, 1993; U. S. EPA, 2005). Especially when a conceptual site model including, for example, a contaminant mass transport model is not available or not sufficient to estimate the spatial attributes, statistical models are often used to make estimates of attribute values and evaluate the estimation uncertainty where no sample is taken (Goovaerts, 1998; Adriaens et al., 2006; Chilès and Delfiner, 1999). The following paragraph provides the basic principles upon which most of the spatial estimation models are formulated.

2.3.1 Spatial Estimation/Uncertainty Evaluation

Spatial estimation is based on the mathematical concept where an estimate is represented by a weighted average of sample values. Usually the similarity/dissimilarity is considered among samples and between sample and estimation locations, so that weights are properly attributed to each sample value according to the similarity

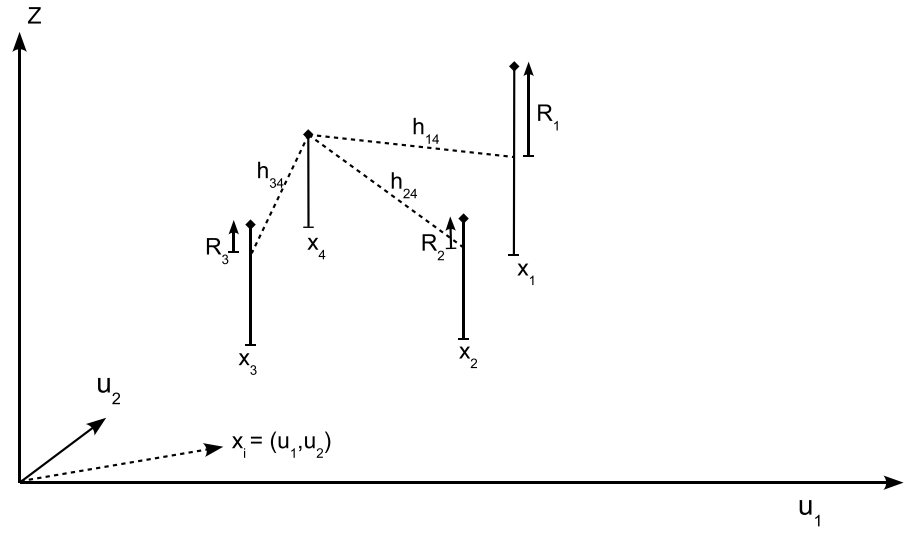

Figure 2.5: Difference between attribute values. x **represents the sample location with spatial coordinates** (u_1, u_2)**, attribute values** Z **are represented as height along the vertical coordinate,** R_{ij} **and** h_{ij} **represent the difference between attribute values and distance between locations.**

(Chilès and Delfiner, 1999; Isaaks and Srivastava, 1989).

One of the simplest forms of expressing dissimilarity is based on using the difference between attribute values (p.76 of Haining 2003), as shown in Fig. 2.5. When the number and sample locations of sample points are considered sufficient to represent the similarity between locations for the entire study area, the values of dissimilarity are grouped into different classes, for example, by distances between locations (pp.74–75 of Haining 2003), as shown in Fig. 2.6. These different values of dissimilarity are considered as the uncertainty when one attribute value is represented by another value at a different location. When estimations are needed at a location where samples are not available, all other sample points can be considered to represent the unknown value, with different uncertainty pertaining to the class of dissimilarity(pp.260–266 of Isaaks and Srivastava 1989). Weights to the different possible representative values are thus attributed to establish a single estimation. The weights are attributed not only to the sample values themselves, but also their corresponding uncertainty, and the combination of the two is a distribution of possible values for the sample location

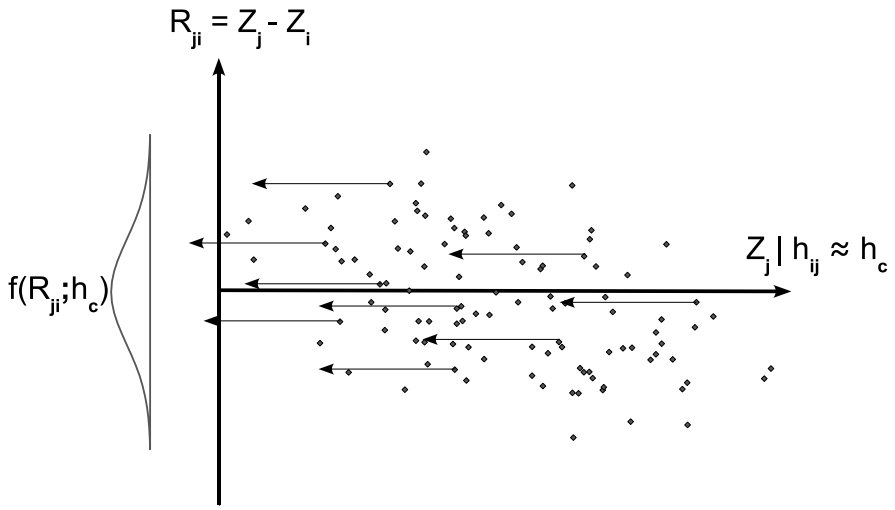

Figure 2.6: Summarizing attribute value differences into distribution of differences. The value differences R_{ji} at a separation distances class h_c are pooled and summarized as the distribution $f(R_{ji}; h_c)$.

(p.249 of Isaaks and Srivastava 1989), as shown in Fig. 2.7. The combination of the weighted sum and uncertainty for a single location can be visualized as Fig. 2.8, the distribution of which is defined as *conditional distribution* for the estimation because value distribution represents the possible true value conditioned on the sampled attribute values $Z(x_i)$ (p.378 of Chilès and Delfiner 1999). For a *spatial* estimation, the estimation and uncertainty evaluation process is repeated throughout locations where the estimation is needed, as shown in 2.9.

It is almost impossible to show the combination of estimation/uncertainty on a single graph, thus representative values, e.g. the mean value, of the value distribution for all locations are connected to form a surface representation (see for example, pp.51–53 of Ripley 2004), as shown in Fig. 2.10. This representative value is usually used as the attribute value based on which the spread of contaminant is delineated.

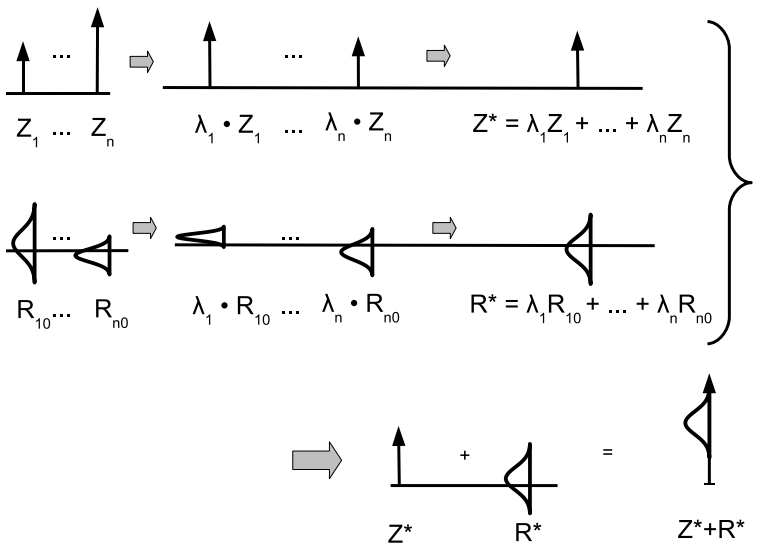

Figure 2.7: Weighted estimation and uncertainty, where $Z_i, i = 1, \ldots, n$ are sample values, R_{i0} are possible difference values between sample and unknown values for Z_i to represent the unknown value. Z^* and R^* are weighted sum of sample values and uncertainty.

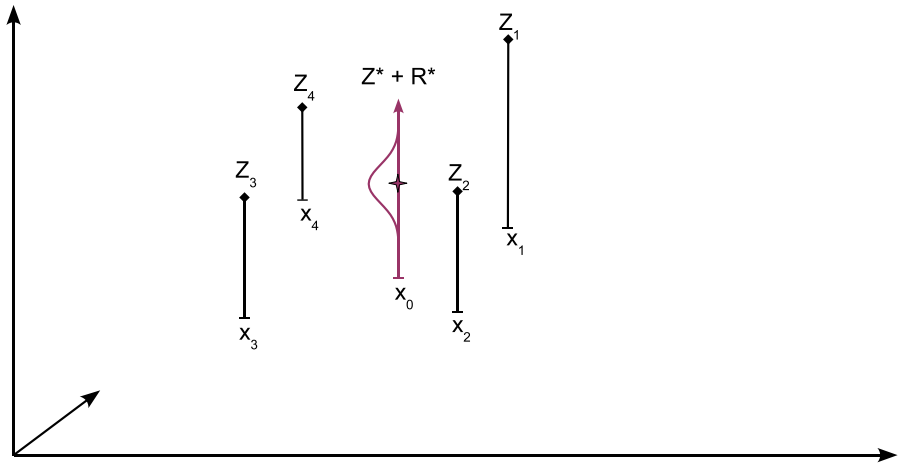

Figure 2.8: Visualization of a local estimation/uncertainty evaluation, where samples are presented as arrows with heads pointing to the value, and the local estimate and its uncertainty is illustrated as a bell shape.

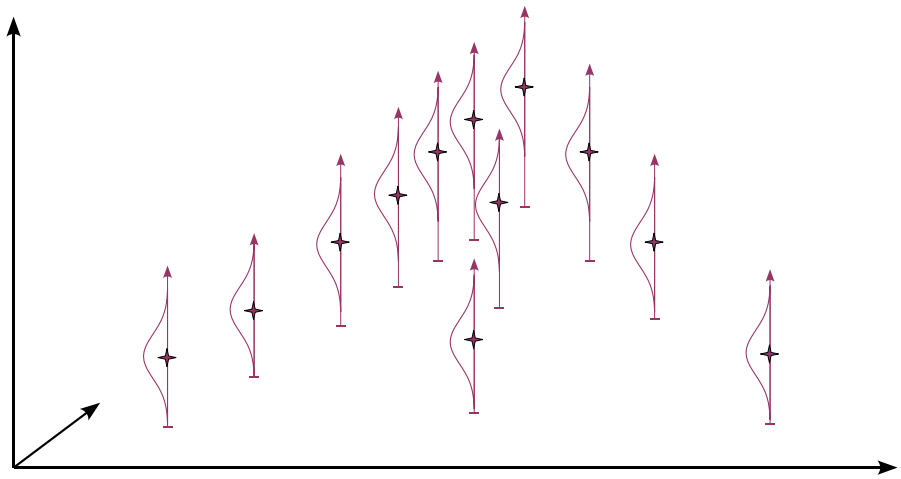

Figure 2.9: Visualization of a spatial estimation/uncertainty evaluation. Estimates and uncertainties are presented at different estimation locations.

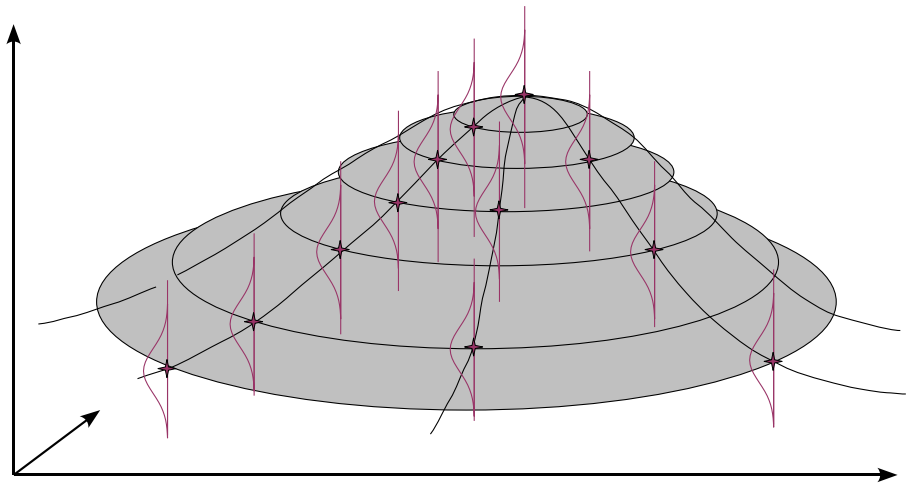

Figure 2.10: Visualization of representative values of a spatial estimation/uncertainty evaluation. The representative values can be used to form an *estimation surface*, presented in the figures as the shaded surface. The most presented estimation surface is the surface of the expected value, although other statistics of the estimation (such as quantiles, upper/lower confidence limits) can also be used.

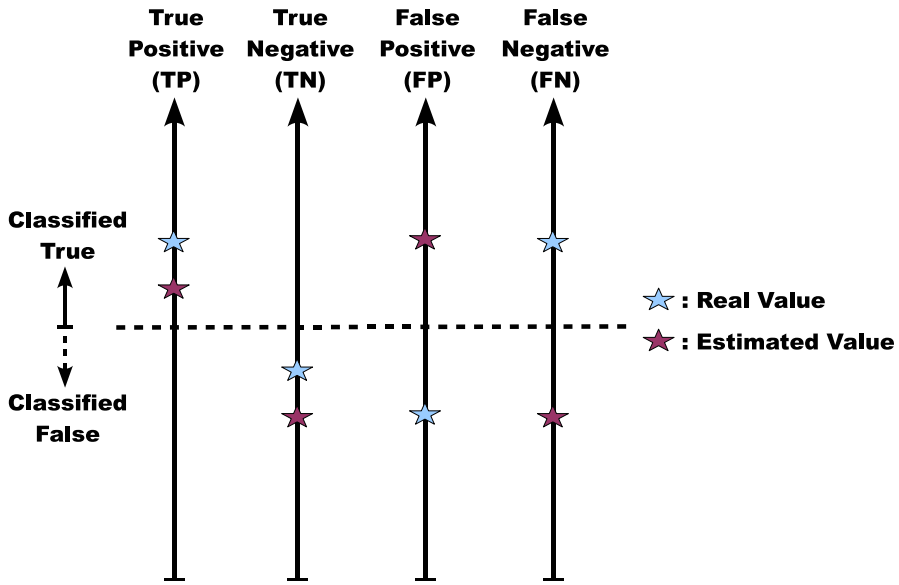

Figure 2.11: Four categories to express quality of classification. For a binary classification like the exceedance/non-exceedance classification, four categories (TP,TN,FP,FN) can be used to express the types of agreement/misclassification of the results.

2.3.2 Classification and Delineation of Threshold Exceedance

Once a representative attribute value is obtained, the value is compared to the regulatory threshold to decide whether the attribute value reaches a detrimental level (Englund and Heravi, 1994). This comparison is called *classification*, and the quality of this classification is usually expressed with four categories (Ramsey et al., 2002), as shown in Fig. 2.11. Usually the classification with larger TP/TN rates and smaller FP/FN rates is considered a better classification.

A collection of all locations where the representative value violates the regulatory threshold is called the *delineation* of contamination in this dissertation, as shown in Fig. 2.12. It is hard to define an "unbiased" delineation of contamination zone in a 2-D or 3-D space, however. As shown by the different cases conceptualized in Fig. 2.13, where dashed line indicate the true extent of contamination, and solid lines indicate

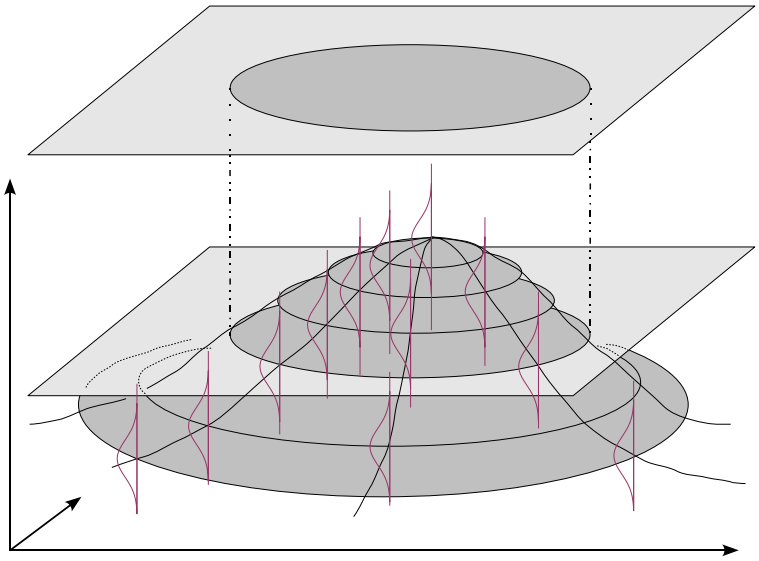

Figure 2.12: Delineation of the extent of contamination. The horizontal plane that cut through the estimation surface represents the regulatory threshold. A delineated extent of contamination is represented as the shaded zone on the projected plane on top of the estimation surface.

the estimated extent of contamination based on the classification at each estimation location. It is clear that (a) is an underestimation of the contamination zone, with much of the FN classification and little FP classification. Similar argument can be made for (c) as an over-estimation of the contamination zone, since the delineation contains much of the FP classification and little FN classification. (see also Squire et al. 2000a) A delineation with similar amount of FP and FN classification, however, may not necessarily define a good estimation. As shown in the same figure, both (b) and (d) are examples of the delineation that has similar amount of FP and FN classification, while it is obvious that (b) is a better delineation than (d) because the amount of both false-positive and false-negative classification is smaller than that of (d). What (b) shows is a more precise estimation that have less tendency toward FP or FN, while (d) have more tendency toward both FP and FN. Imprecise delineations of contaminant may also include those shown by cases (e) and (f), where certain part of the contamination spread is underestimated and another part of the contamination is overestimated. (See also Squire et al. (2000b) for more examples of delineation

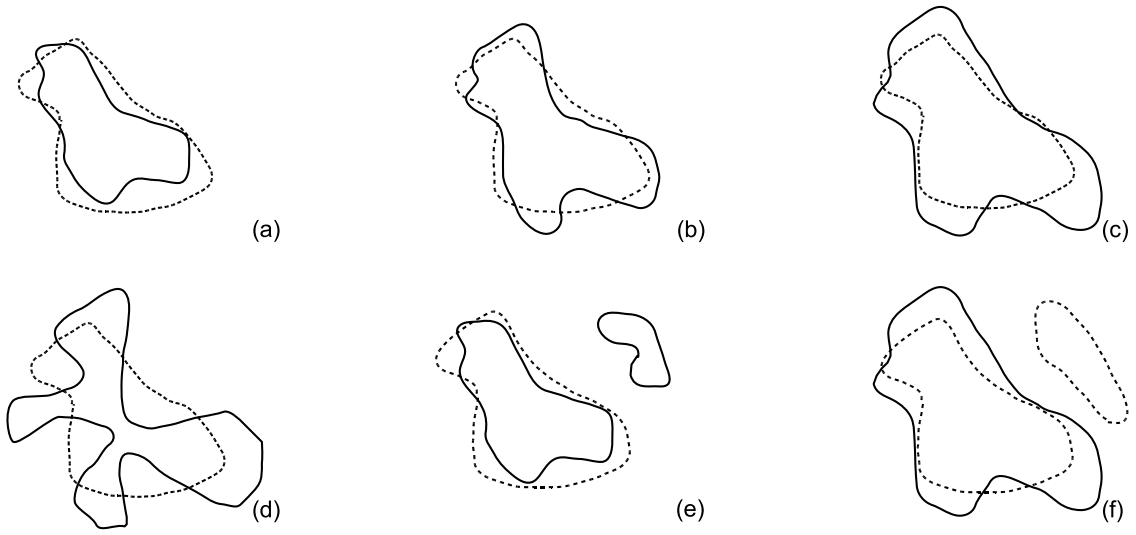

Figure 2.13: Delineation with different precision. Dashed contours represent the actual extent of contamination. Solid contours represent the delineated extent of contamination based on the estimation results.

geometry and their mismatch possibilities with the true contamination zone.)

Although an overestimation of extent of contamination results in a high FP rate, the increase of the estimated extent of contamination usually accompanies a decrease of the FN rate (see for example, the results of FP and FN rates in Englund and Heravi 1993 and Englund and Heravi 1994, and p.54 of Saito 2003). A decision upon quality of delineation, consequently, will depend on the excess cost (Squire et al., 2000b):

$$\text{excess cost} = a(E - i) + b(T - i)$$

where T represents the true area of contamination, E is the estimated area of contamination, and i is the area of overlap of the two regions, T and E. Parameters a and b are used to represent a possible financial penalty in which a false negative classification was estimated to be a certain ratio of a false positive classification. In fact, $E - i$ and $T - i$ represent the areas of FP and FN classification respectively. An excess cost of zero indicates perfect spatial delineation, while in practice the value

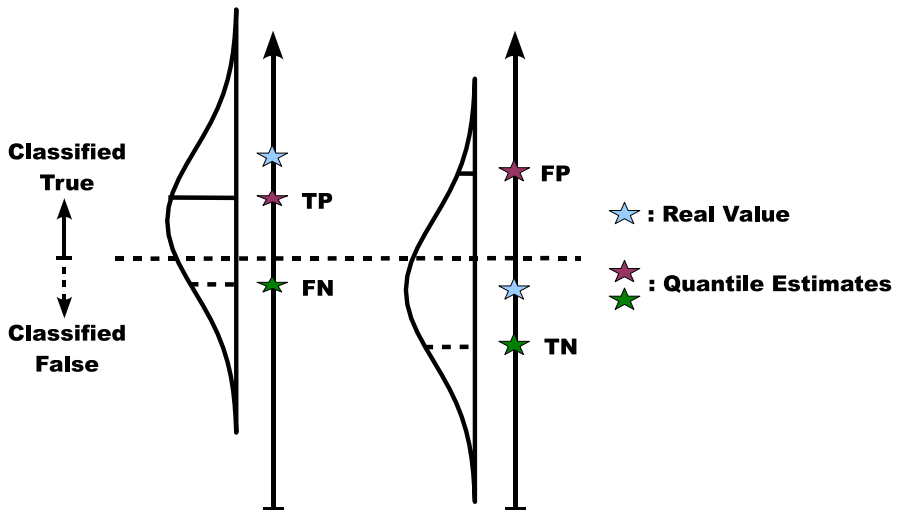

Figure 2.14: Likelihood-based classification. The classification results using different quantiles (likelihood thresholds) may not be the same as the classification of the actual attribute value depending on the choice of quantile. Two cases are shown to demonstrate the possible false-negative and false-positive classification due to the selection of quantiles/likelihood thresholds.

depends on the true extent of contamination, estimated extent of contamination, and the objective of the cleanup of contamination.

2.3.3 Likelihood-Based Estimation and The Uncertainty of Contamination Delineation

Because the classification depends locally on the comparison between the representative attribute and the regulatory threshold, the classification also involves a certain level of uncertainty (Squire et al., 2000b). When the representative value is not assigned to be the mean value, the classification results may vary. As shown in Fig. 2.14, when a quantile value corresponding to a certain likelihood of exceedance is selected as the representative attribute value, the classification may change from true to false or vice versa. For a spatial estimation, consequently, the delineation of contaminant may also vary, as shown in Fig. 2.15. The FP and FN rates will also change according to the change of delineation, resulting in more conservative delin-

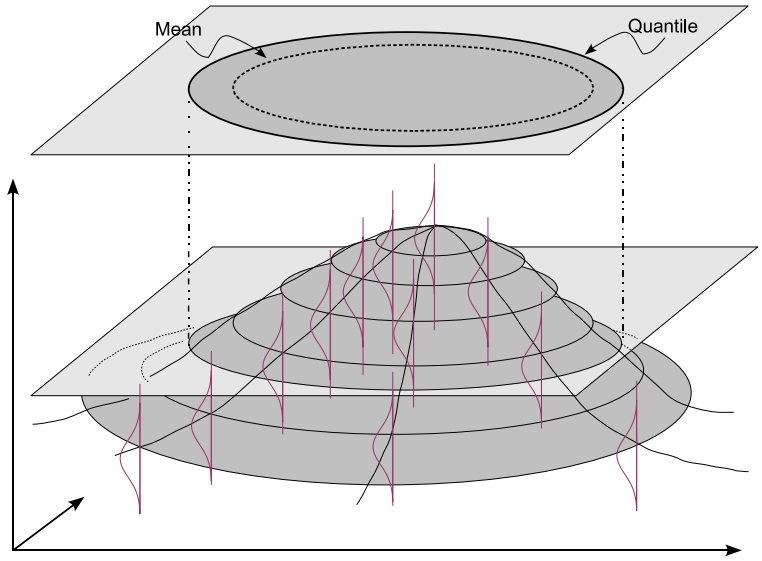

Figure 2.15: Likelihood-based contaminant delineation. The extent of contamination are different depending on the choice of representative values.

eation (larger FP rate) or riskier delineation (larger FN rate). The classification on the basis of quantile is also seen in Ramsey et al. (2002) and Barabás et al. (2001), where a function for expectation of loss is used to make decision on the quantile to use.

2.4 Estimation Models, Their Estimates and Estimation Uncertainty

Although based on similar principles of spatial estimation, different models have been developed to fit different estimation objectives. Kriging approaches have been applied to estimate environmental attributes, such as heavy metal in soil or percent of the water column occupied by submersed vegetation, to achieve the least local estimation errors (Facchinelli et al., 2001; Valley et al., 2005). Parametric and non-parametric methods exist to model conditional distribution of attribute values (Goovaerts, 1998). Ordinary kriging (OK) is a parametric approach generally used for

local estimation which generates a smooth visual representation of the spatial distribution. The smooth estimation map, however, under-represents the spatial variability, a situation usually called *smoothing effect* (Isaaks and Srivastava, 1989). The approach yields good estimates locally that minimize estimation variance, while their estimates exhibit a smoothing effect (Carr, 1990; Olea and Pawlowsky, 1996; Goovaerts, 1997b; Journel et al., 2000; Arik, 2002; Yamamoto, 2005) due to their tendency to produce a central value among sample points, thus reducing the variability of the estimate. This is illustrated by Carr (1990), who described ordinary kriging (OK) as a "low-pass filter", and demonstrated the smoothing effect by plotting the variogram of OK estimates where the semivariance values of all lag distances are underestimated. In other words, kriging approaches were developed to minimize the estimation uncertainty at each estimation location rather than to estimate the variability of attribute values between estimation locations.

Among all the conventional geostatistical approaches, indicator kriging (IK) is a non-parametric approach, with the advantages that the shape of the local exceedance/non-exceedance is derived using spatially varying information rather than assuming a global analytical expression (such as a parametric Gaussian model) (Barabás et al., 2001). The approach provides the same expected mean value of local conditional cdf (E-type estimate) as the estimates of OK and their estimation variance are also identical (Journel, 1983). The merit of IK over OK is the flexibility to evaluate local quantile estimates other than the most conventional E-type estimates, consequently is the flexibility for the delineation to achieve the site-characterization objective, e.g. the least misclassification rate (Barabás et al., 2004). The dilemma for using IK and quantile estimates, however, is that the flexibility to achieve a contaminant delineation is done by assigning the likelihood of exceedance for a certain regulatory threshold (Barabás et al., 2004), which sacrifices the flexibility of having alternatives for stakeholders to negotiate (see Chapter 6 for detailed discussions). Although the E-

type estimate of IK does not involve the problem above, the result coincides with OK, therefore no further benefit is gained for characterizing the contaminant area. Another shortcoming of IK is the lack of capability for generating the likelihood map of block values, as described in Deutsch and Journel (1998) that "we strongly recommend that the block ccdf be approached through small-scale stochastic simulations."

2.5 Spatial Statistical Approaches for the Reproduction of Global and Spatial Variability

To avoid the smoothing effect involved in OK and IK, stochastic simulation is increasingly preferred to kriging for environmental assessment applications such as delineation of contaminated areas (Desbarats, 1996; Goovaerts, 1997c; Kyriakidis, 1997; Naber et al., 1997), or the modeling of groundwater flow in heterogeneous aquifers (Dagan, 1982; Desbarats and Srivastava, 1991; Morgan et al., 1993; Desbarats, 1996; McKinney and Lin, 1996). Conditional simulation generates equiprobable maps that represent spatial variability, and has been recognized as a method that reproduces both the histogram and the variogram (Yamamoto, 2005). The set of alternative realizations generated provides a quantitative measure of spatial uncertainty. Features that appear consistently on most of the simulated maps are deemed certain, as expressed by a corresponding local conditional distribution summarizing the likelihood of the attribute at each simulated grid point. The uncertainty can then be summarized into probability maps, risk maps, or maps of false positives and false negatives (Goovaerts, 1999). Such results can aid in a more scientifically grounded assessment of safety factors and a more realistic and comprehensive formulation of confidence intervals. The problem with conditional simulation is, as described in Journel et al.

(2000),"for expert users, they may have based their decision-making process on multiple equiprobable simulated images." However, even for an estimation approach like the conventional kriging, there is "one advantage in the eyes of many users– that of providing a unique (single) representation/estimation of the unknown reality." In other words, no single map is more representative than the others in a conditional simulation, complicating the decision-making process. Another shortcoming of the conditional simulation is that, even if the approach reproduces spatial variability in each simulated image, the expected exceedance of local values over all simulations is similarly biased, as quantitatively evaluated in Aldworth and Cressie (2003).

Reduction of the smoothing effect has also been attempted in the past by post-processing the kriging results (Olea and Pawlowsky, 1996; Journel et al., 2000; Yamamoto, 2005) such that a single map could reproduce the spatial variability of the original sample points. However, the post-processing approaches also transform the kriging results from their original estimates that follow the constraints designed to fulfill the objectives for the estimation model, such as the unbiasedness constraint of weights attributed to each sample point for the local estimate. Two approaches, known as constrained kriging (CK) (Cressie, 1993) and covariance matching kriging (CM) (Aldworth and Cressie, 2003) are used to generate estimations that reproduce global and spatial variability without ad hoc post-processing. The problem with the two approaches, however, lies in the estimation stability, where estimates may or may not be singular (not evaluated) depending on the strength of covariance (the absolute value of the covariance) between sample and estimation locations.

The major features of the applied geostatistical approaches described in this and the previous sections are compared in Table 2.3. The focus for this dissertation is to develop methodology that enables reproduction of variability and enables single-map estimation. Consequently, the estimation models that aim at reproducing variability

Table 2.3: General comparison of geostatistical approaches. Positive signs indicate the inherence of the feature, negative signs indicate the absence of the feature, and NI represents non-indication of the feature. OK: ordinary kriging; QIK: quantile indicator kriging; SS: stochastic simulation; PP: post-processing approaches; CKCM: general constrained kriging, including constrained kriging and covariance-matching kriging.

Geostats Models	Reproduces Attribute Variability	Flexible Classification Objective	Generates Stable Estimates	Single Estimation Map	Uses Expected Value as Estimate
OK	–	–	+	+	+
QIK	–	+	+	+	–
SS	+	–	+	–	NI
PP	+	+	+	+	–
GCK	+	–	–	+	+

will be further reviewed and compared in the next chapter. Since simulation approaches are essentially not regarded as estimation models, any further comparison falls outside of the scope of this dissertation.

2.6 Data Uncertainty in a Spatial Statistical Approach

Applications that involve the assessment of spatial uncertainty usually consist of the data, the spatial statistical model, and the follow-up model that uses the outcome of the spatial statistical model (p. 369 of Goovaerts 1997b), each of which involve a certain level of uncertainty. As described in section 2.2.1, uncertainties of the estimated attributes (outcome of the spatial statistical model) and the parameters of follow-up model can be expressed in the form of value distributions, subsequently propagated to the assessment output via mathematical analysis or stochastic simulation. It is recognized as well that different spatial statistical models result in different uncertainties depending on the objectives of the statistical model (McBratney et al., 2000). Uncertainty quantification depends on the selection of the model that may or

may not be validated without consideration of the specific needs and appropriateness of the inherent assumptions of the techniques. Examples of this observation could be found in the various types of kriging approaches that generate different uncertainties under different estimation settings, such as the use of estimation objectives for variability reproduction (Cressie, 1993; Aldworth and Cressie, 2003) or the incorporation of secondary information (Goovaerts, 1998). What is also important, in addition to the uncertainties described above, is the uncertainty involved in the input dataset.

Uncertainty in the estimates is a derivative from both the uncertainty in the statistical model as described above, and the uncertainty in the data input. The uncertainty in the *data*, which may come from different sources, is usually expressed in ensemble as the *nugget effect* in a semivariogram. The sources of this data uncertainty include the lack of data (sparse data sampling with micro-scale variability that could not be detected at the minimum sampling distance used) (Schnabel and Tietje, 2003; Western et al., 2004; Su et al., 2006), errors associated with laboratory measurements (Crist, 1998; Holmes et al., 2005; Western et al., 2004), and other procedural errors such as sample dislocation (Holmes et al., 2005; Goovaerts, 1994). Different values for the nugget effect modeled in a semivariogram may influence the result of the estimation (Schnabel and Tietje, 2003). The artificially induced error and micro-scale components should be modeled separately whenever possible (p. 103 of Goovaerts 1997b), while the approaches to distinguish these two components usually involve the comparison of direct- and cross-variograms which requires measurements of more than two attributes (Goovaerts, 1994; Lin et al., 2006). For measurements of a single attribute, however, the nugget effects are typically assumed to involve mainly micro-scale variability (Matheron, 1971), or mainly artificially induced error and discarded or reduced by averaging multiple collocated measurements (Su et al., 2006).

2.7 Spatial Statistics as Tools for Expressing Spatial Relatedness

In addition to providing the basis for spatial estimation and simulation, spatial structures in an estimation model (parameters that express spatial relatedness) also serve as exploratory parameters for scientific explanation of causal/covarying relations among observations in space. The definition of the spatial structures, however, may be expressed in various aspects by the way data are collected and summarized. For example, Crist (1998) used variogram analysis to indicate that spatial dependence in termite occurrence was evident; Jung et al. (2006) used cross-variograms to indicate that some variables (cation exchange capacity, total N, aggregate stability) exhibit correlatedness with electrical conductivity (EC) while no spatial structure is observed when considered separately; Mear et al. (2006) calculated the variogram of fine-grained content in the superficial sediments and stated that a whole effect model of the variogram corresponds to patches of various sizes; Desbarats and Bachu (1994) evaluated the Hurst coefficient and fractal dimensions of hydraulic transmissivity and concluded that the self-similar behavior "reflects a hierarchy of nested correlation structures with increasingly large characteristic length scales." Caniego et al. (2005) used different types of multifractal spectra on certain soil properties (electrical conductivity, organic matter content, soil PH, etc.) measured along transects, to observe whether the corresponding multifractal dimensions have a suitable scaling behavior. These findings expressed different scaling/spatial dependence of the value observed in various perspectives. However, a limited number of studies has explored the covariance/correlation between mean values at different spatial scales, which is possibly the most intuitive measure for the spatial statistics in expressing relationships across scales, with an exception possibly only of the fundamental works by Matheron (1971). Moreover, even though the theory given by Matheron (1971) does parameterize the

covariance among mean values, the development of this work and follow-up research focuses mostly (if not completely) on the point-to-area estimation (block kriging) partially due to the mining practice, where most of the original geostatistical developments were made (Kyriakidis, 2004).

Dale et al. (2002) provide an overview of most of the mathematic tools used for spatial data analysis, and quantitatively explain the differences among the approaches, including the use of variograms among the 28 approaches. It was concluded that, in spite of the diversity of the backgrounds and motivation that give rise to the methods described in their paper, there are some obvious conceptual themes and mathematical similarities that tie them together. While one may not expect that any single method can reveal all the important features of any data set, it should be noticed that the results of different analyses may not be fully independent (Dale et al., 2002). Depending on the specific questions to be answered about spatial characteristics of a given dataset, new approaches are yet to be developed. For example, the use of the covariance between mean values as an exploratory measure may provide a perspective for scientific research focused on the relations among different types of observations in space.

2.8 Knowledge Gaps

2.8.1 Multi-Scale Relatedness of Mean Values as Estimation Parameters

As a scientific approach, spatial structure evaluation provides an opportunity to explore the relationship between measurements at different scales for site attribute

values such as contamination, microbial abundance and activity. The interest for this multi-scale relationship persists among scholars, with different approaches not necessarily involving the use of spatial statistical approaches. However, no quantitative relationships have been developed to describe the strength of relations between samples of different physical sizes, for example by actual mean/average values at different scales.

2.8.2 Variability-Reproducing Estimation on the Basis of Local Mean Values Without Post-Processing

All the estimation and simulation approaches described in the previous sections are decision support tools for site remediation. OK provides a visual tool of the general trend of attribute values, IK further provides the flexibility to assume non-Gaussian attribute distributions. Conditional simulation also provides visual presentation by the equiprobable realizations, while no single map is more representative than another and the expected exceedance is biased for remediation purposes. Post-processing of estimates levels the trade-off between the over-smooth OK and IK approaches for the application of on-site remediation, while deviate from the objectives that were designed for the original estimation approach, e.g. estimates with an objective to minimize their error will not correspond to a minimized error after post-processing. CK and CM look like the best approaches that provide the visual representation of the spatial distribution of site attributes, while the estimates may or may not exist depending on the covariance between sample and estimation locations. No estimation or simulation method serves the need for a single estimation map with estimates at all locations, that describes the spatial variability (e.g. of sediment contamination) in terms of the relatedness between locations, with a criterion that emphasizes preservation of the variability for the estimates, without altering the estimation objectives

such as the minimization of error of estimates.

2.8.3 Delineation of the Extent of Contamination with Corresponding Likelihood Suitable for Decision Making

Delineation of the extent of contamination plays a crucial role in remediation decisions. Critto et al. (2005) used the exceedance of ordinary kriging results to evaluate the risk degree of pollutants stock in the Venice lagoon sediment, which is a derivative of a smoothed map and prone to the underestimation of exceedance (see discussions in p.182 of Goovaerts 1997b). Barabás et al. (2001) used cross-validation of the IK results to decide on a best likelihood threshold for the IK likelihood map in order to minimize the misclassification rates based on the site estimation over the entire research site. The approach, however, sacrifices the flexibility for stakeholders to compromise among alternatives that correspond to different likelihood of exceedance (see Chapter 6 for detailed discussion). Although evidence is shown in Aldworth and Cressie (2003) that CK and CM are unbiased in estimating the likelihood of threshold exceedance for selected locations (estimation location around where the sample is most spatially concentrated and where the sample is least concentrated) within a certain block (triangular sets of estimation grids), no further evidence are shown for the general applicability when the locations and block sizes are not as limited to the specification in that study.

In summary, most decision tools for site assessment and remediation depend on spatial estimation/simulation either using a biased estimation map of the likelihood of exceedance, or by assigning a single decision map corresponding to a given uncertainty criterion. A decision on the basis of the biased likelihood map will be risky when the likelihood map is derived from a smoothed estimation map, while on the other hand

optimizing the misclassification rate by assigning a single contamination zone gives less flexibility to interpret spatial data for decision making.

2.8.4 Quantification of Artificially Induced Error

In spite of the various studies that discuss model development, selection, and error propagation, little is done to study the uncertainty in the sampled data and its impact on the estimates and estimation variance. Accounting for the artificially induced error, however, may totally change the interpretation of the extent of contamination (Ramsey et al., 2002). For studies involving a single attribute, no geostatistical research has been done to evaluate the different components in a measurement that may include the artificially induced error, micro-scale variability not reflected in the dataset, and the structural signal reflected in the dataset. Quantification of artificially induced error vs. actual variability will inform the interpretation of the estimates/estimation variances using the dataset as the input.

2.9 Research Hypothesis

This research attempts to address the knowledge gaps by developing an estimation model which generates a single estimation map that represents the spatial variability on the basis of local means, without the need for post-processing. Quantifying the artificially induced error in the measurement, will further improve the interpretation of the outcomes of the estimation model. The specific hypothesis states that:

> *Effective site assessment and remedial decision making on the basis of delineating the extent of contamination depends on estimation methods and*

uncertainty quantification that reproduce the observed spatial variability of the contaminant concentrations. A better characterization of the extent of contamination can be attained by explicitly accounting for the covariances between multi-scale local means, and quantifying artificially induced error.

The single estimation map generated using the M-Scale model is useful for evaluating clusters of values and parameters derived from the estimation values, with corresponding likelihood map that supports the remediation decisions that could be further improved by quantifying the artificially induced error.

CHAPTER 3

Methods

The previous chapter concluded that a new spatial estimation method is needed that has the ability to produce a single estimation map on the basis of covariance between mean values at different scales, serving both scientific exploration needs and remedial decision support. In fact, a multi-scale estimation method capable of using the covariance between means would, under one single model, facilitate both the explanatory needs for the possible causal relations between different spatial scales by the covariances between means, and estimation purposes that depict the spatial variability over the study area. In this chapter, the M-Scale model is developed to fulfill the needs described above. The estimation uncertainty of the M-Scale model is reduced using Lagrange optimization, which is the same optimization concept used in OK and CK. The explicit incorporation of scaled information for the M-Scale model further enables the reproduction of spatial variability by an additional constraint attributing the largest weight to the scale that is most related (corresponding to the largest covariance) to the target scale, which is expected to reduce the smoothing effect observed in conventional kriging approaches.

3.1 Smoothing and Variability Reproduction

The smoothing effect of conventional kriging approaches is described in the literature in terms of the expected variance of estimates (Goovaerts, 1997b), the change in variogram (Carr, 1990), and in the visual observation of mapped estimates (Istok and Rautman, 1996; Goovaerts, 1998; Carr, 2002; Jung et al., 2006). The smoothing effect corresponds to the reduction of variability relative to the true distribution, which results from the nature of error minimization inherent in most kriging approaches. The reduction of variability implies that the proportions of high and low values in the estimated field do not reflect those in the sample. This is an undesirable feature because the proportion of area classified as exceedance/non-exceedance regarding a certain threshold value indicates the level of contamination for the study area, by which remedial decisions could be made. With a closer look at the equation for error minimization, it is found that the equations consist of a component for minimizing variance of the weighted average of samples (minimizing the impact by sample-to-sample covariances) (Chilès and Delfiner, 1999), and a component for maximizing the covariance between the weighted sample average and the value to be estimated (maximizing the influence between sample-to-estimate covariances) (pp. 299–301 of Isaaks and Srivastava 1989). The semantic meaning of the term "smooth," consequently, differs in the literature and can indicate a reduction of global variability (Goovaerts, 1997b; Yamamoto, 2005) due to the reduced variance of the weighted average (Maravelias et al., 1996), or a reduction of regional variability around sample locations (Kyriakidis, 2004) due to the "redundant" covariance between the weighted average and the variable to be estimated (Carr, 2002). Variability reproduction, consequently, should be examined in the context of its regionalization.

The classical smoothing effect resulting from conventional kriging approaches usually refers to the smoothing in a global perspective, with less variability relative to

the pooled samples when all estimates throughout the estimation domain are collapsed into univariate values regardless of their location. Regionally, however, the conventional kriging estimates usually exhibit of high gradients around sample locations to match the sample values, rendering peaked estimation surfaces that visually do not look smooth (see for example, p.51 in Ripley 2004, Goovaerts 1998, p.132 in Goovaerts 1997b, p.159 in Chilès and Delfiner 1999, and later in this chapter referred to as"concave smoothing" as shown in Fig. 3.2). This type of concave smoothing is also described in p. 439 of Goovaerts (1997b) that "high frequency components are progressively filtered as the location being estimated gets farther away from data locations." On the other hand, the cause of the regional smoothing around sample locations as described in Kyriakidis (2004) has never been discussed, although in some studies this type of regional smoothing is described as being "unsmoothed (Krige and Assibey-Bonsu, 2001)," and referred to as "convex smoothing" as shown in Fig. 3.3 later in this chapter, in contrast to the conventional global smoothing.

Variability reproduction in estimates has been implemented by the use of post-processing of estimates (Yamamoto, 2005; Journel et al., 2000), combining high- and low- pass filter components in kriging estimates (Carr's filtering approach) (Carr, 1990) constrained kriging (CK) (Cressie, 1993) and covariance-matching constrained kriging (CM) (Aldworth and Cressie, 2003). The problem with the post-processing is that these approaches manipulate the estimates ad hoc to achieve the objective of estimation after an expected value is evaluated, a result which should be regarded as a function of the best estimate of the random variable rather than a best estimate itself. Carr's filtering approach evaluates the expected value of a variable with superimposed high-pass and low-pass components for the reproduction of the global (marginal) value distribution, while the problem lies in that the cross-covariance between sample value and the two filter components of the estimate are difficult to evaluate a priori on the basis of solid mathematics. In a paper describing the filtering

approach, Carr used the auto-covariance evaluated from samples of the attribute as a surrogate for the two (supposedly different) cross-covariances (Carr, 1991; Ma, 1993; Carr, 1993). Although CK and CM attempt to reproduce the spatial variability by adding additional constraints on the estimates to reproduce the (co)variance among estimates, problems arise when the covariance between samples and estimates are generally small, resulting in unstable or no estimates at locations farther from the sample locations (Aldworth and Cressie, 1999; Tercan, 2004; Cressie et al., 2005). Aldworth and Cressie (2003) proposed an approach to divide the estimation domain into partitions (sub-regions) to "relax" (Cressie et al., 2005) the constraint to reproduce (co)variances between the sample and estimation locations "within, but not between, partitions, (Tercan, 2004)" and eventually when the estimates become unavailable for these estimation approaches, "we may have to choose instead the universal kriging predictor, (Aldworth and Cressie, 1999)" which means using UK, or OK if a drift model is not defined, as a surrogate estimate. This estimation procedure, unlike the post-processing procedures, does not explicitly violate the concept of finding an expected value as an estimate. The problem however, is that the estimates are subject to inconsistent objectives depending on whether or not the estimation locations fall within the influence range (correlation length of variogram) of the samples.

Area-influenced-kriging (Arik, 2002), or AIK, included the idea of having heavier weights to the closest sample point, a simple implementation that included an objective constraint for variability reproduction. The constraint used in AIK follows a philosophy that "everything is related to everything else, but near things are more related than distant things (Tobler, 1970)," which is generally known as Tobler's first law of geography (Sui, 2004). In fact, the idea indicated in Tobler's law has long been involved in the area of spatial statistics, while the application has been mostly for the characterization of spatial relatedness (Miller, 2004) with little attention for its implication as an objective except a few papers such as Journel et al. (2000); Brown et al.

(2002); Arik (2002); Aldworth and Cressie (2003). AIK is an estimation approach that has consistent objectives among estimation locations. This approach, however, relies on a parameter called "the proportion of area influence" that should be decided by physical parameters of the attribute (for example, by comparing estimates of the tonnage to the theoretical value or historical blast hole data, as indicated in Arik 2002), which is not readily available in most disciplines of research.

3.2 Spatial Pattern Characterization, Multi-Scale Characterization and the Use of Multi-Scale Information

Spatial pattern characterization (describing the spatial pattern of variables, or *structural analysis* in Chilès and Delfiner, 1999) is a major interest of research among ecological researchers. This field of research looks for evidence of regularity, randomness or clumping of an attribute of interest (Kent et al., 2006). The area of research extends from remote sensing (Woodcock and Strahler, 1987; Pax-Lenney and Woodcock, 1997; Zarco-Tejada and Miller, 1999; van Meirvenne and Goovaerts, 2002), glacial surface recognition (Herzfeld, 1999), ecological measurements (Crist, 1998; Nanos et al., 2005; Grego et al., 2006; Kent et al., 2006), to soil (Goovaerts, 1998; Wu et al., 2002; Romshoo, 2004; Ruffo et al., 2005; Su et al., 2006; Jung et al., 2006) and sediment characterization (Longhitano and Nemec, 2005; Adriaens et al., 2006; Mear et al., 2006). The main interest has mostly been the recognition of spatial association. Caniego et al. (2005) reviewed different approaches that characterize spatial patterns of attributes, and concluded that a description of relatedness between different scales remains to be studied. Multi-scale statistics, such as the correlation of mean values between different spatial scales, will facilitate the explanation of association between

different attribute values at spatial scales characterizing different physical behaviors (Brockman and Murray, 1997; Lin et al., 2006). Despite the interest for development and application of the multi-scale characterization, however, little research had been done to incorporate multi-scale information in the reproduction of spatial variability.

Block kriging is a common approach that uses point measurements to estimate the mean value over a target support volume centering at the estimation locations. The approach may be considered as a multi-scale approach, given its flexibility in terms of the size of the target support. However, this approach addresses only point-to-area interpolation, without considering the case for an area-to-point counterpart (Kyriakidis, 2004). In fact, both point-to-area and area-to-point estimation are special cases of the originally formulated spatial estimation using point support or block areal data as inputs (Matheron, 1971). The development of an estimator in Matheron's and later work, however, focus mostly (if not completely) on point-to-area estimation (block kriging) partially due to mining practice, where most of the original geostatistical developments have been made (Kyriakidis, 2004). To address this knowledge gap, a detailed mathematical expression of the area-to-point kriging is described in Kyriakidis (2004). Yoo and Kyriakidis (2006) used this area-to-point kriging to account for situations where the data used for interpolation of point estimates come from different sample volumes (called areal information). Additional to this multi-scale nature of the model designed to use inputs of different spatial scales, they imposed local constraints and inequality limitations informed by external knowledge of their estimate, in order to achieve the reproduction of spatial variability. With the multi-scale inputs and outputs in their model development, however, the applicability to reproduce spatial variability using areal information is not compared to the conventional point-to-point kriging in the case where no external constraining informations are provided (Yoo and Kyriakidis, 2006). In his construction of the framework for area-to-point spatial interpolation, Kyriakidis compared the charac-

Table 3.1: Comparison of variogram-related spatial estimation models. Positive signs indicate the inherence of the feature, and negative signs indicate the absence of the feature. OK: ordinary kriging. PP: post-processing approaches. CF: Carr's filtering approach. CK: constrained kriging. CM: covariance-matching kriging. AIK: area-influenced kriging. BOK: Block ordinary kriging. APK: area-to-point kriging.

Estimation Models	Reproduces Attribute Variability	Consistent Estimation Objectives	Uses Expected Value as Estimate	Generates Areal Covariance	Requires Only Point Samples
OK	–	+	+	–	+
PP	+	+	–	–	+
CF	+	+	+	–	–
CK	+	–	+	–	+
CM	+	–	+	–	+
AIK	+	+	+	–	–
BOK	–	+	+	+	+
APK	+	+	+	+	–

teristics between point-to-point and area-to-point kriging covariances and estimates (Kyriakidis, 2004). It can be observed by the result of this study that the recognition of measurements as areal information improved the reproduction of variability of the estimates with respect to the actual measurements, at least visually in the vicinity of the measurement. Additionally, the estimates are also shown to be "coherent" (as described by Kyriakidis), meaning that the estimates reproduce the mean values of the areal measurements when averaged over the original sampler volume at its original measurement location. Statistics for the reproduction of variability are not presented in this research, although the estimation results of the area-to-point kriging look visually more variable over the estimation domain than the conventional point-to-point estimates because of the use of areal information. In addition, the approach requires areal information to reproduce variability in the estimates. In a sample where only point measurements are available, no area-to-point estimates can be generated. To sum up the characteristics of the spatial estimation models described above, a comparison chart is listed as Table 3.1.

3.3 The M-Scale Perspective

The M-Scale model is proposed here in order to generate estimates that reproduce spatial variability as reflected by the available sample, with neither the conflict of not using expected values as in post-processing approaches, nor the inconsistency of estimation objectives at different locations that CK and CM are subject to. This approach also involves the idea of using areal information for estimation similar to area-to-point kriging. The areal information, however, is derived from point measurements representing mean values at different scales, introducing a different perspective for areal information aside from that of the area-to-point kriging. Unlike AIK, which requires knowledge of the physical nature of the attribute values, the proposed model requires no external information besides the sample measurements, plate sizes of different scales (explained later), and the variogram model. In other words, the M-Scale model is designed to incorporate all features indicated in Table 3.1.

Technically, instead of applying a constraint on the estimate to reproduce an expected variance/covariance like CK or CM, a constraint for the local average is applied. Similar to the AIK, the M-Scale model follows Tobler's Law in applying the constraint for variability reproduction, attributing the heaviest weight to the closest local averages. From the scientific perspective, this approach is conceptually regarded as multi-scale estimation, including the ideas of point-to-area, area-to-point, and area-to-area estimation, since the areal information is derived from point measurements (point-to-area estimation), and the target support of the estimation output can be either a point (resulting in area-to-point estimates) or a block (resulting in area-to-area estimates). From a practical perspective, in addition to the features it provides as shown in Table 3.1, the limiting assumption of areal data as inputs in the area-to-point kriging described is relaxed, providing an alternative to the use of areal information for the reproduction of variability.

With observations of results from Kyriakidis (2004) and Yoo and Kyriakidis (2006), the constraint to preserve local averages is expected to reproduce spatial variability. To validate this conjecture, the estimation results under different application cases will be compared quantitatively to the results generated by OK and CK in later chapters, as a demonstration of the characteristic performance of the M-Scale model for estimation and variability reproduction. The following section, however, begins with reviews of the methods used and developed in this dissertation, including the concepts of variogram-related spatial interpolation models that the M-Scale model inherits. Along with the mathematical/statistical derivation, particular focus is placed on the characteristics that define the precision of delineating threshold exceedance. To facilitate the presentation of the M-Scale model as well as to provide a basis for comparison, the theoretical development of ordinary kriging (OK) (Matheron, 1971) and constrained kriging (CK) (Cressie, 1993) are described first.

3.4 Ordinary Kriging

One of the central concepts of geostatistical analysis is the *variogram*. A variogram is a mathematical expression of dissimilarity described in Chapter 2 as a function of separation distance between any two points in space (Matheron, 1961), and is defined as follows:

$$\gamma(h) = \frac{1}{2} E\left\{[Z(x+h) - Z(x)]^2\right\} \tag{3.1}$$

where $\gamma(h)$ is the variogram value, or *semivariance*, $Z(x)$ and $Z(x+h)$ are two random variables representing attributes separated by a distance h. Fig. 3.1 illustrates a generic variogram, where h is the distance between any two locations, the *influence*

range represents the range of correlation between location pairs, the *sill* and *nugget effect* indicate the possible maximum and minimum semivariance values for long and short separation distances respectively, and the curve in this graph explains the relative strength of interaction in terms of the distance between the sample locations.

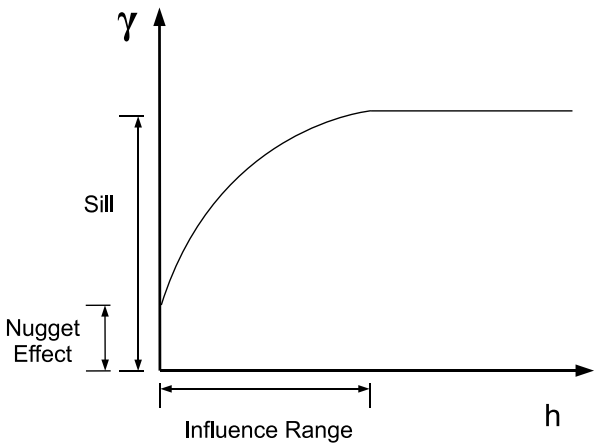

Figure 3.1: Conceptual sketch of a variogram. γ represents the semivariance, and h represents the separation distance. The characteristic parameters of a variogram are illustrated, including the sill, the nugget effect, and the influence range.

Generally, for variogram-based approaches the spatial attribute $Z(\cdot)$ is regarded as a random function that relates location x to the random variable $Z(x)$ representing the local attribute at location x, with a conditional cumulative distribution function (ccdf) $F(x; z|(n))$ at location x, conditional to the n neighboring data $z(x_i)$ (Istok and Rautman, 1996; Barabás et al., 2001; Critto et al., 2005). This ccdf was described in Chapter 2 as the value distribution of weighted sample values under the evaluated uncertainty to represent the real attribute value. The ccdf fully captures the uncertainty at x since it gives the likelihood that the unknown is no greater than any given threshold z (Goovaerts, 1997b).

Suppose $Z_i \equiv Z(x_i)$ represents a possible attribute value at a certain sample location x_i, λ_i represent estimation weights given to Z_i for the linear estimation, and $Z_0 \equiv Z(x_0)$ is the unknown variable to be estimated at the estimation location

x_0. The OK estimator is designed to minimize the estimation variance (Isaaks and Srivastava, 1989):

$$\begin{aligned}\sigma^2_{OK} &= E\left[Z_0 - \sum_i \lambda_i Z_i\right]^2 \\ &= E\left[\sum_i \lambda_i (Z_0 - Z_i) + (1 - \sum_i \lambda_i) Z_i\right]^2 \end{aligned} \tag{3.2}$$

Note that $Z_i, i = 1 \ldots n$ are *possible* values expressed as random variables, by which an estimator for the unknown Z_0 is derived, and the estimated value for Z_0 in terms of $Z_i, i = 1 \ldots n$ can be evaluated when the corresponding actual sample values $z_i, i = 1 \ldots n$ are determined. A signature characteristic of OK is the constraint on the assumption of first-order stationarity $E[Z_i] = E[Z_0] = m$, where $m = E[Z_D] = E\left[\frac{1}{\{D\}} \int_D Z(x)dx\right]$ is the global mean of the population, with D being the spatial domain of size $\{D\}$ (Journel, 1985). The constraint for global unbiasedness is consequently constructed as:

$$\begin{aligned} & E\left[\sum_i \lambda_i Z_i\right] = E[Z_0] \\ & \Rightarrow \sum_i \lambda_i m = m \\ & \Rightarrow \sum_i \lambda_i = 1 \end{aligned} \tag{3.3}$$

Taking account of the global unbiasedness, the estimation variance can subsequently be expressed as

$$\begin{aligned}\sigma^2_{OK} &= E\left[\sum_i \lambda_i (Z_0 - Z_i)\right]^2 \\ &= \sum_i \sum_j \lambda_i \lambda_j E\left[(Z_0 - Z_i)(Z_0 - Z_j)\right] \end{aligned} \tag{3.4}$$

which can be further derived into an equation expressed in terms of semivariances:

$$\begin{aligned}
\sigma_{OK}^2 &= \sum_i \sum_j \lambda_i \lambda_j E\left[Z_0^2 - Z_0 Z_j - Z_i Z_0 + Z_i Z_j\right] \\
&= \sum_i \sum_j \lambda_i \lambda_j \frac{1}{2} E\left[Z_0^2 - 2Z_0 Z_j + Z_j^2 + Z_0^2 - 2Z_i Z_0 + Z_i^2 - Z_i^2 - Z_j^2 + 2Z_i Z_j\right] \\
&= \sum_i \sum_j \lambda_i \lambda_j \left(\gamma_{0j} + \gamma_{0i} - \gamma_{ij}\right) \\
&= 2\sum_i \lambda_i \gamma_{i0} - \sum_i \sum_j \lambda_i \lambda_j \gamma_{ij}
\end{aligned} \tag{3.5}$$

This equation can be understood as quantitatively representing the dissimilarity between the unknown and estimated values (the estimation variance σ_{OK}^2), expressed in terms of the dissimilarity between sample and unknown values γ_{i0}, and the dissimilarity between sample values γ_{ij}.

Another perspective to look at the relation between two local attribute values is the covariance between random variables Z_i and Z_j (Collins and Woodcock, 1999; Kukush, 2005), expressed as

$$C_{ij} = E\left\{[Z_i - m][Z_j - m]\right\} \tag{3.6}$$

Because m is constant over the spatial domain and C_{ij} is a function of Z_i and Z_j like γ_{ij} is, it is possible to express the semivariance γ_{ij} (the dissimilarity) between locations x_i and x_j in terms of the covariances C_{ii}, C_{jj} and C_{ij} (the similarity). In fact,

$$\begin{aligned}
\gamma_{ij} &= \frac{1}{2} E\left[Z_i - Z_j\right]^2 \\
&= \frac{1}{2} E\left[(Z_i - m) - (Z_j - m)\right]^2 \\
&= \frac{1}{2}\left[C_{ii} + C_{jj} - 2C_{ij}\right]
\end{aligned} \tag{3.7}$$

In most cases, second-order stationarity is also assumed for kriging approaches (Matheron, 1971; Legendre and Legendre, 1998; Wu et al., 2002), i.e., the semivariance and covariance between Z_i and Z_j is determined by the separation distance between the corresponding locations x_i and x_j:

$$\begin{aligned} C_{ij} &= C(h = |x_i - x_j|) \\ \gamma_{ij} &= \gamma(h = |x_i - x_j|) \end{aligned} \tag{3.8}$$

Thus

$$\begin{aligned} \gamma_{ij} &= \frac{1}{2}\left[C_{ii} + C_{jj} - 2C_{ij}\right] \\ &= \sigma^2 - C_{ij} \end{aligned} \tag{3.9}$$

where $\sigma^2 = C(h = 0) = C_{ii} = C_{jj}$ is the global variance. This relation between the semivariance and the covariance leads to the following expression for the OK estimation variance:

$$\sigma^2_{OK} = \sigma^2 - 2\sum_i \lambda_i C_{i0} + \sum_i \sum_j \lambda_i \lambda_j C_{ij} \tag{3.10}$$

in which σ^2 is the global variance, C_{i0} is the covariance of the sample and the unknown, and C_{ij} is the covariance between samples. To minimize this estimation variance under the unbiasedness constraint, the method of Lagrange multipliers (Lagrange digitized from the 1760–1761 publication, which is also called Lagrange parameter by Goovaerts 1998) is applied, with the following Lagrangian (see p.133 of Goovaerts 1997b):

$$L = \sigma^2_{OK} - 2\mu\left(\sum_i \lambda_i - 1\right)$$

$$= \sigma^2 - 2\sum_i \lambda_i C_{i0} + \sum_i \sum_j \lambda_i \lambda_j C_{ij} - 2\mu \left(\sum_i \lambda_i - 1 \right) \tag{3.11}$$

The weights $\lambda_i, i = 1 \ldots n$ are evaluated by solving the following matrix equation corresponding to Equation (3.11):

$$\begin{bmatrix} \mathbf{C_{ij}} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \mu \end{bmatrix} = \begin{bmatrix} \mathbf{C_{i0}} \\ 1 \end{bmatrix} \approx \alpha_i^2 \tag{3.12}$$

where $\mathbf{C_{ij}}$ is a $n \times n$ matrix of C_{ij}, $\mathbf{C_{i0}}$ is a $n \times 1$ vector of C_{i0}, $\mathbf{1}$ is a $n \times 1$ vector of ones, $\boldsymbol{\lambda}$ is a $n \times 1$ vector of λ_i, and μ is the Lagrange parameter to ensure the global unbiasedness.

When optimized, the estimation variance can be written as the following equation for the convenience of calculation:

$$\sigma_{OK}^2 = \sigma^2 - \sum_i \lambda_i C_{i0} - \mu \tag{3.13}$$

In most OK applications, the estimation of $C(h)$ is done by modeling $\gamma(h)$ from the experimental variogram calculated from discrete sample points (van Meirvenne and Goovaerts, 2002; Nanos et al., 2005; Lin et al., 2006) because, unlike $C(h)$, $\gamma(h)$ does not require knowledge of the unknown global mean Z_D. The stabilized semivariance for a larger lag distance, indicating no correlation between samples, is taken as the estimated σ^2 because one-half of the variance of difference between two non-correlated variables coincides with the global variance value. The estimation of covariances is subsequently applied to Equation (3.12) to solve for $\boldsymbol{\lambda}$, so that the estimate and estimation variance can be evaluated.

The problem with OK, however, is that only the global unbiasedness constraint is applied in correspondence to the first-order stationarity (Cressie, 1993). No con-

straint to reproduce spatial variability is placed on the Lagrange optimization, although second-order stationarity (the consistency of spatial variability) is assumed so that the variogram/covariance can be estimated from the sample set. The lack of constraint to reproduce spatial variability result in a "concave" smoothing effect, due to which estimates tend toward the global mean as the separation distance to adjacent measurements increases (Yamamoto, 2005), as shown in Fig. 3.2.

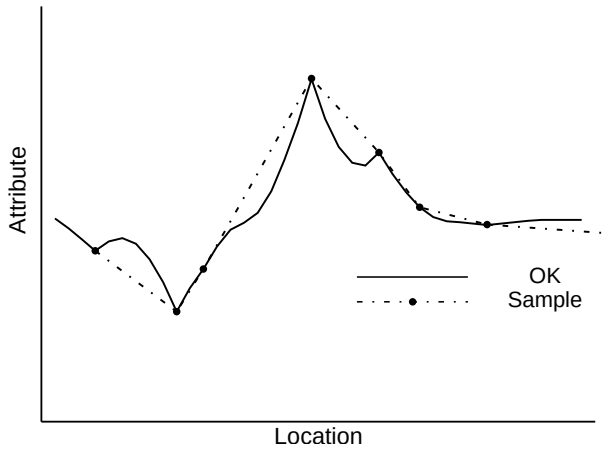

Figure 3.2: Concave smoothing effect of OK estimation. The estimated values tend toward the global mean value.

3.5 Constrained Kriging

The smoothing problem for OK originates from the fact that only one constraint corresponding to the first-order stationarity is applied. With no other constraints, the estimates generally exhibit a tendency toward the global mean to reduce the variance of the estimate $\sum_i \sum_j \lambda_i \lambda_j C_{ij}$. To reduce this type of smoothing effect, Cressie (1993) used another constraint to ensure that the variance of the estimate also matches the

global variance, in other words

$$\sum_i \sum_j \lambda_i \lambda_j C_{ij} = \sigma^2 \tag{3.14}$$

Consequently the estimation variance expressed by Equation (3.10) used for OK becomes the following expression when Euqation (3.14) is valid:

$$\sigma^2_{CK} = 2\sigma^2 - 2\sum_i \lambda_i C_{i0} \tag{3.15}$$

The Lagrangian to obtain the minimum estimation variance under the unbiasedness constraint and the constraint for global variance reproduction (Equation (3.14)) becomes

$$L = 2\sigma^2 - 2\sum_i \lambda_i C_{i0} - 2\mu_1 \left(\sum_i \lambda_i \right) - \mu_2 \left(\sum_i \sum_j \lambda_i \lambda_j C_{ij} - \sigma^2 \right) \tag{3.16}$$

where μ_1 and μ_2 are Lagrangian multipliers to ensure the two constraints are fulfilled. Solutions for $\boldsymbol{\lambda}$, μ_1 and μ_2 can be obtained by the following equations:

$$\begin{aligned}
\boldsymbol{\lambda} =& \frac{1}{\mu_2} \left(\mathbf{C_{i0}} \mathbf{C_{ij}^{-1}} + \mu_1 \mathbf{C_{ij}^{-1}} \mathbf{1} \right) \\
\mu_1 =& \frac{1}{\mathbf{1^T} \mathbf{C_{ij}^{-1}} \mathbf{1}} \left(\mathbf{1^T} \mathbf{C_{ij}^{-1}} \mathbf{C_{i0}} + \mu_2 \right) \\
\mu_2 =& \left[\frac{\mathbf{C_{i0}^T} \mathbf{C_{ij}^{-1}} \mathbf{C_{i0}} \cdot \mathbf{1^T} \mathbf{C_{ij}^{-1}} \mathbf{1} - \left(\mathbf{1^T} \mathbf{C_{ij}^{-1}} \mathbf{C_{i0}} \right)^2}{\mathbf{1^T} \mathbf{C_{ij}^{-1}} \mathbf{1} - 1} \right]^{\frac{1}{2}}
\end{aligned} \tag{3.17}$$

where $\mathbf{1}$, $\boldsymbol{\lambda}$, $\mathbf{C_{i0}}$ and $\mathbf{C_{ij}}$ are the same vectors and matrices with same dimensions for OK estimation. It is observed in Equation (3.15), however, that the variance constraint reduces the optimization capability by setting the sample-to-sample part of the estimation variance to a constant, i.e., by setting $\sum_i \sum_j \lambda_i \lambda_j C_{ij}$ to σ^2. Consequently, the estimation tends to attribute more weights to the samples with larger covariance

between itself and the unknown so that $\sum_i \lambda_i C_{i0}$ is maximized and consequently σ^2_{CK} minimized. The effect of this weight attribution is a "convex" smoothing effect with similar values around each sample location (Aldworth and Cressie, 2003), and large variability between regions around different sample locations, as shown in Fig. 3.3. In other words, the semivariance is under-reproduced for short separation distances, and over-reproduced for long separation distances. When applied to delineation of threshold exceedance, the area of exceedance would be overestimated around samples, and the false-positive rate of classification will be high.

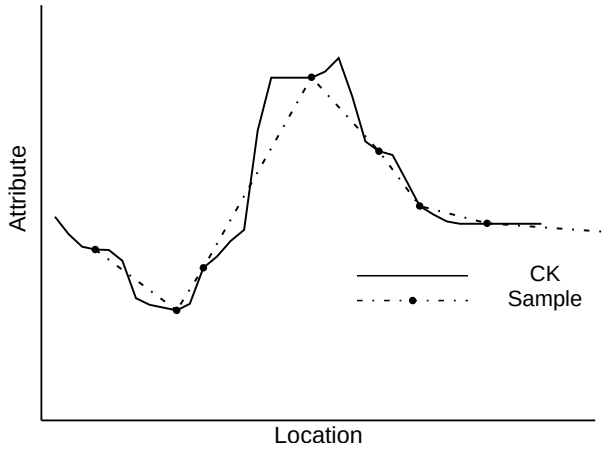

Figure 3.3: Convex smoothing effect of CK estimation. The estimated values tend toward the sample values.

Because CK is designed to reproduce both the global mean and global variance, however, the total amount of area delineated for threshold exceedance over the entire spatial domain is expected to reproduce the actual total area of threshold exceedance particularly when multi-Gaussianity is assumed. In other words, CK is expected to reproduce the total amount of area for threshold exceedance, and if part of the true exceedance is surrounded by samples indicating non-exceedance, the amount of area for threshold exceedance will be complimented by over-estimating the area of exceedance around samples which do indicate exceedance. The false-negative rate of the classification, consequently, may not be high since the over-estimated area already

indicates exceedance. This type of delineation imprecision is illustrated in case (e) of Fig. 2.13. To examine this type of error in delineation, it is suggested that covariograms of the estimates be compared to the covariogram modeled from the sample, and that the false-positive rate be evaluated, so that the smoothness of estimates between locations could be observed, and that the nature of the overestimated area of threshold exceedance could be revealed.

The problem with CK, consequently, is that the area of threshold exceedance is overestimated around sample locations that indicate exceedance, while in order to have global variance reproduction, the effect is balanced by underestimating the area of threshold exceedance where samples do not indicate exceedance of the threshold.

3.6 Attributes and Statistics for The M-Scale Model

Both OK and CK exhibit limitations in the delineation of threshold exceedance. OK estimates tend toward the global mean because of the lack of a constraint to reproduce variability, resulting in underestimation of the extent of threshold exceedance, while CK estimates tend toward sample values, which may overestimate the extent of threshold exceedance in the vicinity of samples that indicate exceedance. To address these limitations, the M-Scale model was developed to generate estimates that tend neither toward the global mean, nor toward the values of nearby sample locations. This model is based on two tenets: (i) calculating average values for different scales at each estimation location, and (ii) attributing different weights to these averages to generate the estimates, as shown in Fig. 3.4. The M-Scale model applies a constraint to reproduce variability, in addition to the unbiasedness constraint, while avoiding overestimation of the extent of threshold exceedance around the sample locations.

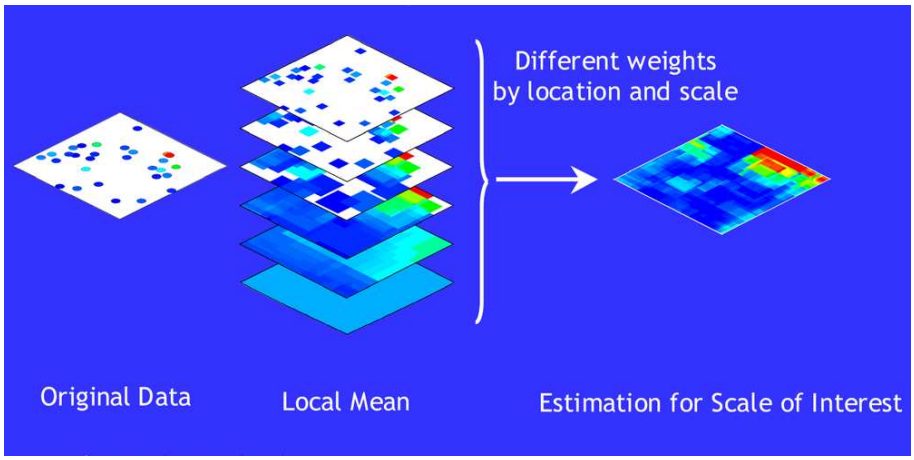

Figure 3.4: Conceptual sketch of the M-Scale model. Original data are used to evaluate the local mean at different spatial scales. The local mean are further attributed different weights and evaluated as the estimates.

Table 3.2: Comparing similar concepts and parameters for spatial attributes between the M-Scale model and conventional kriging approaches.

Spatial Attributes	M-Scale	Kriging
Unknown Value	$Z_{D_t}(x)$	Z_0
Sample Value	$Z^*_{D_a}(x)$	Z_i
Independent Variable	scale a at location x	location x

As a linear spatial estimator, the M-Scale model minimizes estimation variance, and the parameters and statistics are analogous to those used in OK and CK. The concepts and parameters described earlier for OK and CK are compared to those in M-Scale in Tables 3.2 and 3.3.

Table 3.3: Comparing similar spatial covariances between the M-Scale model and conventional kriging approaches.

Spatial Attributes	M-Scale	Kriging
Global Variance of Unknown	$C_{D_tD_t}$	$\sigma^2 = C(0)$
Sample-to-Sample Covariance	$C^*_{D_aD_b}(x)$	C_{ij}
Sample-to-Unknown Covariance	$C_{D_aD_t}$	C_{i0}

3.6.1 Basic Concepts and Statistics for Block-Related Spatial Attributes

The use of average values, instead of local point values, differentiates the M-Scale model from conventional kriging approaches. Conceptually, these average values are the inputs that produce the final estimate, analogous to attribute values at sample locations in a kriging setup. These input values in the M-Scale model, however, characterize significant differences in their statistical features from those of the local attribute values in the following ways:

1. The values depend not only on the estimation/sampling location, but also on the support sizes of the different spatial scales.

2. The average value of a spatial scale represents *the most likely* value within the area of that scale because the area consists of different attribute values at point locations, while the local attribute value represents *the only* value at each point location.

3. Because the average is calculated using attribute values at sample locations to represent the true mean value for all sampled and unsampled locations, uncertainty exists in the average value as an input value.

These characteristic differences are quantitatively defined in this section, with important concepts that guide the idea of multi-scale statistics used in developing the M-Scale model.

Adapted from the notation by Matheron (1971), the local mean of any exhaustive set of attribute values within volume $V(x)$ centering at x is denoted as

$$Z_V(x) = \frac{1}{\{V(x)\}} \int_{V(x)} Z(x')dx' \tag{3.18}$$

where $\{V(x)\} = \int_{V(x)} dx'$ is the measurement of $V(x)$. Two types of variance/covariance measurement between spatial scales are also defined and derived in Matheron (1971), respectively (i) the *extension variance*, denoted here as $C^V_{V'V'}$, and (ii) the covariance between Z_{V_1} and Z_{V_2} within V, called *dispersion covariance* in this research, which is denoted as $C_{V_1V_2|V}$. The extension variance is illustrated in Fig. 3.5, and is defined

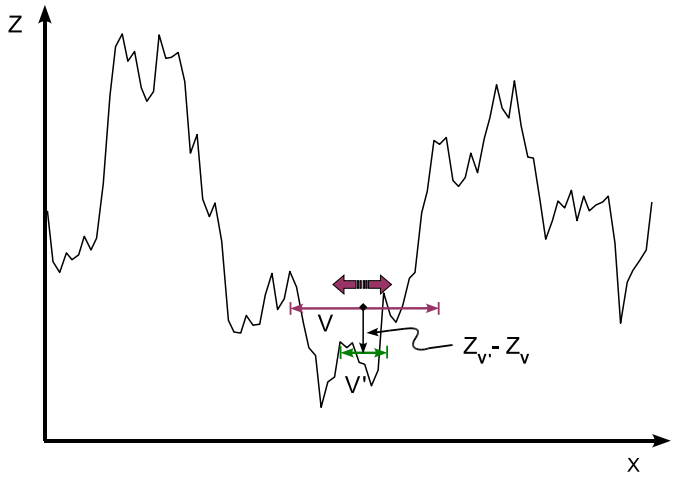

Figure 3.5: Conceptual sketch of extension variance. The green segment represents the mean $Z_{V'}$; The purple segment represents the mean Z_V. The double arrow on top of the purple segment indicates that the mean value changes as the centering location of V' and V changes.

as the uncertainty when a mean value of a smaller scale $Z_{V'}$ at a random location within V is used to represent the mean value of a larger scale Z_V:

$$\begin{aligned} C^V_{V'V'} &= E\left[(Z_{V'} - Z_V)^2\right] \\ &= \frac{2}{\{V'\}\{V\}} \int_{V'}\int_V \gamma(x-y)dxdy - \frac{1}{\{V\}^2}\int_V\int_V \gamma(x-y)dxdy \\ &\quad - \frac{1}{\{V'\}^2}\int_{V'}\int_{V'} \gamma(x-y)dxdy \end{aligned} \tag{3.19}$$

Dispersion covariance represents the block-to-block (between two volumes of certain scales) covariance chosen at random locations but with fixed distance and fixed mutual configuration, as illustrated in Fig. 3.6. Dispersion covariance is an analog of the covariance used in kriging approaches, except that in addition to the fixed separation

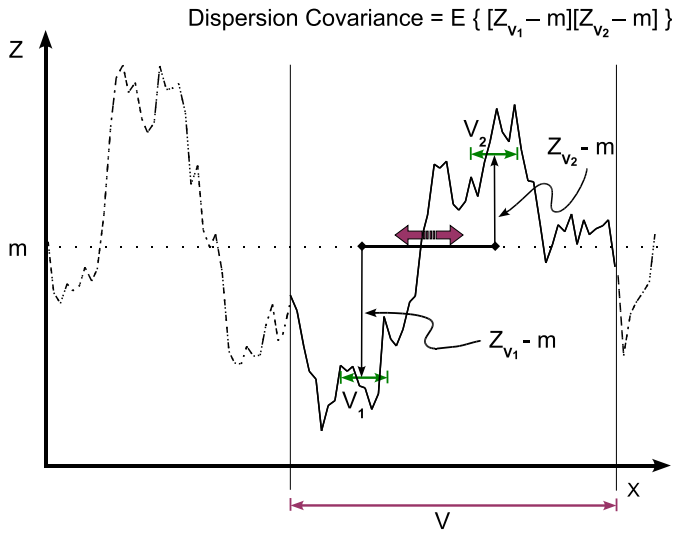

Figure 3.6: Conceptual sketch of dispersion covariance. The purple segment represents the mean $Z_{V'}$; The green segments represent the means Z_{V_1} and Z_{V_2}, and the black segment represents the separation distance between centers of V_1 and V_2.

distance, the fixed shapes, fixed sizes, and fixed relative rotation angle between V_1 and V_2 are also included in the definition. The definition becomes more complex because the two values are no longer attribute values at point locations. In addition to the definition for blocks V_1 and V_2, the random locations for the paired blocks are constrained within the larger block V, a constraint usually not applicable to kriging covariances:

$$\begin{aligned} C_{V_1V_2|V} &= E\left[(Z_{V_1} - m)(Z_{V_2} - m)|V_1, V_2 \subseteq V\right] \\ &= \frac{1}{\{V\}^2}\int_V\int_V \gamma(x-y)dxdy - \frac{1}{\{V_1\}\{V_2\}}\int_{V_1}\int_{V_2}\gamma(x-y)dxdy \end{aligned} \tag{3.20}$$

When the constraining block V expands to the entire domain D (indicating no spatial constraint), a simplified notation $C_{V_1V_2}$ is used for $C_{V_1V_2|D}$ to denote the covariance of Z_{V_1} and Z_{V_2} when no constraining block is used:

$$C_{V_1V_2} = E\left[(Z_{V_1} - m)(Z_{V_2} - m)\right]$$

$$= \frac{1}{\{D\}^2}\int_D\int_D \gamma(x-y)dxdy - \frac{1}{\{V_1\}\{V_2\}}\int_{V_1}\int_{V_2}\gamma(x-y)dxdy$$
$$= \sigma^2 - \frac{1}{\{V_1\}\{V_2\}}\int_{V_1}\int_{V_2}\gamma(x-y)dxdy \tag{3.21}$$

From equations (3.19) and (3.21), the following equivalence is derived:

$$C^V_{V'V'} = C_{V'V'} + C_{VV} - 2C_{V'V} \tag{3.22}$$

which is described in Matheron (1971) as "a perhaps more intuitive expression of the (extension) variance." Equation (3.22) can be understood as "conversion of the extension variance by the global dispersion variances and covariances." In this research, the idea of *extension covariance* is also defined, and its conversion from dispersion variance/covariance expressed as

$$C^V_{V_1V_2} \equiv E\left[(Z_{V_1} - Z_V)(Z_{V_2} - Z_V)\right]$$
$$= C_{V_1V_2} + C_{VV} - C_{V_1V} - C_{V_2V} \tag{3.23}$$

The extension covariance can be understood in the same way as the dispersion covariance, except that Z_V is used instead of m as the reference mean value. Also when V expands to the entire domain, $C^V_{V_1V_2}$ coincides with $C_{V_1V_2}$. To extend the definitions by Matheron (1971), so that the dispersion variance also applies to cases involving point variables, the point-to-point variance $C_{00}(x)$ and point-to-block covariance C_{0V} are also defined in this work as

$$C_{00} \equiv \frac{1}{\{D\}^2}\int_D\int_D \left[Z(x) - m\right]^2 dxdx$$
$$= \sigma^2$$
$$C_{0V} \equiv E\left[(Z(x) - m)(Z_{V'} - m)\middle|\, x, V' \subseteq V\right]$$

$$= \frac{1}{\{D\}^2}\int_D\int_D \gamma(x-y)dxdy - \frac{1}{\{V\}^2}\int_V\int_V \gamma(x-y)dxdy$$
$$= \sigma^2 - \frac{1}{\{V\}^2}\int_V\int_V \gamma(x-y)dxdy \tag{3.24}$$

Note that in this case the volume V is at a fixed distance and has a fixed configuration relative to the point location x. The basic definitions for the mean value, dispersion variance/covariance and extension variance/covariance will later be used in the M-Scale estimation, not only for cases when the target scale is a punctual support, but also when it is a block volume.

3.6.2 Definition of Geometric Components for the M-Scale Model with Their Representative Value and Uncertainty

One important concept in spatial statistics is that "points close in space tends to assume close values" (Chilès and Delfiner, 1999). It is consequently reasonable to group sample values by the separation distance to a target location that may be a sample location or an estimation location. For estimation purposes, therefore, the spatial domain D is split into a set of non-overlapping ring areas centering at any object location x in the entire domain, as shown in Fig. 3.7, so that the values within

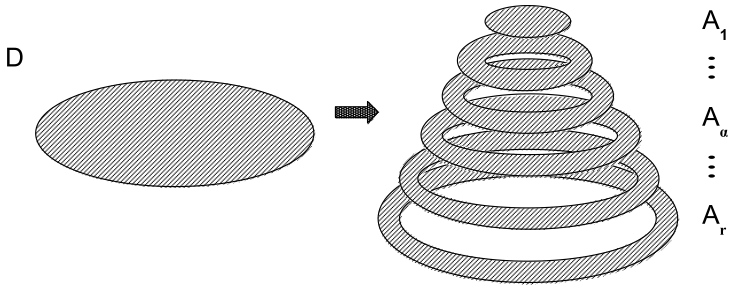

Figure 3.7: Domain splitting. The entire estimation domain is divided into ring areas by the distance from the estimation location. The rings pertains to fixed configurations with centers at the estimation location.

the same ring area have the same degree of similarity with the value(s) of the object point/block:

$$D = A_1(x) \cup \cdots \cup A_\alpha(x) \cup \cdots \cup A_r(x) \tag{3.25}$$

where α is the index for ring A_α, and A_r is the largest ring. The shape and size of each ring area (the geometry of which does not need to be circular) are kept unchanged over different estimation locations.

The representative attribute value for a certain ring area, called the *local mean of ring* $A_\alpha(x)$, can be evaluated with the corresponding dispersion variance for any object location x:

$$\begin{aligned} Z_{A_\alpha}(x) &= \frac{1}{\{A_\alpha(x)\}} \int_{A_\alpha(x)} Z(x')dx' \\ C_{00|A_\alpha}(x) &= \frac{1}{\{A_\alpha(x)\}} \int_{A_\alpha(x)} [Z(x') - Z_{A_\alpha}(x)]^2 dx' \\ &= \frac{1}{\{A_\alpha\}^2} \int_{A_\alpha} \int_{A_\alpha} \gamma(x' - x'')dx'dx'' \end{aligned} \tag{3.26}$$

where x is the center of rings, and x' and x'' are dummy variables for the integral.

In most geostatistical applications, an exhaustive collection of attribute values is generally not available across any specific area, and thus an experimental sample is selected. The experimental value $Z^*_{A_\alpha}(x)$ is used to represent $Z_{A_\alpha}(x)$ and is defined as the *local average of ring* $A_\alpha(x)$, is therefore calculated by

$$Z^*_{A_\alpha}(x) = \frac{1}{n_{A_\alpha}(x)} \sum_{i=1}^{n_{A_\alpha}(x)} Z_{A_\alpha,i} \tag{3.27}$$

which is expressed as a function of several random variables that represent the attribute values at the sample locations. $n_{A_\alpha}(x)$ denotes number of available measure-

ments within area $A_\alpha(x)$, and $Z_{A_\alpha,i}$ is the random variable representing the attribute value at each sample point within $A_\alpha(x)$. When samples are assumed to be randomly distributed over $A_\alpha(x)$, Matheron (1971) also shows that the expected variance of $Z^*_{A_\alpha}(x)$ relative to the true mean value $Z_{A_\alpha}(x)$ can be represented by

$$\sigma^2(Z^*_{A_\alpha}(x)) = E\left\{\left[Z^*_{A_\alpha}(x) - Z_{A_\alpha}(x)\right]^2\right\} = \frac{1}{n_{A_\alpha}(x)} C_{00|A_\alpha} \tag{3.28}$$

which can also be understood as "evaluating the uncertainty $\sigma^2(Z^*_{A_\alpha}(x))$ for $Z^*_{A_\alpha}(x)$ to represent its true mean value $Z_{A_\alpha}(x)$, by the point dispersion variance within A_α (Matheron, 1971)."

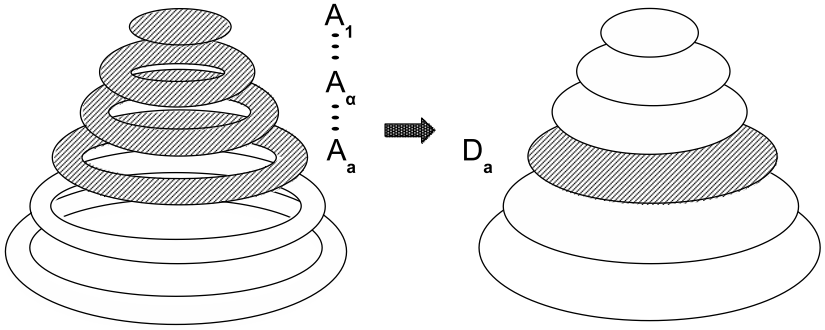

Figure 3.8: Combining for subdomains. The ring areas $A_1 \dots A_a$ at an estimation location are combined into subdomains D_a. The subdomains pertains to fixed configurations with centers at the estimation location.

The M-Scale models uses the average calculated at different scales as the input data. Therefore, the different spatial scales (*plates*), are defined by the subdomains $D_a(x)$ centered at an object location x:

$$D_a(x) = A_1(x) \cup \cdots \cup A_\alpha(x) \cup \cdots \cup A_a(x) \tag{3.29}$$

The conceptual sketch for the plates is shown in Fig. 3.8, where each plate $D_a(x)$ consists of ring areas equal or smaller than $A_a(x)$.

By definition, the subdomain mean $Z_{D_a}(x)$, or *local mean of plate* $D_a(x)$, can be

calculated from $\{A_\alpha(x)\}$ and $Z_{A_\alpha}(x), \alpha = 1 \ldots a$ as follows:

$$Z_{D_a}(x) = \frac{1}{\{D_a(x)\}} \int_{D_a(x)} Z(x')dx' = \frac{\sum_{\alpha=1}^{a} \{A_\alpha(x)\} Z_{A_\alpha}(x)}{\sum_{\alpha=1}^{a} \{A_\alpha(x)\}} \tag{3.30}$$

which can further be abbreviated as

$$Z_{D_a}(x) = \sum_{\alpha=1}^{a} w_\alpha^a(x) Z_{A_\alpha}(x) \tag{3.31}$$

where

$$w_\alpha^a(x) = \frac{\{A_\alpha(x)\}}{\{D_a(x)\}} = \frac{\{A_\alpha(x)\}}{\sum_{\alpha=1}^{a} \{A_\alpha(x)\}} \tag{3.32}$$

Hereafter, the terminology *mean* will be used to indicate the true mean value of $Z(x)$ in the subdomain $A_\alpha(x)$ or $D_a(x)$, and *average* will be used to indicate the corresponding (weighted) average evaluated from sample values to represent the mean values. The different means/averages and their associated variances will be used to define the elements in the spatial covariances.

3.6.3 The M-Scale Covariances

Several types of covariances will be defined in this section. The *plate-to-plate mean covariance for scales a and b*, denoted as $C_{D_a D_b}(x)$, is defined conceptually as the covariance between local means of plates $D_a(x)$ and $D_b(x)$ relative to the global mean of the spatial domain D. From the definitions in Equation (3.20), we can define:

$$\begin{aligned} C_{D_a D_b}(x) =& E\left[(Z_{D_a}(x) - m)(Z_{D_b}(x) - m)\right] \\ =& \frac{1}{\{D\}^2} \int_D \int_D \gamma(|x' - x''|)dx'dx'' \end{aligned}$$

$$- \frac{1}{\{D_a(x)\}\{D_b(x)\}} \int_{D_a(x)} \int_{D_b(x)} \gamma(|x' - x''|) dx' dx'' \tag{3.33}$$

Also with the assumption of second-order stationarity, all block-to-block covariances of the same configuration will have the same values. Hence the M-Scale covariances are independent of the centering location x. The covariance can thus be simplified as $C_{D_a D_b}(x) = C_{D_a D_b}$.

Breaking down $Z_{D_a}(x)$ in terms of the corresponding ring means at location x, we also have the following equations for point-to-plate, ring-to-plate and plate-to-plate mean covariances:

$$\begin{aligned}
C_{0D_b} &= \sum_{\beta=1}^{b} w_\beta^b C_{0A_\beta} \\
C_{A_\alpha D_b} &= \sum_{\beta=1}^{b} w_\beta^b C_{A_\alpha A_\beta} \\
C_{D_a D_b} &= \sum_{\alpha=1}^{a} \sum_{\beta=1}^{b} w_\alpha^a(x) w_\beta^b(x) C_{A_\alpha A_\beta}
\end{aligned} \tag{3.34}$$

where $C_{A_\alpha A_\beta}$ is also defined following Equation (3.20), and C_{0A_β} following Equation (3.24). Again with second-order stationarity, $C_{A_\alpha A_\beta}(x)$ can be simplified as $C_{A_\alpha A_\beta}$. The decomposition of $C_{D_a D_b}$ allows us to evaluate the M-Scale covariances from the covariances of means of rings, which is location independent. The need of repeated integration to solve for $C_{D_a D_b}$ is thus eliminated, adding computational efficiency for estimation at each location.

The covariance between local plate averages, however, requires additional consideration because averages calculated for samples of different sizes correspond to different levels of uncertainty when used to represent the true mean value. Assume

the following expression for evaluating the local averages $Z^*_{D_a}(x)$ and $Z^*_{D_b}(x)$:

$$Z^*_{D_a}(x) = \sum_{\alpha=1}^{a} \nu^a_\alpha(x) Z^*_{A_\alpha}(x); \qquad Z^*_{D_b}(x) = \sum_{\beta=1}^{b} \nu^b_\beta(x) Z^*_{A_\beta}(x) \tag{3.35}$$

where $\nu^a_\alpha(x)$ and $\nu^b_\beta(x)$ are optimized weights for the ring averages to represent plate averages, the value of which will be defined later. It is found that in addition to the population mean covariances $C_{D_a D_a}$, there is another error term due to the expected variance of $Z^*_{A_\alpha}(x)$ relative to the true mean value $Z_{A_\alpha}(x)$:

$$\begin{aligned}
&C^*_{D_a D_b}(x) \equiv E\left\{[Z^*_{D_a}(x) - m][Z^*_{D_b}(x) - m]\right\} \\
&= E\left\{[Z^*_{D_a}(x) - Z_{D_a}(x) + Z_{D_a}(x) - m]\cdot\right. \\
&\qquad \left.[Z^*_{D_b}(x) - Z_{D_b}(x) + Z_{D_b}(x) - m]\right\} \\
&= C_{D_a D_b} + \sum_{\alpha=1}^{min(a,b)} \nu^a_\alpha(x)\nu^b_\alpha(x) \cdot \sigma^2(Z^*_{A_\alpha}(x))
\end{aligned} \tag{3.36}$$

Note that this covariance varies with location due to different $\nu^a_\alpha(x)$, $\nu^b_\alpha(x)$, and $\sigma^2(Z^*_{A_\alpha}(x))$ values.

The covariance between the local average and the local mean is needed in order to complete the set of covariances required for estimation. However, a component of covariance between local means can be separated from the expression for the covariance between the local average and the local mean, i.e.,

$$\begin{aligned}
&E\left\{[Z^*_{D_a}(x) - m][Z_{D_b}(x) - m]\right\} \\
=&E\left\{[Z^*_{D_a}(x) - Z_{D_a}(x) + Z_{D_a}(x) - m][Z_{D_b}(x) - m]\right\} \\
=&E\left\{[Z^*_{D_a}(x) - Z_{D_a}][Z_{D_b}(x) - m]\right\} + E\left\{[Z_{D_a}(x) - m][Z_{D_b}(x) - m]\right\} \\
=&\sum_{\alpha=1}^{a}\sum_{\beta=1}^{b} \nu^a_\alpha\nu^b_\beta E\left\{\left[Z^*_{A_\alpha}(x) - Z_{A_\alpha}(x)\right][Z_{A_\beta}(x) - m]\right\}
\end{aligned}$$

$$+ E\left\{[Z_{D_a}(x) - m][Z_{D_b}(x) - m]\right\} \tag{3.37}$$

in which

$$\begin{aligned}
&E\left\{\left[Z^*_{A_\alpha}(x) - Z_{A_\alpha}(x)\right]\left[Z_{A_\beta}(x) - m\right]\right\} \\
&= E\left\{\left[\left(Z^*_{A_\alpha}(x) - m\right) - \left(Z_{A_\alpha}(x) - m\right)\right]\left[Z_{A_\beta}(x) - m\right]\right\} \\
&= E\left\{\left[Z^*_{A_\alpha}(x) - m\right]\left[Z_{A_\beta}(x) - m\right]\right\} - E\left\{\left[Z_{A_\alpha}(x) - m\right]\left[Z_{A_\beta}(x) - m\right]\right\} \\
&= C_{A_\alpha A_\beta} - C_{A_\alpha A_\beta} = 0
\end{aligned} \tag{3.38}$$

because $Z^*_{A_\alpha}(x)$ is the average of attribute values grouped such that its ring covariance with $Z_{A_\beta}(x)$ is the same as $Z_{A_\alpha}(x)$ with $Z_{A_\beta}(x)$. Consequently,

$$\begin{aligned}
&E\left\{[Z^*_{D_a}(x) - m][Z_{D_b}(x) - m]\right\} \\
&= E\left\{[Z_{D_a}(x) - m][Z_{D_b}(x) - m]\right\}
\end{aligned} \tag{3.39}$$

and therefore the covariance between local averages and local mean is identical to the covariance between local means.

3.7 Estimation Using the M-Scale Model

The covariances defined in the previous sections can be used to define an estimate at a given estimation location and for a given scale of interest, termed the *target scale*, based on available plate averages for the same location at other scales.

3.7.1 Best Linear Unbiased Estimator

The estimator for the M-Scale model is a best linear unbiased estimator (BLUE). As a linear estimator, the local averages are given different weights to estimate the mean at the scale of interest. Supposing the mean value of a target scale t is the unknown mean of interest, we want to use the linear combination $\widehat{Z}_{D_t}(x) = \sum_a \lambda_a Z^*_{D_a}(x)$ to estimate the true mean $Z_{D_t}(x)$. The estimation variance can be evaluated as follows:

$$\sigma^2_{est}(x) = \left[Z_{D_t}(x) - \widehat{Z}_{D_t}(x)\right]^2 = \left[Z_{D_t}(x) - \sum_a \lambda_a Z^*_{D_a}(x)\right]^2 \tag{3.40}$$

The unbiasedness constraint leads to the requirement that:

$$E\left[\sum_a \lambda_a Z^*_{D_a}(x)\right] = E\left[Z_D\right] = m \tag{3.41}$$

which can be satisfied by requiring that

$$\sum_a \lambda_a = 1 \tag{3.42}$$

and the estimation variance becomes (from Equation (3.40)):

$$\sigma^2_{est}(x) = C_{D_t D_t} + 2\sum_a \lambda_a C_{D_t D_a} + \sum_a \sum_b \lambda_a \lambda_b C^*_{D_a D_b}(x) \tag{3.43}$$

3.7.2 Reproducing Variability

In this study, basic concepts by Carr (1990) and Arik (2002) are adopted to form a constraint to reproduce spatial variability, i.e., a constraint that reduces the central

tendency of a linear estimator, which also avoids over-estimation of the spatial variability in the vicinity of sample locations. The approach identifies a scale s, hereafter called the *base scale*, such that the magnitude of $Z^*_{D_s}(x)$ is most representative of the variability of the mean at the target scale $Z_{D_t}(x)$. The base scale is defined as the scale with the smallest extension variance $C^{D_s}_{D_tD_t}$ relative to the target scale, leading to the smallest expected squared difference between $Z_{D_s}(x)$ and $Z_{D_t}(x)$. Therefore, when more weight is attributed to the base scale, the estimate is more likely to reproduce the degree of variability of the real $Z_{D_t}(x)$.

The constraint to reproduce the observed variance is enforced not only by increasing the weights attributed to the base scale, but also by minimizing the expected difference between the value of the target variable $Z_{D_t}(x)$ and the weighted base-scale variable $\lambda_s Z^*_{D_s}(x)$. If this requirement were not enforced, an over-weighting for the base scale could also result in an over-estimation of the extent of threshold exceedance around sample locations, as is observed in the CK approach. The corresponding Lagrangian, by the same concept used in OK and CK, is expressed below for minimizing estimation error under the described constraint:

$$L_s = \left[Z_{D_t}(x) - \lambda_s Z^*_{D_s}(x)\right]^2 = C_{D_tD_t} - 2\lambda_s C_{D_tD_s} + \lambda_s^2 C^*_{D_sD_s}(x) \tag{3.44}$$

The derivative with respect to λ_s is zero for L_s to be minimized, therefore

$$\begin{aligned} &- 2\lambda_s C_{D_tD_s} + 2\lambda_s C^*_{D_sD_s}(x) = 0 \\ &\Rightarrow \lambda_s = \frac{C_{D_tD_s}}{C^*_{D_sD_s}(x)} \end{aligned} \tag{3.45}$$

In short, the variability reproducing constraint is simply a constraint on the weight assigned to the base-scale. The unbiasedness constraint is later incorporated in the final estimation.

3.7.3 M-Scale Estimation: BLUE with Reproduction of Variability

For σ^2_{est} to be minimized under the unbiasedness and variability reproduction constraint, the Lagrangian is defined as follows:

$$L = \sigma^2_{est}(x) + 2\mu_1 \left(\sum_a \lambda_a - 1 \right) + 2\mu_2[\lambda_s - C_{D_tD_s} / C^*_{D_sD_s}(x)] \tag{3.46}$$

The solution is obtained by solving the following set of linear equations:

$$\begin{bmatrix} \mathbf{C^*_{ab}}(x) & \mathbf{1} & \boldsymbol{\delta} \\ \mathbf{1}^T & 0 & 0 \\ \boldsymbol{\delta}^T & 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} \mathbf{C_{ta}} \\ 1 \\ C_{D_tD_s} / C^*_{D_sD_s}(x) \end{bmatrix} \tag{3.47}$$

where $\mathbf{C^*_{ab}}(x)$ is a $n_p \times n_p$ matrix (n_p represents the number of available scales) with plate-to-plate covariances $C^*_{D_aD_b}(x)$, $\mathbf{C_{ta}}$ is a $n_p \times 1$ vector of $C_{D_tD_a}$, $\mathbf{1}$ is a $n_p \times 1$ vector of ones, and $\boldsymbol{\delta}$ is a $n_p \times 1$ vector of Kronecker deltas δ_{as} for all scales a defined as

$$\begin{cases} \delta_{as} = 1 & \text{if } a = s \\ \delta_{as} = 0 & \text{if } a \neq s \end{cases}$$

The corresponding estimation variance becomes:

$$\begin{aligned} \sigma^2_{est}(x) =& C_{D_tD_t} + 2\sum_a \lambda_a C_{D_tD_a} + \sum_a \sum_b \lambda_a \lambda_b C^*_{D_aD_b}(x) \\ =& C_{D_tD_t} - \sum_a \lambda_a C_{D_tD_a} - \mu_1 - \lambda_s \mu_2 \end{aligned} \tag{3.48}$$

Note that the solution to $\boldsymbol{\lambda}$, μ_1 and μ_2 varies with location, because these solutions depend on $\mathbf{C^*_{ab}}(\mathbf{x})$ and $C^*_{D_sD_s}(x)$ in Equation (3.47), both of which vary locally.

Now the only remaining components required for obtaining the desired estimate are the weights ν_α^a in Equation (3.35) for the plate average $Z_{D_a}^*(x)$, which need to be evaluated from ring averages $Z_{A_\alpha}^*(x)$. The objective, consequently, is to minimize the extension variance of $\sum_\alpha \nu_\alpha^a Z_{A_\alpha}(x)$ relative to $Z_{D_a}(x)$. The reason for using $\sum_\alpha \nu_\alpha^a Z_{A_\alpha}(x)$ instead of $\sum_\alpha \nu_\alpha^a Z_{A_\alpha}^*(x)$ is that variability of the attribute value is reproduced without involving the uncertainty associated with using ring averages to represent ring means, so that unwanted signals such as the random error that corresponds to the sampling processes could be excluded from being reproduced (a detailed discussion regarding the impact of the unwanted signal will be given in Chapter 6).

Again defining the Lagrangian under the unbiasedness constraint $\sum_\alpha \nu_\alpha^a = 1$, we have the following:

$$L_m = \sum_\alpha \sum_\beta \nu_\alpha^a \nu_\beta^b C_{A_\alpha A_\beta}^{D_a} - 2\mu_3 \left(\sum_\alpha \nu_\alpha^a - 1 \right) \tag{3.49}$$

where μ_3 is the Lagrange multiplier.

The values for ν_α^a can then be obtained by solving the following equation:

$$\begin{bmatrix} \mathbf{C}_{\boldsymbol{\alpha\beta}}^{\mathbf{a}} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\nu} \\ \mu_3 \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix} \tag{3.50}$$

where $\boldsymbol{\nu}$ is a $n_r \times 1$ vector of ν_α^a (n_r represents number of available rings within D_a), $\mathbf{1}$ and $\mathbf{0}$ are $n_r \times 1$ vectors of ones and zeroes, and $\mathbf{C}_{\boldsymbol{\alpha\beta}}^{\mathbf{a}}$ is a $n_r \times n_r$ matrix containing all $C_{A_\alpha A_\beta}^{D_a}$ values, which can be evaluated using Equation (3.23), with ring-to-plate and plate-to-plate covariances, which, in this equation, are decomposed into sums of ring-to-ring covariances using Equation (3.34). The ring-to-ring covariances can be derived from the variogram model using Equation (3.20) over the entire domain, or (3.24) when the target scale is a point scale.

The overall algorithm for applying the M-Scale model proceeds as follows:

- Spatial covariance evaluation

    1. Evaluate ring-to-ring covariance relative to the domain D using Equation (3.20). Evaluate the point-to-ring covariance using Equation (3.24) if the target scale is a point scale.
    2. Evaluate plate-to-plate (or point-to-plate if the target scale is a point scale) covariance relative to the domain D by Equation (3.34).
    3. Evaluate point dispersion variance within each ring by Equation (3.26) to be used for evaluating the covariance between plate averages.
    4. Evaluate extension covariance $C^{D_a}_{A_\alpha A_\beta}$ by Equations (3.23) and (3.34) to be used for evaluating plate averages.

- Local estimation

    1. Obtain ν^a_α from Equation (3.50) and calculate local plate average $Z^*_{D_a}(x)$ for the estimation location x.
    2. Use Equation (3.36) to evaluate covariance $C^*_{D_a D_b}(x)$ between local plate averages.
    3. Select the base scale s by identifying the scale with the minimum extension variance $C^{D_s}_{D_t D_t}$ and evaluate the base-scale weight $\lambda_s = C_{D_t D_s}/C^*_{D_s D_s}$.
    4. Use Equation (3.47) to obtain estimation weights λ_a.
    5. Calculate local estimate $\widehat{Z}_{D_t}(x) = \sum_a \lambda_a Z^*_{D_a}(x)$.
    6. Evaluate the estimation variance by Equation (3.48).
    7. Repeat steps 1 through 6 for all estimation points x.

3.8 Summary

This chapter presents the mathematical development of the M-Scale model, and reviews basic rules for variogram-based spatial estimation. The developed model uses the covariance between mean values at different scales to describe the spatial structure, and fulfills the need for an exploratory tool for a multi-scale data integration. In addition, the model provides a single estimation map that reproduces the spatial variability without ad-hoc post processing (otherwise resulting in a function of the best estimate rather than a best estimate), or inconsistent objectives that switches among different constraining equations at different estimation locations, such as the case for CK. In the next chapter, the M-Scale model is validated by comparing estimates to those obtained from OK and CK using artificial datasets, by evaluating various quantitative endpoints.

CHAPTER 4

Validation of the M-Scale model Using Artificial Data Sets

As detailed in the previous chapters, ordinary kriging (OK) and constrained kriging (CK) are limited because (i) OK estimates tend toward the global mean due to the lack of mathematical constraints to reproduce variability, which may result in underestimation of the spatial extent and concentration levels of (in this case) sediment attributes (e.g. contamination), and (ii) CK estimates constrain data variability near sampled locations, but may overestimate the contaminated area around samples indicating contamination. However, the discussion of model tendency and limitation described in Chapter 2 was based on theoretical reasoning, and a validation of the rationale is required to re-examine the characteristics of the estimation maps that the models generate.

Hence, the primary objective of this chapter is to objectively compare estimates produced by the three estimation methods; the secondary goal is to validate the attribute precision/reproducibility of the newly developed M-Scale method relative

to OK, the comparison basis for the most commonly adapted estimation model that focus on the precision of estimated attributes, and CK, arguably among the few published estimation methods that focus on preserving global variability of attributes during spatial estimation. In addition to the reproduction of global variability, a demonstration for reproduction of spatial variability is also given, with explanations of the corresponding benefit.

Simulation approaches are described in Chapter 2 as tools for reproducing variability. When models are compared for their general applicability, simulation approaches are also used to generate realizations as equally-likely representations of reality (Meisel and Turner, 1998; Collins and Woodcock, 1999; Bian and Butler, 1999; Kukush, 2005). In this chapter, a sample application with artificially generated data is presented to demonstrate the characteristics of the M-Scale estimates and estimation uncertainty when compared to OK and CK. The reproduction of attribute variability is described globally using pooled statistics over the estimation domain, and spatially as statistics evaluated between locations. In fact, a spatial estimation generally pertains to the estimation goals for unbiasedness and minimal estimation variance (see for example, pp. 260–266 of Isaaks and Srivastava 1989; pp. 46–47 of Saito 2003). In other words, the estimation are usually designed to achieve the following objectives:

1. To reproduce estimates that are most precise in the sense of difference of estimated values against actual values over the entire study area (unbiasedness).

2. To reproduce estimates that are most precise in the sense of squared error of attribute values with respect to the actual values over the entire study area (minimal estimation variance).

which are the premises for the design of OK, that uses the approach of Lagrange multiplier to minimize the expected variance of estimation, under an unbiasedness

constraint (see Chapter 3 for the detailed review including statistical derivations).

Another characteristic to be assessed for a spatial estimation model is its capability to reproduce the value distribution over the study area, i.e., the proportion of area/volume occupied by different ranges of attribute values, representing the *global* variability of the research site. Particularly for remedial decision support, the reproduction of global variability represents the performance to correctly evaluate the volume of contaminated sediments in the research area. In observation of this importance, Cressie (1993) developed an estimation approach called *constrained kriging* (CK) to reproduce global variability in terms of the expected variance of the entire spatial domain, which is useful in generating estimation results that reflects the proportion of area/volume occupied by different ranges of attribute values.

Reproducing global variability, however, may not satisfy the objective of supporting on-site remedial decisions. The reproduction of covariances between local attributes may also be an important characteristic of estimation performance, which represents the reproduction of *spatial* variability that relates the different ranges of attribute values to their relative locations. To illustrate the difference between global and spatial variability, two maps of the same global variability that correspond to different spatial variability are shown in Figs. 4.1 and 4.2, with global and spatial variability represented by the value distributions and covariograms, as shown in Figs. 4.3 through 4.6. With the same global variability (Figs. 4.3 and 4.4), a spatially more variable attribute (Fig. 4.1) and spatially less variable attribute (Fig. 4.2) correspond to different extent of high/low values (indicated by the extent of red/blue tone colors) and different covariograms (Figs.4.5 and 4.6). For an on-site remedial decision support, the reproduction of spatial variability represents the performance to reproduce the extent of contamination, which is particularly important for block remedial techniques such as reactive/non-reactive capping of contaminated sediments.

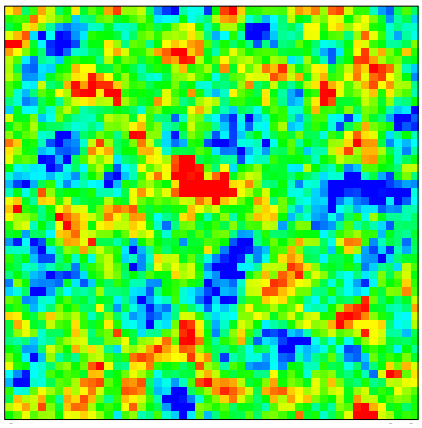

Figure 4.1: Conceptual example of an attribute that is spatially more variable (relative to Figure 4.2).

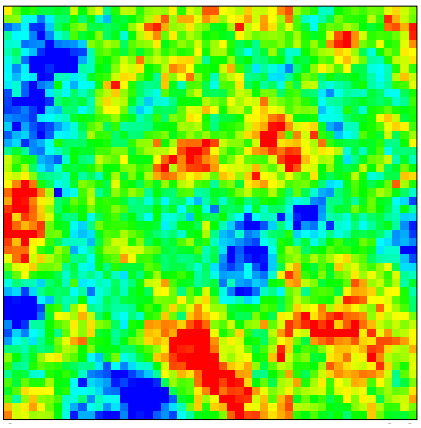

Figure 4.2: Conceptual example of an attribute that is spatially less variable (relative to Figure 4.1).

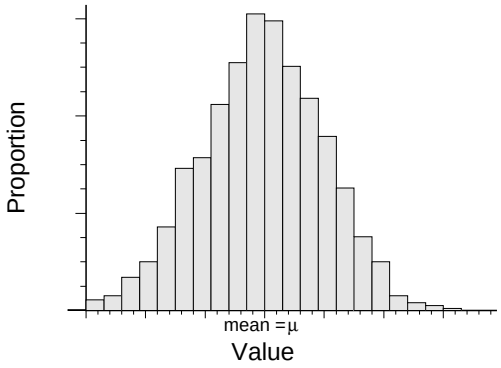

Figure 4.3: Global variability represented by the distribution of values for the attribute presented in Figure 4.1. Note that the distribution is almost identical to Figure 4.4.

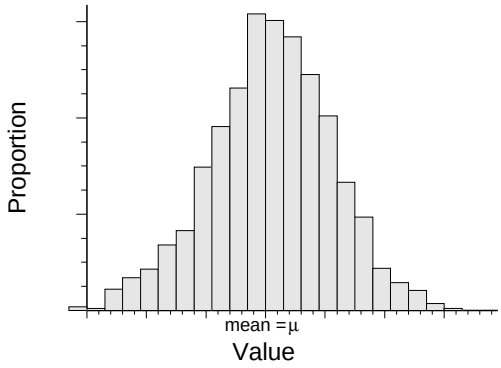

Figure 4.4: Global variability represented by the distribution of values for the attribute presented in Figure 4.2. Note that the distribution is almost identical to Figure 4.3.

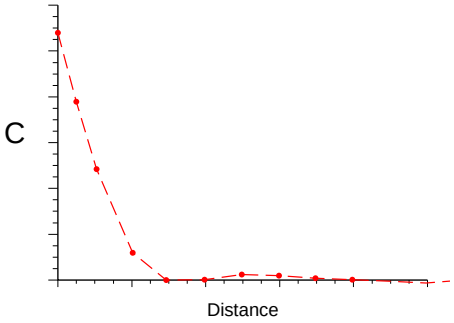

Figure 4.5: Spatial variability represented by the covariogram for for the attribute presented in Figure 4.1.

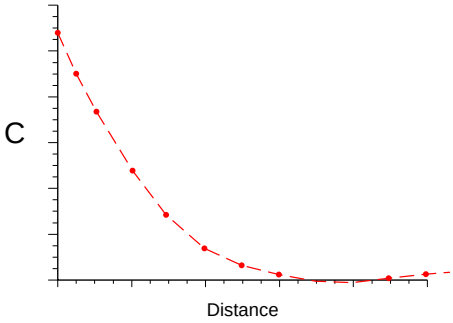

Figure 4.6: Spatial variability represented by the covariogram for for the attribute presented in Figure 4.2.

In order to reproduce spatial variability, Aldworth and Cressie (2003) proposed the use of covariance-matching constrained kriging (CM), which is shown to reproduce covariances between local attributes. The dilemma for using CM is, however, due to the restrictive constraints imposed on CM estimation weights, not all locations are guaranteed to have estimates that incorporate all sampled data and match the covariance-reproduction constraints (see Aldworth and Cressie 2003 and Chapter 3 for the detailed review). The development of CM, nonetheless, revealed the importance for reproducing spatial variability.

As an estimation model developed for the decision support of on-site remedial decisions, consequently, the M-Scale model were compared to OK and CK by the following statistical plots and quantitative endpoints, indicating three different levels/perspectives of variability reproduction:

1. Scatter plot with global standard deviation of the estimates, and Q-Q plot with Anderson-Darling test statistics (A-D test statistics) of the pooled estimates, to indicate the reproduction of global variability.

2. Covariogram (also known as "autocovariance function" in some studies like ver Hoef et al. 2001) with structural variance (also known as "partial sill" in the study of ver Hoef et al. 2001) and influence range of the covariogram for the estimates, to indicate the reproduction of spatial variability.

3. Classification (threshold exceedance/non-exceedance) map with contingency table and Cohen's κ coefficient of agreement, to compare the amount of correct classification, false positive and false negative rates. This analysis serves to examine the applicability of the proposed approach for attribute (e.g. contaminant concentration) delineation, which also provides a different perspective for presenting the precision of the likelihood estimation for threshold exceedance

as given in Aldworth and Cressie (2003).

The diagnostic parameters are evaluated for both a single realization and an ensemble of 92 realizations (100 realizations excluding 8 with singular results in CK estimation). The results serve to indentify/differentiate the model that (i) better describes global variability (ii) better delineates regional features, and (iii) better suits site-characterization needs. All statistics indicate different perspectives of variability reproduction, which feature different types of application for the estimation models to serves as decision-support tools for remedial purposes.

4.1 Selection and Description of the Artificial Data Set

The M-Scale model is compared to OK and CK using an artificial dataset based on a hypothetical variogram with a standardized influence range and sill value. The realization is generated on a 50×50 grid as shown in Fig. 4.7. The realization is generated with a mean value of 4.0 following an idealized spherical variogram model with unit variance (the sum of a nugget effect of 0.1 and a structural variance of 0.9) and an influence range of 10 unit-distance. Although the selection of parameters are standardized, the sill and range parameters can be rescaled to represent the magnitude of variability and extent of spatial association for any attribute of interest, for example the contamination level, without violating the rationale of using a variogram to explain the spatial structure (a similar selection of variogram parameters can be found in p. 254 of Deutsch and Journel 1998). The selection of nugget effect is relatively arbitrary, but it represents the case where the local variability is not negligible, an assumption that is applicable for many environmental studies (see for example,

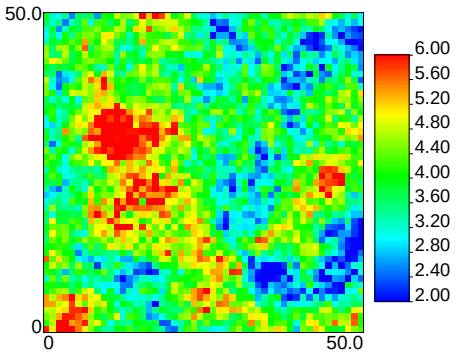

Figure 4.7: The artificial realization generated from the idealized variogram. Red-tone and Blue-tone colors represent high and low attribute values respectively.

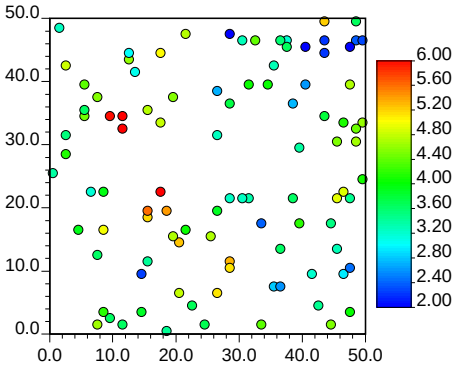

Figure 4.8: Location and value of the selected sample. The same color scales are used corresponding to the artificial realization.

Wu et al. 2002; Schnabel and Tietje 2003; Critto et al. 2005). The magnitude of the nugget effect is decided on a qualitative basis (so that the nugget effect is not negligible) rather than a quantitative selection. A subset of the realization is selected as the experimental sample at 100 locations with coordinates randomly chosen over the 50×50 grid, as shown in Fig. 4.8.

To demonstrate that the samples selected from this dataset are representative of the full realization, a quantile-quantile plot (Q-Q plot) (Deutsch and Journel, 1998) for experimental sample values versus population values is plotted to ensure that the global attribute distribution (value distribution over the entire domain) is reproduced, as shown in Fig. 4.9. Each point on the Q-Q plot represents the quantile pair of the global probability distribution of the two datasets, in this case the population and the sample set. The Q-Q plot provides a visual demonstration of the similarity of two distributions, and the close match to the 45° bisector indicates a close match between the two distributions. In addition, the variograms of both experimental sample and the population values are also compared in Fig. 4.10 to show that the spatial structure of the population is reasonably reproduced by the experimental

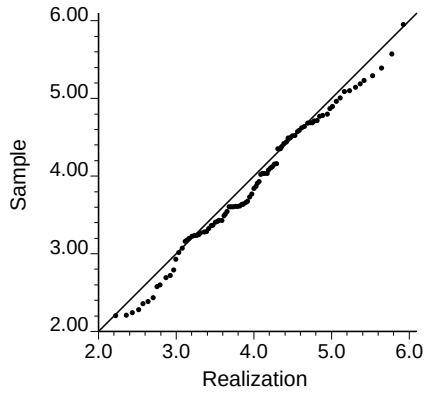

Figure 4.9: Comparison of value distribution by Q-Q plot for the artificial realization and the sample. It is shown in the figure that the sample well represents the value distribution of the artificial realization.

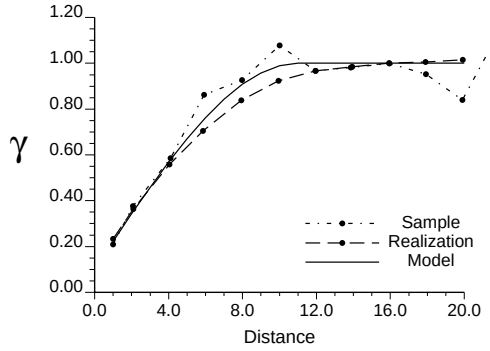

Figure 4.10: Variogram comparison for the sample, the artificial realization, and the variogram model. The comparison indicates that the sample represents the artificial realization reasonably well, both of which are well represented by the modeled variogram.

sample, so that the prescribed theoretical variogram matches both the population and the sample set. The good match of the sill indicates a good reproduction of dissimilarity between attribute values for long separation distances, the good match of the trend and nugget indicate a good reproduction of dissimilarity between attribute values for short separation distances. In this case, the sample generally reproduces dissimilarity for all separation distances, and the variogram model reasonably reflects the dissimilarity of both the realization and the sample.

4.2 Performance of M-Scale Estimation, Ordinary Kriging and Constrained Kriging

M-Scale estimation, ordinary kriging and constrained kriging were performed on the 50×50 grids of the target map. A spherical variogram model is also fitted visually to the experimental variogram for the estimation of all three models, characterized by

a sill of 1.0, a nugget effect of 0.1, and an influence range of 11 unit-distance. For the M-Scale model, spatial supports of diameter 1,2,4,8,16,32, and 100 were pre-selected for the moving subdomains, or plates, as conceptualized in Fig. 3.8. Statistics needed for the M-Scale model, including the different variances/covariances, are described in Chapter 3, and evaluated by numerical integration according to the modeled variogram. The selection of subdomain sizes in this study was arbitrary but captures the separation distance classes for the experimental variogram. Other specific shapes or sizes of the plates can also be used. It is suggested, however, that a smallest scale is included to represent an area only slightly larger than the sampling volume/area so that the nugget effect can be differentiated from variability at other spatial scales. It is also favorable that more scales within the influence range (as indicated by the modeled variogram) be selected relative to the scales outside of the influence range, as long as the ring areas cover an adequate number of sample points over the estimation domain. The implementation process for CK is identical to that of OK, except for the additional constraint for the reproduction of the expected variance.

Using the sampled data and modeled variogram, the estimation maps for the three estimation models are presented in Figs. 4.11 through 4.13. Color scales in the maps indicates the magnitude of the estimated values. Visually, high- and low- value regions (red- and blue- tone areas) are largest in the CK estimation map, and smallest in the OK estimation map. These estimation maps reflect the convex- and concave-smoothing features, where the estimates tends toward the global mean for OK and tends to the sample values for CK estimates (these are features seen in Kyriakidis 2004; Goovaerts 1997b; Yamamoto 2005, as reviewed in Chapter 3).

Standard deviation maps of the model estimates are shown in Figs. 4.14, 4.15 and 4.16, which indicate that, at most of the estimation locations, CK generally has the largest estimation standard deviation among the three estimation approaches. The

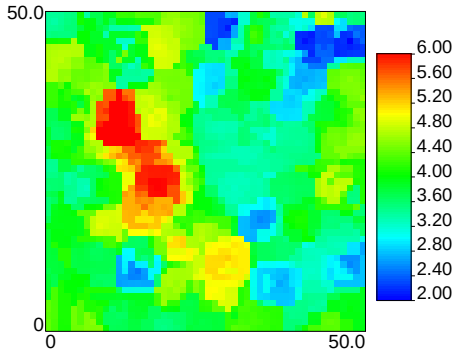

Figure 4.11: Estimation map by the M-Scale model for the example set of artificial data.

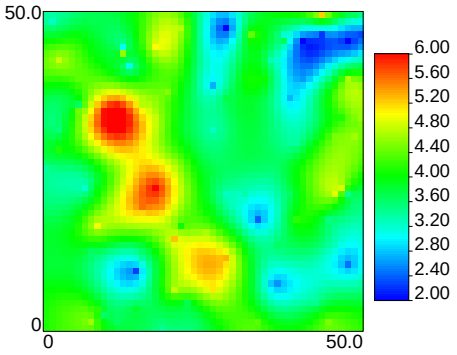

Figure 4.12: Estimation by ordinary kriging for the example set of artificial data.

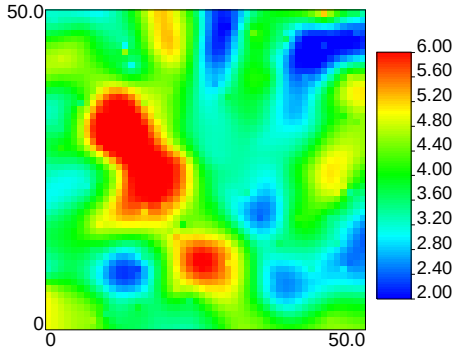

Figure 4.13: Estimation by constrained kriging for the example set of artificial data.

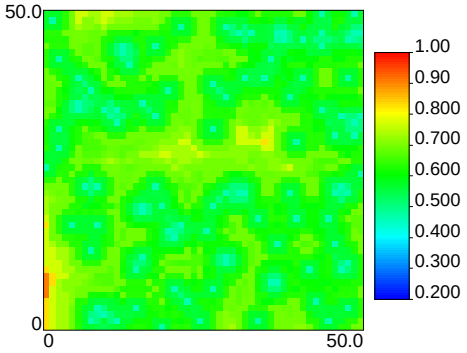

Figure 4.14: Estimation standard deviation by the M-Scale model for the example set of artificial data.

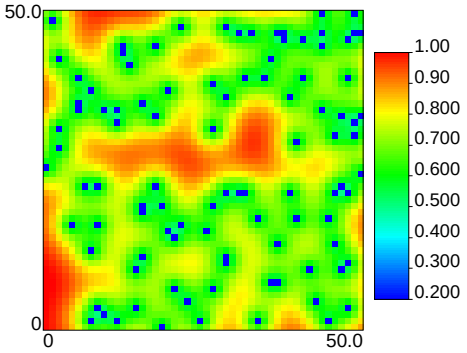

Figure 4.15: Estimation standard deviation by ordinary kriging for the example set of artificial data.

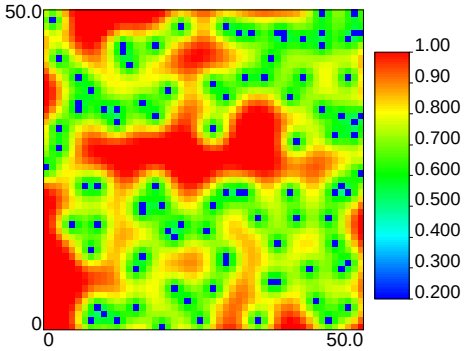

Figure 4.16: Estimation standard deviation by constrained kriging for the example set of artificial data.

large estimation standard deviation results from the constraint imposed to reproduce global variance, that limits the optimization process from reducing the uncertainty by setting the sample-to-sample part of the estimation variance to the global variance (i.e., by setting $\sum_i \sum_j \lambda_i \lambda_j C_{ij}$ to σ^2, see reviews of CK in Chapter 3). Although the M-Scale model imposed additional constraint in the estimation process, the M-Scale estimation standard deviation is lower than the other two models. This can be explained by fact that the M-Scale model considers the *expected* covariance between sample points within each ring area via the dispersion variance of the ring area assuming the samples are randomly located within the ring area disregarding the actual locations. In the cases where sample locations are clustered within ring areas, the dispersion variance may be underestimated since clusters among sample locations usually pertains to the redundancy of samples, resulting in less information and more uncertainty for the sample to represent the estimate (p.188 of Isaaks and Srivastava 1989). Despite the difference in the assumption of sample locations, there appears to be no statistical impact on the confidence interval (characterized by two estimation standard deviations from the estimate) across the estimation map: the M-Scale model generates confidence intervals that cover 91% of the true values of the realization, similar to those based on confidence intervals of OK (92%) and CK (93%) estimation. The insignificant impact on the confidence interval results from the fact that the sample set is in deed selected at random locations from the realization using a random number generator, therefore resemble the ideal case for ring areas around all estimation locations. Additional research will be needed, however, in order to account for the impact on the estimation variance by clusters in data locations.

Four statistical graphs are used to assess the impact of model choice on the data estimates: (i) scatter plots, (ii) Q-Q plots, (iii) covariograms and (iv) classification maps. These measures are quantitatively assessed using the following summary statistics: global standard deviation of estimates for the scatter plot, Anderson-Darling

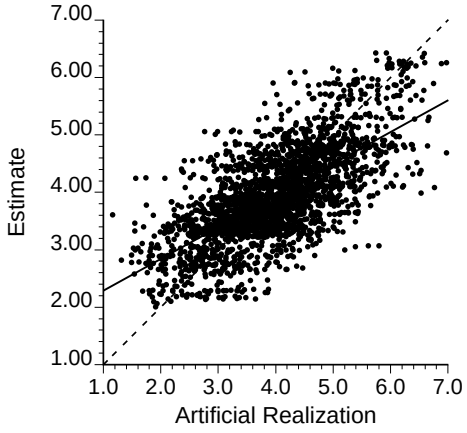

Figure 4.17: Scatter plot of M-Scale estimates vs. the artificial realization.

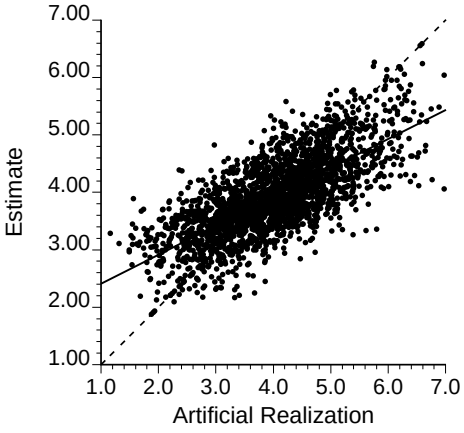

Figure 4.18: Scatter plot of OK estimate vs. the artificial realization

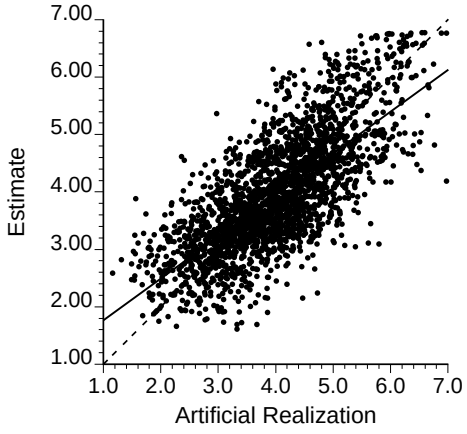

Figure 4.19: Scatter plot of CK estimate vs. the artificial realization

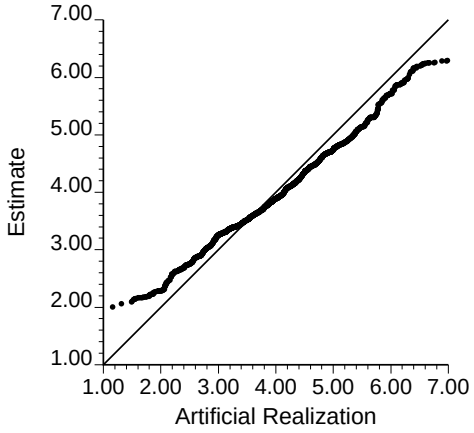

Figure 4.20: Q-Q plot: comparing value distribution of the M-Scale estimates vs. the artificial realization.

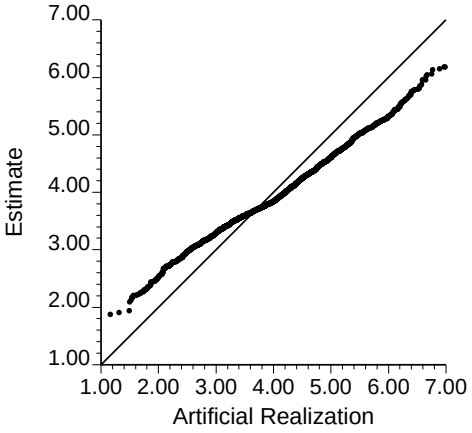

Figure 4.21: Q-Q plot: comparing value distribution of OK estimates vs. the artificial realization.

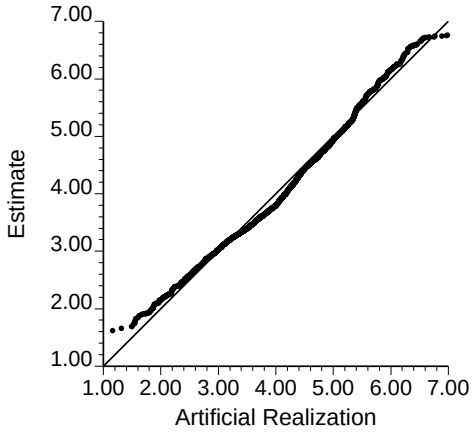

Figure 4.22: Q-Q plot: comparing value distribution of CK estimates vs. the artificial realization.

(A-D) test statistics for the Q-Q plot, influence range and structural variance for the covariogram, and the contingency table with Cohen's κ coefficient for the classification maps.

4.2.1 Scatter Plot and Histogram Analysis

Among the four summary graphs, scatter plots and Q-Q plots capture global variability. Scatter plots for the estimates against the true values are plotted in Figs. 4.17, 4.18 and 4.19; Q-Q plots are shown in Figs. 4.20, 4.21 and 4.22 to compare the overall distribution of pooled values of estimates to the actual realization. In both sets of graphs, a trend to overestimate low values and underestimate high values is observed in OK estimates, while such a trend is not obvious for the M-Scale and CK estimates. Additionally, in the Q-Q plots, more off-diagonal quantiles are observed in OK, fewer off-diagonal quantiles for the M-Scale model, and in the case of CK the histogram is generally reproduced following closely to the diagonal line. This good match of value distribution observed in the Q-Q plot of CK may indicate a good match of quantiles that involve erratic jump in the attributes (local variability) that is not of spatial correspondence, such as the sample values that involve micro-scale variability and unwanted artificially induced error, as illustrated in Fig. 4.23. In fact, as demonstrated by Cressie (1993) in the mathematical development, CK considers the reproduction of global variability without distinguishing the structural variability from local variability. Additionally, no local variability at the estimation location can be inferred using samples unless a sample point is located exactly at the estimation location. The estimates generated using CK, therefore, will appear to be more variable than the structural variability in order to compensate the local variability, as illustrated in Fig. 4.24. Although indicated in Aldworth and Cressie (1999) that the artificially induced error should be excluded from reproduction of global variability,

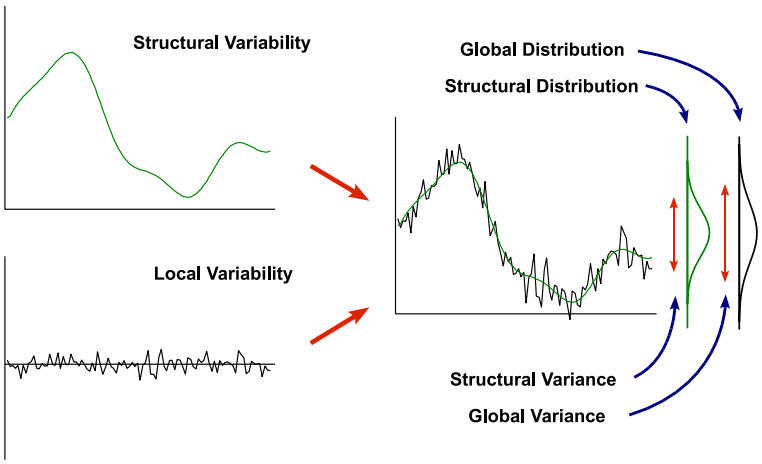

Figure 4.23: Conceptual sketch for an attribute involving local uncertainty. Green curves and distribution shape indicate the structural variability and distribution of the structural variability; black signals indicates local variability; black distribution shape indicates the distribution of global variability.

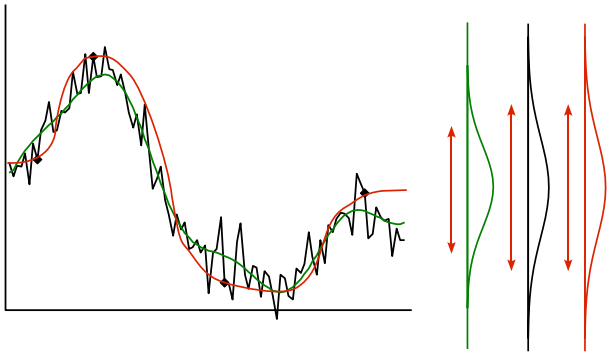

Figure 4.24: Using CK to reproduce global variability. Solid diamond shapes indicate samples; green curve indicates structural variability; black curve indicates the mixture of structural and local variability; red curve indicates the variability reproduced by CK estimation. Value distributions are illustrated on the right according to the colors of the curves.

the approach for distinguishing artificially induced error from micro-scale variability in the completeness of local variability is seldom studied. The CK estimation, consequently, corresponds to either the negligence of the entire local variability in the global variance (including a component of micro-scale variability that is part of the true signal), or the reproducation of the structural variance using the entire local variability as described above (see Chapter 6 for detailed review and discussion for the separation of artificially induced error). In a case where variograms modeled involve

a non-negligible nugget effect, consequently, scatter plots and Q-Q plots may not be sufficient for the examination of variability reproduction since these graphs does not distinguish structural variability from local variability, mixing the reproduction of the variability from what could be inferred from samples and what could not. To meet the end for distinguishing structural variability reproduced by the estimates, scatter plots and Q-Q plots should be demonstrated in parallel with diagnostic graphs such as covariograms, so that the impact of nugget effect on CK estimates could be examined.

Basic regression statistics for the scatter plots are listed in Table 4.1, including the mean, standard deviation and regression slope of estimation values against true values. A lower standard deviation of the estimates relative to that of the population indicates a tendency of estimates to be closer to the global mean, i.e., the tendency to reduce variability of values. This tendency is observed to be more significant in OK than in the M-Scale model and CK estimates. This result indicates that

Table 4.1: Basic statistics for scatter plots comparing results of the M-Scale model, OK and CK to the artificial realization.

Statistics	Realization	M-Scale	OK	CK
Mean	3.995	3.943	3.934	3.941
Std. Dev.	0.992	0.800	0.682	0.979
Slope	-	0.55	0.50	0.72

OK tends to introduce more concave smoothing in the estimates than the M-Scale model and CK in terms of the global variability of the estimate. The standard deviation for CK reproduces the global population standard deviation well, which also includes the reproduction of local variability according to the discussions above for the reproduction of global variability. Regardless of whether the reproduction of global variance is sufficient to indicate good performance of variability reproduction, it is demonstrated that the M-Scale model reproduces global variability better than OK that focus on the precision of the estimates. The global unbiasedness of the estimates

was evaluated by comparing the population mean to the mean of estimated values. Although M-Scale and CK show a lower bias (bias=0.052 and 0.054) relative to OK (bias=0.061), the difference is small relative to the population standard deviation.

The slope of the regression for the scatter plot is a measure of the similarity globally between true attribute values and their corresponding estimates. In this global sense, CK generates estimates closer to the realization values among the three estimation approaches, and OK has the lowest reproduction of the three (Table 4.1). Note that, however, these statistical descriptors reflect perspectives of the global variability reproduction, and spatial statistical descriptors should be used instead if local variability is not part of the characteristics of interest for the attribute under study.

The Anderson-Darling test statistic A^2 (Anderson and Darling, 1954; Ang and Tang, 2007) is used to quantify how the estimates globally reproduce the distribution of the population. This statistic is defined as the weighted average of squared discrepancies between the cumulative distribution function of the estimate and the population, and is interpreted as the "goodness of fit" of two histograms, where all quantiles are treated with equal weights. The results also indicate a better fit for the M-Scale model (A^2=38.8) relative to OK (A^2=91.92). CK (A^2=17.28) again has the best fit of histogram among the three approaches because of the imposed variance constraint.

4.2.2 Covariogram Analysis

Covariograms (Curriero et al., 2002) and threshold indicator maps (Barabás et al., 2001) can be used to compare the spatial variability of the estimated fields relative to that of the realization. Covariograms are used in this example to illustrate the

spatial variability in terms of the covariance between locations. A covariogram, expressed as $C(h) = \sigma^2 - \gamma(h)$ (see Chapter 3 for detailed derivation), is used to indicate the spatial dependence/dissimilarity between local variables (Jackson and Caldwell, 1993), and thus act as a complimentary component with the variogram (the spatial independence/dissimilarity) of the attribute variance σ^2. Although variograms are popularly used to demonstrate spatial relatedness in most geostatistical studies, all separation distances h of the variogram corresponds to a component of local variability/error (the nugget effect) in the semivariance $\gamma(h)$, as shown in Fig. 4.25 (left). In other words, the semivariance evaluated/modeled at each separation distance is

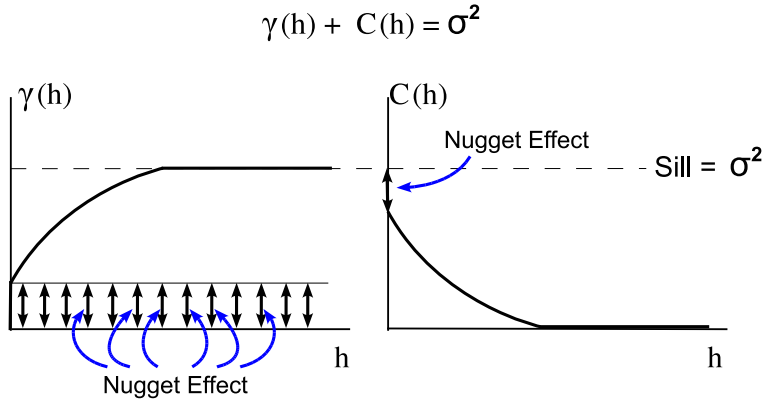

Figure 4.25: The correspondence of variogram vs. covariogram. It is shown in the figure that the nugget effect is involved at all separation distances in the variogram (except at zero separation distance, which is not easy to observe), while appears in the covariogram at only the zero separation distance.

a mixture of a local component expressed by the nugget effect, and a spatial component that is relevant between sample locations. When evaluating the covariogram from the variogram, the local component cancels out the corresponding component involved in the attribute variance σ^2, leaving the spatially-relevant component at each separation distance, except the remaining nugget variance at zero separation distance (see Fig. 4.25, and p.532 of Starks 1986). Note that the semivariance at zero separation distance is essentially zero since $\gamma(0) = E[Z(x) - Z(x)]^2 = 0$, and the curve of variogram jump immediately to the nugget variance at a small separation distance.

For a spatial estimation model that uses measurements at sample locations to inform the value at unsampled locations, the spatial dependence between estimates is usually more interesting (Istok and Rautman, 1996). Hence the covariogram of the estimates are compared to that of the actual covariogram, in order to indicate how the dependence between locations reproduces the (assumed) reality. Structural variance (Robertson et al., 1993; Crist, 1998; Liu et al., 2007), evaluated by subtracting the nugget effect from the sill value, is used as a diagnostic parameter for the covariogram, accompanied by the influence range that indicate the spatial range of relevance between locations. Visually in the covariogram, the structural variance can also be observed by extrapolating the trend of covariance values for the value at zero separation distance.

As shown in Fig. 4.26 and Table 4.2, all the covariograms have identical influence ranges at around 18, indicating that the same data that fall out of the modeled influence range from estimation locations remain included in the three estimation models. This result also indicates that all samples are regarded as informative for the estimation in all three models, since data out of the influence range also provides information about the unknown global mean. The covariogram for the M-Scale estimates has a structural variance of around 0.73, as compared to that for OK and CK estimates at 0.53, and 1.13, respectively (Table 4.2). Although none of the models exactly reproduce the covariogram of the realization, the covariogram of the M-Scale estimate better reproduces the covariogram relative to OK and CK (structural variance = 0.90 (M-Scale), 0.53 (OK) and 1.13 (CK)). In the case of CK, the variability is forced to follow the targeted global variance (involving the nugget effect that indicates local variability) and consequently the lack of reproduction for the structural variance, as discussed in the previous section. Unfortunately, no single standard parameter summaries the goodness of fit between two covariograms, and the comparison are usually based on visual comparison and scientific knowledge of the attribute (van Meirvenne

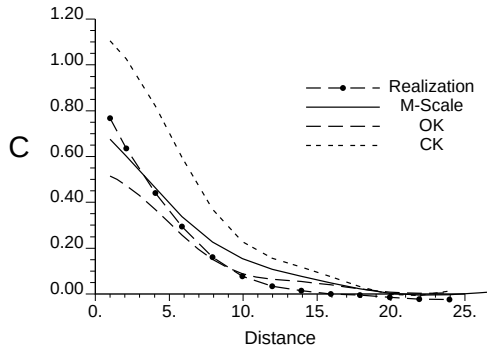

Figure 4.26: Comparison of M-Scale, OK and CK covariograms to the covariogram of the artificial realization to represent their performance in terms of reproduction of spatial correspondence between different locations.

Table 4.2: Structural variance and influence ranges indicating covariogram reproduction.

Statistics	Realization	M-Scale	OK	CK
Structural Variance	0.90	0.73	0.53	1.13
Influence Range	11	18	18	18

and Goovaerts, 2002).

4.2.3 Classification Map Analysis

A classification map represents the discretized variability of the spatial attribute (p.441 of Goovaerts (1997b)). Classification maps for estimates above and below a threshold value, consequently, can be used to demonstrate the spatial variability corresponding to a certain designed threshold (Barabás et al., 2001; Saito, 2003). For example, concave smoothing (see Chapter 3) of estimates corresponds to underestimation of the occurrence of exceedance for high threshold values globally, because these estimates tend toward the global mean. On the other hand, convex smoothing (also see Chapter 3) correspond to overestimation of the occurrence of exceedance regionally in the vicinity of hot spots, because these estimates tend toward the sample value in their close neighbors (see Chapter 3 for statistical explanation of both

Table 4.3: Contingency table for estimation models based on the classification maps.

		M-Scale		OK		CK		Total
		+	−	+	−	+	−	
Target	+	3.9	6.4	2.8	7.5	5.2	5.1	10.3
	−	2.1	87.6	0.9	88.8	4.4	85.3	89.7
Total		6.0	94.0	3.7	96.3	9.6	90.4	100.0

concave and convex smoothing). Specifically for site characterization that supports remedial decision-making, a classification map is important to define areas where attribute values exceed a certain regulatory threshold such that a remedial action is required. Underestimating the occurrence of contamination increases the risk with unrevealed contaminated area, while overestimating the occurrence of contamination may conversely lead to additional remedial cost (Saito and Goovaerts, 2003). Particularly for research areas where the corresponding cost for remediation is linear to the contamination level, a classification map serves as a visual tool for stakeholders to negotiate the area where remedial action is needed.

For illustration purposes, classification maps were generated as shown in Fig. 4.27 for a threshold of 5.25, which is around the 0.9 quantile of the pooled sample measurements, a value that is relatively high comparing to the sampled values. Grey color in the map indicates values above 5.25, and white indicates values below 5.25. The same classification maps are also created for the M-Scale model, OK and CK estimates, as shown in Figs. 4.28, 4.29 and 4.30. It is visually observed that relative to the exceedance of the realization, the exceedance for OK estimates consists of a smaller area (which may underestimate the occurrence of exceedance) around the hot spots, and the exceedance for CK estimates consists of a larger area (which may overestimate the occurrence of exceedance).

The contingency table (also known as the error matrix, see Congalton 1991 and Couto 2003) for the classification maps is listed in Table 4.3 to include the percent-

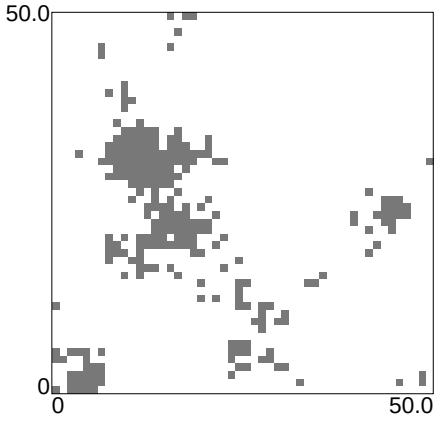

Figure 4.27: Target classification map for threshold 5.25 evaluated using the example set of artificial data.

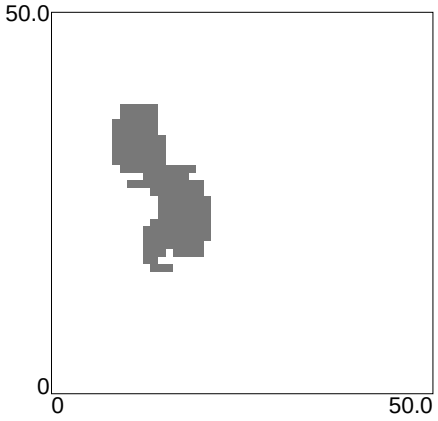

Figure 4.28: Classification map for threshold 5.25 evaluated using M-Scale estimates

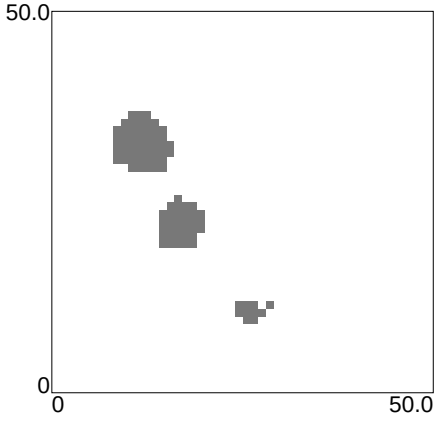

Figure 4.29: Classification map for threshold 5.25 evaluated using OK estimates

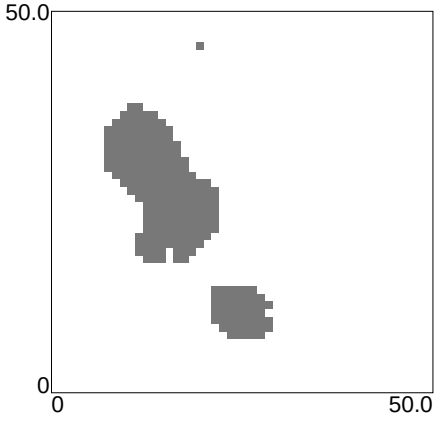

Figure 4.30: Classification map for threshold 5.25 evaluated using CK estimates

age of each classification categories, where "+" indicates values above 5.25 (classified "true"), and "–" otherwise (classified "false"). The result shows that OK exhibits the highest false-negative exceedance estimates, thus underestimating potentially contaminated areas. CK may be an over-conservative assessment approach due to the highest false-positive rate. The false-negative rate for CK, however, is the lowest among the three estimation models, which may due to the fact that the global variance CK is designed to reproduce includes not only the structural variance, but also local variability (the nugget effect). The attribute that involves local variability may correspond to occasional exceedance, and consequently results in occasional agreement for the CK classification, as shown in Fig. 4.31. Among the three approaches,

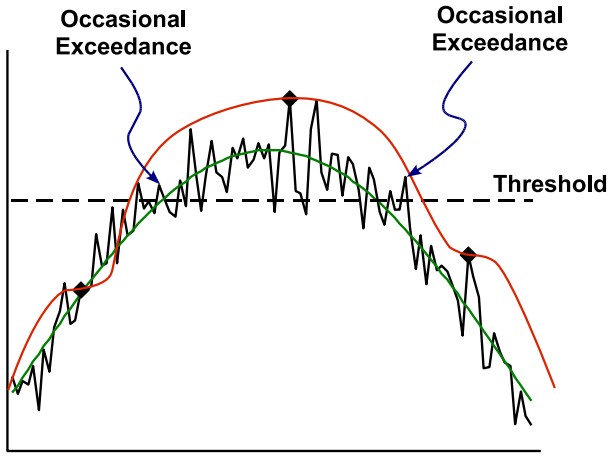

Figure 4.31: Conceptual sketch for occasional exceedance. Green curve indicates structural variability; black curve indicate a mixture of structural and local variability; red curve indicates CK estimates.

the M-Scale exhibits a lower false-negative rate compared to OK, and a lower false-positive rate compared to CK.

Cohen's κ coefficient (Cohen, 1960) is a parameter of the contingency table that examines the agreement of the classification map of estimates against the classification map of the realization. The parameter considers the improvement of agreement relative to a map with random selection of action/no action locations (areas) over the estimation domain. As shown in Fig. 4.32, even by random assignment of

exceedance/non-exceedance that is not relevant to any spatial context, the result of classification may reach a certain level of agreement with the actual realization. The rate of opportune agreement is consequently used as a baseline, extracted from the rate of agreement by a spatial estimation model (as shown in Fig. 4.33) and standardized by the maximum possible improvement of agreement relative to the random assignment. The standardized result is in general known as *Cohen's κ coefficient of agreement*, or κ for convenience, which is expressed mathematically as

$$\kappa = \frac{p_0 - p_c}{1 - p_c} \tag{4.1}$$

where p_0 is the proportion of units correctly classified by the estimation model, and p_c is the proportion of units for which agreement is expected by chance. With intuitive expression, Cohen (1960) described his κ coefficient as "the proportion of agreement after chance agreement is removed from consideration." Originally developed for characterizing level of agreement in general for psychological research, this coefficient is also used in the area of remote-sensing for the classification agreement of NDVI (Normalized Difference Vegetation Index) (Pax-Lenney and Woodcock, 1997), in ecological research for the classification agreement of tree species cover (Naesset, 1996), and in environmental studies for the classification of contamination level (Saito, 2003).

Cohen's κ is a standardized coefficient that takes value from -1 to 1, which represents the goodness of classification for an estimation model ranging from complete disagreement (-1) to perfect agreement (1). The rate of agreement is interpreted by Landis and Koch with six categories of the strength of agreement(Landis and Koch, 1977; Naesset, 1996):

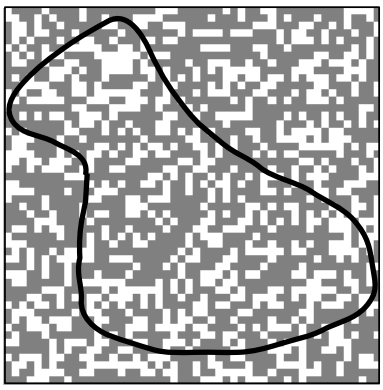

Figure 4.32: Random assignment of exceedance/non-exceedance (with κ=0). Solid line represents the perimeter of actual exceedance.

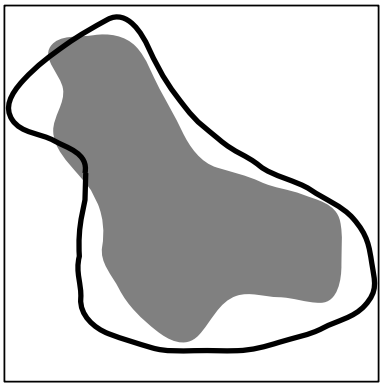

Figure 4.33: Classification based on estimation (with κ=0.68). Solid line represents the perimeter of exceedance.

Kappa Value Range	Strength of Agreement
< 0.00	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost Perfect

which provides an intuitive expression of how the estimation reproduces the actual realization, or "useful benchmarks for the discussion" as describe in Landis and Koch (1977).

The resulting κ coefficients for the three estimation models are listed in Table 4.4, in which OK has the largest κ coefficient (best agreement) among the three models.

Table 4.4: Cohen's κ coefficient for the three estimation models evaluated by comparing the classification results of estimates to the classification results of the artificial realization.

	M-Scale	OK	CK
κ	0.53	0.54	0.48

Table 4.4 also indicates that CK estimates results in less agreement in the classification of exceedance, possibly due to the false reproduction of local variability using structural variability. All three estimation models are categorized as "moderate" agreement by the definition of Landis and Koch (1977), while the difference between the M-Scale model and OK is much closer than that between CK and OK. The result, at least for this example, indicates a possibility that the M-Scale model and OK performs at the same level, and CK performs worst when overall precision is expressed by the results of classification. A simulation using multiple realizations, will be used to validate the performance characteristics indicated in the example realization.

4.3 Performance Comparison Using Multiple Realizations

In order to examine the consistency of model performance, the statistics described in the previous paragraphs were calculated for an additional 99 simulated realizations (a total of 100 realizations) with the same random field generator, each corresponding to samples taken at the same sampling configuration used for the example in the previous section. Eight of the simulations were excluded from the sensitivity analysis due to the singularity of estimates for CK, where the covariances between sample and locations become too small for the estimate to reproduce the global variance from sample values (Aldworth and Cressie, 2003). The resulting variability parameters are individually averaged over the 92 realization used, and the average values are listed in Table 4.5.

Results of the consistency check indicate that the standard deviation of the estimates is best reproduced, and the population probability distribution is best fitted for CK. The false negative rate is worst for OK, while the false positive rate is worst for

Table 4.5: Average values for parameters of variability reproduction.

	Avg. Estimate	Std. Dev.	A-D Stats	False +	False –	κ
Realization	3.982	1.000	–	–	–	1.00
M-Scale	3.983	0.753	66.66	2.30	7.74	0.436
OK	3.983	0.671	107.45	1.21	8.22	0.473
CK	3.984	0.989	18.69	5.70	5.66	0.361

CK. The false negative rate looks best for CK, due to the over-estimation of hot-spot regions, which coincides with scattered exceedance not corresponding to the sample set (also observed in the example classification maps). Cohen's κ indicated that OK classification maps are most precise, followed by the M-Scale model and CK has the worst classification precision among the three estimation models. These results confirm that the M-Scale model performs similarly well in the precision of estimates relative to OK, the conventional model that focus on the precision of the estimates, in terms of the classification agreement using Cohen's κ coefficient. Standard deviation of estimates and A-D test statistics indicate the best reproduction by CK, followed by the M-Scale model and worst by OK. This result confirms that the M-Scale model performs similarly well in the reproduction of global variability relative to CK, the novice model that focuses on the reproduction of global variability, in terms of the global variance and A-D test statistics. It is not obvious which model performs best in the category of absolute bias, nevertheless, which is not hard to understand because the constraint of the global unbiasedness is implemented in all three models.

The covariograms of the estimates are compared to sample covariograms as an indication of the reproduction of *sample* covariograms. The reproduction of sample covariogram is equivalent to the comparison of covariograms between the estimates and the population when the sample covariogram reproduces the covariogram of the population. The reproduction of population covariogram by the sample covariogram, however, is not the focus of this dissertation and require future research for its impact on the resulting estimation map.

Nevertheless, the reproduction of covariogram cannot be indicated by a single diagnostic parameter. Comparison of the 92 covariograms of the estimates to sample covariograms, consequently, needs to be conducted visually. To meet this end, the covariances at each separation distance for the 92 samples are pooled, and the ranges (expressed by two standard deviations from the mean) of the covariances are plotted on the covariogram. The ranges of sample covariogram (ranges of covariances plotted for all separation distances) are compared to the mean covariogram (mean covariances plotted for all separation distances) of the estimation maps to indicate the general tendancy of covariogram reproduction for the estimation models, as shown in Figs. 4.34 through 4.36. It is observed that OK mean covariogram lie closer to the lower limit, CK mean covariogram lie closer to the upper limit, and the M-Scale mean covariogram generally lie between the limits of the sample covariogram. The structural variances for the sample covariogram are found to be 1.10 for the upper limit, and 0.56 for the lower limit. The structural variances for the mean covariogram of the estimation maps are 0.58 for the M-Scale, 0.46 for OK, and 0.99 for CK. The ranges of the mean covariances are found to be 17.0 for all estimation models, because the same set of sample is used by the estimation models to generate estimation maps of the population.

Based on the structural variances and visual comparison of the mean covariances of the estimation maps to the ranges of sample covariances at all separation distances, it is observed that OK generally underestimates spatial variability, and CK overestimates spatial variability. The M-Scale model, with a range that lies closer to the modeled covariograms, indicates the best reproduction of spatial variability. This result is consistent with the example given using one realization. This conclusion, however, is made by an assumption that the spatial variability of the sample is representative of the spatial variability of the population, which needs additional support by studies that focus on the reproducibility of population variability by samples.

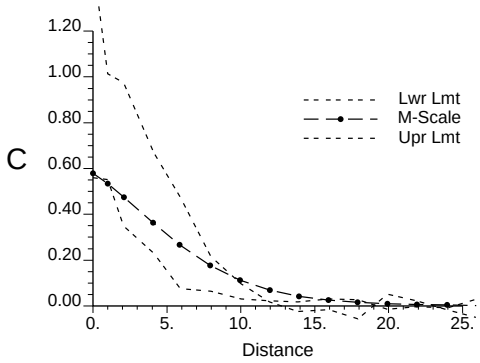

Figure 4.34: Comparison of M-Scale mean covariogram of the 92 estimation maps to the upper and lower limits of the 92 sample covariograms to represent their performance in terms of reproduction of spatial variability.

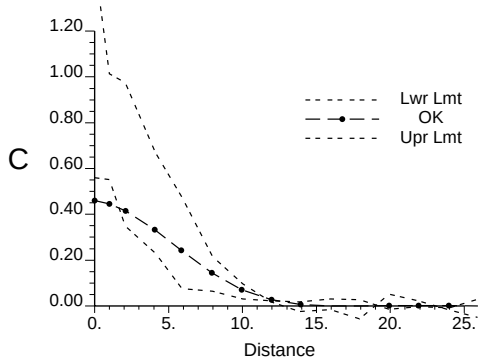

Figure 4.35: Comparison of OK mean covariogram of the 92 estimation maps to the upper and lower limits of the 92 sample covariograms to represent their performance in terms of reproduction of spatial variability.

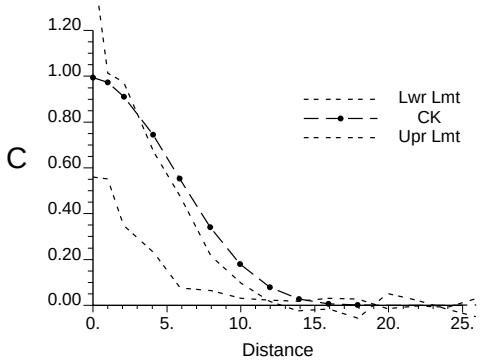

Figure 4.36: Comparison of CK mean covariogram of the 92 estimation maps to the upper and lower limits of the 92 sample covariograms to represent their performance in terms of reproduction of spatial variability.

4.4 Summary

It is concluded in this chapter that the M-Scale model across all statistical endpoints performs as well as OK in classification precision without over-smoothing the estimation map, and as well as CK in reproducing variability without falsely reproducing local variability by increasing structural variability. In addition, the spatial variability appears to be best reproduced by the M-Scale model, as visually observed in the covariogram reproduction and structural variance quantified for the covariogram. Hence, this proposed spatial estimation approach that explicitly incorporates multiple scales of samples appears to be a robust method, that potentially lends itself to applications where relationships between variable spatial supports need exploring, and where the remedial decisions depend on the spatial variability of the estimation map.

Regardless of the smoothing effect that is consistently observed in conventional kriging estimation, OK does best in terms of classification precision, as indicated by Cohen's κ coefficient. Although standardized as a measure of agreement, Cohen's κ treats false positives (FP) in the same way as false negatives (FN), whereas in a case of actual contamination, a false-negative may be regarded as more dangerous than false-positive classification. The four metrics in a contingency table, consequently, will assist the choice of estimation model, if the classification objective depends on a particular classification performance such as the false-negative rate. It is observed that the absolute bias of the pooled estimates corresponds to classification precision in terms of the sum of FP and FN rates, which is equivalent to the comparison using Cohen's κ coefficient. In fact, a smaller bias of the OK estimates indicates that, on average, the estimates are closer to the actual value, consequently a better chance of being classified correctly. Particularly in a multi-Gaussian case, this indicates that most of the middle-range values, which take up the highest percentage of the counts

of values, are closer to the actual values. Regardless of the overall small misclassification rates for OK, the concave smoothing of OK corresponds to a small FP rate but is compensated by a large FN rate, which is not favored as an estimation model for site remediation (p. 421 of Isaaks and Srivastava 1989). On the other hand, although CK results indicated the lowest false-negative rate, the model is at a lower level for classification precision as indicated by Cohen's κ coefficient. This result for CK also confirms the conclusion by Carr (2002), who described that an estimation approach that "yields a sharper image" may "result in diminished estimation quality" for data which "indicate a substantial nugget value." He further described this substantial nugget value as an implication that "smaller scale spatial variability is not well represented in the sample data," and an estimation approach that attributes higher weights for the data during estimation "as if these smaller scales are present" in order to produce higher spatial variability, will result in higher error in the estimation.

Results for the M-Scale model, although indicating a lower κ coefficient relative to OK, has its advantage in the low false negative rate relative to the OK result. This characteristic implies that the M-Scale model is better suited than OK as a decision support tool for site remediation, since FN classification "usually carries a heavier penalty, in real-life (Turney, 1995)," which "might lead to substantial financial penalties ... due to consequent litigation or from extra delays in site redevelopment (Ramsey et al., 2002)." The M-Scale model, consequently, can be understood as a better estimation model for remedial purposes relative to OK, while at the same precision level as OK. The decision-making process, however, depends on how different stakeholders negotiate on the importance of the classification results, which will be site-specific and subject to change in time.

The lower κ coefficient for the M-Scale model relative to OK should not be regarded simply as a loss of reproducibilty of attribute reality, however, because the

reproduction of map attributes could be described from different perspectives, including value precision, covariogram reproduction, and classification precision. The results presented confirms the description in Journel et al. (2000) that "semivariogram reproduction cannot be obtained without sacrificing local accuracy." Moreover, Krige and Assibey-Bonsu (2001) further stated in their comment on Journel et al. (2000) that "the possibility of it (deriving a single estimation map that reproduces semivariogram without sacrificing accuracy) ever being successfully met is a mathematical impossibility," and "the common-sense reason for this is that no mathematical technique can overcome the lack of sufficient data at the local level." Unlike CK, nonetheless, the effect of the nugget value on the M-Scale model is not as prominent, and its precision as indicated by the κ coefficient is controlled at the same level as in OK.

Contrary to the results indicated by Cohen's κ coefficient, however, the estimation variance of the example in this chapter indicates smaller values for the M-Scale model at most estimation locations compared to OK, which may be due to the fact that the M-Scale model regards samples in the ring area (increments between spatial supports) as randomly located with little redundancy among data, thus underestimating the *estimation variance* when the samples are actually clustered within ring areas. The resulting estimation variance for the M-Scale model in the example realization has similar coverage rate (proportion of times that the actual value is included in the confidence level evaluated) as OK, because the sampling scheme is close to a selection of purely random locations so that the "underevaluation" of the estimation variance is not of significance.

To sum up, reproducing global variability indeed helps to lower the false negative rates, as shown in the CK and the M-Scale results. When considering the ensemble of performances based on the reproduction of spatial variability, the estimation precision, and the applicability for remedial decisions, the M-Scale model is found to be a

robust model among the three estimation models presented, compared to OK which focuses only on the estimation precision, and to CK that emphasizes on the reproduction of global variability. Additionally, the good performance shown in the example realization indicate a possibility for the M-Scale model to outperform OK and/or CK in the precision of remedial cost that uses the estimated contamination level as the input, particularly when the associating function is unknown. To evaluate the applicability for different cases of remedial decisions, however, additional risk-benefit analysis, such as the *excess cost* described in Chapter 2, should be conducted to balance the cost originated from FP and FN, as described in Weber and Englund (1992) that "society pays a cost for all contaminated areas, either as a remediation cost for each block cleaned, or as a less easily defined group of costs (health effects, ecological damage, etc.)"

The evaluation of applicability is challenging because of the lack of comparison basis with actual measurements, and the uncertainties associated with comparative risk-benefit analysis based on different estimators. Since no remedial plan can proceed without a thorough sampling of the contaminated site (p. 351 of Isaaks and Srivastava 1989), a solid diagnostic comparison of performance of the estimation model is required. This diagnostic approach will be described in the next chapter, based on a 3-dimensional site characterization of the Passaic River (NJ) for remedial decision-making regarding sediment dioxin contamination.

CHAPTER 5

Validation on Rich Data Set for Point and Block Estimation

The M-Scale model was previously shown to be useful for scientific exploration by characterizing covariances between mean values at different scales in Chapter 3, and was further validated by using an artificial dataset to be a remedial decision support tool in Chapter 4. To demonstrate the performance of the M-Scale model in a real-world application for delineating contamination in the sediment of a waterway, the Passaic River data set was used as an example. Also demonstrated is the performance comparison for block estimation for the M-Scale model, OK and CK, because the target scale of interest (i.e., the area or volume over which the mean value is of interest) tends to not be a single point, but a larger area. To provide the base-line knowledge of statistical applications for surface water sediments, a review is provided of relevant literature.

Most research on statistical spatial analysis for sediments is limited to applications such as validating spatial correspondence or finding the strength of spatial related-

ness. Iriarte et al. (1997), for example, described the spatial variability of microbiological measurements in the Urdaibai estuary of Northern Spain by the observation of measurements among sampling stations, and concluded that the variability may be a consequence of tidal flushing. Raghukumar et al. (2006) studied seasonal and kilometer-scale spatial variability for biological measurements in deep-sea sediments in the Central Indian Basin, including the variability of bacterial counts, total organic matter, extracellular enzyme activities, adenosine triphosphate (ATP), and protein. ANOVA between locations was performed to indicate the spatial variability in the different attributes. This study concluded that whereas no significant variability exists vertically (within cores), significant variability was observed horizontally (between cores), and the variability is most significant temporally (between sampling stages). Point et al. (2007) observed the measurements of trace metal, organotin concentration and macrofaunal species abundance in the Thau Lagoon (France) by comparing the local coefficient of variation (CV) of sample triplicates within each station (local variability) to the global CV of the triplicate average (possible spatial variability) so that the existence of spatial variability in the local mean values could be confirmed. This study also compared the incidence of spatial variability of trace metal and organotin to the macrofaunal species abundance to infer the causes of the observed variability. Steven and Ekermo (2003) explained bathymetric variation of sediment depth in Goöteborg Harbour (Sweden) by comparing the variograms of sediment depth, and determined that the cause of variability in sediment was associated with area size of erosion and accumulation area. Middelboe et al. (2006) collected viral and bacterial abundance measurements in the sediments of Sagami Bay, Japan, and concluded that the change sign of Moran's I (Moran 1950, a statistic that studies spatial association at different separation distance, similar to the use of covariogram) corresponds to the patch size of up to 150 m in diameter for the abundance distribution. They further described the importance of studying spatial structure "to evaluate and interpret

differences in abundance between locations or over time." However, no indication of the importance of studying spatial correspondence for mapping attribute values and uncertainty evaluation was included in their discussion.

As discussed in Chapter 2, the application of geostatistical estimation methods for contaminant delineation is relatively recent, and few studies exist. Ouyang et al. (2002) studied DDT (and normalized DDT by total organic carbon, TOC) in river sediment, and used kriging to "characterize" its spatial distribution. In addition to conventional geostatistical structural analysis by Mear et al. (2006) of the fine-grained content of the superficial sediments in eastern Bay of the Seine (France) that characterizes a variogram model of the hole-effect type as an indication of patches of various sizes, these authors also performed conditional simulation of the same attribute (fine-grained content). Locally calculated mean values of the 80 conditional realizations were used to generate the estimation map, an approach with results identical to kriging results without further benefit, when the realization is generated using a multiGaussian approach (p. 341 of Goovaerts 1997b; p.154 of Chilès and Delfiner 1999). In Ouyang et al. (2003), heavy metal concentrations (and normalized heavy metal by Aluminum concentration) were estimated in river sediment using kriging. The paper concluded that kriging is useful "especially with regard to the potential river sediment dredging" but offered no implication for classification for remedial purposes. In fact, among the few geostatistical applications, even fewer offer approaches towards decision-making, such as classification of contamination levels, except for the study of Barabás et al. (2001). All literature information indicated the importance for characterizing relatedness between spatial scales and mapping attribute values, while no literature is found in the studies of under-water sediment attributes, that uses a single expectation map to reproduce spatial variability and describe the spatial distribution as a supporting tool for remedial application. Indeed, it makes intuitive sense to use spatial relatedness for the explanation of the cause of spatial variability,

such as the association of microbial activity distribution to the microbial abundance distribution over the area of study. A precise estimation approach like kriging, in addition to characterizing spatial relatedness, generates estimation maps that gives visual aid for the map users to observe the general trend of spatial variability, e.g. a map of microbial activity, with confidence intervals that indicate the precision of the estimate individually at each location. What is more important, other than the precision of the estimation map, is how the variability is reproduced, and consequently the precision of the ultimate result that uses the estimates generated as inputs. For example, to assess the total amount of dechlorination by microbial activity, a map that reproduces the variability of microbial activity is needed. In addition, a classification map that represents the discretized variability of the spatial attribute (p.441 of Goovaerts (1997b)) serves as decision support tool when the decision is made on the basis of the exceedance/non-exceedance of the attribute, or when the decision is made upon a cost model that is linear to the contamination level. Few studies of sediments focus on decision support using classification map, and no literature indicates the importance of reproducing spatial variability in an estimation map.

To demonstrate the performance of the M-Scale model in a real-world application for reproducing spatial variability and delineating threshold exceedance in the sediments of a waterway, the Passaic River dataset is used as an example. Estimates are generated at two resolutions:

(i) the punctual/point estimation, with estimates representing the value at each point, and

(ii) the block estimation, with estimates of average values over a certain support size.

Since the interest of site characterization and decision support for contamination in-

vestigation is rarely at a point scale, block estimations (estimating at a scale larger than the sampling units) are often used in practice. A major benefit for using block estimation is the reduction of uncertainty when the spatial resolution for the representative estimates does not need to be as fine as the measurement unit (i.e. sample). The propositional topic for block estimation in this chapter, consequently, is different from most studies that use block estimation. In the research that involves block estimation, the key topic usually pertains to the impact of block sizes on the variogram (see Clark 1977), the smoothing of estimates under different block sizes (see p.158 of Goovaerts 1997b), and the reduction of the corresponding estimation variances for the block estimates (p.311 of Journel and Huijbregts 1978). This chapter focuses on the study of a more fundamental difference between using point estimation (estimating at the same scale as the sample scale) and block estimation (estimating at a scale larger than the sample scale), that is, the impact of the nugget effect. Although popularly modeled in geostatistics as a characteristic parameter, the nugget effect is only effective for point estimation due to the following reason: according to Journel and Huijbregts (1978) (see p.311), the target scale (size of the estimation support) is often much larger than the size of the sample in practice, and thus the impact of the nugget effect on the sample-to-estimate and estimate-to-estimate covariances is negligible when compared with the nugget effect at the point scale. In fact, the sample-to-estimate and estimate-to-estimate covariances defines the impact of nugget effect on the estimates/estimation variance (see pp.299–301 and p.326 of Isaaks and Srivastava 1989 for the components affecting estimation weights, and the formula for calculating estimation variance under block estimation) because the nugget effect is inherent in the sample-to-sample covariance matrix (Aldworth and Cressie, 1999). With an intuitive perspective, block estimates substantially reduce the effect of artificially induced error and micro-scale variability (the impact of the nugget effect) because a block average is essentially the average of point variables

within an area/volume (see for example, p.328 of Isaaks and Srivastava 1989), by which the spatially independent micro-scale variability and artificially induced error are averaged out. Since the impact of nugget effect is only discernible for point estimation, a comparison between the point estimation (estimating at the same scale as the sample scale) and a block estimation (estimating at a scale much larger than the sample scale) at a "small" block (at a diameter that is regarded as small relative to the distances between sample locations) will reveal the difference between the use of point- and block- estimations due to the impact of the nugget effect. The comparison between block estimation and point estimation in this chapter, consequently, is expected to differentiate the impact of the nugget effect in terms of the diagnostic graphs/parameters stated in the previous chapter. In addition, because this comparison signifies the impact of the nugget effect, it will also provide the examination of whether an estimation model is sensitive/subject to the modeled nugget effect, when a point estimation is performed.

A comparison of estimates to the reality, however, is not readily available for many research sites. The main issue for comparing estimation models using real datasets is not only the lack of idealized mechanistic (e.g. fate and transport) models that the data should follow, but also the unavailability of an exhaustive measurement over the entire area/volume as the comparison basis. The exception is when the validating sets are available exactly at the sampling locations, and when precision/reproducibility of estimates at other locations are not the objective of the study (Schweizer and Kronholm, 2007; Schnabel and Tietje, 2003; Mear et al., 2006). For a spatial estimation model, however, the interest for validation is to choose among models to be used as a decision support tool for spatial assessment of mapped attribute values. In this case, the samples are usually not re-sampled at the exact sample locations. Cross-validation is an approach widely accepted for comparing among estimation models (Maravelias et al., 1996; Goovaerts, 1998; Su et al., 2006) although the approach is

also applied ad-hoc for the selection of certain estimation parameters (an example is presented by Barabás et al. 2001, who used cross-validation to select the likelihood threshold for delineation of contamination). The principle of this approach is to estimate the attribute value at each sample point as if the sampled value is unknown (p.111 of Chilès and Delfiner 1999), which enables the comparison of estimates to actual values at each sample location.

Cross-validation is usually used to compare model performance locally since this approach generates pairs of true and estimated values only at sample locations. However, since the estimation models for reproducing variability such as CK and the M-Scale model aim at reproducing variability by separate estimates at each location, the result for cross-validation would also reflect the performance of variability reproduction by these models. Re-estimation of sample points will be conducted to compare model performance, under the point estimation case as well as the block estimation case. Particularly, the cross-validation of the point estimation case is used to reconfirm the characteristics of estimation described in Chapter 4 for the three estimation models. From a remedial decision-making perspective, the comparison also provides an approach for model selection by exploratory examination, especially for studies where non-negligible nugget effects are observed and the comparison involves estimation models that are sensitive to the nugget effect.

The objective of this chapter is to compare the performance of the M-Scale model vs. OK and CK in the delineation of dioxin contamination in Passaic River sediments using the same statistical endpoints employed in Chapter 4. As a measure of assessing estimation performance and examining model sensitivity against the nugget effect, cross validation will be conducted using point and block estimates of contamination.

5.1 The Passaic River Data Set

The Lower Passaic River Restoration Project is a comprehensive study of the 17-mile tidal stretch of the Passaic River from Dundee Dam to Newark Bay in northern New Jersey. The study is being carried out by a partnership of federal and state of New Jersey agencies. During the course of the study, the sediments in the lower eight miles of the river have been identified as a major source of contamination to the 17-mile stretch of the Passaic River and to Newark Bay. The partner agencies have developed a companion Focused Feasibility Study (FFS) to evaluate a range of alternatives that might be implemented as an early action to control this major source of pollution. This action is intended to take place in the near term, while the comprehensive 17-mile study is on-going.

The specific study area is a 10 km reach of the Passaic River in New Jersey, under investigation by the EPA as part of its evaluation of the Diamond Alkali Superfund site (Barabás et al., 2001). Polychlorinated dibenzo-p-dioxin and furan (PCDD/F) data for 2,3,7,8-substituted congeners were acquired from the U.S. EPA and from a database of New Jersey sediment data compiled by the NOAA (U. S. EPA, 2002; NOAA, 2002). The data set consists of 94 surface and 444 subsurface samples (a total of 538) from 94 sediment cores taken along 27 approximately equally spaced (˜370 m) transects in 1995 (Figure 5.2, left). Most transects consist of a mid-stream and two bank-side cores about 48 m apart. Observations represent depth-averaged concentrations from approximately 30 cm core intervals.

The dataset used in this work is the most toxic congener, 2,3,7,8-TCDD, in order to delineate the contaminated areas and determine the adequacy of the data in terms of the uncertainty distribution in space. The sample histogram for the 10-based logarithm of 2,3,7,8-TCDD data are displayed on Figure 5.1.

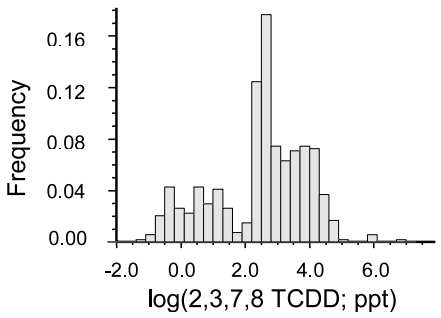

Figure 5.1: Sample histogram (value distribution) for the 2,3,7,8-TCDD data (figure used with permission by Barabás et al. 2001).

In geostatistics, dissimilarity between data is represented as a function of Euclidean distance between points. Euclidean distance can be meaningless in a meandering river as it could relate to points over intervening land. For implementation of the spatial estimation, the study area is consequently converted into a rectangular volume, mathematically "unbending" the meandering river as some of them would necessarily be measured over land. The physical boundary was first defined into four segments that correspond to the four sides of a rectangle. For this study site, the two banks correspond to two opposite lengths of the rectangle, while the lines marking the northern and southern ends of the river section correspond to the other two opposite sides. An equal number of grid points were then laid out along each of the pairs of facing boundaries. Sampling units in the data set are approximately 30 cm core intervals with coordinates of sampling locations transformed prior to analysis. Fig. 5.2 shows the sampling points in the original and transformed coordinate systems, with coordinates generated by Barabás et al. (2001). The samples in the three-dimensional space are shown in Fig. 5.3 (adopted from Barabás et al. (2001)). Notice that samples are generally sparser in the deeper sediments.

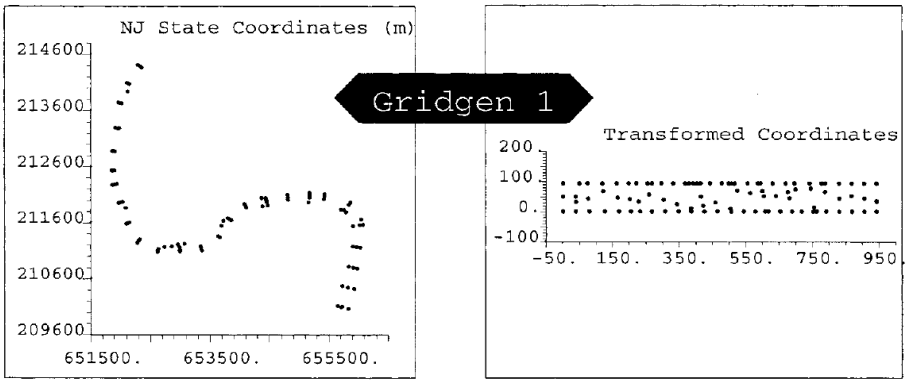

Figure 5.2: Coordinate transformation for the Passaic River dataset. One unit distance in the transverse direction (y-direction) corresponds to 0.6-4.0 m depending on the local width of the channel, and in the flow direction (x-direction) one unit distance corresponds to 19-21 m. Units in the vertical direction (depth direction, z-direction) were kept unchanged (measurements in centimeters) since the direction does not involve any meandering nature (figure used with permission by Barabás et al. 2001).

5.2 Model Application

For all three estimation models, including the M-Scale model, OK and CK, the same variogram model is used. The modeled variogram is an exponential model with a sill value of 1.0 (including a nugget effect of 0.2 and structural variance of 0.8), and with an effective range of 90 at the flow (x) and transverse (y) directions. The geometric anisotropy ratio is observed in directional variograms to be 2.0 at the vertical (z) direction so an effective range of 180 is used at that direction. The experimental variograms in the x, y, and z directions are shown in Fig. 5.4 with the corresponding directional variogram models visually fitted to the experimental variograms. OK and CK are performed according to the model variogram, while the M-Scale estimation is performed using sphere volumes of diameters 1, 20, 40, 80, 120, 240, 480 and 1000 units (converted from original units of coordinates) at the x and y directions. The same anisotropy for the variogram is also considered for the M-Scale model so for all scales the diameters in the z direction are doubled.

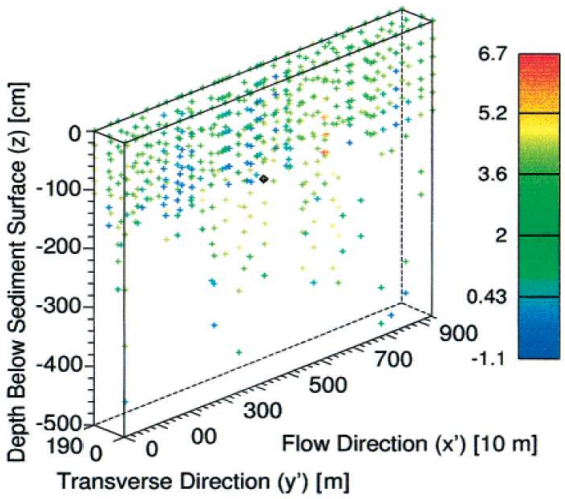

Figure 5.3: Samples in the transformed coordinate. Color scales represents the amount of log-TCDD concentration (figure used with permission by Barabás et al. 2001).

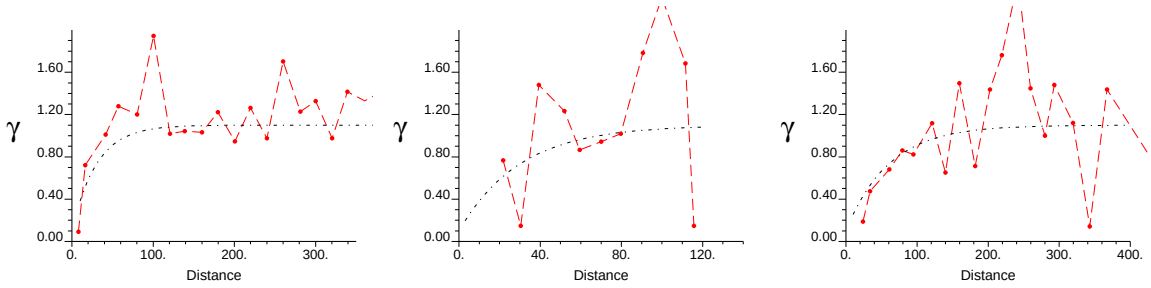

Figure 5.4: Experimental variograms with the variogram model for the Passaic River study area. Left to right: flow direction (x), transverse direction (y) and vertical direction (z).

5.3 Results for Point Estimation

5.3.1 Visual Comparison by Estimation Map

The estimation maps and the corresponding estimation standard deviation created for the x-z cross-section at y=1 using the M-Scale model, OK and CK are illustrated in Figs. 5.12, 5.13 and 5.14, where the variability of different sample concentrations is visually observed in the M-Scale and CK estimates, while due to the smoothing

effect with estimates tending toward the global mean, the variability of estimates in the OK estimates are visually smaller. The estimation standard deviation is smallest for the M-Scale model and largest for CK, an observation which is consistent with the observation when applied using artificial realizations. This observation indicates that the performance of variability reproduction for the three estimation models is independent of the absolute value for the modeled sill and range parameters. In other words, in spite of the strength of relatedness characterized for the individual research sites (the range of influence spatially, and the magnitude of the associated fluctuation between locations), the concave and convex smoothing for OK and CK persists as long as OK estimation includes no further constraint other than the unbiasedness, and CK estimation involves a variogram with a nugget effect that is not negligible relative to the sill value.

5.3.2 Cross-Validation and Diagnostic Parameters for Reproduction of Global and Spatial Variability

As mentioned in the previous chapter, a diagnosis approach is needed for model selection. Three statistical summary graphs and corresponding diagnostic parameters are presented in this study for the cross-validation results, respectively scatter plot with global standard deviation, quantile-quantile (Q-Q) plot with Anderson-Darling (A-D) test statistics, and covariograms with structural variances and influence ranges. As mentioned in the beginning of the chapter, although cross-validation is usually used to compare model performance locally at sample locations, CK and the M-Scale model aim at reproducing variability by separate estimates at each location, hence the result by re-estimating individual datum removed by turn would also reflect the performance of variability reproduction using these models. Note that data are removed one point at a time instead of one core (the entire column in the z-direction) at a time because

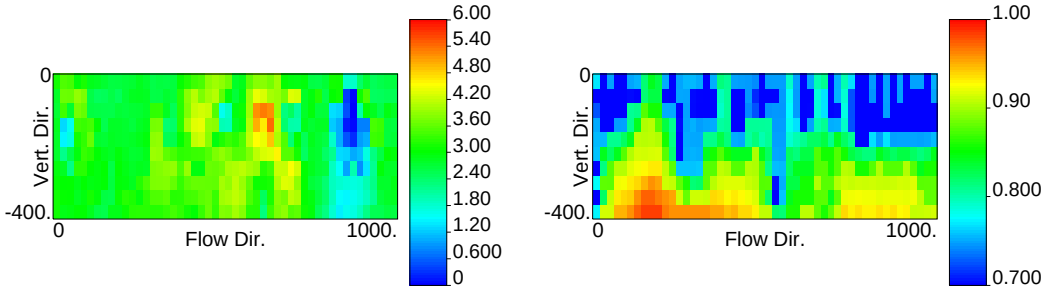

Figure 5.5: M-Scale point estimation using the Passaic River dataset. Left: Estimation map. Right: estimation standard deviation.

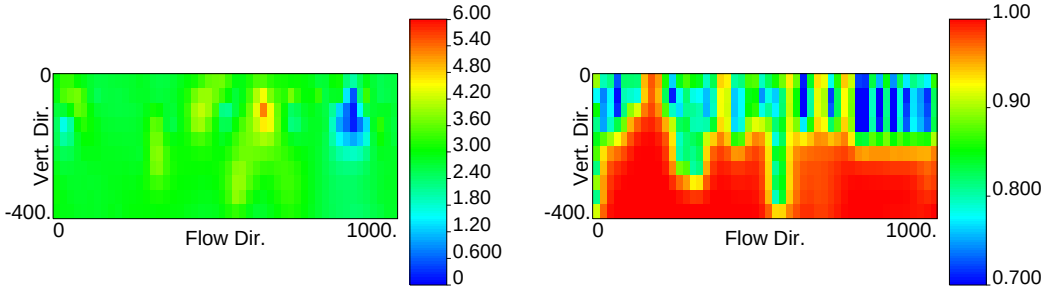

Figure 5.6: OK point estimation using the Passaic River dataset. Left: Estimation map. Right: estimation standard deviation.

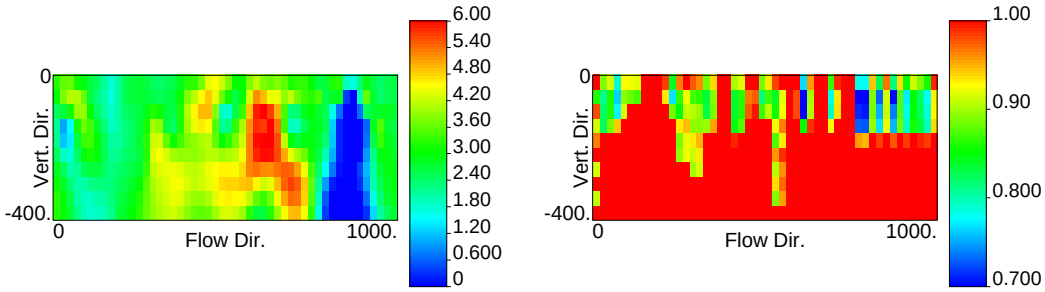

Figure 5.7: CK point estimation using the Passaic River dataset. Left: Estimation map. Right: estimation standard deviation.

the validation is done for the estimation in a three-dimensional space instead of the mean values over blocks covering the z-columns.

Scatter plots and Q-Q plots for the re-estimates of the cross-validation are shown for the M-Scale model, OK and CK (Fig. 5.8 through Fig. 5.10). Covariograms of the re-estimates are also shown in Fig. 5.11, where the covariograms are demonstrated for the flow direction. Diagnostic parameters for the cross-validation results are also listed in Table 5.1. Standard deviation of the re-estimates indicates the extent of values and a reduction of standard deviation indicates smoothing. Smaller A-D test statistics indicate a better fit of the global histogram. Reproduction of spatial variability is represented by the structural variances and influence ranges. Smaller structural variances and larger influence range indicate more (concave) smoothing for the estimates. It is observed that CK best reproduces the global variability according to the stan-

Table 5.1: Diagnostic parameters for the precision (the mean value) and variability reproduction (others) of the cross-validation performed for the point estimation using the Passaic River dataset.

Statistics	Target	M-Scale	OK	CK
Mean	2.90	2.91	2.92	3.00
Standard Deviation	1.00	0.84	0.60	1.14
Slope	–	0.43	0.33	0.61
A-D test statistic	–	8.57	26.08	2.48
Structural Variance	0.80	0.65	0.40	1.00
Influence Range (Flow Direction)	90	120	120	120

dard deviation (comparing to the M-Scale model and OK, produces closest standard deviation of the re-estimates relative to the standard deviation of the validation sample), slope of regression (closest to 1.00), and A-D test statistics (closest to 0.00). the M-Scale model best reproduces spatial variability according to the structural variance (closest value to the validation sample), and OK performs worst in the reproduction of both global and spatial variability by the statistics listed in the table except the mean which does not indicate reproduction of variability. The characteristics of performances for variability reproduction for the three estimation models are consistent

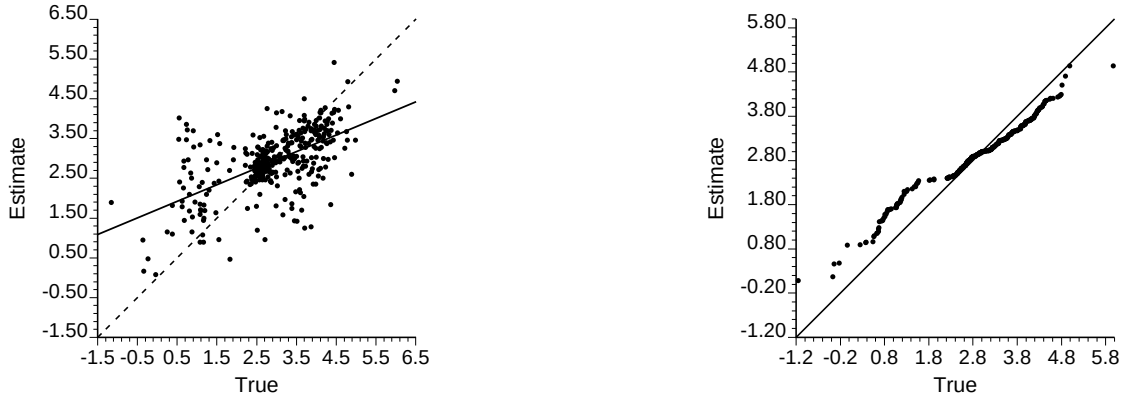

Figure 5.8: M-Scale scatter plot and Q-Q plot for the point re-estimates and with respect to the validation set using the Passaic River dataset. Left: scatter plot. Right: Q-Q plot.

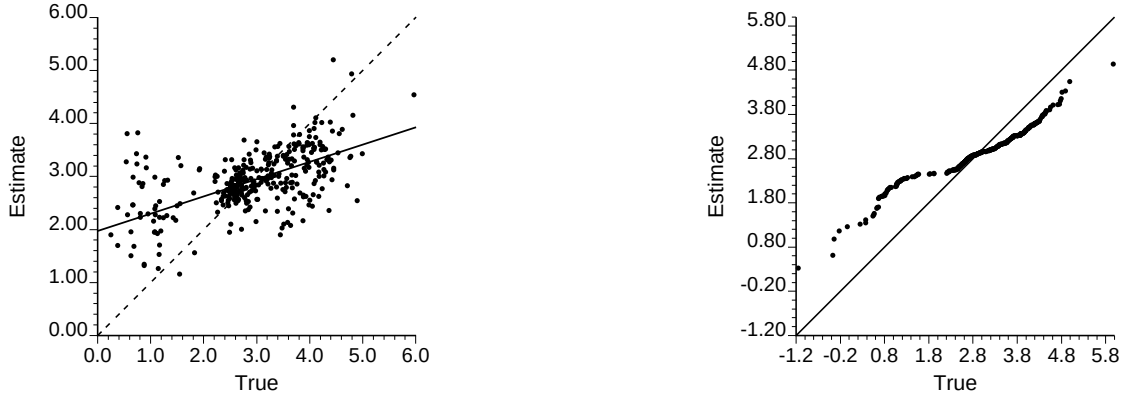

Figure 5.9: OK scatter plot and Q-Q plot for the point re-estimates with respect to the validation set using the Passaic River dataset. Left: scatter plot. Right: Q-Q plot.

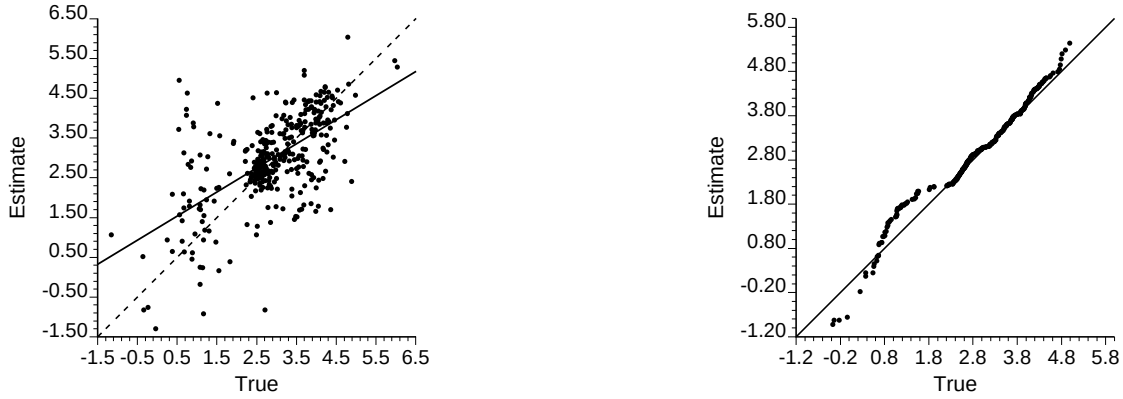

Figure 5.10: CK scatter plot and Q-Q plot for the point re-estimates with respect to the validation set using the Passaic River dataset. Left: scatter plot. Right: Q-Q plot.

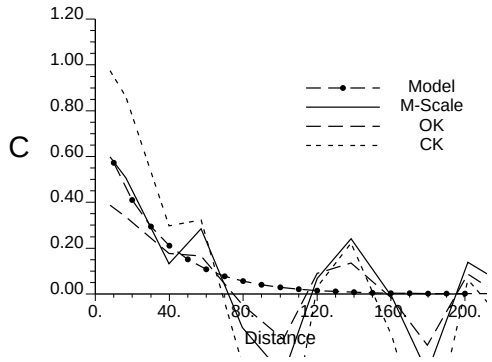

Figure 5.11: Covariogram reproduction for the point re-estimates with respect to the validation set using the Passaic River dataset. All re-estimates of the 346 sample locations are included. Notice that the M-Scale covariogram of the re-estimates lies closer to the model covariogram. At short separation distances, covariances of OK estimates are lower than the model covariogram, and covariances of CK estimates are higher than the model covariogram.

with the characteristics of results described in Chapter 4, in that CK generates the closest standard deviation of the re-estimates relative to the standard deviation of the validation sample, the slope of regression closest to 1.0, and smallest A-D test statistics (see diagnostic statistics in Section 4.2.1). It should be noted, however, that the results for global variability, in addition to the spatial variability that associates attribute values between locations, include a combination of artificially induced error and micro-scale variability in the validation sample, which may not represent the true variability of the population (see p.52 of Chilès and Delfiner (1999) and reviews of the nugget effect in Chapter 6). Thus, a good reproduction of global variability may not necessarily represent good estimation performance (see for example, p.8 of Aldworth and Cressie 1999). The resulting covariogram indicates that the best spatial variability is reproduced by the M-Scale model as observed visually in the covariogram comparison (Fig. 5.11), and this performance is further quantified by the structural variance that has the closest value relative to that generated by the validation sample (Table 5.1). The same influence ranges for all three models indicate that all measurements except the left-out measurement are used for re-estimation, which is a similar characteristic seen in Chapter 4 for the influence range (Table 4.2).

In addition to the three diagnostic graphs and statistics, the contingency table for the classification of the re-estimates is also created as shown in Table 5.2, where the criterion for classification of exceeding 10000 ppt (4.0 on the log-scale, about the average plus one standard deviation of the sample values) is selected as an example. Although literature of dioxin studies exemplifies the use of 10 and 25 ppt as thresholds for 2,3,7,8-TCDD contamination for the Passaic River study (Barabás et al., 2001; Iannuzzi et al., 1995), these values fall around the 0.09 and 0.14 quantiles of the sample, which are low quantiles that are likely to fall below the background value, expected to correspond to high likelihood of exceedance. Moreover, according to the site-characterization performed by Barabás (2003), the threshold exceedance of the two thresholds using kriging as the estimation model indicates more than 95% of the research area for both thresholds (see p.106 of Barabás 2003). As indicated in SPAWAR Systems Center (SSC) San Diego and Battelle Science & Technology International (2003), "cleanup levels must not be lower than ambient (i.e., background) levels," otherwise "sediments from elsewhere in the water body can be transported to and deposited at the site through natural processes," and "result in a cleanup that effects little or no ecological improvement." The remediation for the research site under this threshold of exceedance, consequently, would be a suspension for the cleanup action, on the basis of either the pooled sample (quantile statistics without any spatial context), or the use of spatial estimation (indicator kriging, for exmaple). In other words, a spatial estimation may not provide more support for the negotiation among stakeholders, relative to an analysis based on the univariate analysis of the pooled sample. The selection of exceedance threshold in this study, although it appears to be arbitrary, exemplifies the case where *selective remediation* (see for example, Smith and Williams 1996a,b) is used as the clean-up strategy, i.e., clean up only the hot spots within the site to lower the pollutant level of the entire site (see also p.26 of Saito 2003), which is also suggested by Barabás et al. (2001) "to clean

the most risky locations" for the Passaic River study. Unlike the example in Chapter 4, however, no classification map is presented to observe the clustering of exceedance because no such map exists for the reality over the entire research area. Cohen's

Table 5.2: Contingency table for the classification of point re-estimates with respect to the validation set using the Passaic River dataset.

		M-Scale		OK		CK		Total
		+	−	+	−	+	−	
Target	+	4.4	11.8	2.6	13.6	8.4	7.8	16.2
	−	2.0	81.8	0.3	83.5	9.5	74.3	83.8
Total		6.4	93.6	2.9	97.1	17.9	82.1	100.0

κ coefficients are also presented as shown in Table 5.3, expressing the standardized precision of classification for the exceedance/non-exceedance at re-estimation locations. The contingency table indicates that CK is a significantly more conservative

Table 5.3: Cohen's κ coefficient of the point re-estimates with respect to the validation set for the three estimation models using the Passaic River dataset.

	M-Scale	OK	CK
κ	0.49	0.49	0.36

approach for point estimation compared to the other two approaches (higher positive rates for the estimation model, consequently higher true positive/false positive than true negative/false negative rates relative to other estimation models). This tendency of conservative estimation comes from the fact that the constraint for the reproduction of global variance involves the reproduction of both micro-scale variability and artificially induced error, consequently covers the range of *occasional* exceedance (Cohen, 1960) resulting from local variability. Regardless of the better TP rate, however, the result for Cohen's κ coefficients indicate the same classification precision for the M-Scale model when comparing to OK, both better than that of CK. The same characteristics of classification precision is also observed in the artificial data sets, as shown in Table 4.3 and 4.4 that indicates smaller κ coefficient for the M-Scale model and OK relative to CK, and a conservative tendency of CK relative to OK and

the M-Scale model. Coverage rates for the confidence intervals indicated that the M-Scale covers 85% of the true values, comparing to OK at 94% and CK at 95% is an underestimation of the range of estimation uncertainty. As explained in Chapter 4, clusters in the samples results in underevaluation of estimation variance for the M-Scale model. The underevaluation of estimation variance, however, is obvious in comparison to the underevaluation of estimation variance demonstrated for the example in Chapter 4 (M-Scale: 91%, OK: 92%, CK: 93%), due to the fact that the samples in the Passaic River study are intensively clustered towards river bed and around locations of the core columns, as shown in Fig. 5.3, deviating from the ideal case of a purely random selection of locations as approximated in the example given in Chapter 4. As also indicated in Chapter 4, further research will be required for the M-Scale model to account for the effect of sample clusters.

5.4 Results of the Block Estimation

A point target scale may not always make sense for remedial decisions, as described in Schumacher et al. (1998) that even when "sampling were done with tablespoons," for remedial applications "backhoes, rather than tablespoons, would more likely be used to remediate the area." In this study, the target scale is chosen to have a diameter of 20 length units in the flow direction as a demonstration, approximately a block of 400m×80m×40cm in the original x, y, and z coordinates, which is comparable with respect to the smallest distance between sampling locations (the smallest distance between sample locations is around 10 unit distances, therefore a block centering at the sample location may include about one other sample point depending on where the center location is), and regarded as large in comparison to the sample units (the core samples are collected using using 2" PVC pipes (according to the EPA sampling

protocol, see U. S. EPA 2007) cut at lengths of approximately 30 cm.) The choice of this target scale is expected to yield results similar to those generated using point estimation if the estimation model is not sensitive to the impact of the nugget effect. A result that does not resemble the result of the point estimation will characterize the difference between the use of block and point estimations, without complicating the discussion for the choice among different sizes for the target scale. In fact, the block averaging involved in a block estimation corresponds to a substantial, if not complete, reduction in the impact of the nugget effect (Starks, 1986; de Roo et al., 1992; Atkinson and Kelly, 1997). The impact of nugget effect in this example can be regarded as completely reduced, since block averaging alters the nugget effect by the inverse ratio of the block volume (see p.78 of Matheron 1971). In a real application, however, a pre-defined remediation unit of the target scale needs to be decided according to the clean-up criteria and management approaches used as the result of cost-benefit analysis. The performance characteristics of CK in the point estimation are also compared to the CK block estimation, which serves to observe the sensitivity of CK estimates against the nugget effect that is conceptually illustrated in the previous chapter.

5.4.1 Visual Comparison by Estimation Map

The block estimation was conducted on a block size with a diameter of 20 length units. The estimation maps and the corresponding estimation standard deviations created for the x-z cross-section at y=1 using the M-Scale model, OK and CK are illustrated in Figs. 5.12, 5.13 and 5.14. As compared to point estimation maps, the block estimate for OK and the M-Scale model is visually similar to the point estimates across the study area, a result due to the fact that the selection of the target scale corresponds to a diameter not significantly larger diameter than the smallest distance

between samples. In fact, because OK and the M-Scale model are not designed to reproduce local variability (the nugget effect), reduction of the nugget effect by using a small target scale (although considered a large area relative to the sampling units) impacts the results of the estimate by only a negligible variability that is visually not observable. The CK estimates, however, looks significantly less variable than CK point estimation map because the target global variance to be reproduced is largely reduced with the reduction of artificially induced error and micro-scale variability. This comparison of point- and block- estimation for CK indicates that CK is sensitive to spatially uncorrelated components of the global variance (the variance from artificially induced error and micro-scale variability). The maps for estimation standard deviation are generally smaller in all three models when compared to their point estimation counterpart, because the estimate is a block mean value, and the variability of which is considerably smaller than the individual point attribute values due to block averaging.

5.4.2 Cross-Validation of Block Estimates by Rescaled Samples for Reproduction of Global and Spatial Variability

Unlike the cross-validation for point re-estimates where the validation set is readily available as the sample values, block re-estimates at the target scale should compare to a value representing actual mean value over the target scale, which is unknown. A rescaling is thus required for the original sample to represent the block mean value of the target scale at the exact sampling locations before the cross-validation can be performed. In fact, cross-validation approaches are seldom performed for block estimation due to the lack of the actual block mean value. Pardo-Igúzquiza (1998) studied the areal average climatological rainfall mean (i.e., the block average of the

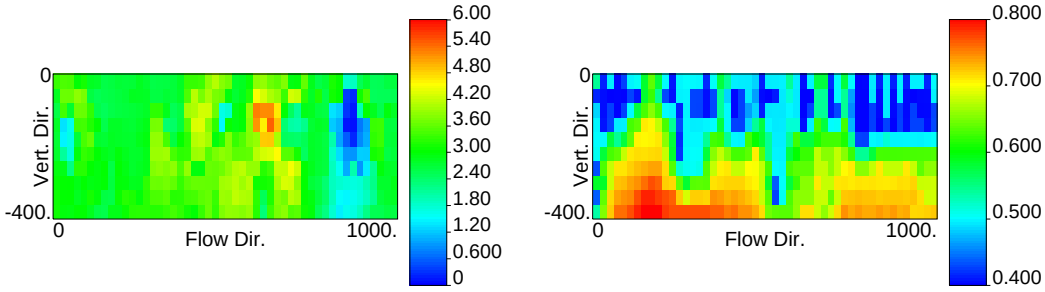

Figure 5.12: M-Scale block estimation for the Passaic River dataset. Left: Estimation map. Right: estimation standard deviation.

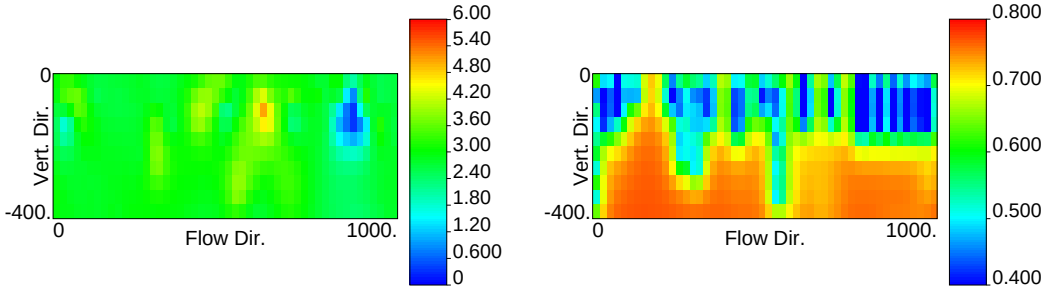

Figure 5.13: OK block estimation for the Passaic River dataset. Left: Estimation map. Right: estimation standard deviation.

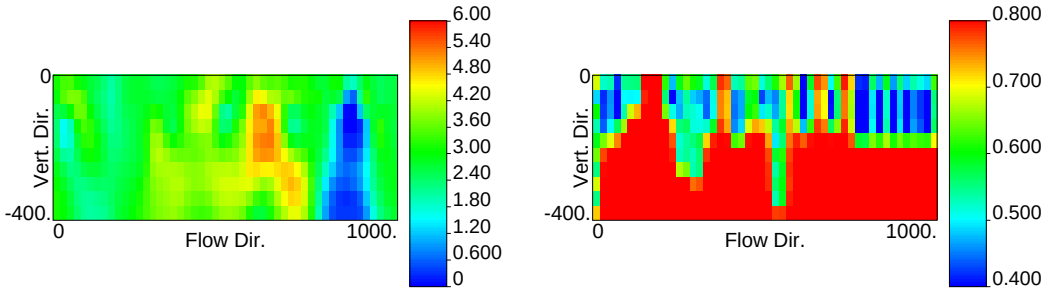

Figure 5.14: CK block estimation for the Passaic River dataset. Left: Estimation map. Right: estimation standard deviation.

climatological rainfall) in the Guadalhorce River basin (Spain), and indicated that "cross-validation must be used with caution, small differences in the cross-validation results may be given just by sampling variability." However, no sampling variability, i.e., the nugget effect, is reduced in this study and the cross-validation for the block estimates is done with a validation set of point measurements. There are rare cases where cross-validation can be done with a hard-evidence validation set. For example, in the description of Stark et al. (2003), kriging is used to study the yield of corn. Cross-validation of the block estimate is performed by using the mean of exhaustive samples within the estimation block as the validation set. However, an exhaustive sample for each block volume around the sample location is usually not attainable in most geostatistical applications. A block kriging performed on a sample location without leaving out the sample value at that location, to this end, fits the need for the rescaling of samples that provides the validation set without the need of an exhaustive sample. In fact, the block kriging estimates at the sample locations without leaving out the sample values is considered to have the least estimation error at sample locations among all linear estimation approaches (Chilès and Delfiner, 1999), consequently best represents the real block mean value at the sample location. As block surrogates of the measured data, subsequently, it is proposed in this study that the rescaled samples are compared to the re-estimates generated by the estimation models, with one sample point left out each time at the corresponding sample locations. The results of the cross-validation for block estimates are compared to the results of the point estimates to indicate the impact of using block estimation relative to point estimation. The performance statistics for the block estimation of the Passaic River data is not compared to the performance statistics for point estimation of the artificial data sets, because the comparison involves different objectives using a different dataset, which does not characterize the model performance in either the change of target scale, or the general applicability for the different data sets.

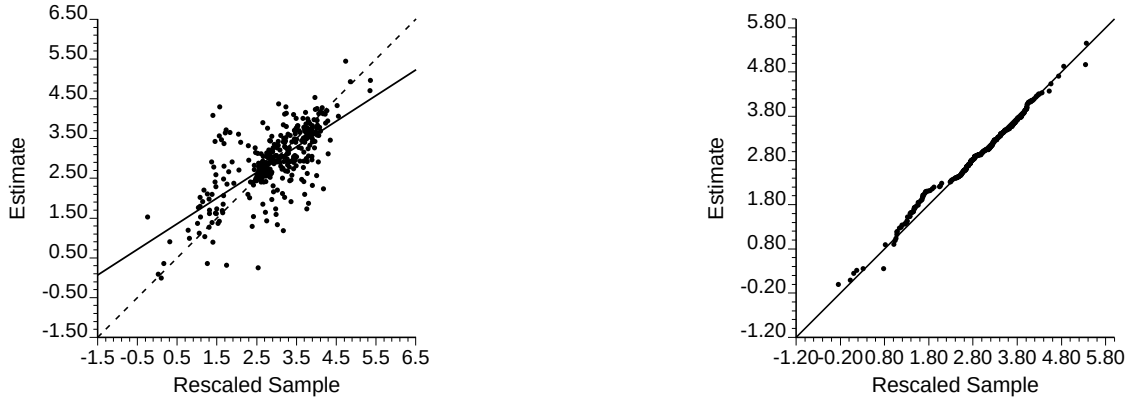

Figure 5.15: M-Scale scatter plot and Q-Q plot for the block re-estimates with respect to the validation set using the Passaic River dataset. Left: scatter plot. Right: Q-Q plot.

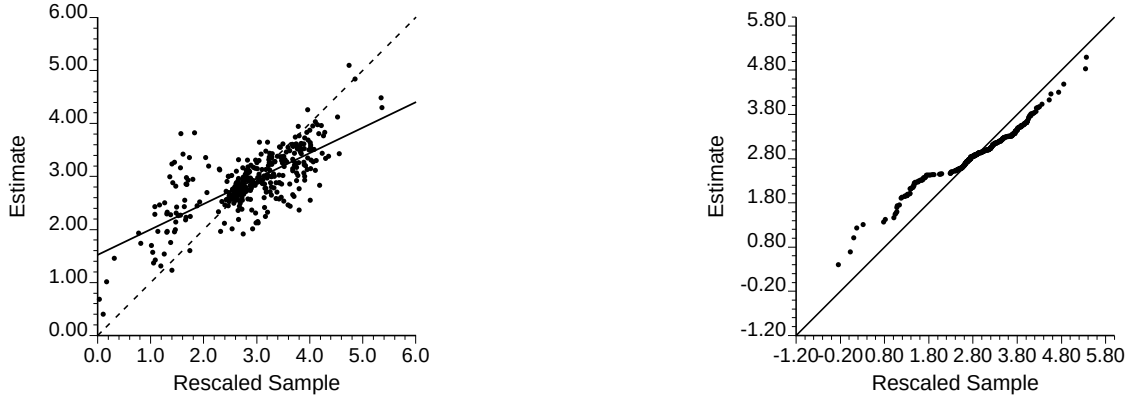

Figure 5.16: OK scatter plot and Q-Q plot for the block re-estimates with respect to the validation set using the Passaic River dataset. Left: scatter plot. Right: Q-Q plot.

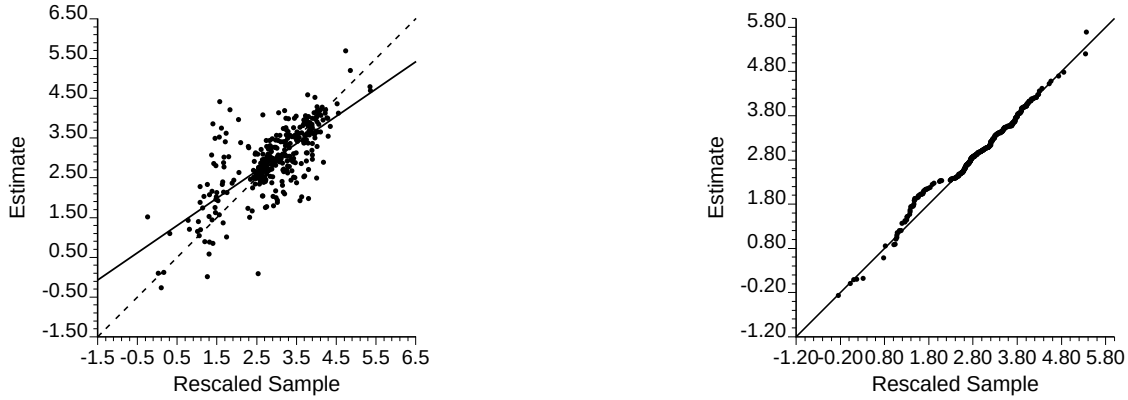

Figure 5.17: CK scatter plot and Q-Q plot for the block re-estimates with respect to the validation set using the Passaic River dataset. Left: scatter plot. Right: Q-Q plot.

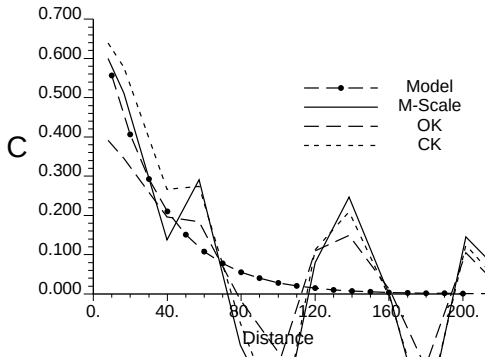

Figure 5.18: Covariogram reproduction for the block re-estimates with respect to the validation set using using the Passaic River dataset. All re-estimates at the 346 sample locations are included. Notice that the M-Scale covariogram of the re-estimates lies closer to the model covariogram. At short separation distances, covariances of OK estimates are lower than the model covariogram, and covariances of CK estimates are slightly higher but lie reasonably close to the model covariogram.

Scatter plots and Q-Q plots for the re-estimates of the cross-validation are shown for the M-Scale model, OK and CK in Fig. 5.15 through Fig. 5.17. Covariograms of the re-estimates are also shown in Fig. 5.18, where the covariograms are demonstrated for the flow direction. Diagnostic parameters for the cross-validation results are listed in Table 5.4. Visually by the scatter plots and Q-Q plots, it is observed that the estimate-validation pairs lie closer to the 45° bisector for the M-Scale model and CK, an indication of the reproduction of global variability relative to OK. By the covariograms, it is also observed that both the M-Scale model and CK lies closer to the ideal block variogram model, and indication of better reproduction of the spatial variability relative to that of OK. The performance is not as clearly different between the M-Scale model and CK. Unlike point estimation, the A-D test statistics also indicate a similar reproduction of histogram for the M-Scale model and CK. These results, in comparison to the results of point estimation (Table 5.1), indicates that when local variability (the nugget effect) is reduced by block averaging, CK and the M-Scale model have similar performance in reproducing variability, both globally and spatially, at least for the current application of this study.

Table 5.4: Diagnostic parameters for the precision (the mean value) and variability reproduction (others) of the cross-validation performed for the block estimation using the Passaic River dataset.

Statistics	Target	M-Scale	OK	CK
Mean	2.92	2.91	2.92	2.95
Standard Deviation	0.86	0.84	0.58	0.85
Slope	–	0.65	0.48	0.68
A-D test statistic	–	1.25	13.65	1.33
Structural Variance	0.67	0.65	0.40	0.66
Influence Range (Flow Direction)	90	120	120	120

Table 5.5: Contingency table for the classification of block re-estimates with respect to the validation set using the Passaic River dataset.

		M-Scale		OK		CK		Total
		+	–	+	–	+	–	
Target	+	4.0	4.5	1.8	6.7	4.8	3.7	8.5
	–	3.5	88.0	0.2	91.3	4.7	86.8	91.5
Total		7.5	92.5	2.0	98.0	9.5	90.5	100.0

The contingency table for the classification of the re-estimates is again created for block estimation as shown in Table 5.5, with respect to the same exceedance threshold of 4.0 used in point estimation. CK remains the most conservative among the three estimation models for block estimation, as is also indicated by the results of point estimation in Table 5.2. When comparing the κ coefficient (Table 5.6) to the point estimation (Table 5.3), however, it is observed that relative to the significant difference between CK and the other two models (both differences 0.13 relative to the κ of 0.36 for CK), the difference between CK and the other two models (0.03 and 0.10 relative to the κ of 0.46 for CK) is not as significant as in the results of point estimation. Again the difference between point- and block- estimations confirms that CK estimates are sensitive to artificially induced error and micro-scale variability.

Table 5.6: Cohen's κ coefficient of the block re-estimates with respect to the validation set for the three estimation models using the Passaic River dataset.

	M-Scale	OK	CK
κ	0.49	0.56	0.46

5.4.3 Sensitivity Analysis by Conditional Realization of Rescaled Samples

Although simple diagnostics can be evaluated using cross-validation for block estimation, the diagnostics involve a certain level of uncertainty because the rescaled samples are products of block kriging estimations themselves. The sensitivity analysis for this uncertainty of rescaling is consequently required to ensure the result of cross-validation (see p.93–94 of Deutsch and Journel 1998 for a discussion of the uncertainty corresponding to block estimates). However, the estimation variance/estimation standard deviation evaluated at separate estimation locations does not provide uncertainty statistics for parameters of spatial association (see for example, pp. 493–494 of Journel and Huijbregts 1978 and pp. 369–370 of Goovaerts 1997b), therefore a simulation is needed in order to parameterize the uncertainty in the spatial statistics such as the uncertainty in the covariograms.

Simulation approaches have long been used for comparison of model performance between models under different equi-probable realizations Meisel and Turner (1998); Collins and Woodcock (1999); Bian and Butler (1999); Curriero et al. (2002); Kukush (2005) and the main reason for applying simulation approaches is that spatial variability is reproduced within each simulated realization Istok and Rautman (1996); Goovaerts (1998); Lin et al. (2000); Proce et al. (2004), so that certain parameters of interest that characterize spatial correspondence, e.g. the gradients between locations or length between locations, could be evaluated for their possible values. Unlike the applications described above, the equi-probable realizations are used in this study not as the input for estimation models or target parameter of interest, but as a comparison basis on the impact of uncertainty of the rescaled validation set, i.e., the sensitivity analysis of the rescaling.

For the sensitivity analysis with respect to the uncertainty of rescaling, two hun-

dred realization of the TCDD concentration distribution are generated. Average values of the target scale at each sample locations are calculated for each of the 200 realizations, regarded as the equi-probable rescaled samples. Q-Q plots and variograms of the equi-probable rescaled samples are shown in Figs. 5.19 through 5.21, with solid lines in the Q-Q plot indicating the 90% upper and lower limits of each quantile for the 200 realizations. Also in the covariograms, solid lines represent covariances of the re-estimates, and dash lines with plots indicates the 90% upper and lower bounds of the covariance at different separation distances of the equi-probable rescaled samples along the flow direction. In the Q-Q plots, it is observed that the M-Scale model and CK histograms better reproduce the histogram of the equi-probable rescaled samples since both the upper and lower limits for the Q-Q plots generated under a comparison between OK and the rescaled validation sets deviate from the 45° bisector on the Q-Q plot. When different covariograms of block estimation are compared to the summarized covariogram for the reproduction of the equi-probable rescaled samples, it is visually observed that a better reproduction is achieved by the M-Scale model, followed by CK and OK. It is also observed in the figures that the fluctuation of covariance between separation distances is larger than the fluctuation of covariance between realizations at the same separation distance (more variability between locations than between rescaled realizations of the same location). The results indicate that uncertainty corresponding to rescaling (unknown attribute values within the block centering at a location with known sample value) is smaller than the uncertainty originated from the re-estimating of data (assumed unknown sample value). In other words, for the characterization of covariogram between blocks, local uncertainty of the block is generally smaller than uncertainty between locations. The results for Q-Q plots and covariograms indicate that the uncertainty due to sample rescaling is not a significant factor in the cross-validation for the study site, i.e., the result of cross-validation is not sensitive to rescaling, at least for this particular case

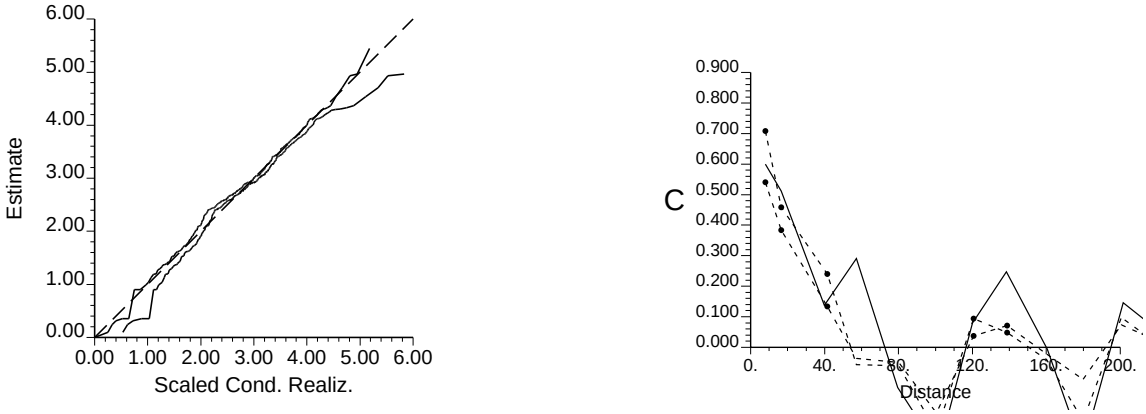

Figure 5.19: Q-Q plot and covariogram for the 200 rescaled realizations comparing to M-Scale block re-estimates using the Passaic River dataset.

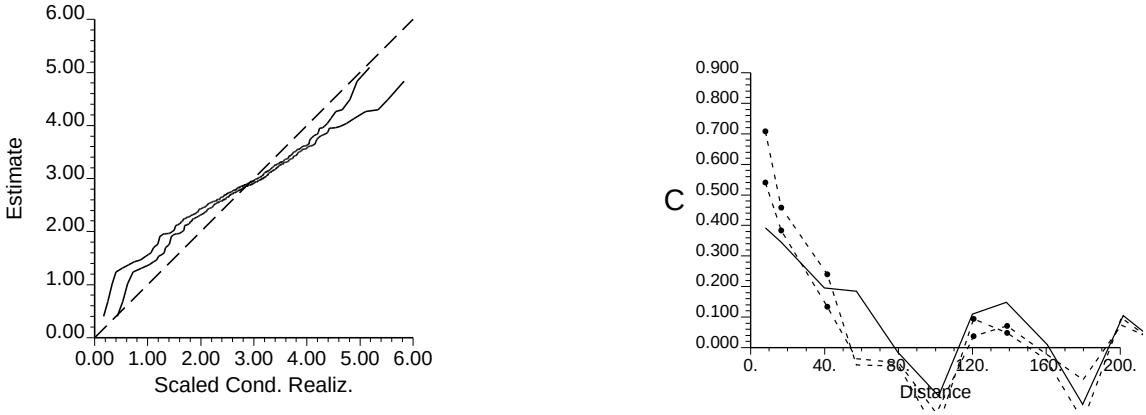

Figure 5.20: Q-Q plot and covariogram for the 200 rescaled realizations comparing to OK block re-estimates using the Passaic River dataset.

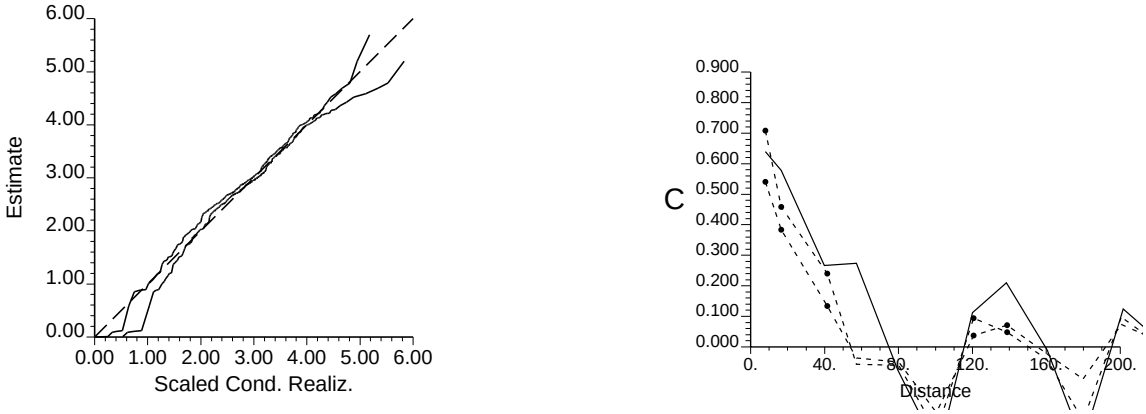

Figure 5.21: Q-Q plot and covariogram for the 200 rescaled realizations comparing to CK block re-estimates using the Passaic River dataset.

Table 5.7: Summarized diagnostic parameters for the block re-estimates of the three estimation models with respect to the 200 rescaled realizations using the Passaic River dataset. The range of values indicates the 90% lower and upper bounds of the diagnostic parameters.

Statistics	Target	M-Scale	OK	CK
Mean	2.90–2.94	2.91	2.92	2.95
Standard Deviation	0.83–0.89	0.84	0.58	0.85
Slope	–	0.67–0.72	0.50–0.53	0.71–0.77
A-D test statistic	–	0.84–2.04	12.38–16.77	0.98–2.56
Structural Variance	0.60–0.75	0.65	0.40	0.66
Influence Range (Flow Direction)	90	120	120	120

under the target spatial scale.

Diagnostic parameters are calculated for each equi-probable sample, and the summarized parameters are listed in Table 5.7. It can be observed that both the M-Scale model and CK reproduce global and spatial variability better than OK, as indicated by the standard deviations, A-D test statistics, and structural variances. The difference is not significant for the ranges of values between the M-Scale model and CK, which is consistent with the results of a single set of rescaled samples presented in Table 5.4. In other words, these diagnostics are similar to the results presented without performing the sensitivity analysis, indicating that the cross-validation presented here for block estimation is not sensitive to uncertainty due to sample rescaling in this study

The contingency table for the 200 conditional realizations is also listed in Table 5.8, with Cohen's κ coefficients calculated for each realization and summarized in Table 5.9. The results are also consistent with the cross-validation results for a single set of rescaled samples, and further confirms quantitatively by the ranges of possible κ coefficients that the classification precision are not significantly different among the three models, which implies that all three models could serve as precise estimation models for scientific exploration under a block target scale.

Table 5.8: Summarized contingency table for the block re-estimates of the three estimation models with respect to the 200 rescaled realizations using the Passaic River dataset. The range of values indicate the 90% lower and upper bounds of the diagnostic parameters.

		M-Scale		OK		CK	
		+	−	+	−	+	−
Target	+	3.5–4.6	3.5–5.5	1.4–2.0	5.5–7.9	4.0–5.5	2.9–4.6
	−	2.9–4.0	87.0–89.0	0.0–0.6	90.0–92.5	4.0–5.5	85.8–87.6

Table 5.9: Summarized Cohen's κ coefficient for the block re-estimates of the three estimation models with respect to the 200 rescaled realizations using the Passaic River dataset. The range of values indicate the 90% lower and upper bounds of the diagnostic parameters.

	M-Scale	OK	CK
κ	0.42–0.56	0.50–0.61	0.38–0.53

5.5 Summary

This chapter compares the estimation maps generated using a data set from a field site that has a different spatial structure to the artificial data set used in the previous chapter, with focus on the reproduction of both global and spatial variability. Cross-validation is performed to observe the impact of the estimation model on spatial variability. Very different model performances were observed between point estimation and block estimation for CK due to its sensitivity to the nugget effect.

Based on the result of cross-validation for point (punctual) estimates, the performance of the three models is consistent with the observations using the artificial dataset in Chapter 4, in which OK is found to reduce both global and spatial variability. CK reproduces global variability that includes both the actual variability and attribute-irrelevant artificially induced error. Covariograms with structural variances indicated that the M-Scale model best represents the spatial variability, without excess sensitivity to the nugget effect, and Cohen's κ coefficient confirms the same precision level for the M-Scale model and OK, both exceed CK.

For block estimations, however, CK performs at the same level as the M-Scale model for the reproduction of spatial variability. This is observed by the covariograms of the re-estimates against rescaled samples, where CK and M-Scale show similar reproduction of covariograms with respect to the covariogram of the rescaled samples. The standard deviation of the estimates also indicates the best reproduction of global variability for CK and the M-Scale model, and the results of Cohen's κ coefficient further indicates the same level of classification precision for CK relative to the M-Scale model, although both are slightly inferior to OK. In fact, with local variability of point attributes reasonably reduced by block averaging, the sensitivity of CK to nugget effect diminishes. The results imply that under specific cases such as a block target scale, CK may also serve as a suitable model to reproduce variability as well as a precise estimator for threshold exceedance. The target scale for the block estimation is chosen to demonstrate the impact of the nugget effect, which is the most significant difference in concept between applying point estimation and block estimation. The impact of the block size on reducing global/spatial variability may gradually change among block estimation performed at different target scales, while the results should not indicate different conclusions for the sensitivity of CK against the nugget effect, since the impact of nugget effect is considered negligible in most cases of block estimation in practice (as described in the introductory section of this chapter, see also p. 311 of Journel and Huijbregts 1978, pp.299–301 and p.326 of Isaaks and Srivastava 1989, and Aldworth and Cressie 1999).

The rescaling of samples for the cross-validation of block estimates, however, introduces uncertainty. Sensitivity analysis by the conditional realization is proposed in this study to provide a way to observe the uncertainty of the rescaled sample. It is observed by the result of the sensitivity analysis that the performance comparison using the single set of only the rescaled samples is validated.

In conclusion, using real samples collected from the Passaic River research area, the M-Scale model is the most robust model among the three for *point* spatial estimation that reproduces spatial variability and maintains the same level of classification precision with respect to OK. For *block* spatial estimation, however, CK may also be used as a suitable estimation model, in consequence of the reduction of local variability that CK is most sensitive to. As stated in Chapter 4, each estimation model may be suitable in certain particular cases depending on the objectives of estimation. Nevertheless, it is further observed in this chapter that cross-validation is essential prior to generating estimation maps for the comparison of performance among candidate models, not only to decide on the type of estimation model to be used, but also for the target spatial scale at which the estimation model is performed.

One of the characteristic differences of the Passaic River dataset in this chapter from the artificial dataset in the previous chapter consists of the preferential sampling scheme, which is commonly seen in applied geostatistics (see for example, Smith and Williams 1996a; Ouyang et al. 2002; Schnabel and Tietje 2003). This clustered sampling scheme results in the under-evaluation of the estimation variance/standard deviation for the M-Scale model in the case of the Passaic dataset, as observed in the lower coverage rates (proportion of actual values included in the confidence intervals evaluated) using the M-Scale model. Cross-validation is required, consequently, to observe the applicability of the M-Scale model if the estimation variance/standard deviation is a crucial component in the process of decision making, particularly when a preferential sampling scheme is used. Further development needs to be conducted to improve the quantification of the estimation variance/standard deviation in order to achieve a better coverage rate.

It should be noticed, however, that the result of cross-validation reflects only the model performance at the sample locations. The performance may or may not be

reflected in the cross-validation results depending on the richness/sparseness of the sample points, among other factors that may impact estimation results. In addition, it is not clear how the estimation variance, in other words the uncertainty of the estimates, are useful for the decision making of the contaminant management. Approaches will be described in the next chapter for use of estimation variance, and how the estimation error will affect its performance. A sparse dataset with samples collected from different scales in the Anacostia River study will be used as an example, which also shows the applicability of using cross-validation approaches on a study with spatially sparse samples.

CHAPTER 6

Application to Sparse Data Set and the Effect of Local Uncertainty

Spatial estimation is a procedure that uses the sample set to represent the attribute of interest, generating an estimate at each estimation location, and an uncertainty parameter such as estimation variance corresponding to this estimate. When a spatial estimation model is applied to generate estimation maps, the uncertainty of the estimates would enable the decision process to take into account a range of possible outcomes with respect to their likelihood of occurrence, so that the most cost-effective remedial solution can be chosen (p.44 of Barabás 2003). Moreover, as is described in Goovaerts (1998), "ignoring this uncertainty may, for example, lead one to declare safe a contaminated location on the basis of a wrong estimate of pollutant concentration which is slightly below the regulatory threshold."

As described in Chapter 2, the estimation procedure uses variability at/between sample locations (sample-to-sample variance/covariance) and variability between sample and estimation locations (sample-to-estimate covariance) to represent possible

deviation of actual values from the estimate at the estimation location (estimation variance). The conditional cumulative distribution function (ccdf) of the estimates can subsequently be evaluated by adding the possible deviations to the estimated value (see Chapter 2 for detailed reviews with illustrations). In an ideal case where samples are perfectly measured (without artificially induced error), the ccdf evaluated using variability at/between sample locations and between sample and estimation locations reflect the exact uncertainty (variability/range) of actual values. When measurements involve artificially induced error, however, the variability at sample locations are increased by the error variance of this artificial signal. Consequently, the estimation uncertainty, as a function of variability at/between sample locations and between sample and estimation locations, also becomes larger because the measurements that involve artificially induced errors are usually incorporated as part of the attribute values when parameters of variability (such as the variance) of the artificial error is missing in the statistical analysis. If the artificially induced error can be evaluated, however, this part of variance can be subtracted from the the function that evaluate the uncertainty of the estimates. Variability of the actual values relative to the estimates, consequently, can be correctly reflected without being overestimated by the variability of artificial errors. In other words, the estimation uncertainty can be reduced if the variability of the artificially induced error could be characterized.

Characterizing the variability of the artificially induced error, consequently, plays an important role in facilitating decision making. This unwanted artificial error can be quantified in geostatistics and is included in a parameter of the variogram called the *nugget effect* which characterizes the total of variability contributed by the artificial error that comes from the sampling process, and the micro-scale variability of the attribute at scales smaller than the measurement scale. Little research has been done, however, regarding how the nugget effect could be apportioned between the artificial error and micro-scale variability based on scientifically measured data, to

improve upon the uncertainty of spatial estimates.

The nugget effect also becomes important when a model like CK is used as a candidate for decision support models. As indicated in Chapter 5 by the comparison of point estimates to block estimates, CK estimates are strongly sensitive to the nugget effect, which may or may not be part of the real attribute value. On the contrary, as mentioned in Schnabel and Tietje (2003), for OK and other conventional kriging approaches the sensitivity to nugget effect is not significant. An alternative for the CK approach to become applicable for estimation is the reduction of local variability, because the variance constraint in CK tends to falsely reproduce local variability of the population by the spatial variability of its estimates, as illustrated (Fig. 4.24) and explained in Chapter 4. Without information of the artificially induced error, however, removing the nugget effect for CK means that the micro-scale variability corresponding to actual attributes is not considered at all, which may result in the underestimation of the estimation error at the sample locations. To evaluate the estimation error, and in some cases to reduce sensitivity to the nugget effect, the quantification of the artificially induced error becomes an important consideration for estimation tools in support of on-site remedial decisions. Although studies have suggested to detect the existence of a true nugget effect by observing the cross-variogram among different attributes (Lin et al., 2006; Holmes et al., 2005), this nugget effect may be a result of unwanted random signals associated with laboratory measurements of similar cause, e.g., unstable measurements due to the change in lab temperature. The "true" micro-scale variability observed in the cross-variograms, consequently, may not correspond to the true micro-scale variability for the attribute of interest.

An alternative for quantifying the artificially induced error is proposed in this study using measurements taken at site-scale and micro-scale, by which the amount

of micro-scale variability evaluated for the site-scale attribute is not subject to the unwanted random signals. In addition, in support of the conclusion made in Chapter 5 for the sensitivity of CK to the nugget effect, the point estimates before and after nugget effect adjustment (filtering of artificially induced error) are also compared.

What is more important about uncertainty reduction for an on-site remediation project, however, is its impact on the remedial decision making (Goovaerts, 1997a). Consequently, in addition to quantifying the amount of reducible uncertainty, the likelihood of exceedance for certain thresholds should also be examined to assess the impact of this uncertainty reduction on remedial decisions. Although numerous methods have been applied to produce uncertainty measures corresponding to their estimates (see for example, Maravelias et al. 1996; Collins and Woodcock 1999; Curriero et al. 2002; Schnabel and Tietje 2003), little has been done to evaluate the likelihood of exceedance for certain regulatory thresholds. These uncertainty measures, consequently, offer only the information regarding where future samples should be taken to improve the precision of the estimates (see Smith and Williams 1996a, for example). The change in classification maps associated with uncertainty reduction, moreover, is never expressed in the estimation approaches that offer uncertainty measures, making it hard for project managers to understand the benefit of uncertainty reduction on cleanup actions and costs.

The impact of uncertainty reduction is expected to be more apparent for the data in this chapter due to the small sample size, since more uncertainty is involved in estimates based on small sample sizes, and the weight attributed to each datum generally increases when the sample size decreases. The consequence of the increased impact for each datum is two-fold: (i) any artificially induced error or unwanted noise involved in the datum will increase its impact. (ii) with one datum removed, a large proportion of information contributed by the datum is also removed (which may

impact the applicability of cross-validation – see later in this chapter). Regardless of the large uncertainty encountered in small sample sizes, however, the origins of this uncertainty needs to be defined in order to clarify its impact. To facilitate discussions in the following sections, consequently, it is necessary to describe the components of wanted/unwanted information in the attribute variable and in the data set.

6.1 Components of Uncertainty

Sampling plans for site characterization and remedial decision-making are informed by the conceptual site model (CSM) that details not only the physical description of the site, but also the main processes thought to play a role in the fate and transport of contaminants (American Society for Testing and Materials, 1995). The inputs to these models tend to be dependent on both deterministic and random analytical uncertainties (Figure 6.1) that need to be managed to improve site management options. As the figure describes, analytical uncertainty is made up of errors associated with the input data (measurements, parameter choice and data aggregation), the physical models (formulation and application) used to describe the site and stochastic variability. Among the different uncertainties, this study focuses on the one termed "stochastic variability" in Fig. 6.1, a component which is most impacted by geostatistical models, and takes into account the artificially errors that is categorized under input error. The stochastic uncertainty itself is composed of uncertainties associated with the different spatial scales where measurements are taken, as shown in Fig. 6.2. The random error involved may originate from the limit of the instrument, the error due to sampling disturbance, or sample transportation/preservation (Goovaerts, 1997b, p.102). These errors are usually not evaluated and are incorporated as meaningful data in geostatistical applications (Matheron, 1971). The impact

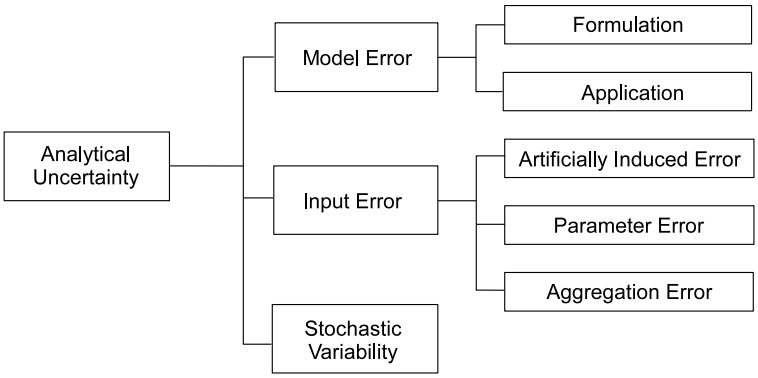

Figure 6.1: Conceptual sketch of sources of uncertainties.

of including this "background" variability (as shown in Fig. 6.2 C) as real data results in an increase of the estimation variance, leading to larger likelihood of misclassification for the true attribute value. In other words, the artificially induced error may increase the risk for false positive (FP) and false negative (FN) designations, thus impacting the results of site characterization. On the other hand, if the nugget effect is assumed to comprise solely the artificially induced error, the estimate may lead to inappropriate designations as well, because localized real variability (as shown in Fig. 6.2 B) is ignored in the measurement.

6.2 Estimation Uncertainty for Remedial Decision

In previous chapters, the delineation of the contaminant is characterized by the classification of exceedance/non-exceedance for a certain estimate. The uncertainty of the estimate is usually evaluated as the expected value of the squared residual, termed "estimation variance," between the true and estimated values. This quantitative uncertainty is useful when the remediation objective offers a level of flexibility to optimize between the risk of underestimating the extent of contamination, and

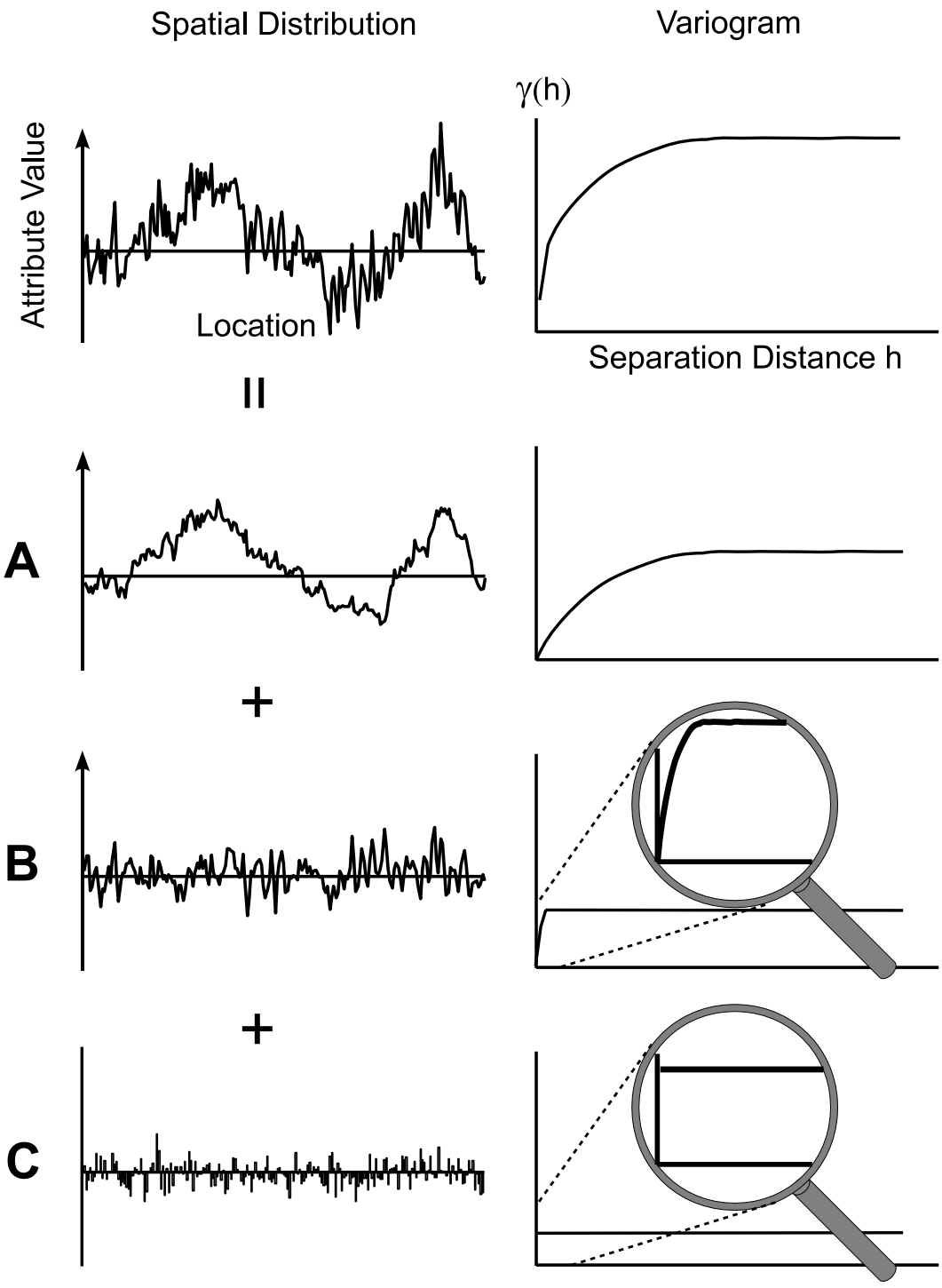

Figure 6.2: Components of a realization (left) and the corresponding variograms (right). The added components are: (A) site-scale variability, (B) micro-scale variability and (C) random error.

the expense for cleaning up unnecessary areas (Ramsey et al., 2002). In the current application, the classification of exceedance will be applied not to delineate contaminant levels, but to inform the spatially distributed likelihood of microbial activity (gas production) impacting the effectiveness of capping strategies (via physical destabilization of the caps). By using spatially measured values of microbial abundance and microbial respiratory competence (see section 6.5.1) as proxy parameters for gas production potential, geostatistical tools will be applied to selected cut-off values to probe where excessive gas production might become a problem for the remedial objective (long term in situ sediment capping).

6.2.1 Conditional Distribution, Confidence Interval and the Likelihood of Exceedance

When stochastic spatial estimation is performed, the estimation uncertainty is usually considered as the *conditional distribution* describing the probability/likelihood distribution of the parameter values at a given estimation location. This is usually represented by a cumulative distribution function called *conditional cumulative distribution function*, or *ccdf* (Istok and Rautman, 1996; Schweizer and Kronholm, 2007). A confidence interval can consequently be evaluated to indicate the possible bounds of the real attribute value under a certain confidence level that represents the proportion of times when the bounds correctly contains the real value. In other words, the confidence level is a model-dependent value that indicates the likelihood of a population value being inside of the confidence interval and can be used to observe the uncertainty of estimates at each estimation location to adjust future sampling plans.

For remedial decisions, however, the confidence *interval* offers no support for re-

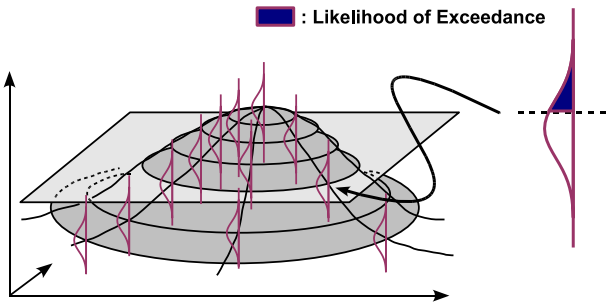

Figure 6.3: Conceptual sketch for likelihood of exceedance. On the left is the estimation surface with uncertainty and exceedance threshold as also illustrated in Chapter 2. On the right is the illustration for the likelihood of exceedance for one of the estimation location.

medial decision because it does not indicate the possibility for the true value to exceed a certain regulatory threshold. It is by characterizing the ccdf that the likelihood of exceedance corresponding to a regulatory threshold can be evaluated, providing the possibility for attribute values at each location to exceed the threshold, which is useful in the risk-assessment models in a decision process (Barabás et al., 2001; Critto et al., 2005).

6.2.2 Remedial Decision Making on the Basis of Likelihood of Exceedance

When a threshold for the likelihood of exceedance is used for remedial decision, the corresponding quantile for the likelihood threshold deviates from the fixed expected value depending on the threshold value used, as shown in Fig.6.3. This type of decision making gives the flexibility to trade the level of confidence for certain remediation objective, e.g., for optimizing the misclassification rate as described in Barabás et al. (2001).

The balance has to be made, however, between the benefit of classification (risk-reduction or cost-saving) on the basis of the likelihood threshold, i.e., the "willingness

of the analyst to accept the risk of an incorrect (positive) decision (Istok and Rautman, 1996)." In other words, the benefit of interest for different stakeholders may conflict with the willingness of the analyst to accept the risk of an incorrect decision. For example, if a clean-up action for an exceedance threshold is based on an extreme 10% likelihood of exceedance in order to prevent the risk of residual contaminant levels, the contamination zone thus delineated will have a large area so that the benefit of risk-reduction is attained, while the corresponding cost for the large clean-up area will be high (the analyst will only be "10% willing to accept that the positive classification is wrong" because the classification extensively covers the locations even where the likelihood of exceedance is only 10%). On the other extreme, if the clean-up action is based on a 90% likelihood of exceedance, the remediation zone will be small and the benefit of cost-saving is attained, while the risk of having residual contaminants will be high because areas that are less than 90% likely to exceed the exceedance threshold are not considered contaminated (the analyst will be "90% willing to accept that the positive classification is wrong" because the classification does not cover locations where the likelihood of exceedance is lower than 90%). Note that the likelihood threshold can be designed for each exceedance threshold of the attribute. Barabás et al. (2001) for example, used likelihood thresholds of 54% and 52 % to achieve the minimum misclassification rate of exceedance/non-exceedance classification, for the exceedance thresholds of 10 and 25 ppt of the TCDD concentration in the Passaic River dataset (the same dataset presented in the last chapter), while in this case the likelihood thresholds are not extreme. Although the decision does not directly depend on the likelihood thresholds chosen for different alternatives in the subsequent risk-benefit analysis, an extreme design would be less favorable for stakeholders to reach a compromise.

Reducing estimation uncertainty derived from the random noise induced during the sampling process will provide greater accuracy in the estimation. The estimation

variance, after discarding the effect of the noise, will be smaller, which is expected to correspond to more precise estimation with the reduction of both the false positive and false negative rates. Research has addressed the *reduction* of artificially induced error directly in the *data* by physical approaches such as mixing samples (Su et al., 2006) or by taking average values of multiple samples (Starks, 1986), however the impact of the reduction of local variability on the changes in the classification map has not been reported. Unlike these approaches, however, the approach used in this chapter *quantifies* the amount of artificially induced error in the data, and reduces the impact of artificially induced error in the *estimates* by attributing only the structural and micro-scale variability in the estimate-to-estimate variance and sample-to-estimate covariances. Similar to the use of block estimation, reducing the impact of artificially induced error does not change the values of sample-to-sample covariances since the artificially induced error is inherent in the measurement. In addition, this chapter will also address the impact of reducing estimation uncertainty on the change in estimated area of exceedance. In the following sections, reducing the impact of artificially induced error on the estimates will be expressed as *adjusting* the nugget effect for convenience, since the artificially induced error is subtracted from the nugget effect in the sample-to-estimate and estimate-to-estimate covariances to represent the actual variability of the attribute, but not from the sample-to-sample covariances. To facilitate the understanding of possible ways for reducing impact from artificially induced error, a review for the nugget effect (that includes micro-scale variability and the artificially induced error) is provided in the following section.

6.3 Nugget Effect in Geostatistics

In geostatistics, observed local variance is termed as the "nugget effect," which conceptually consists of two components: the artificially induced error irrelevant to the true distribution of the attribute of interest, and the micro-scale variability that corresponds to the true local variability of the attribute values. The nugget effect has been described and treated in different ways in the scientific literature.

The source of the nugget effect is described in different studies. (Crist, 1998) observed a large nugget effect in the termite occurrence, and indicated its sources to result "from sampling error or spatial dependence ... at finer spatial scales than the sampling resolution." Schnabel and Tietje (2003) indicated that the nugget effect observed in the heavy metal concentration in soil "might be caused by a lack of data (sparse data sampling) or high local heterogeneity of the variable." Western et al. (2004), in a study on spatial relatedness of soil moisture, described the cause of the nugget effect, indicating that "several factors, such as sampling error and short scale variability, may cause sample values separated by extremely small distances to be quite dissimilar ... The vertical jump from the value of 0 at the origin to the value of the variogram at extremely small separation distances is called the nugget effect." Holmes et al. (2005) indicated that the nugget effect of the variograms for several soil nutrients "accounts for both micro-scale variability in soil properties and laboratory and other procedural errors," and "a variable with no spatial dependence would have a variogram that is pure nugget (a horizontal line), ..., the geostatistical approach yields the same results as those obtained through classical statistical analysis."

Although sources of nugget effect are described in the literature, few studies involve detecting/differentiating micro-scale variability from artificially induced error as components of the nugget effect. Mear et al. (2006) examined actual detailed sam-

ple measurements of fine-grained content in the superficial sediments around certain locations, stating that "no explanation could be given for this nugget effect (errors of analysis or the use of different analytical methods)." In other words, the entire variability is contributed by the micro-scale variability in their study, which "underlines the existence of morphological phenomena (bathymetry and sedimentology) with dimensions lower than the average sampling distance." Lin et al. (2006) studied biogeochemical measurements in groundwater, and compared the nugget effect of the cross variogram with that of direct variograms, indicating that "the nugget effect on the direct variogram results mainly from micro-scale variation," because "any errors associated with laboratory measurements or georeferencing are likely to be independent and cancel each other out, so the nugget in the cross-variogram better represents the true micro-scale variability."

How the components could be further processed for the benefit of estimation, however, is seldom discussed. Among the few available studies, the reduction of nugget effect was done by physically mixing local samples or by taking average values of multiple samples. For example, Su et al. (2006) took five soil samples on two diagonals at 15 cm depth using a soil auger (diameter: 5 cm) and bulked the samples to obtain a composite sample, and thus eliminate the micro-scale variability. In most cases, however, the interest for nugget-effect reduction was to reduce the component caused by artificially induced error. Characterizing artificially induced error in the nugget effect is not common in geostatistical research except for scientific exploration, mainly because the impact of nugget effect for conventional kriging is not obvious, as described by Schnabel and Tietje (2003). The nugget effect influences the estimation weights attributed to sample values "in such a way that more distant estimation points carry a slightly larger weight than they would do without a nugget effect (Schnabel and Tietje, 2003)." Moreover, the nugget effect is usually taken solely as the micro-scale variability in classical kriging approaches, assuming the estimate will reproduce

the sample value exactly without estimation error at the sample locations Matheron (1971). The rationale for the assumption of having purely micro-scale variability as the nugget effect was not explained.

6.4 The Analysis for the Impact of Nugget-effect Adjustment on Likelihood-Based Classification Maps

To reduce the impact of the artificially induced error, the artificially induced error should be separated to achieve the biggest benefit without having an over-confident risk assessment. The literature reviewed in the last section either characterized the nugget effect as simply micro-scale variability (Mear et al., 2006; Lin et al., 2006) or used the nugget effect as a parameter that has no correspondence in characterizing spatial dependence and thus make no further discussion for how the nugget effect could affect the precision of estimates or impact the decision making process (Crist, 1998; Schnabel and Tietje, 2003; Western et al., 2004; Holmes et al., 2005). In this chapter, however, the impact of nugget effect will be studied, with an approach proposed for distinguishing the different components.

The analysis to separate the artificially induced error from micro-scale variability involves the use of two variograms from measurements taken at different spatial scales, respectively the site scale that describes the spatial variability in general as illustrated in the top of Fig. 6.2, and a scale far smaller than the site scale for the assessment of the micro-scale variability as illustrated in Fig. 6.2 (B). When characterized, the micro-scale variability will be subtracted from the nugget effect of the site-scale variogram to inform the amount of artificially induced error.

With artificially induced error subtracted from the site-scale variogram, the structural variogram (Ouyang et al., 2003) is obtained. The impact of nugget-effect adjustment is consequently evaluated by generating likelihood-based classification maps using variograms before and after the adjustment of the nugget effect where (i) the entire nugget effect is attributed to micro-scale variability vs. (ii) where a portion of the nugget effect is attributed to artificially induced error. The analysis follows the steps below:

1. Fit variogram models using datasets at both the site scale and the micro-scale.
2. Evaluate the micro-scale variability using the lab-scale variogram. Calculate the artificially induced error by subtracting the micro-scale variability from the full nugget effect.
3. With the site scale variogram including *and* excluding the artificially induced error, generate the likelihood maps of exceedance for the threshold of interest before and after adjustment of the nugget effect.
4. For various likelihood thresholds, quantify the impact of nugget-effect adjustment by calculating the difference in the estimated area of exceedance evaluated in step 3.

It should be noted for step 2, however, that the micro-scale variability also depends on the practical definition of "local" samples, as illustrated in Fig. 6.4. No micro-scale variability is smoothed out in the point samples (Fig. 6.4 A). When only a single block sample is collected locally for the site-scale at each sample location, the micro-scale variability is characterized by the dispersion variance and covariance within the volume of the site-scale sampler (Fig. 6.4 B). When multiple repetitions of block samples are collocated at each sampling location and values of the repetitions averaged

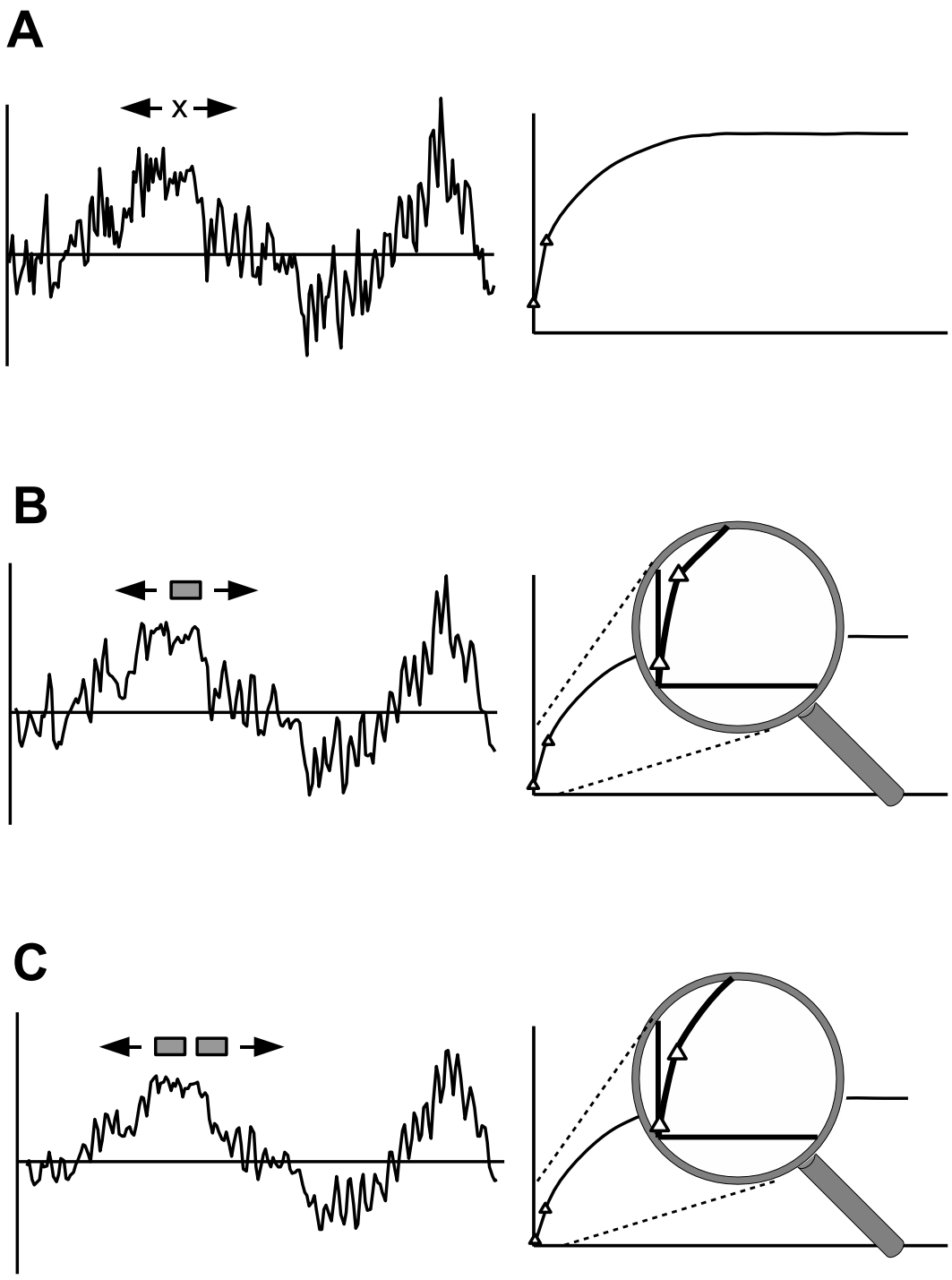

Figure 6.4: Conceptual sketch for the effect of sampling volume/configuration on variograms. The figure demonstrates exhaustive samples for illustration of the smoothing of variability. (A) point samples, (B) single-block samples (C) collocated-block samples (average values of two single-block samples). The segment in the variograms with triangular marks at both ends indicate the range of micro-scale variability. Note the smoothed out variability from (A) through (C) correspond to both smaller micro-scale variability and smaller artificially induced error.

as one sample value, the micro-scale variability in the variogram should be evaluated by the average of block-semivariances between all possible pairs of repetitions of the block variograms (Barnes, 1991) in which the block size is defined by the volume of the site-scale sampler (Fig. 6.4 C). The variograms for single-block samples can be evaluated from the point/punctual variogram using Equation (3.21), and the nugget effect of the variogram for the collocated block samples, following Barnes (1991), can be derived from the variogram for single-block samples using Equation (6.1),

$$C_0 = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \gamma_V(|x_i - x_j|) \tag{6.1}$$

where C_0 is the nugget effect for the collocated block samples, and the right-hand side of the equation refers to the variogram $\gamma_V(|x_i - x_j|)$ for blocks of size V between locations x_i and x_j, averaged over all possible location pairs within the configuration of the collocated sample. In fact, this nugget effect C_0 characterizes the micro-scale variability of the collocated block samples since the block averaging substantially reduces the artificially induced error of the punctual variogram.

6.5 The Anacostia River Dataset

This study uses a set of measurements taken at two different scales to evaluate artificially induced error in the nugget effect: (i) a benchtop laboratory experiment aimed at testing the impact of sediment capping strategies on contaminant effluxes from the Anacostia River (Washington, D.C.), and (ii) measurements collected during a field sampling campaign three months after the caps were installed at a pilot scale.

The Anacostia River is a freshwater tidal system draining an urban watershed encompassing 176 square miles in Maryland and the District of Columbia (Figure

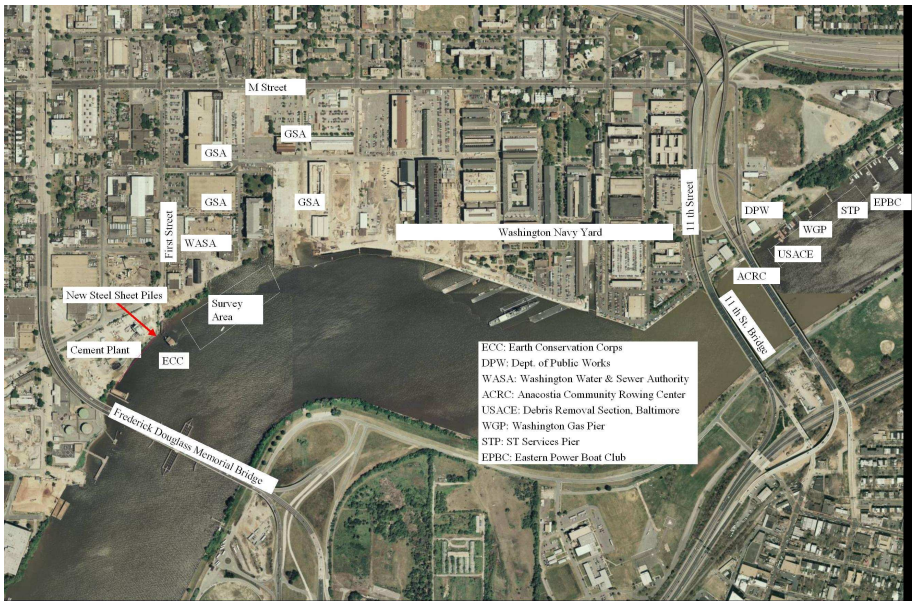

Figure 6.5: Bird's eye view of the Anacostia River study site.

6.5). The Anacostia River watershed is a subwatershed within the Potomac River Drainage Basin, which in turn empties into the Chesapeake Bay. Substantial deforestation and agricultural development, intense and continuous urbanization, industrial development, and significant loss of fringe wetlands and marshes have resulted in significant degradation of water quality in the Anacostia River. Elevated levels of hazardous substances, including PCBs, pesticides such as chlordane, lead and other heavy metals, and PAHs have been found in sediment throughout the 8.4-mile run of the Anacostia River (Syracuse Research Corporation and National Oceanic and Atmospheric Administration, 2000; Velinsky and Ashley, 2001). Hazardous substances such as PCBs have been found in fish at concentrations exceeding the Food and Drug Administration (FDA) Action Levels. The District of Columbia declared several fish consumption health advisories in the 1990s, with restrictions on consuming bottom-feeding species and game fish and a fish consumption ban for pregnant women and children.

In addition, the U.S. EPA identified the Anacostia River as one of the most con-

taminated rivers in the Chesapeake Bay watershed; it is one of three Regions of Concern recognized by the U.S. EPA's Chesapeake Bay Program as posing a significant risk to aquatic life from sediment contamination.

The U.S. Navy Space and Warfare Systems Center (SPAWAR) conducted the in-situ measurements of aqueous flux, or seepage, across the river-bottom interface to the overlying water column at six locations in the Anacostia River (Chadwick et al., 2001). These results indicated that the mean seepage rate ranges from a weak (and negative) measurement of -0.049 cm/d, to a moderate measurement of 1.1 cm/d. The results showed a weak to moderate tidal influence, with tidal amplitude ranging from -0.049 cm/d to 2.7 cm/d. The phase shift of the seepage meter readings was similar showing strongest discharge seepage just after high water, continuing a trend of smaller phase shifts for the down-river sites. Both the tidal and mean seepage signals at this site ranged from weak to moderate among other sites along the river. These results indicate that there is only a weakly active groundwater migration to the Anacostia River, in spite of the porous sediments observed in the deployment area.

In 2004, four capping strategies were implemented in the field (Fig. 6.6), including coke breeze, apatite, a clay polymer (Aquablok^{TM}), and sand. The objective of these cells was to test the following:

- AquaBlok: Evaluate tidal seepage control, and potential for uplift during tidal range
- Coke: Evaluate PAH sequestration/retardation, and placement strategies of laminated mats
- Apatite: Evaluate metal sequestration/retardation, and effectiveness of direct placement

- Sand: Current capping strategy

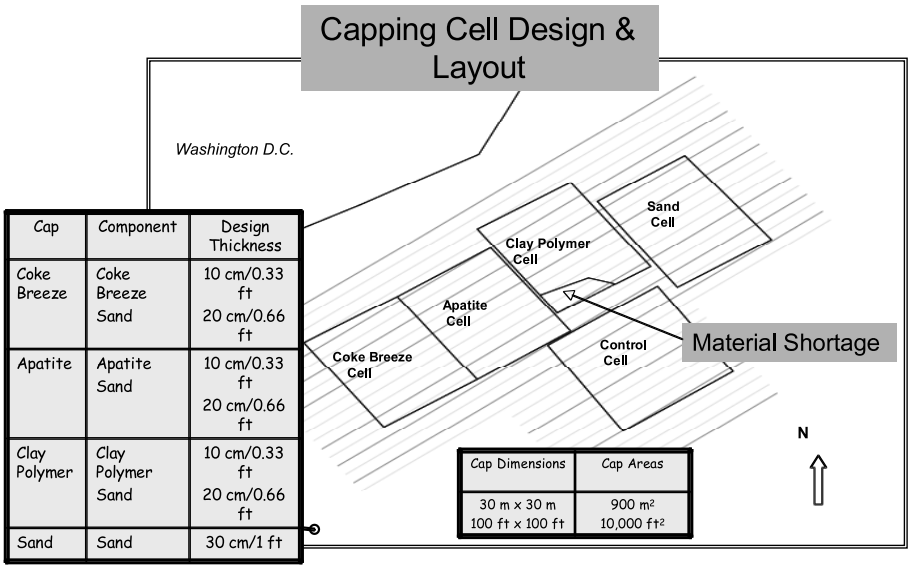

Figure 6.6: Illustration of the four capping strategies in the Anacostia River study site.

In support of the field efforts, the University of Michigan was contracted to assess the potential impact of microbial activity on in situ gas generation from sediment organic carbon, as this may impact the long term capping strategies. Two types of data were collected that form the core of the spatial analysis described in this chapter: microbial abundance, and microbial activity (as measured by respiratory competence, see section 6.5.1).

Samples at two scales were used in this task: (i) paired collocated core samples, separated around 1 foot (30 cm) apart with 2 inch (5 cm) diameter for each core, collected from the uncapped control, sand cap and Aquablok capped sediments (Fig. 6.7 A), and (ii) spatially distributed samples collected from the uncapped control used in a bench-top flux chamber experiments (Fig. 6.7 B). The field samples analyzed in this study represented both surficial (5 cm below the cap) and deeper segments from the core (see Table 6.1). The samples collected from the field and flux chamber were designed to capture both short and long distance covariances in microbial attributes.

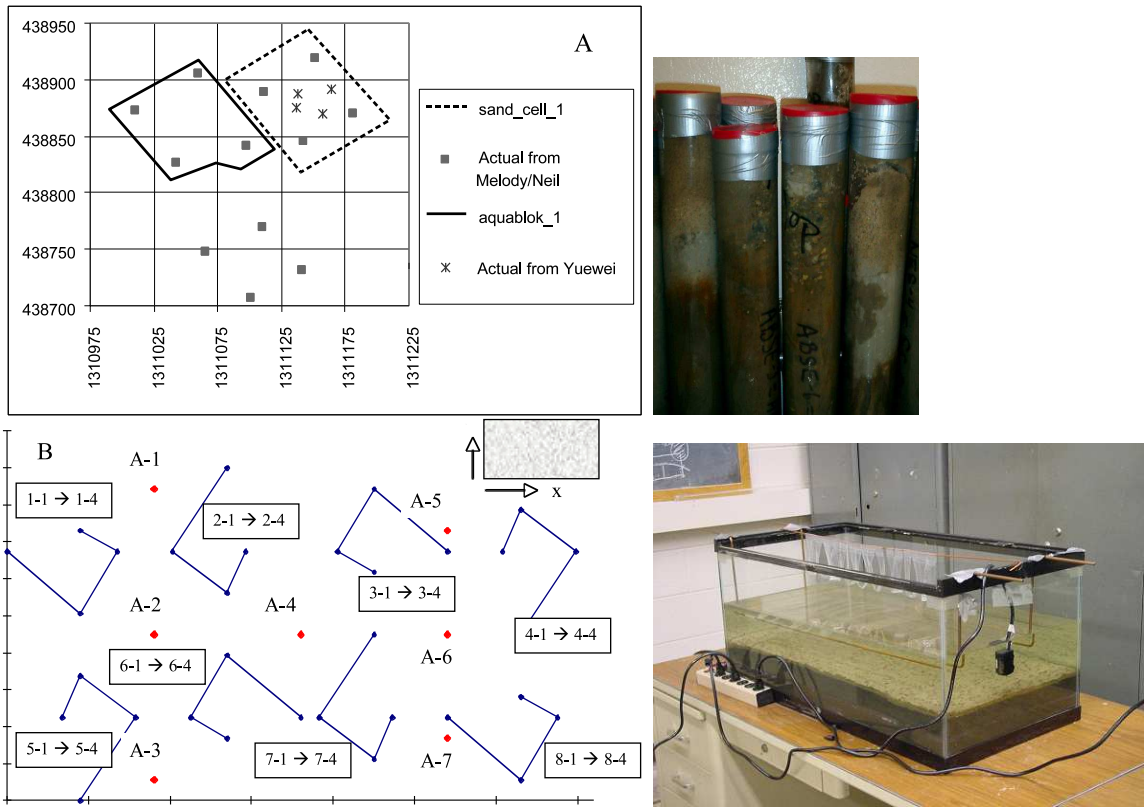

Figure 6.7: Sampling strategies for field (A) and flux chamber (B). The photograph of the field sample cores and the photograph of the flux chamber are shown on the right of the figures of sampling strategies.

Although the available spatial measurements from the Anacostia River are sparse, they provide important information on separating unwanted variance of error signals from actual local variability in a spatially scattered measurement set. By modeling variograms for the site-scale and tank-scale measurements, the nugget effect can be separated into components for micro-scale variability (local variability corresponding to real signal) and artificially induced error (local irrelevance independently included in each measurement), providing the rationale to reduce the estimation uncertainty. The benefit of this uncertainty reduction applies not only to the M-Scale model, but also to all other estimation models that generate estimation variances as a measure of the estimation uncertainty.

6.5.1 Microbial Abundance and Activity Measurement

Microorganisms were eluted, using an established protocol (Barkovskii and Adriaens, 1996, 1998) from Anacostia River sediments. Using dye experiments, sediment microorganisms were enumerated and respiratory competence measured (Gruden et al., 2003).

Field Scale Microbiology

The results from the dual stain microscopy data are shown in Fig. 6.8, illustrated for background samples, and those collected from the sand cap and Aquablok cap. Aquablok is a proprietary clay-encapsulated granular material that expands and seals once in contact with water. The difference in spatial distribution of microbial activity can be observed from Fig. 6.9, which shows an epifluorescent microscopic photograph for samples with 0.7 and 65% activity. The total microbial numbers range was 1.4×10^5

Table 6.1: Designation of samples collected from field pilots (see Fig. 6.7).

Core ID	Date	S(x)	N(y)	Water Depth (m)
No cap samples				
CNW-3	9/24/2004	1311065.98	438747.34	14.08
CNE-3	9/24/2004	1311110.85	438769.83	15
CSW-4	9/24/2004	1311100.60	438706.90	15
CSE-4	9/24/2004	1311140.97	438730.67	14.5
Aquablok^{TM} samples				
ABSE-6(4)	9/21/2004	1311097.64	438841.50	12.75
ABSE-6(5)	9/22/2004	1311097.64	438841.50	12.75
ABNE-6	9/22/2004	1311059.63	438905.34	7
AS_5	9/20/2004	1311010.53	438872.22	
Sand cap samples				
SES-5	9/23/2004	1311181.11	438869.50	9.75
NWS-5	9/23/2004	1311111.93	438888.49	8.25
NES-4	9/22/2004	1311151.06	438918.34	6.75
LT-3	10/13/2004	1311164.00	438891.00	

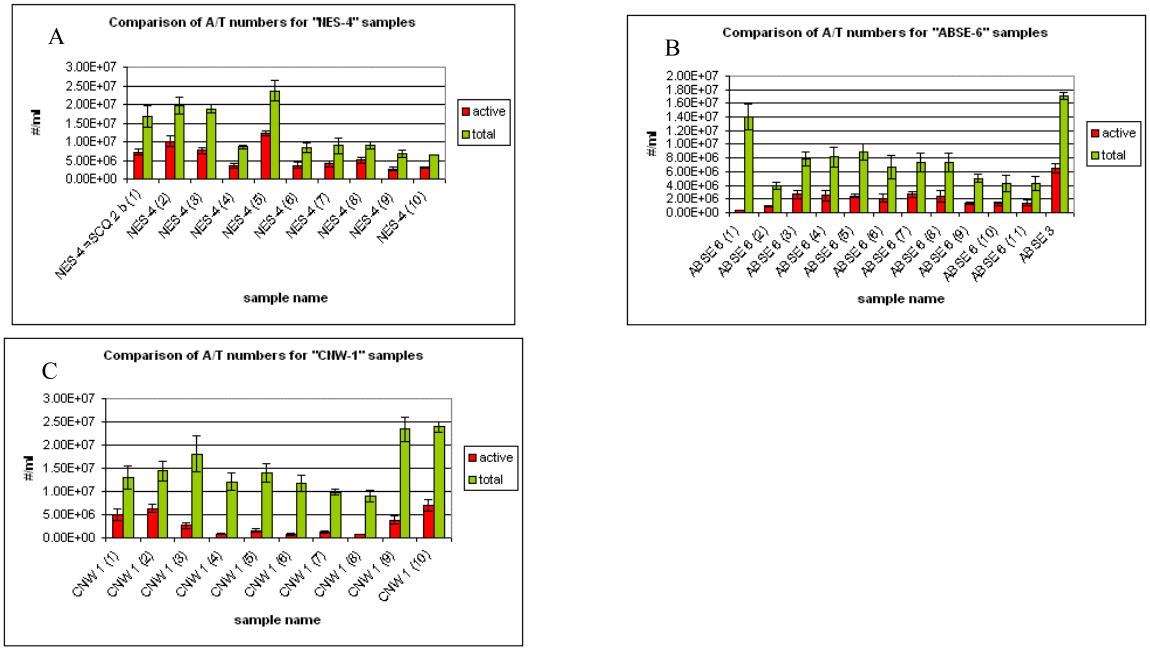

Figure 6.8: Total and active microorganisms in sand (A), Aquablok^{TM}(B) and uncapped (C) sediments.

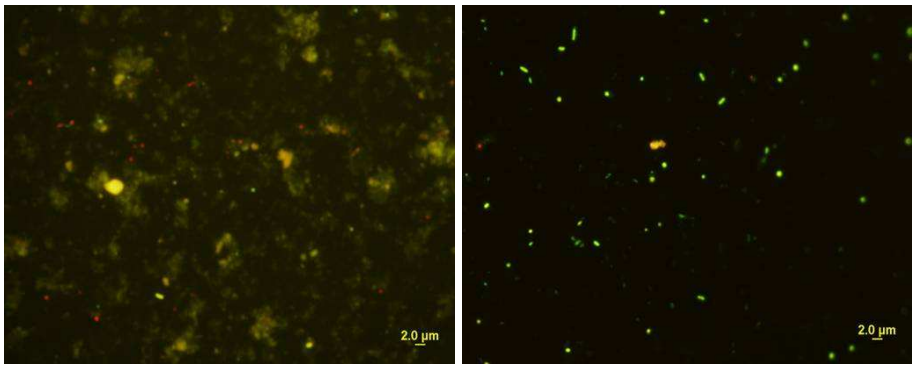

Figure 6.9: Microbial activity as measured using CTC for 65% (left) and 0.7% (right) active sediment samples.

to 2.8×10^7 per gram of sediment.

The generally high activity numbers are on par with anaerobic digesters and are well above those measured in other riverine and estuarine sediments, and indicate that there should be an ample supply of labile organic carbon to sustain microbial activity. The values for capped and uncapped sediments are shown in Fig. 6.8.

The following trends can be discerned: (i) capped sediments exhibit lower total microbial numbers than uncapped sediments, but a higher fraction of the organisms is active; (ii) the total bacterial count appears to peak at 15–20 cm under the cap or river bottom; and (iii) the active fraction of organisms is fairly constant with depth under caps, but decreases by up to 90% in uncapped sediments.

Flux Chamber Microbiology

The results from sediment testing for microbial abundance and metabolic competence in the unamended flux chamber are shown in Fig. 6.10. The data representing cores A1-A7 indicate that, despite the extensive sediment mixing, there is a great degree of spatial heterogeneity across the tank in active numbers of microorganisms, ranging from 0 to about 10% of total organisms. The total microbial abundance is

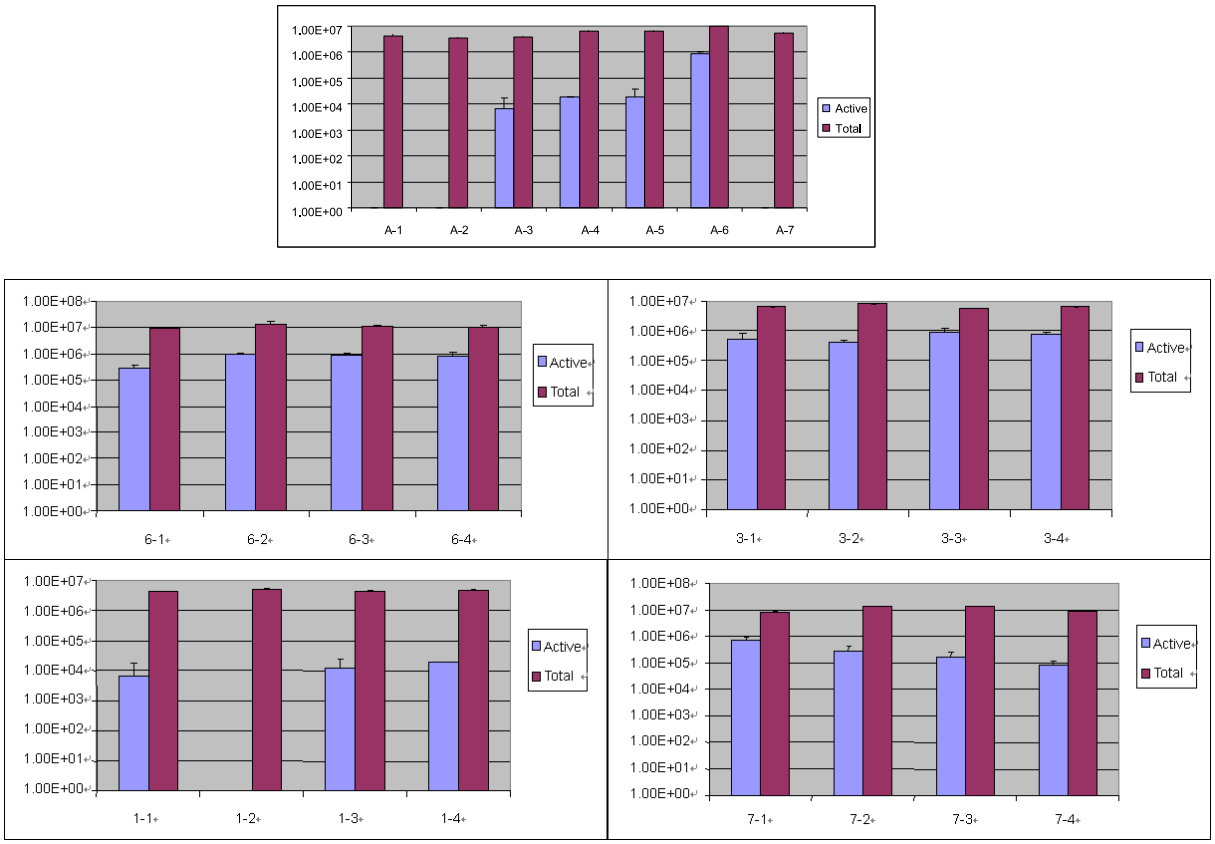

Figure 6.10: Results for tank sample abundance and activity.

similar to that observed in the field cores, but the fraction of respiratory competent organisms is much lower than in the cores (6–45%). This observation may be due to the extensive manipulation of the sediments prior to the laboratory test, including: the sediments were collected with a backhoe and deposited in 55 gal. drums; the sediments were exposed to aerobic conditions for mixing prior to deposition in the tank; and the sediments were exposed to an artificially-induced "ebullition" flux of air.

At the fine-scale spatial resolution (within sample clusters), much greater homogeneity was observed in active microorganisms, as exemplified for sample clusters 1.1–1.4, 3.1–3.4, 6.1–6.4 and 7.1–7.4 (for location, see Fig. 6.7). Again, the abundance of total microorganisms was on the order of 10^7 microorganisms/g, while active organisms ranged from $10^4 - 10^6$ per gram sediment, except for sample 1.2 (Fig. 6.10). This represents an active fraction of organisms on the order of 1–16% between these four sample clusters (with a tighter distribution within a cluster), which is lower than in the field cores, but similar to that observed at large spatial resolution. Interestingly, sampling clusters in the neighborhood of samples A-1 and A-7 (which showed no active microorganisms) indicate the presence of 1–10% active microorganisms, suggesting that even at the small scale, inhomogeneities exist.

The field-scale samples and micro-scale samples are summarized into maps with point attributes for the purposes of spatial estimation, as shown in Figs. 6.11 and 6.12.

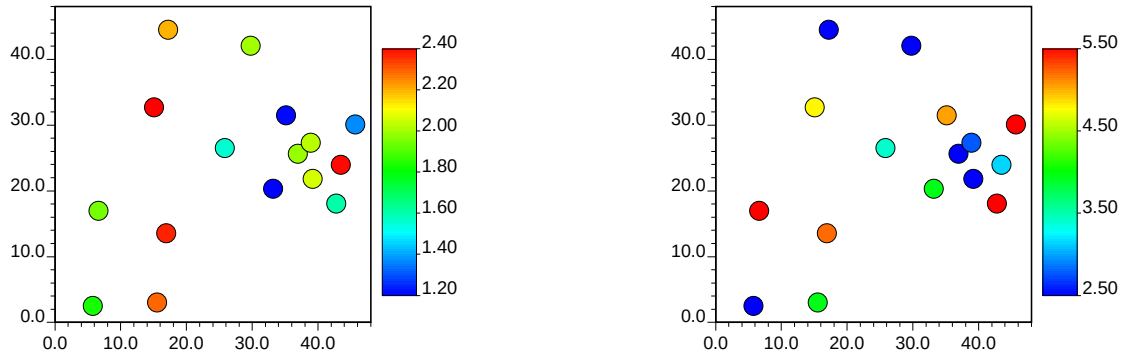

Figure 6.11: Site-scale sample locations with values indicated in color scales. Left: microbial abundance ($\times 10^7$ microorganisms/g) Right: microbial activity ($\times 10^6$ microorganisms/g). Units in distance: m.

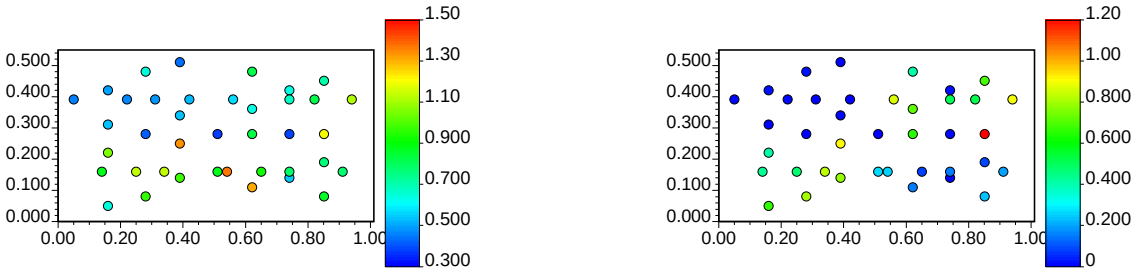

Figure 6.12: Micro-scale sample locations with values indicated in color scales. Left: microbial abundance ($\times 10^7$ microorganisms/g) Right: microbial activity ($\times 10^6$ microorganisms/g). Units in distance: m.

6.6 Analysis of the Nugget Effect for the Anacostia River, DC Dataset

In this study, and indeed in many laboratory-scale experiments, samples taken at the tank scale are implicitly assumed to be representative of the field scale, and thus the micro-scale variability present on-site. This assumption may not precisely reflect the reality, since the lab analysis involves sediments dredged up using a backhoe volume, homogenized and re-cultured in room temperatures inside of a confined flux chamber. The resulting micro-scale variability may be underestimated in this study, consequently, with artificially induced error overestimated. The uncertainty will thus be over-reduced, corresponding to a risky design with underevaluated estimation variance that covers less variability than what actually is between the true attribute value and the estimate. A lab-simulated micro-scale sample may not always under-represent the on-site micro-scale variability, e.g. by using an initial injection of contaminant in a flux chamber to represent the variability of a stablized plume on-site that corresponds to less spatial variability due to diffusive mass transfer. The uncertainty will thus be over-represented, corresponding to a conservative design with overevaluated estimation variance that covers more variability than reality. Using lab experiment to simulate the micro-scale reality on-site, consequently, needs more scrutiny regarding its reproducibility of variability. This study, however, provides an example to characterize micro-scale variability when the micro-scale samples are closely related to the micro-scale variability on-site, e.g., by taking samples in a microcosm installed on-site with negligible disturbance.

The micro-scale variability of the site-scale measurements, consequently, is evaluated using the variogram of the tank-scale measurements as part of the site-scale nugget effect, and the artificially induced error variance is assumed to represent the remainder of the observed nugget effect of the site-scale variogram. The resulting

variogram of the site-scale variability is used to generate estimates and estimation variance for the three estimation models.

Variograms for both the site-scale sample set and the micro-scale sample set are constructed as shown in Figs. 6.13 and 6.14, and are represented by the following equations:

Variograms for microbial abundance:

$$\text{site-scale:} \quad \gamma(h) = 0.176 + 0.244 \cdot \text{Exp}\left(\frac{h}{25.00^m}\right)$$

$$\text{micro-scale:} \quad \gamma(h) = 0.020 + 0.080 \cdot \text{Exp}\left(\frac{h}{0.38^m}\right)$$

Variograms for microbial activity:

$$\text{site-scale:} \quad \gamma(h) = 1.980 + 3.800 \cdot \text{Exp}\left(\frac{h}{9.00^m}\right)$$

$$\text{micro-scale:} \quad \gamma(h) = 0.020 + 0.150 \cdot \text{Exp}\left(\frac{h}{0.38^m}\right)$$

where Exp(·) denotes the exponential variogram model. The nugget effect for the site-scale variogram is calculated directly using the 1 foot-apart collocated core sample pairs and averaged over all sample locations. Although the variogram of the flux chamber abundance and activity samples also feature a nugget effect, the amount of this part relative to that of the core sample is found to be negligible in this analysis, particularly after the block averaging (see the conceptual sketch for the effect of block averaging in Fig. 6.4 B and C). Block variograms from the tank-scale microbial abundance and activity were subsequently evaluated for the analysis of micro-scale variability, as shown in Fig. 6.15. In this study, the micro-scale variability is defined as the block-to-block semivariance of the core sample pairs because the block-to-block semivariance equals the expected variance of the core sample pairs. For microbial abundance data, the value is 0.068 as calculated using Equation (6.1) and dispersion

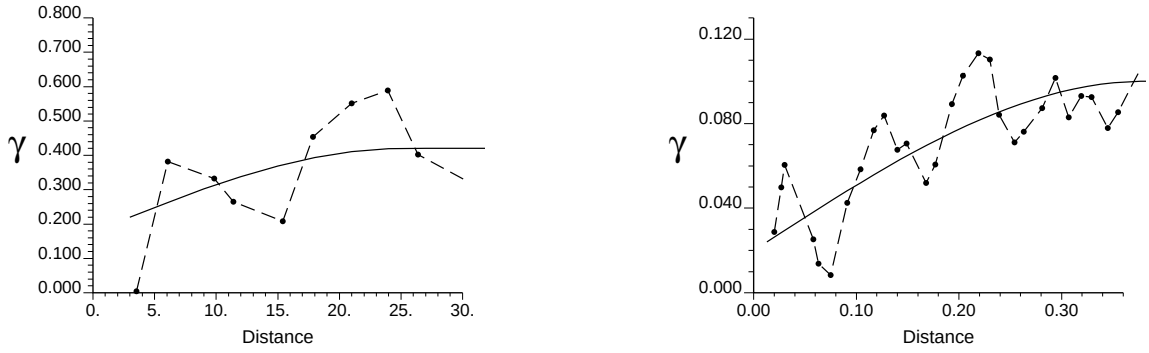

Figure 6.13: Variogram fitting for site- and micro- scale datasets of microbial abundance. Solid curves indicate the experimental variogram of the samples. Dash curves indicate the modeled variograms. Left: site-scale. Right: micro-scale. Unit in distance: m.

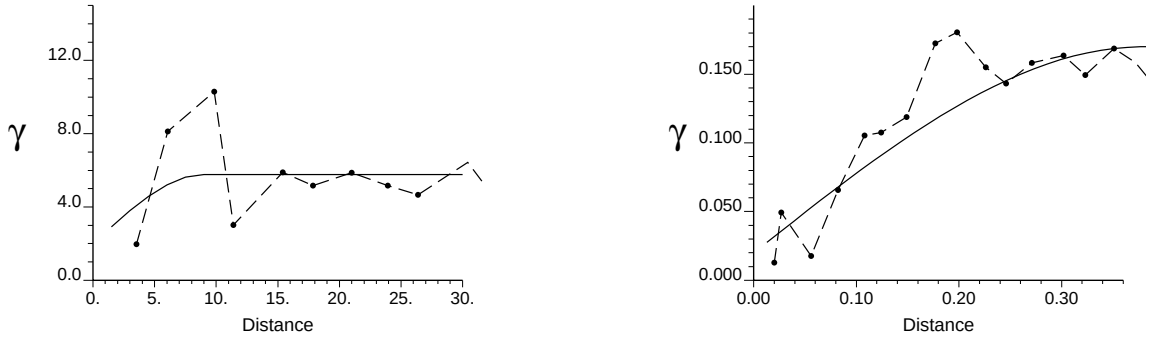

Figure 6.14: Variogram fitting for site- and micro- scale datasets of microbial activity. Solid curves indicate the experimental variogram of the samples. Dash curves indicate the modeled variograms. Left: site-scale. Left: site-scale. Right: micro-scale. Unit in distance: m.

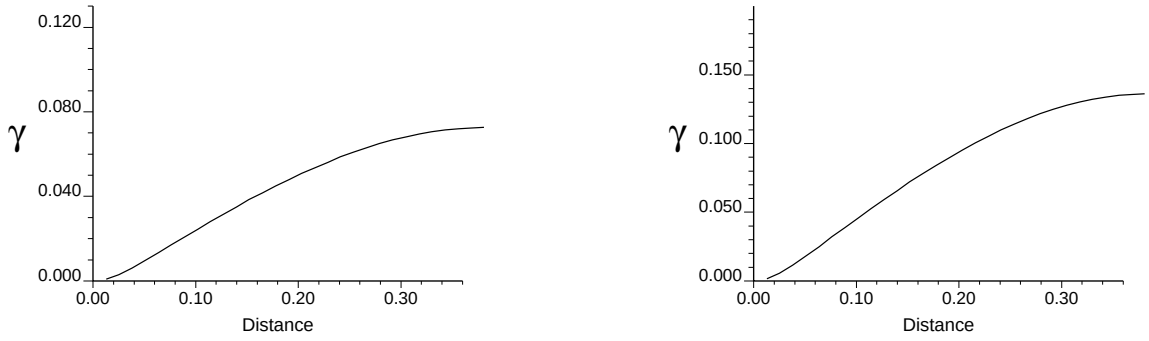

Figure 6.15: Block-to-block variogram by the micro-scale punctual variograms. Left: abundance Right: activity. Unit in distance for both figure: m.

covariance of the tank variogram model (Equation (3.21)), and for microbial activity using the same approach the value is 0.102. Subsequently, the artificially induced error is found by subtracting the amount of micro-scale variability from the value of the nugget effect. The artificially induced error is found to be $0.176 - 0.068 = 0.108$ for the microbial abundance, and $1.980 - 0.128 = 1.852$ for the microbial activity.

6.7 Impact of Nugget-Effect Adjustment on Model Performance

The M-Scale model, OK and CK are used as estimation models to observe the impact of nugget-effect adjustment. Six maps of three sets are generated for each estimation model, with two maps of each set representing the results for retaining and adjusting the nugget effect, respectively. The three sets of maps are respectively the estimation map, the estimation variance map, and the likelihood map for exceeding a certain attribute threshold, as shown in Figs. 6.16 through 6.21, where the black pixels of the figures for CK indicate singular solutions due to the lack of covariance between sample and estimation locations to reproduce variability.

Since the remedial option under investigation in the Anacostia River is in situ capping, this strategy is potentially impacted by, among others, microbial abundance and microbial respiratory activity (particularly methanogenic activity) which, as stated previously, serve as proxy parameters for microbial gas production. Since the cut-off values for abundance and activity at which gas production becomes disruptive to the in situ capping strategy have not been determined at this time, illustrative values are used to demonstrate the impact of the nugget effect on spatial variability. Target threshold values are selected as 2.2×10^7 (representing the 0.75 quantile of the pooled sample) for the total microbial abundance, and 4.5×10^6 (representing the 0.6 quantile of the pooled sample) for the microbial activity (respiratory competence) as examples, to demonstrate the generality of estimation characteristics for the three estimation models, in addition to the demonstration by using the 0.86 quantile (the mean value plus one standard deviation) as the threshold in Chapter 5.

As shown in the figures, the adjustment of the nugget effect generally has a minimal impact on the estimation maps for OK and the M-Scale model (comparing top graphs of Figs.6.16, 6.19, 6.17, and 6.20), while the impact is obvious on the estimation variance for all three models (comparing middle graphs of Figs.6.16 through 6.21). This result for OK and the M-Scale model is consistent with the observation by Schnabel and Tietje (2003) for the slight impact of the nugget effect on estimation maps. However, the impact of the nugget effect on the estimation map is larger in CK (comparing top graphs of Figs. 6.18 and 6.21), due to the constraint of estimate variance to reproduce variability. In all estimation models the estimation variance is lower with nugget effect adjusted, which is obvious since the estimation variance is mainly impacted by the variance of the estimate and the weighted mean of covariances between sample and estimate pairs. The likelihood of exceedance visually shows more contrast (comparing bottom graphs of Figs.6.16 through 6.21), with higher estimates to be more likely to exceed the threshold, and lower estimates to be less likely to

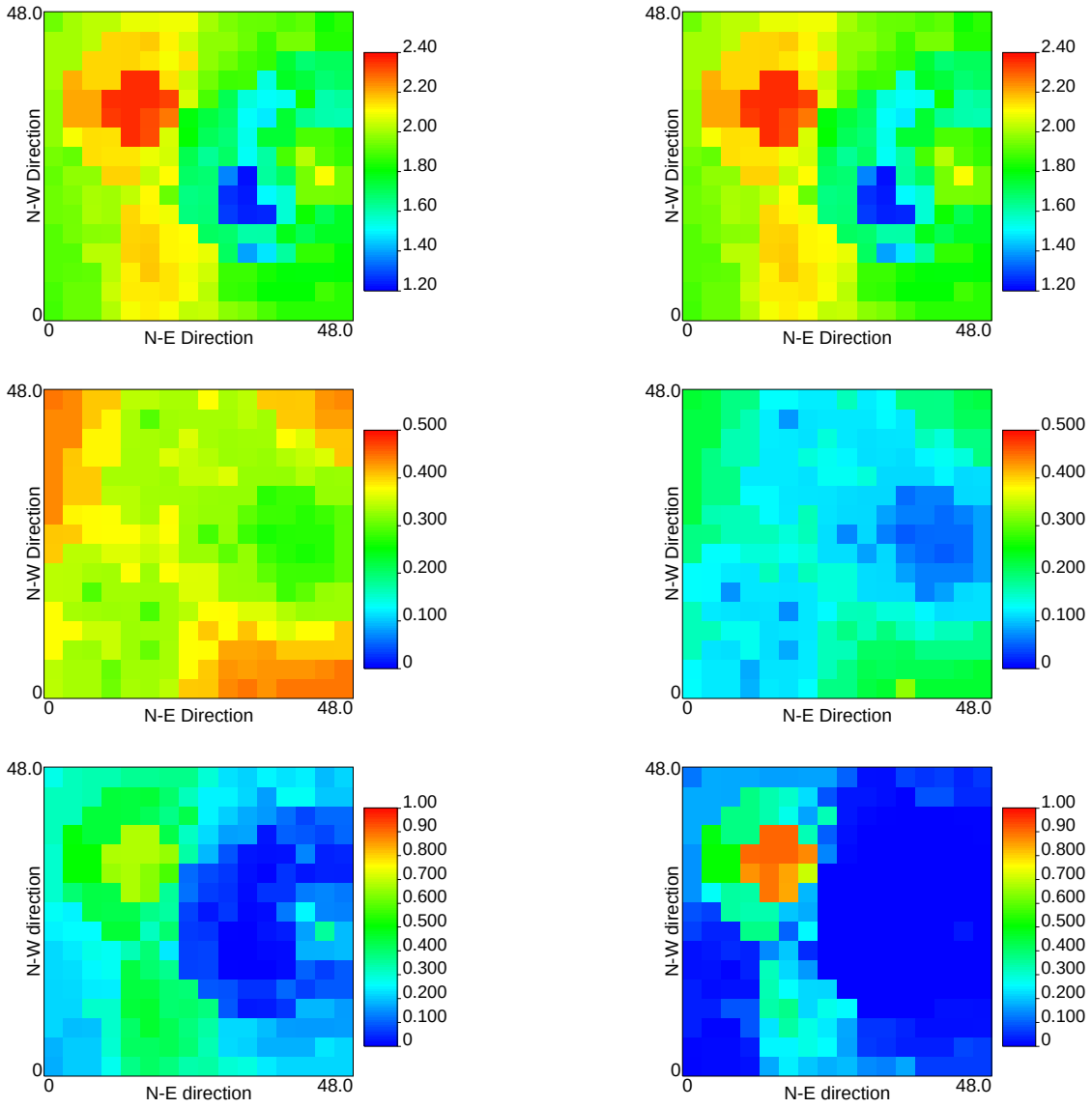

Figure 6.16: Microbial abundance estimation by the M-Scale model. Top to bottom: estimation map, estimation variance, likelihood of the estimate to exceed 2.2×10^7. Left: nugget effect retained. Right: nugget effect adjusted.

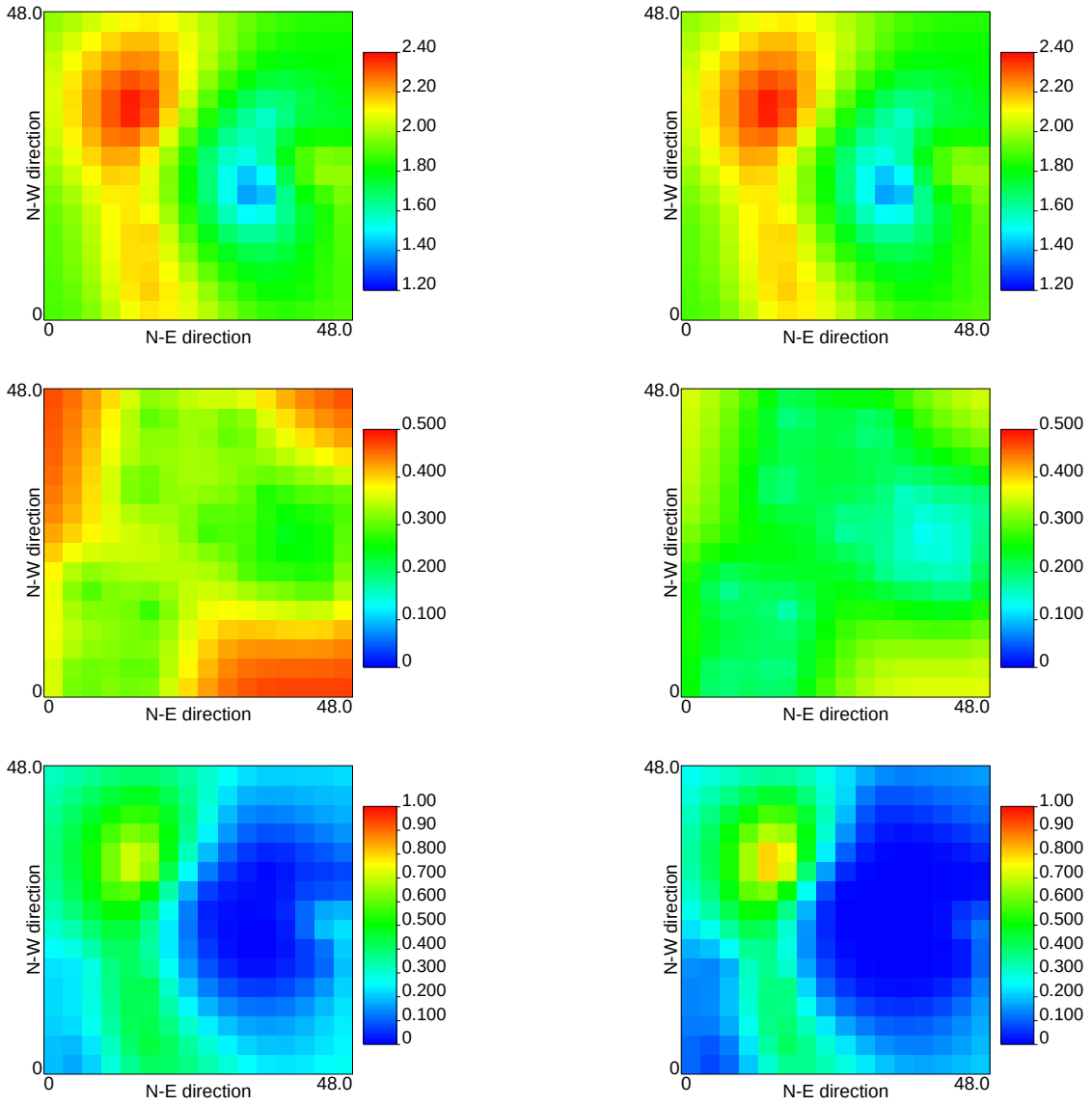

Figure 6.17: Microbial abundance estimation by OK. Top to bottom: estimation map, estimation variance, likelihood of the estimate to exceed 2.2×10^7. Left: nugget effect retained. Right: nugget effect adjusted.

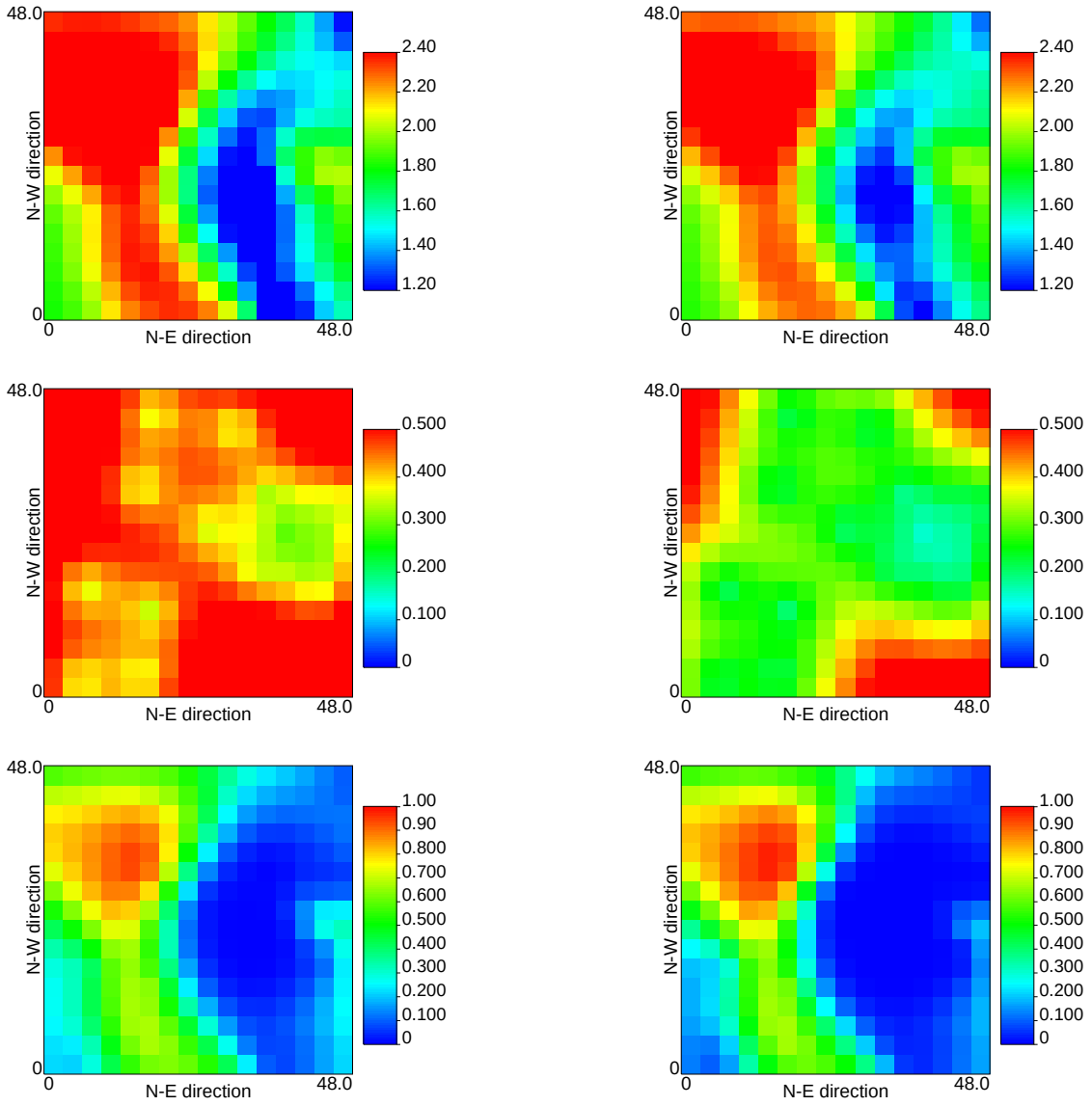

Figure 6.18: Microbial abundance estimation by CK. Top to bottom: estimation map, estimation variance, likelihood of the estimate to exceed 2.2×10^7. Left: nugget effect retained. Right: nugget effect adjusted.

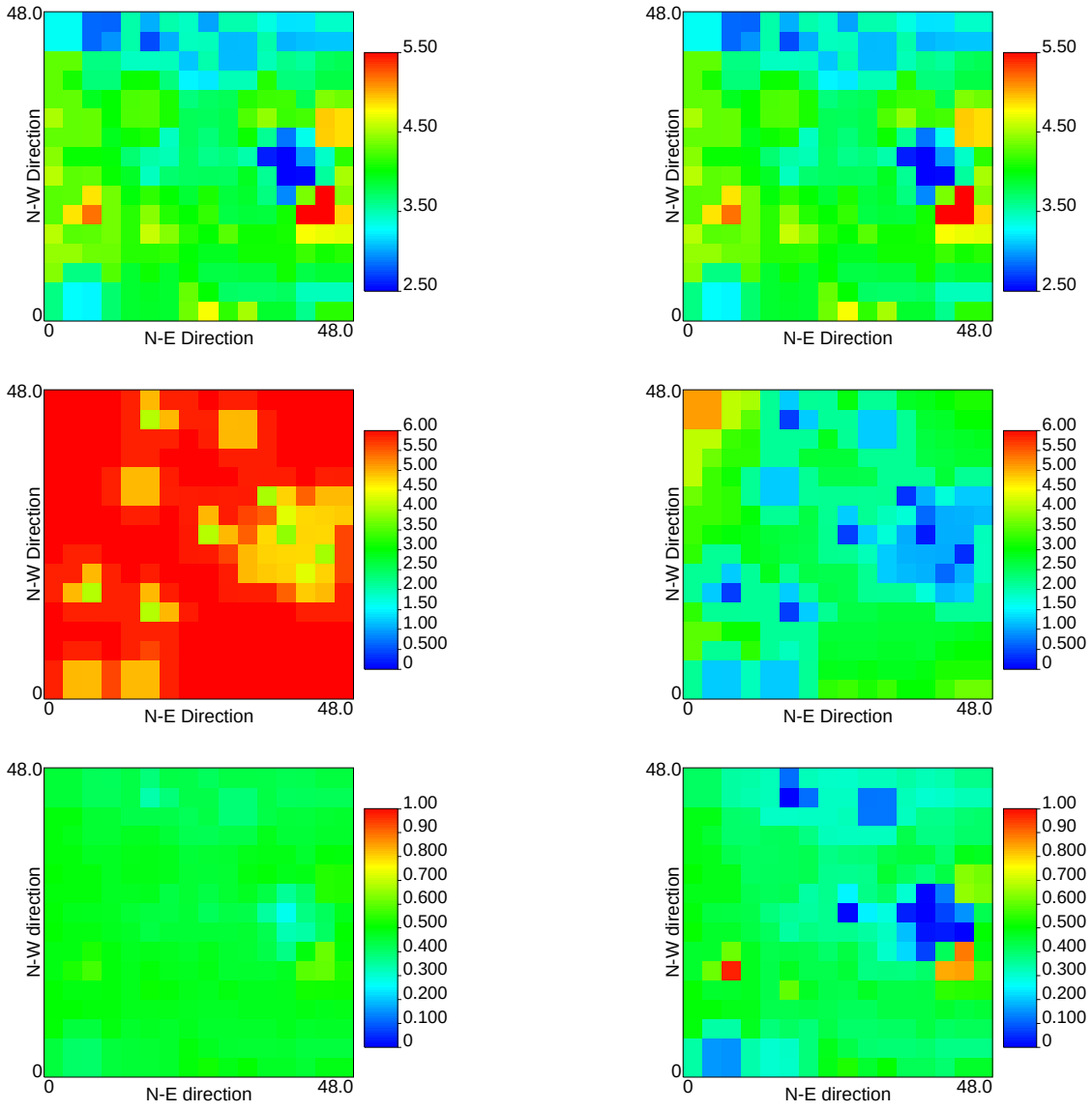

Figure 6.19: Microbial activity estimation by the M-Scale model. Top to bottom: estimation map, estimation variance, likelihood of the estimate to exceed 4.5×10^6. Left: nugget effect retained. Right: nugget effect adjusted.

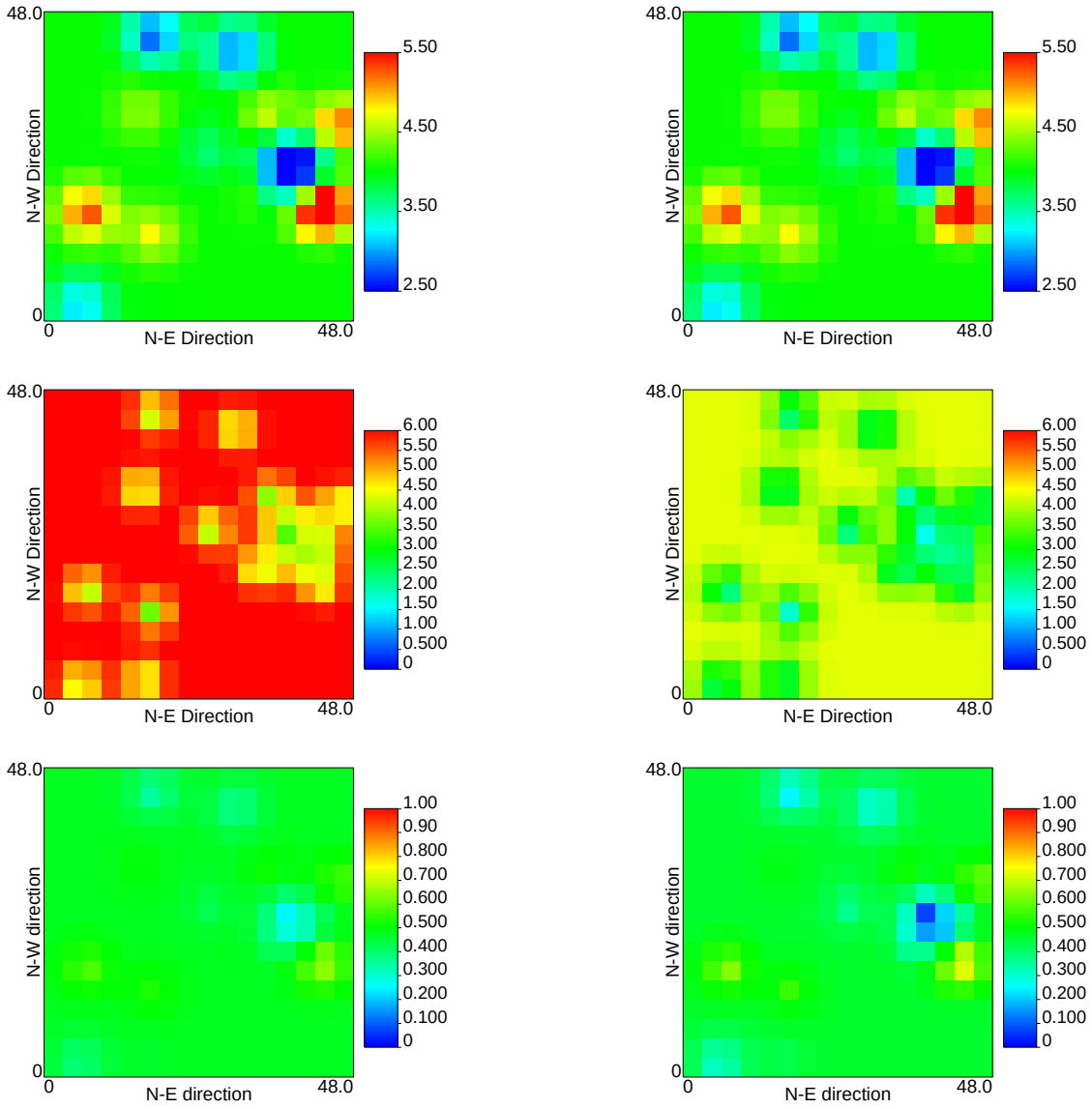

Figure 6.20: Microbial activity estimation by OK. Top to bottom: estimation map, estimation variance, likelihood of the estimate to exceed 4.5×10^6. Left: nugget effect retained. Right: nugget effect adjusted.

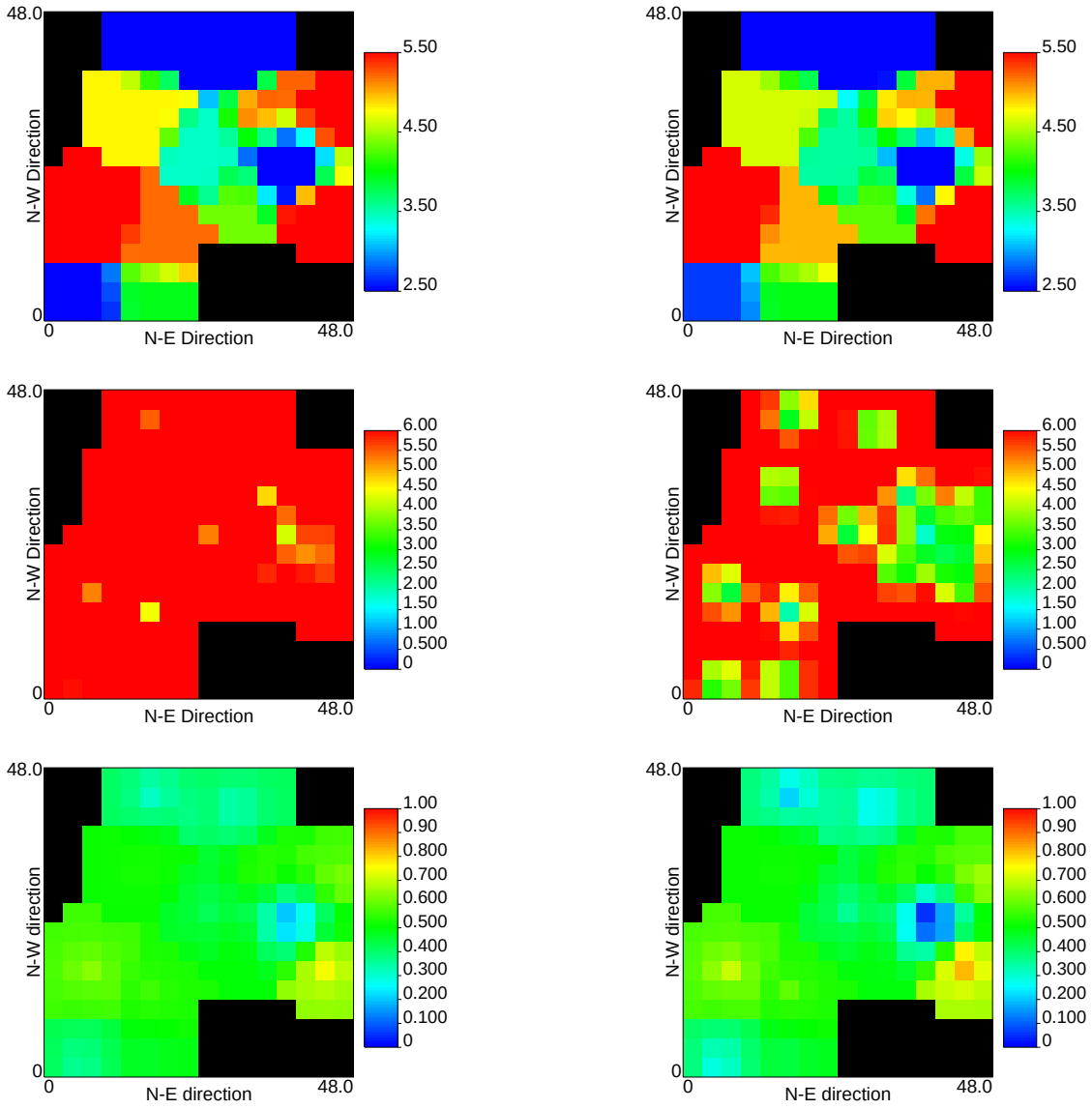

Figure 6.21: Microbial activity estimation by CK. Top to bottom: estimation map, estimation variance, likelihood of the estimate to exceed 4.5×10^6. Left: nugget effect retained. Right: nugget effect adjusted.

exceed the threshold. This result indirectly confirms the statements by Istok and Rautman (1996) for the likelihood of exceedance in that the likelihood of exceedance for higher estimates (non-exceedance for lower estimates), which represents the likelihood to fulfill the "willingness of analyst to accept the risk of an incorrect decision," is increased with the improvement of precision, meaning that the possibility for an incorrect decision is decreased.

In addition to the characteristics in the map of likelihood exceedance, the impact of nugget-effect adjustment is also examined for the map of likelihood-based classification, as an indication of improvement for the decision support of remedial applications. The probability of exceedance is illustrated for microbial abundance and microbial activity, two sediment attributes thought to have relevance to microbial gas production and thus stability of sediment caps or contaminant transport. For illustration purposes, microbial abundance of 2.2×10^7 and microbial activity of 4.5×10^6 were chosen as cut-off values for these attributes. The percentage areas for the delineated microbial endpoint (exceeding the assumed threshold of microbial endpoints) within the estimation domain are listed for microbial abundance and activity respectively in Tables 6.2 and 6.3. It is observed that to achieve a higher likelihood thresholds (high willingness to accept an incorrect decision, i.e., low-cost objective in exchange of high risk), a larger area of exceedance is delineated after nugget-effect adjustment, consequently the risk of non-cleanups is reduced. On the other hand, for a design with lower likelihood thresholds (low willingness to accept an incorrect decision, i.e., high cost objective in exchange of low risk), a smaller area of exceedance is delineated, consequently the expenses for unnecessary cleanups is saved. Therefore, by observing the range of percent areas under different likelihood thresholds, the table further indicates smaller differences between the ends of objectives (cost-saving/high threshold of likelihood vs. risk-reduction/low threshold of likelihood), an indication that a compromise between stakeholders becomes easier to reach.

Table 6.2: Percent area of high microbial abundance ($\geq 2.2 \times 10^7$) classified over the estimation domain under different likelihood threshold.

	M-Scale		OK		CK	
Likelihood	Nugget Effect		Nugget Effect		Nugget Effect	
Threshold	Retained	Adjusted	Retained	Adjusted	Retained	Adjusted
0.2	60.55	14.84	59.77	39.45	59.77	48.05
0.4	10.94	8.20	19.92	13.28	41.41	36.33
0.6	4.69	6.64	3.12	3.52	25.00	23.83
0.8	0.00	3.52	0.00	0.39	7.81	10.16

Table 6.3: Percent area of high microbial activity ($\geq 4.5 \times 10^6$) classified over the estimation domain under different likelihood thresholds.

	M-Scale		OK		CK	
Likelihood	Nugget Effect		Nugget Effect		Nugget Effect	
Threshold	Retained	Adjusted	Retained	Adjusted	Retained	Adjusted
0.2	100.00	90.62	100.00	98.83	76.56	75.00
0.4	92.58	49.61	95.31	89.45	61.33	54.30
0.6	0.00	4.30	0.39	1.56	6.25	10.16
0.8	0.00	1.56	0.00	0.00	0.00	0.39

However, a reduction of area is observed with the adjustment of nugget effect for the classification of microbial abundance by CK estimation using a 0.6 likelihood threshold. This could be explained by the fact that CK estimates are sensitive against the nugget effect, consequently the nugget effect impacts the ccdf evaluated by CK for the likelihood-based classifications not only on the variance of the ccdf (reduces the size of distribution), but also on the expected value of the ccdf (shifts the position of the distribution). Because CK falsely reproduce local variability of the samples by spatial variability of the estimates (see Fig. 4.24 for illustration), the benefit of nugget-effect reduction is not only the reduction for estimation variance, but also the adjustment of expectation from falsely reproduced variability.

6.8 Cross-Validation and Diagnostic Parameters for Reproduction of Global and Spatial Variability

With the sparse dataset collected in the Anacostia River study area, the nugget-effect adjustment was shown to have substantial impact on the estimation variance, likelihood of exceedance, and delineation of the extent of contamination. This large impact observed, in fact, indicates the sparseness of the dataset is usually associated with a substantial uncertainty in the estimates. In other words, a large dissimilarity exists between the estimates and the true attribute values for a sparse dataset.

The use of cross-validation for model comparison does not necessarily involve the observation of absolute values of the parameters of performance between the re-estimates and the validation set, since the comparison is usually done *relatively* between performance parameters (see Maravelias et al. 1996; Su et al. 2006 and examples in Chapters 4 and 5, for example) and the performance parameters represents only those evaluated for the sample locations. The comparison using cross-validation may lose its power, however, when the dataset is either very rich so that each model performs equally well, or very sparse so that each model performs equally poor. The sparse dataset collected in the Anacostia River study area is examined on the applicability of the cross-validation approach. To demonstrated a general case without information of the artificially induced error, the nugget effect is not adjusted in the estimation.

In order to assess the applicability of cross-validation using a sparse sample set, re-estimates are generated and compared to each removed datum. Scatter plots and Q-Q plots for the re-estimates of the cross-validation are shown for the M-Scale model, OK and CK as in Fig. 6.22 through Fig. 6.24. Diagnostic parameters for the cross-

Table 6.4: Diagnostic parameters for the precision (the mean value) and variability reproduction (others) of the cross-validation performed using data of the microbial amount.

Statistics	Target	M-Scale	OK	CK
Mean	1.89	1.90	1.89	1.92
Standard Deviation	0.50	0.16	0.14	0.32
A-D test statistic	–	2.63	2.65	0.84
Correlation Coeff.	–	0.18	-0.08	0.12

validation results are also listed in Table 6.4. It is observed in all three scatter plots that with the removal of one datum, the re-estimation is poorly correlated to the true measurement value with low correlation coefficient (indicated in Table 6.4). The Q-Q plot, standard deviation and A-D test statistic indicates a better reproduction of global variability for CK, while the variability originates from a larger difference between the true value and the re-estimate. In other words, with little correlation between the re-estimate and the remaining data, the variability reproduced by CK comes mainly from artificially induced error and micro-scale variability. Unless the global variance is the only basis of comparison regardless of the estimation precision, cross-validation does not provide a significant method to help select a suitable model in this case.

The same cross-validation graphs and diagnostic parameters are also shown for microbial activity measurements in Figs. 6.25 through 6.27 and Table 6.5. The results provides less indication for model performance than the results for microbial amounts because eight of the CK re-estimation for microbial activity indicates singularity due to the lack of covariance between the re-estimation location and the locations of the remaining data. The cross-validation result for CK, consequently, are performed with the remaining eight re-estimates (a total of 16 points, excluding the 8 points without re-estimates). The lack of covariance can be observed in that the influence range of the variogram in the case of microbial activity is shorter than that of microbial abundance (see left graphs of Figs. 6.14 and 6.13, where the influence range is around

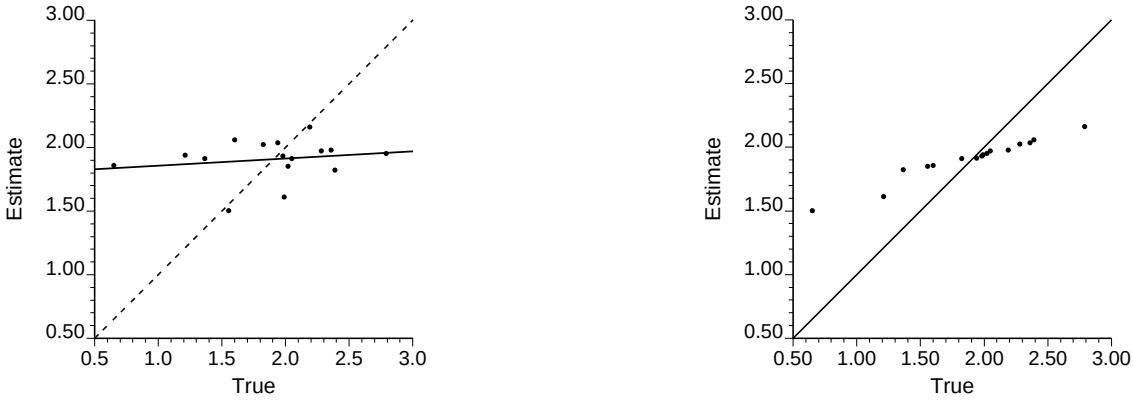

Figure 6.22: M-Scale scatter plot and Q-Q plot for the re-estimates with respect to the validation set using data of the microbial amount.

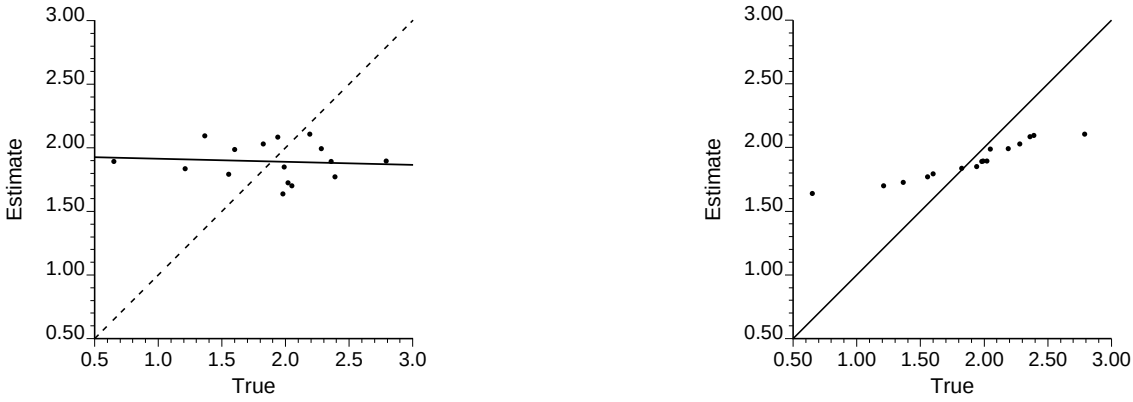

Figure 6.23: OK scatter plot and Q-Q plot for the re-estimates with respect to the validation set using data of the microbial amount.

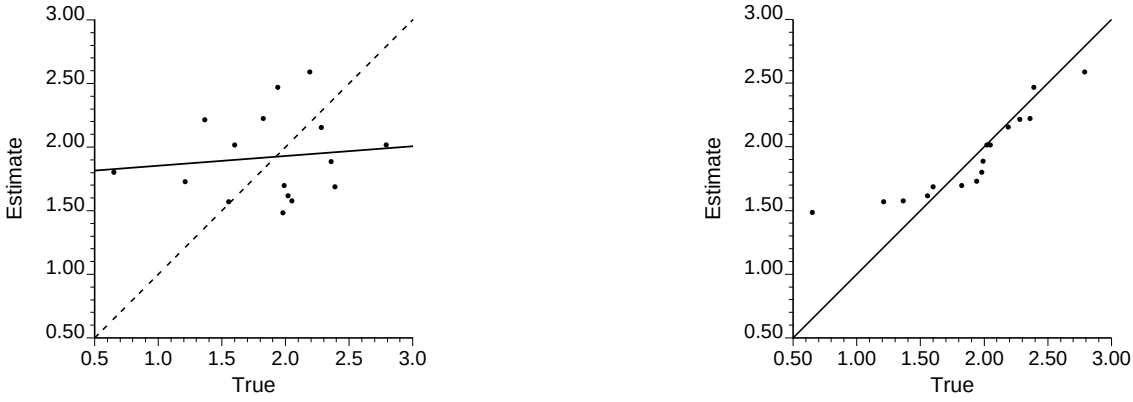

Figure 6.24: CK scatter plot and Q-Q plot for the re-estimates with respect to the validation set using data of the microbial amount.

Table 6.5: Diagnostic parameters for the precision (the mean value) and variability reproduction (others) of the cross-validation performed using data of the microbial activity.

Statistics	Target	M-Scale	OK	CK
Mean	3.84	3.71	3.80	2.26
Standard Deviation	1.94	0.58	0.37	1.09
A-D test statistic	–	3.19	3.99	17.26
Correlation Coeff.	–	-0.18	-0.13	-0.05

25 m for the microbial activity and around 9 m for microbial abundance). The mean and A-D test statistic for CK consequently indicates the worst performance among the three estimation models. The global standard error of the estimates for CK, however, remains reproducing well, reconfirming its sensitivity to data variability.

6.9 Summary

This chapter includes general observations on the impact of sparse data sets on spatial estimation. An approach to discern artificially induced error from micro-scale variability is proposed, by which the impact of nugget-effect adjustment is shown to be substantial due to the sparseness of the dataset. Additionally, by examining the performance of estimation using cross-validation on the sparse dataset, the limitation of cross-validation is also demonstrated.

The impact of attributing a portion of the nugget effect to the artificially induced error to is represented as the area change in the classification map, so that the assessment of this impact can be made in terms of cost/benefit in a contaminated sediment management. In addition, it is also observed that the CK classification map is impacted by the nugget effect not only by the estimation variance, but also the estimates. The adjustment of nugget effect generally has a minimal impact on the estimation maps for OK and the M-Scale model, while the impact is significant on

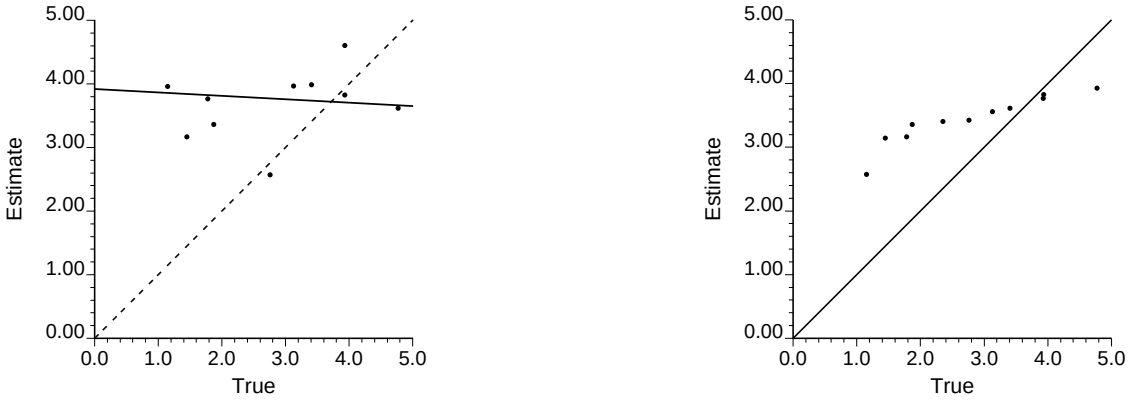

Figure 6.25: M-Scale scatter plot and Q-Q plot for the re-estimates with respect to the validation set using data of the microbial activity.

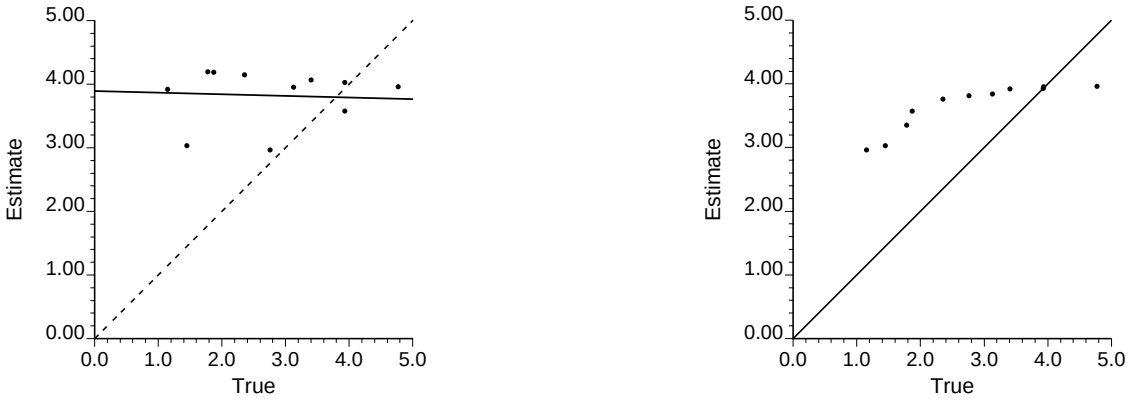

Figure 6.26: OK scatter plot and Q-Q plot for the re-estimates with respect to the validation set using data of the microbial activity.

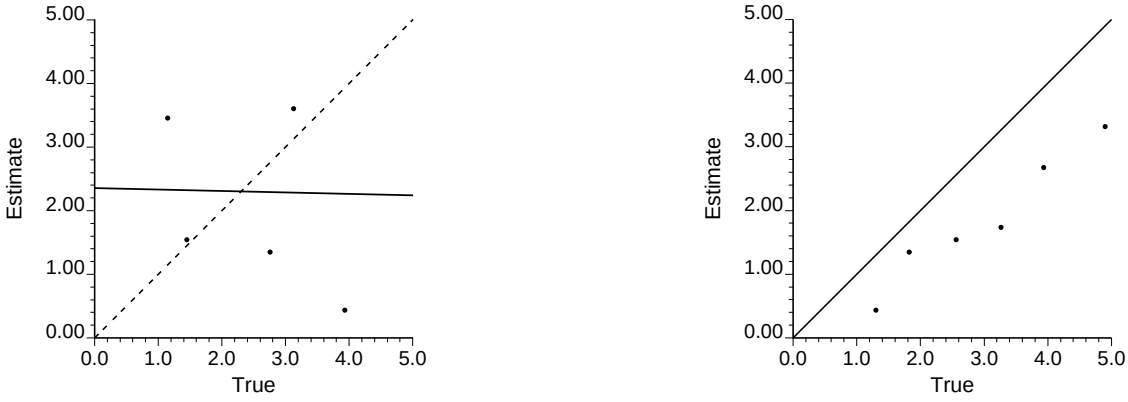

Figure 6.27: CK scatter plot and Q-Q plot for the re-estimates with respect to the validation set using data of the microbial activity.

the estimation variance for all three models.

Based on the Anacostia River dataset, it is concluded that the reduction of artificially induced error improves not only the apparent estimation variance, but also the confidence of estimation expressed in terms of the likelihood of exceedance. Following these improvements, the benefit of nugget effect adjustment is that classification between categories becomes more distinguishable. For a binary case like the exceedance/non-exceedance classification, the artificially induced errors randomly increase/decrease the value of attributes, subsequently increase the occasional likelihood of non-exceedance for locations where the expected value exceeds the exceedance threshold. In other words, artificially induced error impacts the estimation by decreasing the likelihood of exceedance for locations where the expected value exceeds the exceedance threshold. Similarly, artificially induced error also impacts the estimation by decreasing the likelihood of non-exceedance for locations where the expected value is below the exceedance threshold. By adjusting the nugget effect (attributing part of the nugget effect to the artificially induced error), more units are suggested to be cleaned up when the likelihood threshold is high (a cost-saving threshold), and less units are required to clean up when the likelihood threshold is low (a risk-reduction threshold), consequently reducing the risk of non-cleanup for a cost-saving design, and reducing unnecessary cost of over-cleanup for a risk-reduction design. In addition, by observing both the designs of higher (above 0.5) and lower (below 0.5) likelihood thresholds, another benefit is apparent for reducing artificially induced error. The difference between alternatives of different perspectives (reducing risk or reducing clean-up cost) becomes smaller, so that a compromise is easier to reach. The reduction of artificially induced error, consequently, not only provides the benefit of risk/cost reduction for each design of likelihood threshold, but also facilitates the site managers/stakeholders to make a decision on selecting the likelihood design.

For an estimation model like CK that involves reproduction of global variability, however, the classification may or may not be improved depending on the choice of confidence level. In fact, when the estimate of the model is adjusted to match the global variability, the center of the ccdf is away from the best estimate comparing to the cases without the variance constraint. The strong influence of artificially induced error, consequently, acts not only on the variance, but also on the expected value of the ccdf. Because CK estimates are most sensitive to the nugget effect, the characteristics of their ccdf are most different from those of OK and the M-Scale model in the case without nugget-effect adjustment. Although it is discussed that the nugget effect can be adjusted ad hoc for CK when the sample involves solely artificially induced error (Aldworth and Cressie, 1999), such adjustment may not be reasonable if the nugget effect involves an amount of micro-scale variability that is not negligible. An actual signal would be ignored if the nugget effect is not attributed for the part of micro-scale variability. Conversely, even if the artificially induced error could be attributed as part of the nugget effect, CK treated micro-scale variability as part of structural covariance (the covariance associated with attribute values of other locations), thus impacting both the estimates and the estimation uncertainty. In fact, the variability characterized by the nugget effect should impact only locally regardless of the attribute values at other locations. As an estimation model that indicates the likelihood of exceedance, CK estimation may need more scrutiny with respect to the amount of nugget effect involved, so that the ccdf reflects the actual distribution of the estimation model instead of the impact from artificially induced error. Unlike CK, OK and the M-Scale model reflect the impact of nugget effect mainly by the amount of local uncertainty, without the conflict of the conceptual correspondence between the samples and estimation outcomes, both of which are applicable for likelihood-based classification.

Lastly, as a demonstration, cross-validation may not be an effective tool for eval-

uating estimation performance when the sample size is small, because removing one datum removes a large portion of information from the data set, in comparison with an analysis for spatially rich sample set.

CHAPTER 7

Conclusions and Future Research

7.1 Conclusions

Unlike estimation models that have been developed, the M-Scale model provides an alternative that solves problems at a target scale by characterizing multi-scale covariances of means. The goal of this dissertation was to develop, validate and demonstrate the use of the M-Scale model for site characterization, including the analysis of spatial structure and the delineation of the extent of contamination. The following hypothesis was tested:

> *Effective site assessment and remedial decision making on the basis of delineating the extent of contamination depend on estimation methods and uncertainty quantification that reproduce the observed spatial variability of the contaminant concentrations. A better characterization of the extent of contamination can be attained by explicitly accounting for the covariances between multi-scale local means, and quantifying artificially induced error.*

The specific objectives were (1) To provide a spatial statistical approach for the characterization of the spatial structure and spatial distribution of a certain attribute, such as contaminant concentration or microbiological parameters (2) To investigate the applicability of the developed model to field data relevant to contaminated sediments using performance diagnostics of various estimation objectives (3) To explore the sensitivity of M-Scale and other methods to nugget effect characteristics (artificially induced error and microscale variability) using laboratory and field data from the Anacostia River (NJ).

The approach for the first objective involved the M-Scale model development (Chapter 3) and validation using an artificial dataset under a standardized variogram (Chapter 4). The M-Scale model provides an approach to evaluate spatial structures by quantifying covariances between mean values at different scales. The estimator thus developed integrates benefits from ordinary kriging (precision of the estimates) and constrained kriging (preservation of attribute variability) estimators. Validation was accomplished by a graphical comparison of reproduced population covariograms, and quantitative comparison of the structural variance. The comparisons indicate that the M-Scale model, relative to OK and CK, best reproduces *spatial* variability. The approach also provides a balance between the reproduction of spatial variability and the precision for delineating the extent of contamination. As indicated by Cohen's kappa coefficient, which represents the standardized precision of exceedance/non-exceedance classification, the M-Scale model has a high level of classification precision in comparison to OK, while keeping the reproduction of spatial variability as described above. The improvement of false-negative classification rates over OK further indicates that the M-Scale is a suitable approach that serves to support the remedial decisions, because risk-reduction is usually a preferred characteristic. Even though CK has the lowest false-negative rate among the three estimation models,the reproduction does not differentiate between the structural variance and

the nugget effect in the constraint imposed on the data interpolation. Hence, the smaller false-negative rate is a result of occasional agreement of the classification, which is compensated by an even larger false-positive rate among the three estimation models, resulting in the worst precision of classification as indicated by Cohen's kappa coefficient, as illustrated and explained in Chpater 4.

Based on my work, I argue that a robust estimation model that balances the different objectives is a better approach, unless an optimized ratio for the false-positive/false-negative classification can be decided a priori, or a funtion that specifies the association between remedial cost and contamination level can be defined. Based on the description above, the M-Scale model serves as a robust estimation model among the three estimation models presented in this dissertation, with regard to the following characteristics. The model:

- Evaluates spatial structure in the form of covariances between mean values at different scales, providing a measure of the information that each scale provides about other scales.
- Generates estimates that reflect the structural variability characterized by the sample. In other words, provides an intuitive reference of the variability of spatial attributes that could be explained by the samples, without impacts from the nugget effect that is inherent in samples but not informative for the estimates.
- Generates classification map for the extent of contamination, which is applicable as a decision-support tool for remedial purposes.

For the second objective, cross-validation was performed using dioxin data from the Passaic River (NJ) to investigate the performance of different estimation models

under point and block estimation scenarios for a rich sample set (Chapter 5). An analysis was performed on the Anacostia river dataset (microbial parameters) to investigate the applicability of cross-validation for a sparse sample set (Chapter 6). A comparison of point and block estimations in Chapter 5 reveals the sensitivity of CK estimation to the nugget effect. For point estimation, the results of scatter plots, Q-Q plots, standard deviation of estimates, and the A-D test statistics indicate that CK provides the best reproduction of *global* variability, while the covariogram and structural variance indicate that the M-Scale model provides the best alternative to reproduce *spatial* variability. The contingency table and Cohen's κ coefficient, which reflect the model performance for the classification maps, indicate worse classification precision for CK, better classification precision for OK and the M-Scale model. The low false-negative rate for CK is compensated by a high false-positive rate, and the M-Scale model has a lower false-negative rate relative to that of OK with the same level of precision as that of OK. All descriptions above for the cross-validation are consistent with the descriptions of performance stated for the different estimation models as validated by the same diagnostic parameters using artificial data sets. For block estimation, however, the nugget effect between samples and estimates is reduced due to block averaging. Consequently, the results of block estimation indicate the same performance of spatial-variability reproduction for CK and the M-Scale model, and the same level of precision is attained for all three models with a slightly better performance for OK. The conclusion for block estimation, therefore, is that all three models perform similarly, while the reproduction of spatial variability is better for CK and the M-Scale model with lower false-negative rates and worst for OK with higher false-negative rates for the delineation of the extent of contamination. Although cross-validation for the block estimation uses rescaled sample values that involve uncertainty from the validation set, results of conditional simulation indicates this uncertainty is relatively small relative to the uncertainty due to the lack of data.

Consequently, the result of cross-validation on the basis of the rescaled sample is regarded as valid. The conclusion of using cross-validation on a sparse dataset is that the information from each data point is essential in the estimation process, and the removal of one single measurement in the dataset implies the removal of a big proportion of information in the input of the re-estimation process. Diagnostic parameters evaluated for the reproduction of the validation data, consequently, are rendered invalid since limited relatedness remains between the re-estimate and the remaining data points.

For the third objective, an approach is introduced to adjust the nugget effect, i.e., to attribute portions of the observed nugget effect to micro-scale variability and artificially induced error. By adjusting the nugget effect, the precision enhancement in the estimate is generally reflected in the change of estimation variance, regardless of whether the estimated values are also impacted. The impact of nugget-effect adjustment, which results in the reduction of the estimation variance, is also reflected in the ccdf at each estimation location. The ccdf with nugget effect adjusted and retained is further derived into the likelihood of exceedance (the likelihood to fulfill the "willingness of the analyst to accept the risk of an incorrect decision (Istok and Rautman, 1996)."), and subsequently the difference in the area of threshold exceedance for each designed likelihood threshold on the basis of likelihood-based classification. The benefit of reducing artificially induced error is a gain in the likelihood of correct classification. Another observation is that the difference in areas of exceedance between risk-reduction and cost-saving objectives becomes smaller after reducing artificially induced error. A compromise between stakeholders, consequently, is easier to reach after the impact of artificially induced error is excluded from the estimation variance.

Among the three estimation models, the benefit of reducing artificially induced error in CK estimation is found to conflict with the statements above for certain

likelihood thresholds, due to the sensitivity of estimates against the nugget effect when the artificially induced error is involved. The inconsistency comes from the two-fold benefit of reducing artificially induced error, including not only the reduction of estimation variance, but also the correction of the estimates from falsely reproducing local variability by spatial variability in the estimates. When CK is selected as the site characterization model, the attribution of the nugget effect into portions of artificially induced error and micro-scale variability becomes an important procedure prior to the estimation.

Summarizing the development and investigation performed in previous chapters, the following conclusions can be formulated:

1. The M-Scale model generally provides a robust model for reproducing spatial variability and the delineation of the extent of contamination. For scientific exploration, the approach is useful for characterizing relatedness between mean value at different spatial scales, and reproducing spatial variability in the estimation map. For the support of remedial decisions, consequently, the approach is expected to provide a robust alternative in estimating remedial cost because the extent of contamination depends on the reproduction of spatial variability. The M-Scale model also provides similarly precise classification as OK, with lower false-positive rates compared to OK estimation. The performance of the M-Scale model, however, may depend on the objective of estimation, which includes the decision of the use of point or block estimation.

2. Cross-validation using the M-scale model indicates that it best reproduces spatial variability, and result in moderate levels of both false-positive and false-negative rates in the exceedance classification. In the Passaic River example presented in Chapter 5, the M-Scale model is found to be the most suitable alternative for the characterization of the site when the objective of estimation

depends on the reproduction of spatial variability, or when a balance is needed between the risk of false-negative assignments and the precision of the resulting classification map.

3. Given analysis of site-scale and micro-scale (based on laboratory experiments) variograms, the artificially induced error can be evaluated, and the nugget effect adjusted to enhance estimation precision. The benefit of enhancing estimation precision for a likelihood-based remedial decision is expressed in terms of the change in the area of threshold exceedance, which is informative for the subsequent risk-benefit analysis and stakeholders negotiation. This benefit of nugget effect adjustment is found in OK and the M-Scale model for the example given using the Anacostia River dataset. The nugget-effect adjustment indicates a different impact for CK, due to the sensitivity of CK estimates to the nugget effect, resulting in a change of the ccdf not only in its variance, but also in its expected value of the distribution. More scrutinized quantification needs to be performed for the nugget effect when CK is selected as a candidate model for site characterization.

7.2 Implication of the Results for Contaminated Sediment Management

The major challenges related to sediment management are: (i) the uncertainty associated with characterization of large contaminated areas with low level diffuse contamination; and (ii) the uncertainties related to long term in situ remediation strategies. This dissertation was aimed at developing and validating a method capable of quantifying the spatial uncertainty by capturing the information collected at various spatial scales.

Different estimation models feature different characteristics in their estimates as well as their estimation uncertainty. The M-Scale model was developed to provide a robust model that generates estimates reproducing spatial variability characterized by the samples. The extent of contamination can be delineated using the estimates generated, or by the likelihood derived from the estimates and the associated estimation variance. However, the selection of estimation model should depend not only on the estimation characteristics but also on the nature of the problem (e.g., site characterization or decision support for remedial decision), the definition of problem (e.g., the regulatory threshold and likelihood threshold), objective of remediation (e.g., risk-reduction or cost-reduction), target spatial scale (e.g., point or block estimation), and spatial configuration and density of sample locations. Cross-validation is recommended prior to the application of any estimation model, given that the sample size is sufficiently large to provide such validation. Adjustment of the nugget effect enhances the estimation results when a likelihood-based classification is applied, indicating a benefit that a compromise is easier to reach between the objectives favoring risk-reduction and cost-saving. A sampling scheme is needed for both the site-scale and the micro-scale if elimination of the impact by artificially induced error from the estimated uncertainty is favored for easier compromises between stakeholders.

The implications of this work are best illustrated using the project framework that funded this work, as shown in Fig. 7.1. The target application of this work was to quantify the spatial uncertainty associated with in situ capping strategies subjected to advection and microbial gas ebullition. Funded under auspices of the Strategic Environmental Research and Development Program (SERDP), and part of a larger combined experimental and modeling project, the challenge was to integrate data collected at laboratory (flux chamber, flume) and field (Anacostia River) scales to improve decision-making (i.e. to narrow the predictive uncertainty bounds on experimental data). Aside from scale-specific experimental uncertainty estimations (mainly

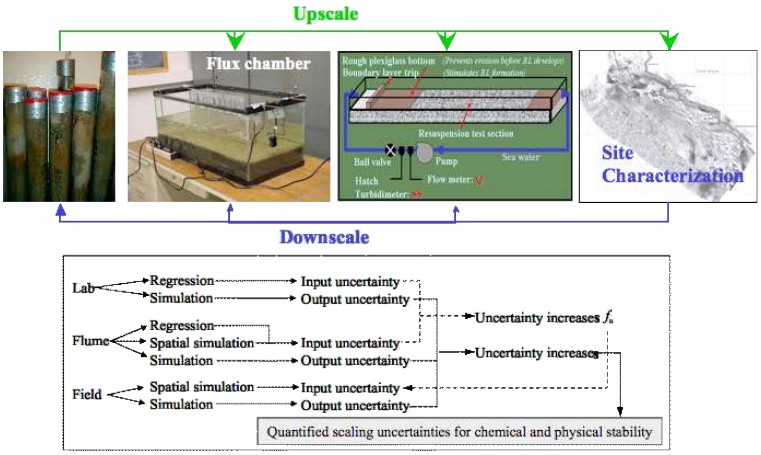

Figure 7.1: Project framework for uncertainty-based sediment management in Anacostia River sediments.

through regression analyses), there was a need to be able to translate data from various scales to the field. With limited experimental and field sampling data, and after quantifying the artificially induced error and microscale variability component of the variogram's nugget effect, M-Scale was capable of reducing the estimation uncertainty of the spatial delineation of sediment microbial characteristics which were shown to serve as proxy parameters for microbial gas production. Since gas production destabilizes the sediment bed and increase contaminant fluxes out of the sediment into the overlaying water column, the M-scale estimates can then be used to inform spatial likelihood of cap failure and future requirements for improvement in cap design.

7.3 Limitations of the M-Scale Model and Future Research

Although the M-Scale model provides a robust alternative framework for the site characterization and remedial decision-making, a few limitations for the model implementation were also observed during the research presented in this dissertation:

1. Dependence of performance on sampling design: It is demonstrated in Chapter 4 that the estimation variance may or may not be under-evaluated depending on the configuration of samples (affecting the redundancy between sample points) in each ring area.

2. Selection of shape and size of different scales: It is generally recognized that different levels of subjectivity can be involved in spatial statistics. However, rules for deciding the number, shape and size of scales in the M-Scale estimation are not fully explored. For the current stage of development of the M-Scale model, the selection will depend on the scientific judgment of the spatial analyst.

In order to overcome current limitations of the M-Scale model and to extend its capability to different objectives for site characterization, the following future research needs are identified:

7.3.1 Effect of Sample Redundancy on the M-Scale Model

The M-Scale model assumes that the configuration of sample locations in each ring area is random. The result of this assumption may be an under-evaluation of the estimation variance, which has the strongest impact on the likelihood-based classification. Research to incorporate the sample configuration for the variance of the sample ring average relative to the population ring mean is recommended, in order to have an unbiased estimate for the estimation variance. The conceptual sketch is illustrated in Fig. 7.2. If the clustered sample locations in B are falsely regarded as random sample locations, the actual amount of uncertainty shown in B will be regarded as the uncertainty shown in A (an underevaluation of uncertainty) since the two correspond to the same sample size. The improvement will need to be evaluated using the coverage rate of the estimation confidence interval described in Chapter

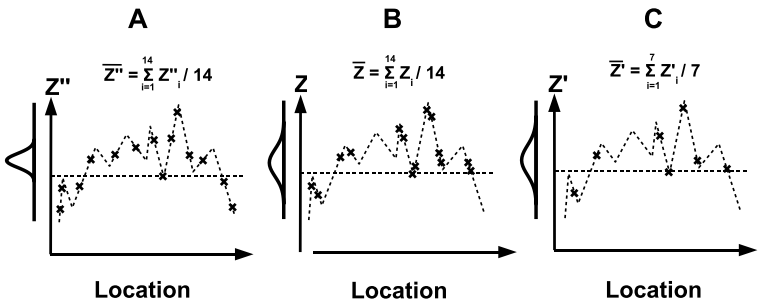

Figure 7.2: Conceptual sketch of the sample redundancy in a one-dimension sample. Dashed curves and horizontal lines indicate the unknown reality and its mean value. Bell shapes on the left of each graph indicates the uncertainty of using sample mean to represent the mean value of reality. A: a sample at random locations with sample size = 14. B: a sample at clustered locations with sample size = 14. C: a sample at non-clustered locations with sample size = 7. Note that uncertainty shown in B is similar to that in C because each sample point in the clusters in B provides similar information, similar to information provided by one single point.

4, and further described as the change in area of exceedance under the likelihood-based classification described in Chapter 6, to provide indication of its applicability for decision-support of the contamination management.

7.3.2 Selection of Size and Shape for Different Scales

As described in the limitation of the M-Scale model, the effect of shape and size, as well as the number of scales to include for the M-Scale estimation is yet to be explored. The M-Scale model, however, has the flexibility to define the shape of rings/plates so that the influence of samples on estimates varies not only by the separation distance but also by the relative direction, which may be site specific. The study on the size, shape and number of scales will not only improve the M-Scale model for a more robust estimation model, but also extend its capability to model the stochastic part of contamination flow mechanism, such as the direction, distance and anisotropy of influence, as shown in Fig.7.3.

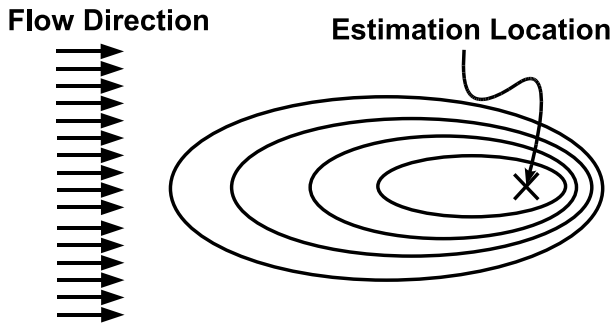

Figure 7.3: Example of the selection of shape and size for the M-Scale model. In this example, influence from the upstream is stronger than the influence from the downstream, and influence from the flow direction is stronger than the influence from the transverse direction.

7.3.3 Integration of Additional Information

The M-Scale model is currently designed to perform estimation using samples of a single attribute. Including, but not limited to variogram-based estimation approaches, however, it is reported that incorporating additional information in the estimation will enhance the precision of estimation (Schnabel and Tietje, 2003). A detailed description for the types of additional information is listed in Saito (2003) (pp. 12–14), which includes (i) more densely sampled secondary attributes (e.g., expert judgment, remote-sensing data, or inexpensive surrogate measurements) (ii) historical records (e.g., locations of suspected sources of contamination) (iii) statistical properties (e.g., mean or variance of the primary attribute obtained prior to the current sample set) (iv) areal information (e.g., soil types or land use maps) (v) physical/chemical relations for modeling the trend of the primary attribute (e.g., predominant wind direction and distance to the sources, or periodic behavior in geomagnetic data). Particularly for (iv), the areal information may also include samples collected in multiple scales, for which the spatial structure could be characterized by the covariances between mean values using the M-Scale model, and further used for the estimation at different target scales. Direction and distance information for the predominant wind direction indicated in (v) can be achieved by the selection of shape

and size as indicated in Section 7.3.2.

For a sparse sample configuration, this secondary information is particularly important since the information will not only enhance the precision of the estimate, but also further assist the reproduction of variability in addition to the contributions given by the average value of neighboring samples. The incorporation of secondary information, similar to the use of kriging with a trend model (p.313 of Journel and Huijbregts 1978; p.164 of Chilès and Delfiner 1999) or cokriging (p.203 of Goovaerts 1997b;Wang et al. 2006), is recommended for the future development of the M-Scale model.

7.3.4 Extended Multi-Scale Applications

The M-Scale model is developed in this dissertation as a spatial estimation model for the decision support of on-site remediation. However, the definition of the M-Scale model is conceptual and general, involving no specification of the attribute of interest, and no definition of the objective parameter. Consequently, the M-Scale model can be applied to attributes such as the extent of rainfall intensity that serves as the input of a runoff evaluation model. The objective parameter, in addition to the estimates and estimation uncertainty, may also involve the weights attributed to the different local means, which is indicative of the best scale to perform estimation locally. Nevertheless, examining the applicability to the different objective parameter or different attributes of interest requires knowledge of the physical process for the particular attributes, and thus further research needs to be conducted via interdisciplinary cooperations and appropriate selection of the validation approach.

BIBLIOGRAPHY

P. Adriaens, A. L. Barkovskii, and I. D. Albrecht. Fate of chlorinated aromatic compounds in soils and sediments. In D. Adriano, J.-M. Bollag, W. Frankenberger, and R. Sims, editors, *Bioremediation of Contaminated Soils*, pages 175–212, Madison, WI, 1999. Soil Science Society of America / American Society of Agronomy Monograph, Soil Science Society of America Press.

P. Adriaens, C. Gruden, and M. L. McCormick. Biogeochemistry of halogenated hydrocarbons. In B. Sherwood-Lollar, editor, *Environmental Geochemistry*, volume 9 of *Treatise on Geochemistry*, page 511V541. Kluwer, Amsterdam, The Netherlands, 2003.

P. Adriaens, M.-Y. Li, and A. M. Michalak. Scaling methods of sediment bioremediation processes and applications. *Engineering in Life Sciences*, 6(3):217–227, 2006.

J. Aldworth and N. Cressie. Sampling designs and prediction methods for Gaussian spatial processes. In S. Ghosh, J. Srivastava, et al., editors, *Multivariate Analysis, Design of Experiments, and Survey Sampling*, pages 1–54. Marcel Dekker, 1999.

J. Aldworth and N. Cressie. Prediction of nonlinear spatial functionals. *Journal of Statistical Planning and Inference*, 112(1-2):3–41, 2003.

American Society for Testing and Materials. Standard guide for developing conceptual site models for contaminated sites. In *Annual book of ASTM standards. Vol. 04.08 E1689-95*. ASTM, West Conshohocken, PA, 1995.

T. W. Anderson and D. A. Darling. A test of goodness of fit. *Journal of the American Statistical Association*, 49(268):765–769, 1954.

A. H.-S. Ang and W. H. Tang. *Probability Concepts in Engineering: Emphasis on Application in Civil and Environmental Engineering*. John Wiley & Sons, Inc., 2 edition, 2007. ISBN 978-0-471-72064-5.

S. Apitz, J. Davis, K. Finkelstein, D. Hohreiter, R. Hoke, R. Jensen, J. Jersak, V. Kirtay, E. Mack, V. Magar, D. Moore, D. Reible, and

R. Stahl. *Critical Issues for Contaminated Sediment Management*, 2002. http://meso.spawar.navy.mil/docs/MESO-02-TM-01.pdf.

S. E. Apitz, J. W. Davis, K. Finkelstein, D. W. Hohreiter, R. Hoke, R. H. Jensen, J. Jersak, V. J. Kirtay, E. E. Mack, V. S. Magar, D. Moore, D. Reible, and J. Ralph G. Stahl. Assessing and managing contaminated sediments: Part I, developing an effective investigation and risk evaluation strategy. *Integrated Environmental Assessment and Management*, 1(1):2–8, 2005.

S. E. Apitz and E. A. Power. From risk assessment to sediment management: An international perspective. *Sediment Management*, 2(2):1–6, May 2002.

A. Arik. Area influence kriging. *Mathematical Geology*, 34(7):783–796, 2002.

P. M. Atkinson and R. E. J. Kelly. Scaling-up point snow depth data in the UK for comparison with SSM/I imagery. *International Journal of Remote Sensing*, 18(2): 437–443, 1997.

N. Barabás. *Quantitative Evaluation of Dechlorination Signatures in Contaminated Sediments Using Modified Polytopic Vector Analysis with Uncertainty Assessment*. PhD thesis, University of Michigan, 2003.

N. Barabás, P. Adriaens, and P. Goovaerts. Modified polytopic vector analysis to identify and quantify a dioxin dechlorination signature in sediments. 1. Theory. *Environmental Science and Technology*, 38(6):1813–1820, Mar 2004.

N. Barabás, P. Goovaerts, and P. Adriaens. Geostatistical assessments and vlidation of uncertainty for three-dimentional dioxin data from sediments in an estuarine river. *Environmental Science and Technology*, 35(16):3294–3301, 2001.

A. L. Barkovskii and P. Adriaens. Microbial dechlorination of historically present and freshly spiked chlorinated dioxins and diversity of dioxin-dechlorinating populations. *American Society for Microbiology*, 62(12):4556–4562, 1996.

A. L. Barkovskii and P. Adriaens. Impact of model humic constituents on microbial reductive dechlorination of polychlorinated dibenzo-p-dioxins. *Environ. Toxicol. Chem*, 17:1013–1021, 1998.

R. J. Barnes. The variogram sill and the sample variance. *Mathematical Geology*, 23 (4):673–678, 1991.

G. E. Batley, G. A. Burton, P. M. Chapman, and V. E. Forbes. Uncertainties in sediment quality weight-of-evidence (WOE) assessments. *Human and Ecological Risk Assessment*, 8:1517–1547, 2002.

W. M. Berelson, T. Townsend, D. Heggie, P. Ford, A. Longmore, G. Skyring, T. Kilgore, and G. Nicholson. Modelling bio-irrigation rates in the sediments of Port Phillip Bay. *Marine and Freshwater Research*, 50(6):573–579, 1999.

S. L. Bertelsen, A. D. Hoffman, C. A. Galliant, C. M. Elonen, and J. W. Nichols. Evaluation of log KOW and tissue lipid content as predictors of chemical partitioning to fish tissues. *Environmental Toxicology and Chemistry*, 17:1447–1455, 1998.

L. Bian and R. Butler. Comparing effects of aggregation methods on statistical and spatial properties of simulated spatial data. *Photogrammetric engineering and remote sensing*, 65(1):73–84, 1999.

P. Bonin, E. R. Ranaivoson, N. Raymond, A. Chalamet, and J. C. Bertrand. Evidence for denitrification in marine sediment highly contaminated by petroleum products. *Pol. Bull.*, 28:89–95, Mar 1994.

F. J. Brockman and C. J. Murray. Subsurface microbiological heterogeneity: Current knowledge, descriptive approaches and applications. *FEMS Microbiology Reviews*, 20(3-4):231–247, July 1997.

D. G. Brown, P. Goovaerts, A. Burnicki, and M.-Y. Li. Stochastic simulation of land-cover change using geostatistics and generalized additive models. *Photogrammetric Engineering and Remote Sensing*, 68(10):1051–1061, 2002.

J. Cable, W. Burnett, J. Chanton, and G. Weatherly. Estimating groundwater discharge into the northeastern Gulf of Mexico using radon-222. *Earth and Planetary Science Letters*, 144(3-4):591–604, 1996.

J. E. Cable, W. C. Burnett, J. P. Chanton, D. R. Corbett, and P. H. Cable. Field evaluation of seepage meters in the coastal marine environment. *Estuarine, Coastal and Shelf Science*, 45(3):367–375, 1997.

F. J. Caniego, R. Espejo, M. A. Martin, and F. San Jose. Multifractal scaling of soil spatial variability. *Ecological Modelling*, 182(3-4):291–303, Mar. 2005.

D. G. Capone and R. P. Kiene. Comparison of microbial dynamics in marine and freshwater sediments: Contrasts in anaerobic carbon catabolism. *Limnology and Oceanography*, 33(4):725–749, 1988.

J. R. Carr. Application of spatial filter theory to kriging. *Mathematical Geology*, 22 (8):1063–1079, 1990.

J. R. Carr. Reply to "letter to the editor" by Hans Wackernagel. *Mathematical Geology*, 23(7):985–986, 1991.

J. R. Carr. Response to comment by Yuan Z. Ma. *Mathematical Geology*, 25(3): 405–406, 1993.

J. R. Carr. On visualization for assessing kriging outcomes. *Mathematical Geology*, 34(4):421–433, 2002.

P. Casper, S. C. Maberly, G. H. Hall, and B. J. Finlay. Fluxes of methane and carbon dioxide from a small productive lake to the atmosphere. *Biogeochemistry*, 49:1–19, 2000.

D. Chadwick, C. Katz, J. G. Groves, A. Carlson, C. Smith, R. Paulsen, D. O'Rourke, and N. Gahr. Anacostia River seepage and porewater survey report – Draft. Technical report, Marine Environmental Quality Branch, SPAWAR Systems Center, San Diego, CA 92152, Feb 2001.

B. V. Chang, S. W. Chang, and S. Y. Yuan. Anaerobic degradation of polycyclic aromatic hydrocarbons in sludge. *Advances in Environmental Research*, 7:623–628, 2003.

P. M. Chapman, B. G. McDonald, and G. S. Lawrence. Weight-of-evidence issues and frameworks for sediment quality (and other) assessments. *Human and Ecological Risk Assessment*, 8:1489–1515, 2002.

N. S. Cheng. Effect of concentration on settling velocity of sediment particles. *ASCE Journal of Hydraulic Engineering*, 123(8):728–731, 1997a.

N.-S. Cheng. Simplified settling velocity formula for sediment particles. *ASCE J. Hydr. Eng.*, 123(2):149–152, 1997b.

J.-P. Chilès and P. Delfiner. *Geostatistics: Modeling Spatial Uncertainty*. Wiley Series in Probability and Statistics- Applied. John Wiley & Sons, Inc., 1999.

I. Clark. Regularization of a semivariogram. *Computers and Geosciences*, 3:341–346, 1977.

J. U. Clarke and V. A. McFarland. Uncertainty analysis for an equilibrium partitioning-based estimator of polynuclear aromatic hydrocarbon bioaccumulation potential in sediments. *Environmental Toxicology and Chemistry*, 19:360–367, 2000.

G. E. Claypool and K. A. Kvenvolden. Methane and other hydrocarbon gases in marine sediments. *Ann. Rev. Earth Planet. Sci.*, 11:299–327, 1983.

J. Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37, 1960.

J. B. Collins and C. E. Woodcock. Geostatistical estimation of resolution-dependent variance in remotely sensed images. *Photogrammetric engineering and remote sensing*, 65(1):41–50, 1999.

B. Conant, J. A. Cherry, and R. W. Gillham. A PCE groundwater plume discharging to a river: Influence of the streambed and near-river zone on contaminant distributions. *Journal of Contaminant Hydrology*, 73(1-4):249–279, 2004.

B. Conant Jr. Delineating and quantifying ground water discharge zones using streambed temperatures. *Ground Water*, 42(2):243–57, 2004.

R. G. Congalton. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.*, 37:35–46, 1991.

P. Couto. Assessing the accuracy of spatial simulation models. *Ecological Modelling*, 167(1):181–198, 2003.

N. Cressie. Aggregation in geostatistical problems. In A. Soares, editor, *Geostatistics Tróia '92*, volume 1, pages 25–36, Dordrecht, the Netherlands, 1993. Kluwer Academic Publishers.

N. Cressie, J. Zhang, and P. F. Craigmile. Geostatistical prediction of spatial extremes and their extent. In P. Renard, H. Demougeot-Renard, and R. Froidevaux, editors, *Geostatistics for Environmental Applications: Proceedings of the Fifth European Conference on Geostatistics for Environmental Applications*, pages 27–37, 2005.

T. O. Crist. The spatial distribution of termites in shortgrass steppe: A geostatistical approach. *Oecologia*, 114(3):410–416, Apr. 1998.

A. Critto, C. Carlon, and A. Marcomini. Screening ecological risk assessment for the benthic community in the Venice lagoon (Italy). *Environment International*, 31(7): 1094–1100, Sept. 2005.

A. Cullen. Comparison of measured and predicted environmental PCB concentrations using simple compartmental models. *Environmental Science & Technology*, 36(9): 2033–2038, 2002.

F. C. Curriero, M. E. Hohn, A. M. Liebhold, and S. R. Lele. A statistical evaluation of non-ergodic variogram estimators. *Environmental And Ecological Statistics*, 9 (1):89–110, Mar. 2002.

B. S. Cushing. *State of Current Contaminated Sediment Management Practices*. Applied Environmental Management, Inc., 1999.

G. Dagan. Stochastic modeling of groundwater flow by unconditional and conditional probabilities. 1. Conditional simulation and the direct problem. *Water Res. Research*, 18:813–833, 1982.

M. R. T. Dale, P. Dixon, M.-J. Fortin, P. Legendre, D. E. Myers, and M. S. Rosenberg. Conceptual and mathematical relationships among methods for spatial analysis. *Ecography*, 25:558–577, 2002.

J. W. Davis, T. Dekker, M. Erickson, V. Magar, C. Patmont, and M. Swindoll. Framework for evaluating the effectiveness of monitored natural recovery (MNR) as a contaminated sediment management option – draft. Technical report, Remediation Technology Development Forum, June 2004.

M. A. de Angelis and M. I. Scranton. Fate of methane in the Hudson River and estuary. *Global Biogeochemical Cycles*, 7(3):509–524, 1993.

A. P. J. de Roo, L. Hazelhoff, and G. B. M. Heuvelink. Estimating the effects of spatial variability of infiltration on the output of a distributed runoff and soil erosion model using Monte Carlo methods. *Hydrological Processes*, 6(2):127–143, 1992.

T. J. Dekker, N. Barabás, and J. Wolfe. Incorporating uncertainty analysis into decision-making at contaminated sediment sites – Case studies. In *The Fourth International Conference on Remediation of Contaminated Sediments*, Savannah, Georgia, January 2007.

K. E. Dennett, T. W. Sturm, A. Amirtharajah, and T. Mahmood. Flume studies on the erosion of cohesive sediments. In *1st Intnl. Conf. on Water Resources*, pages 199–203, San Antonio, TX, 1995.

A. Desbarats and S. Bachu. Geostatistical analysis of aquifer heterogeneity from the core scale to the basin-scale – A case-study. *Water Resources Research*, 30(3): 673–684, Mar. 1994.

A. Desbarats and R. Srivastava. Geostatistical characterization of groundwater-flow parameters in a simulated aquifer. *Water Res. Research*, 27:687–698, 1991.

A. J. Desbarats. Modeling spatial variability using geostatistical simulation. In S. Rouhani, R. Srivatsava, A. Desbarats, M. Cromer, and A. Johnson, editors, *STP 1283 Geostatistics for Environmental and Geotechnical Applications*, pages 32–48. American Society for Testing and Materials, Philadelphia, PA, 1996.

C. V. Deutsch and A. G. Journel. *GSLIB: Geostatistical Software Library and User's Guide*. Applied Geostatistics Series. Oxford University Press, 2 edition, 1998.

I. G. Droppo and E. D. Ongley. Flocculation of suspended sediment in rivers of southeastern canada. *Water Research*, 28(8):1799–1809, 1994.

B. Elberling. Gas phase diffusion coefficient in cemented porous media. *J. Hydrol.*, 178:93–108, 1996.

A. W. Elzerman and J. T. Coates. Hydrophobic organic compounds: Equilibria and kinetics of sorption. In R. Hites and S. Eisenreich, editors, *Sources and Fate of Aquatic Pollutants*. American Chemical Society, Washington, D.C, 1987.

E. J. Englund and N. Heravi. Conditional simulation: Practical application for sampling design optimization. In A. Soares, editor, *Geostatistics Tróia '92*, volume 2, pages 613–624, Dordrecht, the Netherlands, 1993. Kluwer Academic Publishers.

E. J. Englund and N. Heravi. Phased sampling for soil remediation. *Environmental and Ecological Statistics*, 1:247–263, 1994.

M. J. Erickson, C. L. Turner, and L. J. Thibodeaux. Field observation and modeling of dissolved fraction sediment-water exchange coefficients for PCBs in the Hudson River. *Environ. Sci. Technol.*, 39(2):549–556, 2005.

A. Facchinelli, E. Sacchi, and L. Mallen. Multivariate statistical and GIS-based approach to identify heavy metal sources in soils. *Environmental Pollution*, 114(3): 313–324, 2001.

N. J. Fendinger, D. D. Adams, and D. E. Glotfelty. The role of gas ebullition in the transport of organic contaminants from sediments. *The Science of the Total Environment*, 112:189–201, 1992.

A. Fioole, E. J. Houwing, and L. M. van der Heijdt. SURFIS: A tool for designing and optimizing dredging schemes. *Water Science And Technology*, 37(6):103–107, 1998.

Q. S. Fu, A. L. Barkovskii, and P. Adriaens. Dioxin cycling in aquatic sediments: The Passaic River estuary. *Chemosphere*, 43(4):643–648, 2001.

M. Gerino, R. C. Aller, C. Lee, J. K. Cochran, J. Y. Aller, M. A. Green, and D. Hirschberg. Comparison of different tracers and methods used to quantify bioturbation during a spring bloom: 234-thorium, luminophores and chlorophyll a. *Estuarine, Coastal and Shelf Science*, 46(4):531–547, 1998.

B. Gevao, J. Hamilton-Taylor, C. Murdoch, K. C. Jones, M. Kelly, and B. J. Tabner. Depositional time trends and remobilization of PCBs in lake sediments. *Environmental Science and Technology*, 31:3274–3280, 1997.

G. D. Ginsburg and V. A. Soloviev. Methane migration within the submarine gas-hydrate stability zone under deep-water conditions. *Marine Geology*, 137:49–57, 1997.

Y. Gong and J. V. Depinto. Desorption rates of two PCB congeners from suspended sediments–II. model simulation. *Water Research*, 32(8):2518–2532, 1998.

P. Goovaerts. Study of spatial relationships between two sets of variables using multivariate geostatistics. *Geoderma*, 62:93–107, 1994.

P. Goovaerts. Accounting for local uncertainty in environmental decision making process. In E. Baafi and N. Schofield, editors, *Geostatistics Wollongong '96*, pages 929–940, Dordrecht, 1997a. Kluwer Academic Publishers.

P. Goovaerts. *Geostatistics for Natural Resources Evaluation.* Applied Geostatistics Series. Oxford University Press, 1997b.

P. Goovaerts. Kriging vs. stochastic simulation for risk analysis in soil contamination. In A. Soares, J. Gmez-Hernndez, and R. Froidevaux, editors, *geoENV I – Geostatistics for Environmental Applications*, pages 247–258. Kluwer Academic Publishers, Dodrecht, the Netherlands, 1997c.

P. Goovaerts. Geostatistical tools for characterizing the spatial variability of microbiological and physico-chemical soil properties. *Biology And Fertility Of Soils*, 27 (4):315–334, Sept. 1998.

P. Goovaerts. Geostatistics in soil science: State-of-the-art and perspectives. *Geoderma*, 89:1–45, 1999.

C. R. Grego, S. R. Vieira, and A. L. Loureno. Spatial distribution of Pseudaletia sequax Franclemlont in triticale under no-till management. *Scientia Agricola*, 63: 321–327, 2006.

C. L. Gruden, A. Khijniak, and P. Adriaens. Activity assessment of microorganisms eluted from sediments using 5-cyano-2,3-ditolyl tetrazolium chloride: a quantitative comparison of flow cytometry to epifluorescent microscopy. *Journal of Microbiological Methods*, 55(3):865–874, 2003.

R. C. Gularte, W. E. Kelly, and V. A. Naeeri. Erosion of cohesive sediment as a rate process. *Ocean Engineering*, 7(4):539–551, 1995.

R. Haining. *Spatial Data Analysis: Theory and Practice*. Cambridge University Press, 2003.

B. Hartman and D. E. Hammond. Gas exchange rates across the sediment-water and air-water interfaces in south San Francisco Bay. *Journal of Geophysical Research*, 89(C3):3593–3603, 1984.

F. M. Henderson. *Open channel flow*. Collier-Macmillan Publishing Co. Inc., 1966.

U. C. Herzfeld. Geostatistical interpolation and classification of remote sensing data from ice surfaces. *International Journal Of Remote Sensing*, 20(2):307–327, Jan. 1999.

K. W. Holmes, P. C. Kyriakidis, O. A. Chadwick, J. V. Soares, and D. A. Roberts. Multi-scale variability in tropical soil nutrients following land-cover change. *Biogeochemistry*, 74(2):173–203, June 2005.

W. Huang, P. Peng, Z. Yu, and J. Fu. Effects of organic matter heterogeneity on sorption and desorption of organic contaminants by soils and sediments. *Applied Geochemistry*, 18(7):955–972, 2003.

H. Huls, M. Costello, and R. Sheets. Gas, NAPL and PAH flux assessment in sediments. In M. Pellei and A. Porta, editors, *Proceedings of the Second International Conference on Remediation of Contaminated Sediments*, Venice, Italy, 2003.

T. J. Iannuzzi, N. L. Bonnevie, and R. J. Wenning. An evaluation of current methods for developing sediment quality guidelines for 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Archives of Environmental Contamination and Toxicology*, 28(3):366–377, 1995.

Y. A. Ibad-zade and S. P. Ghosh (translator). *Movement of Sediment in Open Channels*, volume 49 of *Russian Translations*. A. A. Balkema, Rotterdam, The Netherlands, 1992.

A. Iriarte, I. de Madariaga, F. Diez-Garagarza, M. Revilla, and E. Orive. Primary plankton production, respiration and nitrification in a shallow temperate estuary during summer. *Journal Of Experimental Marine Biology And Ecology*, 208(1-2): 127–151, Jan. 1997.

E. H. Isaaks and M. M. Srivastava. *An Introduction to Applied Geostatistics*. Oxford University Press, 1989.

J. D. Istok and C. A. Rautman. Probabilistic assessment of ground-water contamination .2. Results of case study. *Ground Water*, 34(6):1050–1064, Nov. 1996.

R. B. Jackson and M. M. Caldwell. Geostatistical patterns of soil heterogeneity around individual perennial plants. *The Journal of Ecology*, 81(4):683–692, 1993.

R. Jepsen, J. McNeil, and W. Lick. Effects of gas generation on the density and erosion of sediments from the grand river. *Journal of Great Lakes Research*, 26(2): 209–219, 2000.

A. G. Journel. Nonparametric estimation of spatial distributions. *Mathematical Geology*, 15(3):445–468, 1983.

A. G. Journel. The deterministic side of geostatistics. *Journal of the International Association for Mathematical Geology*, 17(1):1–15, 1985.

A. G. Journel and C. J. Huijbregts. *Mining Geostatistics*. The Blackburn Press, Caldwell, NJ, 1978.

A. G. Journel, P. C. Kyriakidis, and S. Mao. Correcting the smoothing effect of estimators: A spectral postprocessor. *Mathematical Geology*, 32(7):787–813, 2000.

W. K. Jung, N. R. Kitchen, K. A. Sudduth, and S. H. Anderson. Spatial characteristics of claypan soil properties in an agricultural field. *Soil Science Society Of America Journal*, 70(4):1387–1397, July 2006.

A. Kandiah. *Fundamental Aspects of Surface Erosion of Cohesive Sediments*. PhD thesis, University of California - Davis, 1974.

S. W. Karickhoff. Organic pollutant sorption in aquatic systems. *Journal of Hydraulic Engineering*, 110(6):707–735, 1984.

S. W. Karickhoff, D. S. Brown, and T. A. Scott. Sorption of hydrophobic pollutants on natural sediments. *Water Research*, 13:241–248, 1979.

M. Kent, R. A. Moyeed, C. L. Reid, R. Pakeman, and R. Weaver. Geostatistics, spatial rate of change analysis and boundary detection in plant ecology and biogeography. *Progress In Physical Geography*, 30(2):201–231, Apr. 2006.

J. P. C. Kleijnen. Sensitivity analysis and related analyses: a review of some statistical techniques. *Stat. Comput. Simul.*, 57:111–142, 1997.

D. G. Krige and W. Assibey-Bonsu. Comment on "correcting the smoothing effect of estimators: A spectral postprocessor" by A.G. Journel, P.C. Kyriakidis, and S. Mao. *Mathematical Geology*, 33(6):761–764, 2001.

R. B. Krone. Flume studies of the transport of sediment in estuarial shoaling processes. Final report. Technical report, Hydraulic Engineering Laboratory and Sanitary Engineering Research Laboratory, University of California, Berkeley., 1962.

R. B. Krone. Effects of bed structure on erosion of cohesive sediments. *J. Hydraulic Engineering*, 125(12):1297–1301, 1999.

A. Kukush. Kriging and prediction of nonlinear functionals. *Statistics*, 34(2):175–184, 2005.

P. Kyriakidis. Selecting panels for remediation in contaminated soils via stochastic imaging. In E. Baafi and N. Schofield, editors, *Geostatistics Wollongong '96*, volume 2, pages 973–983. Kluwer Academic Publishers, Dordrecht, the Netherlands, 1997.

P. C. Kyriakidis. A geostatistical framework for area-to-point spatial interpolation. *Geographical Analysis*, 36(3):259–290, 2004.

J. L. Lagrange. Essai d'une nouvelle methode pour determiner les maxima et les minima (digital version). In *Miscellanea Taurinensia*, pages 335–362. Göttinger Digitalisierungszentrum, digitized from the 1760–1761 publication. In French, directory in the digital version: Oeuvres de Lagrange — Oeuvres de Lagrange (Volume Tome 1) — Memoires Extraits des Recueils de L'Academie de Turin.

J. R. Landis and G. G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174, 1977.

Y. L. Lau. Influence of antecedent conditions of critical shear stress of bed sediments. *Water Research*, 34(2):663–667, 2000.

S.-C. Lee and A. J. Mehta. Cohesive sediment erosion. Technical report, Dredging Research Program (DRP) 94-6, US Army Corps of Engineers, 1994.

P. Legendre and L. Legendre. *Numerical Ecology*, volume 20 of *Developments in Environmental Modelling*. Elsevier, Amsterdam, the Netherlands, 2 edition, 1998.

S. A. Levin. The problem of pattern and scale in ecology. *Ecology*, 73:1943–1967, 1992.

W. Lick, H. Huang, and R. Jepsen. Flocculation of fine-grained sediments due to differential settling. *Journal of Geophysical Research*, 98(C6):10–279, 1993.

Y. B. Lin, Y. P. Lin, C. W. Liu, and Y. C. Tan. Mapping of spatial multi-scale sources of arsenic variation in groundwater on ChiaNan floodplain of Taiwan. *Science Of The Total Environment*, 370(1):168–181, Oct. 2006.

Y. P. Lin, C. C. Lee, and Y. C. Tan. Geostatistical approach for identification of transmissivity structure at Dulliu area in Taiwan. *Environmental Geology*, 40(1-2): 111–120, Dec. 2000.

I. Linkov, D. Burmistrov, J. Cura, and T. S. Bridges. Risk-based management of contaminated sediments: Consideration of spatial and temporal patterns in exposure modeling. *Environmental Science and Technology*, 36:238–246, 2002.

Z. Liu, Q. Shu, and Z. Wang. Applying pedo-transfer functions to simulate spatial heterogeneity of cinnamon soil water retention characteristics in western Liaoning province. *Water Resources Management*, 21(10):1751–1762, 2007.

R. Lohmann, E. Nelson, S. J. Eisenreich, and K. C. Jones. Evidence for dynamic air-water exchange of PCDD / Fs: A study in the Raritan Bay / Hudson River estuary. *Environ. Sci. Technol.*, 34:3086–3093, 2000.

E. R. Long. Spatial extent of sediment toxicity in US estuaries and marine bays. *Environ. Monitor. Assess.*, 64:391–407, 2000.

S. G. Longhitano and W. Nemec. Statistical analysis of bed-thickness variation in a Tortonian succession of biocalcarenitic tidal dunes, Amantea Basin, Calabria, southern Italy. *Sedimentary Geology*, 179(3-4):195–224, Aug. 2005.

D. R. Lovley, E. J. Phillips, and D. J. Lonergan. Enzymatic versus nonenzymatic mechanisms for Fe(III) reduction in aquatic sediments. *Environmental Science and Technology*, 25(6):1062–1067, Jun. 1991.

Y. Z. Ma. Comment on "application of spatial filter theory to kriging". *Mathematical Geology*, 25(3):399–403, 1993.

D. R. Maidment, editor. *Handbook of Hydrology*. McGraw-Hill, Inc., New York, NY, 1992.

C. D. Maravelias, D. G. Reid, E. J. Simmonds, and J. Haralabous. Spatial analysis and mapping of acoustic survey data in the presence of high local variability: Geostatistical application to North Sea herring (Clupea harengus). *Canadian Journal Of Fisheries And Aquatic Sciences*, 53(7):1497–1505, July 1996.

C. S. Martens and J. V. Klump. Biogeochemical cycling in an organic-rich coastal marine basin – I. Methane sediment-water exchange processes. *Geochimica et Cosmochimica Acta*, 44(3):471–490, 1980.

G. Matheron. Precision of exploring a stratified formation by boreholes with rigid spacing. In *International symposium on mineral resources*, pages 407–423, Paris, February 1961.

G. Matheron. *The Theory of Regionalized Variables and Its Applications*, volume 5 of *Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau*. École Nationale Supérieure des Mines de Paris, 1971.

A. B. McBratney, I. O. A. Odeh, T. F. A. Bishop, M. S. Dunbar, and T. M. Shatar. An overview of pedometric techniques for use in soil survey. *Geoderma*, 97(3-4): 293–327, Sept. 2000.

M. S. McBride and H. O. Pfannkuch. The distribution of seepage within lake beds. *United States Geological Society Journal of Research*, 3:505–512, 1975.

D. C. McKinney and M. D. Lin. Pump and treat ground-water remediation system optimization. *Journal of Water Resources Planning and Management-ASCE*, 122: 128–136, 1996.

J. C. Means, S. G. Wood, J. J. Hassett, and W. L. Banwart. Sorption of polynuclear aromatic hydrocarbons by sediments and soils. *Environmental Science & Technology*, 14(12):1524–1528, 1980.

Y. Mear, E. Poizot, A. Murat, P. Lesueur, and M. Thomas. Fine-grained sediment spatial distribution on the basis of a geostatistical analysis: Example of the eastern Bay of the Seine (France). *Continental Shelf Research*, 26(19):2335–2351, Dec. 2006.

J. E. Meisel and M. G. Turner. Scale detection in real and artificial landscapes using semivariance analysis. *Landscape Ecology*, 13(6):347–362, Dec. 1998.

M. Middelboe, R. N. Glud, F. Wenzhofer, K. Oguri, and H. Kitazato. Spatial distribution and activity of viruses in the deep-sea sediments of Sagami Bay, Japan. *Deep-Sea Research Part I-Oceanographic Research Papers*, 53(1):1–13, Jan. 2006.

H. J. Miller. Tobler's first law and spatial analysis. *Annals of the Association of American Geographers*, 94(2):284–289, 2004.

W. B. Mills, J. D. Dean, D. B. Porcella, S. A. Gherini, and R. J. M. Hudson. Water quality assessment: A screening procedure for toxic and conventional pollutants, parts 1 and 2. Technical Report EPA-600/6-85-002a and b, U.S. Environmental Protection Agency, Athens, GA, 1985.

P. A. P. Moran. Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2): 17–23, 1950.

D. R. Morgan, J. W. Eheart, and A. J. Valocchi. Aquifer remediation design under uncertainty using a new chance constrained programming technique. *Water Resources Research*, 29:551–561, 1993.

E. M. Murphy and J. A. Schramke. Estimation of microbial respiration rates in groundwater by geochemical modeling constrained with stable isotopes. *Geochim. Cosmochim. Acta*, 62(21):3395–3406, 1998.

C. J. Murray, H. J. Lee, and M. A. Hampton. Geostatistical mapping of effluent-affected sediment distribution on the Palos Verdes Shelf. *Continental Shelf Research*, 22:881–897, 2002.

T. E. Myers, R. P. Gambrell, and M. E. Tittlebaum. Design of an improved column leaching apparatus for sediments and dredged material, miscellaneous paper D-91-3. Technical report, U.S. Army Corps of Engineers Waterways Experiment Station, Vicksburg, MS, 1991.

S. J. Naber, B. E. Buxton, M. J. Bertoni, T. D. Scheibe, and J. Warren. U.S. EPA computer system for simulating site characterization activities at superfund cleanup sites. In E. Baafi and N. Schofield, editors, *Geostatistics Wollongong '96*, volume 2, pages 1066–1074. Kluwer Academic Publishers, Dordrecht, the Netherlands, 1997.

E. Naesset. Use of the weighted kappa coefficient in classification error assessment of thematic maps. *International Journal Of Geographical Information Systems*, 10 (5):591–603, July 1996.

N. Nanos, F. Pardo, J. A. Nager, J. A. Pardos, and L. Gil. Using multivariate factorial kriging for multiscale ordination: a case study. *Canadian Journal Of Forest Research-Revue Canadienne De Recherche Forestiere*, 35(12):2860–2874, Dec. 2005.

National Research Council. *Contaminated Sediments in Ports and Waterways: Cleanup Strategies and Technologies*. National Academy Press, Washington, DC, 1997.

NOAA. http://response.restoration.noaa.gov/cpr/watershed/watershedtools.html. National Oceanic and Atmospheric Administration / National Ocean Service, Office of Response and Restoration., Accessed Nov 2002.

B. L. Nowicki. The effect of temperature, oxygen, salinity, and nutrient enrichment on estuarine denitrification rates measured with a modified nitrogen gas flux technique. *Est. Coast Shelf Sci.*, 38:137–156, 1994.

R. A. Olea and V. Pawlowsky. Compensating for estimation smoothing in kriging. *Mathematical Geology*, 28(4):407–417, 1996.

B. Ollivier, P. Caumette, J.-L. Garcia, and R. A. Mah. Anaerobic bacteria from hypersaline environments. *Microbiological Reviews*, 58:27–38, 1994.

Y. Ouyang, J. Higman, J. Thompson, T. O'Toole, and D. Campbell. Characterization and spatial distribution of heavy metals in sediment from Cedar and Ortega rivers subbasin. *Journal Of Contaminant Hydrology*, 54(1-2):19–35, Jan. 2002.

Y. Ouyang, P. Nkedi-Kizza, R. S. Mansell, and J. Y. Ren. Spatial distribution of DDT in sediments from estuarine rivers of central Florida. *Journal Of Environmental Quality*, 32(5):1710–1716, Sept. 2003.

M. R. Palermo, T. A. Thompson, and F. Swed. White Paper No. 6B – In-situ capping as a remedy component for the lower Fox River – Response to a document by the Johnson Company: Ecosystem-based rehabilitation plan – An integrated plan for habitat enhancement and expedited exposure reduction in the lower Fox River and Green Bay. Technical report, Bureau for Remediation and Redevelopment, Wisconsin Department of Natural Resources, Dec. 2002.

C. E. Papadopoulos and H. Yeung. Uncertainty estimation and Monte Carlo simulation method. *Flow Measurement and Instrumentation*, 12:291–298, 2001.

T. M. Parchure and A. J. Mehta. Erosion of soft cohesive sediment deposits. *J. of Hydraulic Engineering*, 111(10):1308–1326, 1985.

E. Pardo-Igúzquiza. Comparison of geostatistical methods for estimating the areal average climatological rainfall mean using data on precipitation and topography. *International Journal of Climatology*, 18(9):1031–1047, 1998.

M. Pax-Lenney and C. E. Woodcock. The effect of spatial resolution on the ability to monitor the status of agricultural lands. *Remote Sensing of Environment*, 61(2): 210–220, 1997.

K. Petersen, E. Kristensen, and P. Bjerregaard. Influence of bioturbating animals on flux of cadmium into estuarine sediment. *Marine Environmental Research*, 45(4-5): 403–415, 1998.

J. J. Pignatello, F. J. Ferrandino, and L. Q. Huang. Elution of aged and freshly added herbicides from a soil. *Environmental Science & Technology*, 27(8):1563–1571, 1993.

D. Point, M. Monperrus, E. Tessier, D. Amouroux, L. Chauvaud, G. Thouzeau, F. Jean, E. Amice, J. Grall, A. Leynaert, J. Clavier, and O. F. X. Donard. Biological control of trace metal and organometal benthic fluxes in a eutrophic lagoon (Thau Lagoon, Mediterranean Sea, France). *Estuarine Coastal And Shelf Science*, 72(3): 457–471, Apr. 2007.

C. J. Proce, R. W. Ritzi, D. F. Dominic, and Z. X. Dai. Modeling multiscale heterogeneity and aquifer interconnectivity. *Ground Water*, 42(5):658–670, Sept. 2004.

C. Raghukumar, B. N. Nath, R. Sharma, P. A. L. Bharathi, and S. G. Dalal. Long-term changes in microbial and biochemical parameters in the central Indian basin. *Deep-Sea Research Part I-Oceanographic Research Papers*, 53(10):1695–1717, Oct. 2006.

M. H. Ramsey, P. D. Taylora, and J.-C. Lee. Optimized contaminated land investigation at minimum overall cost to achieve fitness-for-purpose. *J. Environ. Monit.*, 4:809–814, 2002.

T. M. Ravens and P. M. Gschwend. Flume measurements of sediment erodibility in boston harbor. *Journal of Hydraulic Engineering*, 125(10):998–1005, 1999.

J. W. Readman, R. F. Mantoura, and M. M. Rhead. A record of polycyclic aromatic hydrocarbon (PAH) pollution obtained from accreting sediments of the Tamar Estuary, UK: Evidence for non-equilibrium behaviour of PAH. *Science of The Total Environment*, 66:73–94, 1987.

L. N. Reddi and R. S. Govindaraju. Particle mobilization in sand-clay mixtures and facilitation of contaminant removal. In Y. B. Acar and D. E. Daniel, editors, *Geoenvironment 2000: Characterization, Containment, Remediation, and Performance in Environmental Geotechnics*, pages 1222–1236, New York, NY, Feb. 1995. ASCE.

D. Reible, D. Hayes, C. Lue-Hing, J. Patterson, N. Bhowmik, M. Johnson, and J. Teal. Comparison of the long-term risks of removal and in situ management of contaminated sediments in the Fox River. *Soil and Sediment Contamination*, 12(3):325–344, 2003.

D. D. Reible, K. T. Valsaraj, and L. J. Thibodeaux. Chemodynamic models for transport of contaminants from sediment beds. In *Handbook of Environmental Chemistry*, pages 187–228. Springer-Verlag, Heidelberg, 1991.

G. Y. Rhee, R. C. Sokol, B. Bush, and C. M. Bethoney. Long-term study of the anaerobic dechlorination of Aroclor 1254 with and without biphenyl enrichment. *Environmental Science & Technology*, 27(4):714–719, 1993.

D. C. Rhoads and L. F. Boyer. The effects of marine benthos on physical properties of sediments: A successional perspective. In P. McCall and M. Tevesz, editors, *Animal-Sediment Relations*, pages 3–52. Plenum Presss, 1982.

B. D. Ripley. *Spatial Statistics*. John Wiley & Sons, Inc., 2004.

J. A. Robbins, J. R. Krezoski, and S. C. Mozley. Radioactivity in sediments of the Great Lakes: Post-depositional redistribution by deposit-feeding organisms. *Earth and Planetary Science Letters*, 36(2):325–333, 1977.

G. P. Robertson, J. R. Crum, and B. G. Ellis. The spatial variability of soil resources following long-term disturbance. *Oecologia*, 96(4):451–456, 1993.

M. Robinson, D. Gallagher, and W. Reay. Field observations of tidal and seasonal variations in ground water discharge to tidal estuarine surface water. *Ground Water Monitoring and Remediation*, 18(1):83–92, 1998.

S. A. Romshoo. Geostatistical analysis of soil moisture measurements and remotely sensed data at different spatial scales. *Environmental Geology*, 45(3):339–349, Jan. 2004.

R. E. Rossi, P. W. Borth, and J. J. Tollefson. Stochastic simulation for characterizing ecological spatial patterns and appraising risk. *Ecological Applications*, 3:719–735, 1993.

F. Rothfuss and R. Conrad. Effect of gas bubbles on the diffusive flux of methane in anoxic paddy soil. *Limnol. Ocean.*, 43:1511–1518, 1998.

M. L. Ruffo, G. A. Bollero, R. G. Hoeft, and D. G. Bullock. Spatial variability of the illinois soil nitrogen test: Implications for soil sampling. *Agronomy Journal*, 97(6): 1485–1492, Nov. 2005.

D. W. Rutherford, C. T. Chiou, and D. E. Kile. Influence of soil organic matter composition on the partition of organic compounds. *Environmental Science & Technology*, 26(2):336–340, 1992.

H. Saito. *Geostatistical Data Fusion to Characterize Contaminated Sites for Remediation*. PhD thesis, University of Michigan, 2003.

H. Saito and P. Goovaerts. Selective remediation of contaminated sites using a two-level multiphase strategy and geostatistics. *Environ. Sci. Technol.*, 37:1912–1918, 2003.

L. C. Schaffner, R. J. Diaz, C. R. Olsen, and I. L. Larsen. Faunal characteristics and sediment accumulation processes in the James River estuary, Virginia. *Estuarine, Coastal and Shelf Science*, 25(2):211–226, 1987.

U. Schnabel and O. Tietje. Explorative data analysis of heavy metal contaminated soil using multidimensional spatial regression. *Environmental Geology*, 44(8):893–904, Nov. 2003.

B. Schumacher, J. J. van Ee, and E. J. Englund. Soil sampling at hazardous waste sites. In R. A. Meyers, editor, *Encyclopedia of Environmental Analysis and Remediation*, pages 4498–4517. John Wiley & Sons, Inc., 1998.

J. Schweizer and K. Kronholm. Snow cover spatial variability at multiple scales: Characteristics of a layer of buried surface hoar. *Cold Regions Science And Technology*, 47(3):207–223, Mar. 2007.

M. L. Smith and R. E. Williams. Examination of methods for evaluating remining a mine waste site. Part I. Geostatistical characterization methodology. *Engineering Geology*, 43(1):11–21, 1996a.

M. L. Smith and R. E. Williams. Examination of methods for evaluating remining a mine waste site. Part II. Indicator kriging for selective remediation. *Engineering Geology*, 43(1):11–21, 1996b.

SPAWAR Systems Center (SSC) San Diego and Battelle Science & Technology International. *Implementation Guide for Assessing and Managing Contaminated Sediment at Navy Facilities (UG-2053-ENV)*. Washington, DC, Mar. 2003.

S. Squire, M. H. Ramsey, and M. J. Gardner. Collaborative trial in sampling for the spatial delineation of contamination and the estimation of uncertainty. *Analyst*, 125:139–145, 2000a.

S. Squire, M. H. Ramsey, M. J. Gardner, and D. Lister. Sampling proficiency test for the estimation of uncertainty in the spatial delineation of contamination. *Analyst*, 125:2026–2031, 2000b.

S. B. Stark, R. G. Mohanty, and J. M. ver Hoef. Methods and systems for plant performance analysis. U.S. Patent # 6662185, Dec 2003.

T. H. Starks. Determination of support in soil sampling. *Mathematical Geology*, 18 (6):529–537, 1986.

L. J. Steinberg, K. H. Reckhow, and R. L. Wolpert. Characterization of parameters in mechanistic models: A case study of a PCB fate and transport model. *Ecological Modeling*, 97:35–46, 1997.

R. L. Steven and S. Ekermo. Sedimentation and erosion in connection with ship traffic, Goteborg Harbour, Sweden. *Environmental Geology*, 43(4):466–475, Feb. 2003.

C. E. Stivers and R. Sullivan. Restoration and capping of contaminated sediments. In *The 2nd International Conference on Dredging and Dredged Material Placement. Part 2 (of 2)*, pages 1017–1026, Lake Buena Vista, FL, Nov. 1994.

Y. Z. Su, Y. L. Li, and H. L. Zhao. Soil properties and their spatial pattern in a degraded sandy grassland under post-grazing restoration, inner Mongolia, northern China. *Biogeochemistry*, 79(3):297–314, July 2006.

D. Z. Sui. Tobler's first law of geography: A big idea for a small world? *Annals of the Association of American Geographers*, 94(2):269–277, 2004.

G. W. Suter. Statistics and risk assessment. In *1999 Lukacs Symposium: Frontiers of Environmental and Ecological Statistics for the 21st Century – Synergistic Challenges, Opportunities and Directions for Statistics, Ecology, Environment, and Society*, Bowling Green, Ohio, Apr. 1999. Department of Mathematics and Statistics, Bowling Green State University.

Syracuse Research Corporation and National Oceanic and Atmospheric Administration. Interpretive summary of existing data relevant to potential contaminants of concern within the Anacostia River watershed. SRC No.: FA292. Technical report, (SRC and NOAA), Jun. 2000.

A. E. Tercan. Global recoverable reserve estimation by covariance matching constrained kriging. *Energy Sources*, 26(12):1177–1185, Oct. 2004.

L. J. Thibodeaux. *Environmental Chemodynamics: Movement of Chemicals in Air, Water, and Soil*. John Wiley & Sons, Inc., New York, second edition, 1996.

L. J. Thibodeaux and K. T. Duckworth. The effectiveness of environmental dredging: A study of three sites. *Remediation Journal*, 11(3):5–33, Aug 2001.

W. R. Tobler. A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46:234–240, 1970.

P. D. Turney. Cost-sensitive classification: Empirical evaluation of a hybrid genetic decision tree induction algorithm. *Journal of Artificial Intelligence Research*, 2: 369–409, 1995.

U. S. EPA. *EPA's contaminated sediment management strategy*. U. S. Government Printing Office, Washington, DC, 1998.

U. S. EPA. *Guidance for Data Quality Assessment: Practical Methods for Data Analysis (QA/G-9QA00 Update)*. U. S. Government Printing Office, Washington, DC, 2000.

U. S. EPA. http://www.epa.gov/r02earth/superfnd/sedsamp.htm. Accessed Nov 2002, 2002.

U. S. EPA. *Contaminated Sediment Remediation Guidance for Hazardous Waste Sites*. Washington, DC, 2005.

U. S. EPA. http://www.epa.gov/storet/metadata.html. Accessed Dec 2007, 2007.

R. D. Valley, M. T. Drake, and C. S. Anderson. Evaluation of alternative interpolation techniques for the mapping of remotely-sensed submersed vegetation abundance. *Aquatic Botany*, 81(1):13–25, Jan. 2005.

M. M. R. van der Loeff, L. G. Anderson, P. O. J. Hall, A. Iverfeldt, A. B. Josefson, B. Sundby, and S. F. G. Westerlund. The asphyxiation technique: An approach to distinguishing between molecular diffusion and biologically mediated transport at the sediment-water interface. *Limnology and Oceanography*, 29(4):675–686, 1984.

M. van Meirvenne and P. Goovaerts. Accounting for spatial dependence in the processing of multi-temporal SAR images using factorial kriging. *International Journal Of Remote Sensing*, 23(2):371–387, Jan. 2002.

T. C. E. van Weering, G. T. Klaver, and R. A. Prins. Gas in marine sediments – An introduction. *Marine Geology*, 137:1–3, 1997.

D. J. Velinsky and J. T. F. Ashley. Deposition and spatial distribution of sediment-bound contaminants in the Anacostia River, District of Columbia. report no. 01-30d. Technical report, Patrick Center for Environmental Research, The Academy of Natural Sciences, Philadelphia, PA, 2001.

J. M. ver Hoef, N. Cressie, R. N. Fisher, and T. J. Case. Uncertainty and spatial linear models for ecological data. In C. Hunsaker, M. Goodchild, M. Friedl, and T. Case, editors, *Spatial uncertainty for ecology: Implications for remote sensing and GIS applications*, chapter 10, pages 214–237. Springer-Verlag, 2001.

X. J. Wang, R. M. Liu, K. Y. Wang, J. D. Hu, Y. B. Ye, S. C. Zhang, F. L. Xu, and S. Tao. Application of multivariate spatial analysis in scale-based distribution and source study of PAHs in the topsoil: An example from Tianjin, China. *Environmental Geology*, 49(8):1208–1216, Apr. 2006.

D. Weber and E. Englund. Evaluation and comparison of spatial interpolators. *Mathematical Geology*, 24(4):381–391, 1992.

A. W. Western, S. L. Zhou, R. B. Grayson, T. A. McMahon, G. Bloschl, and D. J. Wilson. Spatial correlation of soil moisture in small catchments and its relationship to dominant spatial hydrological processes. *Journal Of Hydrology*, 286(1-4):113–134, Jan. 2004.

R. A. Wheatcroft, I. Olmez, and F. X. Pink. Particle bioturbation in Massachusetts Bay: Preliminary results using a new deliberate tracer technique. *Journal of Marine Research*, 52(6):1129–1150, 1994.

J. Widdows, M. D. Brinsley, P. N. Salkeld, and M. Elliott. Use of annular flumes to determine the influence of current velocity and bivalves on material flux at the sediment-water interface. *Estuaries*, 21(4):552–559, 1998.

T. C. Winter. Subaqueous capping and natural recovery: Understanding the hydrogeologic setting at contaminated sites. In *DOER Technical Notes Collection (TN DOER-C26)*. U. S. Army Engineer Research and Development Center, Vicksburg, MS, 2002.

T. C. Winter, J. W. Harvey, O. L. Franke, and W. M. Alley. *Ground Water and Surface Water: A Single Resource*. U. S. Geological Survey (USGS), 1998.

C. E. Woodcock and A. H. Strahler. The factor of scale in remote sensing. *Remote Sensing of Environment*, 21(3):311–332, 1987.

J. Wu, W. A. Norvell, D. G. Hopkins, and R. M. Welch. Spatial variability of grain cadmium and soil characteristics in a durum wheat field. *Soil Science Society Of America Journal*, 66(1):268–275, Jan. 2002.

S. C. Wu and P. M. Gschwend. Sorption kinetics of hydrophobic organic compounds to natural sediments and soils. *Environmental Science & Technology*, 20(7):717–725, 1986.

J. K. Yamamoto. Correcting the smoothing effect of ordinary kriging estimates. *Mathematical Geology*, 37(1):69–94, 2005.

E. H. Yoo and P. C. Kyriakidis. Area-to-point kriging with inequality-type data. *Journal of Geographical Systems*, 8(4):357–390, 2006.

P. J. Zarco-Tejada and J. R. Miller. Land cover mapping at BOREAS using red edge spectral parameters from CASI imagery. *Journal of Geophysical Research*, 104(D22):27921–27934, 1999.

J. L. Zhou, T. W. Fileman, S. Evans, P. Donkin, J. W. Readman, R. F. C. Mantoura, and S. Rowland. The partitioning between flouranthrene and pyrene between suspended particles and dissolved phase in the Humber Estuary: A study of the controlling factors. *Science of the Total Environment*, 243-244:305–321, 1999.