Comparison of Spatial Interpolation Methods for Water Quality Evaluation in the Chesapeake Bay

Rebecca R. Murphy[1]; Frank C. Curriero[2]; and William P. Ball, M.ASCE[3]

Abstract: Spatial interpolation methods are frequently used to estimate values of physical or chemical constituents in locations where they are not measured. Very little research has been conducted, however, to investigate the relative performance of different interpolation methods in surface waters. The study reported here uses archived water quality data from the Chesapeake Bay to compare three spatial interpolation methods: inverse distance weighting, ordinary kriging, and a universal kriging method that incorporates output from a process-based water quality model. Interpolations were performed on salinity, water temperature, and dissolved oxygen "snap shots" (cruise-based data sets) taken between 1985 and 1994 at 21 different depths for multiple locations in the mainstem Bay, using data compiled by the prototypical Chesapeake Bay Environmental Observatory. The kriging methods generally outperform inverse distance weighting for all parameters and depths. Incorporating output from the water quality model through universal kriging appears to improve some of the interpolations by specifically accounting for some physical and biogeochemical features of the estuary. Such integration of process-based information with statistical interpolation warrants further study.

DOI: 10.1061/(ASCE)EE.1943-7870.0000121

CE Database subject headings: Chesapeake Bay; Dynamic models; Spatial analysis; Statistics; Water quality.

Author keywords: Kriging; Interpolation; Water quality modeling; Salinity; Temperature; Dissolved oxygen; Chesapeake.

Introduction

Spatial interpolation, or prediction, is used in water bodies such as the Chesapeake Bay to estimate the value of a constituent of interest in a location where it is not measured based on measurements taken at other locations. Many different spatial interpolation methods have been developed, including both stochastic (or statistical) methods such as kriging and nonstatistical methods such as inverse distance weighting (IDW). The differences between some of these methods, including the ones used here, are described by Cressie (1989). Commonly, stochastic methods are referred to as spatial prediction; whereas nonstatistical methods frequently are called spatial interpolation. For ease of comparison in this study, all spatial predictions or interpolations will be, for the most part, referred to as interpolations.

In many applications spatial interpolation is used as an intermediate tool in scientific study, but the study of interpolation methods is rarely an end goal in itself. For instance, research studies have used various spatial interpolation methods to address specific hypotheses about Chesapeake Bay water quality such as historical dissolved oxygen (DO) levels (Hagy et al. 2004) and blue crab abundance (Jensen et al. 2006). Current research efforts are using interpolation of historical data to investigate the trend in the volume of Chesapeake Bay water that is hypoxic, or has depleted DO, and these efforts have helped to motivate the research presented here (CBEO Project Team 2008). In addition, the Chesapeake Bay management community uses spatial interpolation of observations to assess water quality issues such as depleted DO levels, reduced water clarity, and reduced submerged aquatic vegetation. For example, spatially interpolated maps of water quality parameters are a necessary input for application of EPA's adopted cumulative frequency diagram (CFD) approach to characterize the spatial and temporal extent of criteria exceedances (USEPA 2007). Currently, the Chesapeake Bay and Tidal Tributary Interpolator tool (Bahner 2006) that uses an IDW algorithm is recommended for use in the CFD water quality criteria assessment. IDW is commonly used and is easy to automate, and this particular tool has been carefully developed to work efficiently for this application; however, EPA recognizes that kriging could be a more accurate option and should be tested more for use in the Bay (USEPA 2007).

Numerous studies outside the Chesapeake Bay have compared the performance of kriging and IDW, both in site-specific applications and in application to synthetic data. A list of some of these studies is provided by Zimmerman et al. (1999) who performed their own evaluation using synthetic data that showed kriging to outperform IDW for various types of data sets. Mueller et al. (2004) also provide a summary of other interpolation comparison studies, and, in their own comparison of methods for interpolating soil samples, generate varied results as to which interpolators per-

[1]Ph.D. Student, Dept. of Geography and Environmental Engineering, Johns Hopkins Univ., 3400 N. Charles St., Baltimore, MD 21218 (corresponding author).

[2]Associate Professor, Dept. of Environmental Health Sciences and Dept. of Biostatistics, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe St., Baltimore, MD 21205.

[3]Professor, Dept. of Geography and Environmental Engineering, Johns Hopkins Univ., 3400 N. Charles St., Baltimore, MD 21218.

Note. This manuscript was submitted on November 6, 2008; approved on June 29, 2009; published online on July 1, 2009. Discussion period open until July 1, 2010; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Environmental Engineering***, Vol. 136, No. 2, February 1, 2010. ©ASCE, ISSN 0733-9372/2010/2-160–171/$25.00.

form best. In these and other studies comparing IDW to kriging, results appear to be very study and site specific, with as yet no consensus as to a superior or preferred method. To our knowledge, no comparisons of IDW and kriging have been heretofore conducted for the Chesapeake Bay.

Estuaries, including the Chesapeake Bay, present some unique challenges for the implementation of spatial interpolation methods. Complex hydrodynamics and biochemistry create a system where observations from a relatively small sample set might not capture the range of conditions in the Bay. The Chesapeake Bay is a partially stratified estuary, meaning that it is fairly common for a density gradient to develop between less dense fresher water flowing toward the ocean on the top of the water column and denser, saltier water flowing in from the ocean on the bottom of the water column (Schubel and Pritchard 1987). This density gradient, or pycnocline, varies in depth throughout the Chesapeake Bay and is but one example of the complex physical dynamics that can affect the chemical concentrations of constituents of concern such as DO, chlorophyll, and nutrients (Day et al. 1989).

Considering the characteristics of the Chesapeake Bay, it is reasonable to expect that interpolation based solely on the static measurements taken during a "snapshot" of the Bay will not perform well alone, and that there will be value in incorporating knowledge about the Bay's physical and biochemical dynamics. Within this context, deterministic numerical models of the dynamics of Chesapeake Bay water flow and quality have been developed and applied over many decades and for a wide variety of purposes. In particular, numerical models of Chesapeake Bay hydrodynamics (Johnson et al. 1993; Wang and Johnson 2000) and water quality (Cerco and Cole 1993; Cerco and Meyers 2000; Cerco and Noel 2004) have been developed, calibrated, and updated through collaborative efforts between the U.S. EPA and the Army Corps of Engineers. These numerical models have been run over many years to connect information about watershed, oceanic, and atmospheric boundary conditions with complex physical, chemical, and biological dynamics and interactions in the Chesapeake Bay to simulate water quality with respect to numerous state variables, including DO as affected by algal growth and subsequent decay. Incorporation of the output from these models that represents the spatially varying effects of the hydrodynamics, bathymetry, and biochemisty of the Chesapeake Bay into interpolation of static water quality values could potentially improve the interpolation of measurements. At the same time, the outputs from the hydrodynamic and water quality models do not precisely match the instantaneous water quality observations of interest. (These are highly complex models with long run times, such that calibration is an inexact science that uses only selected data sets.) Thus, the observations are still the best measures of reality, although the model can serve to better inform the interpolation between the sometimes distant points of observation.

The idea of incorporating dynamic model output into interpolation of observations has been explored in other applications. Goovaerts et al. (2008) used the deposition output from an air quality model, which takes into account meteorological conditions, topography, and source characteristics, to aid in the interpolation of soil concentrations of dioxin. Rivest et al. (2008) found that incorporation of output from a groundwater model, which takes into account the physics of groundwater flow, improved "the precision and the realism" of the hydraulic head interpolations for both a synthetic case study and an application in a real earth dam. Shlomi and Michalak (2007) incorporated transport information from a groundwater model into kriging of a groundwater plume. In a surface water quality application, chlorophyll *a* from a water quality model (WQM) was used to aid in the interpolation of chlorophyll *a* measurements throughout a system of reservoirs and resulted in reduced uncertainty in interpolation estimates (LoBuglio et al. 2007). Although each of these studies applied different methods of interpolation, the external trend kriging method used by Rivest et al. (2008) is equivalent to the universal kriging (UK) method used for the work presented here.

Specifically, this paper describes our own comprehensive comparison of two approaches of kriging with each other and with the more commonly applied IDW method using data sets collected over a 10-year period in the Chesapeake Bay for salinity, water temperature, and DO. Our study also incorporated some recently archived and stored dynamic model output in order to provide interpolations using a UK method designed to test the hypothesis that improved estimates can be obtained through the combined use of both water quality measurements and predictive output from a deterministic WQM. One of the challenges for implementing kriging in flowing water bodies is meeting the assumption of stationarity—that is, that we assume a constant mean for our parameter of interest (salinity, water temperature, or DO) and model all variability as fluctuations around this constant mean with a spatially varying stochastic process that is a function only of location. In UK, the constant mean assumption is replaced with a linear trend and the assumption of stationarity is then imposed on the residuals. In this context, we note that the necessary requirement is for "intrinsic stationarity," which is a requirement more easily met than that of "second-order stationarity" (see Cressie 1993 for details). As in the majority of geostatistical applications, it is sometimes difficult to prove this assumption of intrinsic stationarity of the stochastic random field with only one partial realization of the spatial process (Cressie 1993). Other researchers have implemented kriging in surface water applications and have addressed the assumption of stationarity by either splitting the data into regions where stationarity can more reliably be assumed (Simard et al. 1992) or fitting a spatial trend to the data to remove nonstationarity (Georgakarakos and Kitsiou 2008). We addressed the stationarity assumption in this study with both of these approaches, by splitting the data by depth of collection and by fitting a spatially varying trend in the data through UK.

Some of the unique aspects of this study are its application in an estuary; its large scope of comparisons across time, depth, and parameters; its comparison of methods currently being considered for management applications; and its integration of increasingly available dynamic model output to aid in analysis of observed water quality. Results presented here provide a starting point for performing further research and making informed decisions on spatial interpolation in the Chesapeake Bay.

Methods

Study Area and Water Quality Data

The Chesapeake Bay, a large partially mixed estuary on the eastern coast of the United States, extends from the Susquehanna River approximately 320 km to the Atlantic Ocean near Virginia Beach, Va. Its width ranges from 5.5 km near Aberdeen, Md., to 56 km near the mouth of the Potomac River [Chesapeake Bay Program (CBP) 2008a]. This study focuses on the mainstem of the Bay that was defined using segments from the CBP 2003 Analytical Segmentation Scheme (USEPA and Region III Chesapeake Bay Program Office 2005). The analysis area is shown as

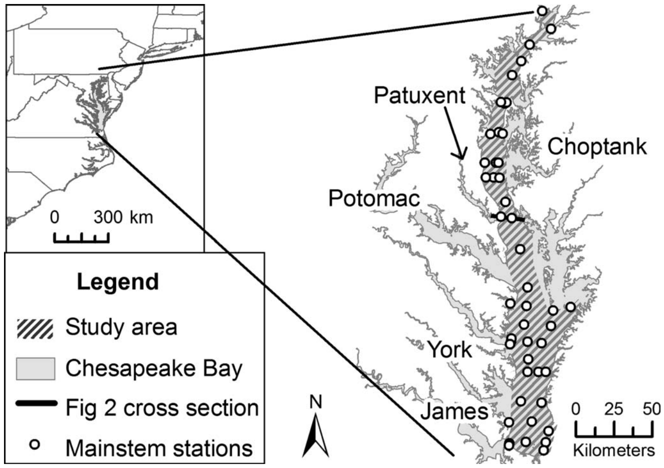

Fig. 1. Chesapeake Bay, study area, and stations

the hatched region of Fig. 1. The mainstem Bay was selected to avoid complications of spatially interpolating within nonconvex, or irregularly shaped, regions (Curriero 2007). This latter topic is also of considerable interest but is beyond the scope of our current study. In a similar way, the present study also does not focus on the aforementioned issue of strong anisotropy due to stratification, which is critically important to properly account for when making 3D interpolations. Therefore, 2D interpolations were studied here, using data collected from bins that were either 1 or 2 m in depth, selected to match the vertical sampling density. Fig. 2 shows the vertical segmentation of the Bay used in this study at a lateral transect, indicated on Fig. 1, near the mouth of the Patuxent River. This vertical segmentation of samples helped to account for stationarity in the data sets because the mean values of these parameters changed greatly with depth.

Measurements of salinity, water temperature, and DO from 1985 to 1994 that have been compiled and stored by the CBP were used in this study (CBP 2008b). The data compiled by the CBP (a regional partnership between Bay states, the federal government, and other stakeholder groups) is collected by participating agencies and universities in Maryland and Virginia (CBP 1993). For this ongoing program, salinity, water temperature, and DO are all measured in a vertical profile at a set of fixed station

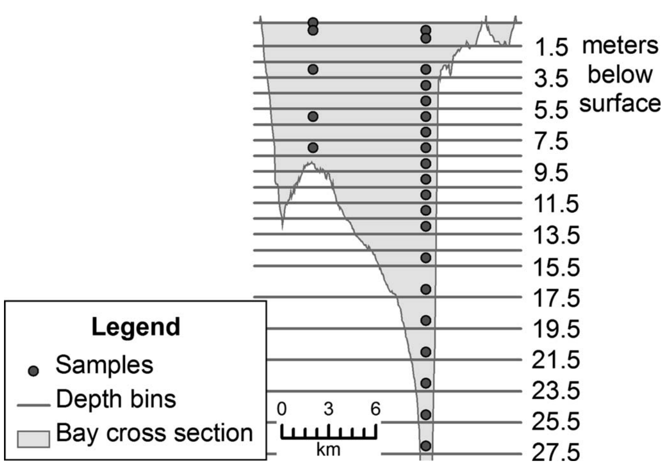

Fig. 2. Lateral cross section at Patuxent River mouth showing delineations used to split data sets by depth and corresponding sample depths at this location

locations (see Fig. 1) using an in situ probe (see CBP Water Quality Data Dictionary for probe information, CBP 2008b). In addition, the DO probe-based measurements are validated once a day with a Winkler titration (CBP 1993). Other water quality data have been collected throughout the life of this program, but these three parameters were selected for the current study because of their comprehensive coverage and range of geochemical complexity from simple (temperature and salinity) to highly complex (DO). DO in particular was selected because of its importance as a water quality parameter and the especially strong need for interpolation in the context of meeting water quality criteria, as discussed earlier (USEPA 2007).

CBP data collection is organized into "cruises" or periods of time with coordinated sampling among the states. Specific cruise dates are documented online (CBP 2008b) and consist of one cruise each month from November to February, and two cruises per month in each of the other 8 months. In total, data from 200 cruises were used in this study. Samples were taken at approximately the same locations in every cruise during the years of interest, although there were some exceptions. For most cruises included in this study, there were about 40 locations sampled within the mainstem monitoring segments. The number of stations used for each interpolation was an average of 40 at shallow depths, 20 at middepths, and 11 at bottom depths. The difference in sample number with depth is predominantly the result of bathymetry of the Bay—most of the Bay is fairly shallow and deep locations simply do not exist for many stations. Sample sets for a particular cruise and depth were not analyzed in this study if there were fewer than 11 samples taken at that depth during the cruise. Eleven samples is a very small number for a spatial interpolation. However, the data sets with this small number of samples were almost always at the deepest locations in the Bay where the values of the water quality parameters being interpolated did not show large variability.

Chesapeake Bay Hydrodynamic and Water Quality Models

The Chesapeake Bay hydrodynamic model and WQM have been developed, refined, and used for almost two decades through collaboration between the U.S. Army Corps of Engineers and the CBP. The hydrodynamic model (CH3D-WES) computes 3D velocities, surface elevation, effective (turbulent) vertical viscosity and diffusivity, temperature, salinity and density on a grid covering the mainstem Bay, and the tidally influenced regions of its major tributaries (Johnson et al. 1993; Wang and Johnson 2000). This information, as well as input from a watershed model (USEPA 2008), is fed into the WQM (CE-QUAL-ICM) to simulate nutrient cycling, algal biomass, DO, and many other water quality parameters. The WQM has been developed in phases, as documented by Cerco and Cole (1993), Cerco and Meyers (2000), and Cerco and Noel (2004). For this study, water quality and hydrodynamic predictive output from the 12,920 grid cell version of the WQM was obtained from the Army Corps of Engineers for a 10-year calibration run from 1985–1994. Along with many other parameters, the output contains estimates of salinity, water temperature, and DO values every 2 h in each model grid cell. Cells are an average 6.4 km^2 in surface area and are all 1.5 m deep, with the exception of cells in the surface layer, whose depth varies (e.g., with tides and storm events). Cells containing a fixed monitoring station were identified, and the WQM's estimated salinity, water temperature, and DO were averaged over the two hour outputs (for the dates corresponding to a data collection cruise) so that the fixed station data and WQM output would share the same spatial and temporal resolution.

The WQM output and CBP observational data were accessed for this study through the Chesapeake Bay Environmental Observatory (CBEO) SQL server testbed. The CBEO is a prototype project funded by the 2005 National Science Foundation program "Cyberinfrastructure for Environmental Observatories: Prototype Systems to Address Cross-Cutting Needs (CEO:P)." A principle goal of this project has been to compile and use disparate data sets on the Chesapeake Bay to address scientific hypotheses (CBEO Project Team 2008). The use of the CBEO testbed made it possible to iteratively query and interpolate the data and model results stored together.

Interpolation Methods

The water quality data in the Chesapeake Bay were interpolated using three methods in this study: IDW and two versions of the geostatistical kriging method—ordinary kriging (OK) and UK. For notational purposes we let $\{y(s_1),\ldots,y(s_n)\}$ represent a set of observed water quality data at the fixed station locations denoted by $s_1,\ldots,s_n$. Furthermore, we let $Y(s_0)$ denote corresponding water quality at an unsampled (unmonitored) location s_0 for which an interpolated (predicted) value is sought. The three methods considered produce interpolations that are weighted averages of the observed data (i.e., linear)

$$\hat{Y}(s_0)=\sum_{i=1}^{n} w(s_i)y(s_i) \tag{1}$$

with $w(s_i)$ denoting the weights and $\hat{Y}(s_0)$ representing the interpolated value for $Y(s_0)$. For this study, interpolations using each method were produced at the centroid of each of the 12,920 WQM grid cells to spatially align interpolations with WQM output. All methods tested give more weight to those observations $[y(s_i)]$ closer to the location of the estimate, s_0. This is intuitive since most environmental data, and many other spatial data for that matter, exhibit positive spatial dependence, defined as observations closer together in space being more similar in value than those further away. Methods of interpolation differ, however, in regard to how the weights in Eq. (1) are determined.

For the nonstatistical IDW method of spatial interpolation considered here, the weights in Eq. (1) are fixed a priori to be a particular function solely of distance. A common weighting scheme employed in many geographic information systems is that of inverse distance squared, so the IDW interpolator can be written as in Eq. (1) with the following weights:

$$w(s_i)=\frac{d(s_0,s_i)^{-2}}{\sum_{i=1}^{n} d(s_0,s_i)^{-2}} \tag{2}$$

where $d(s_0,s_i)$ represents the distance between the location to be interpolated s_0 and the sample location s_i. For the case where $d(s_0,s_i)$ is zero, meaning that a sample location is coincident with an estimation location, the $w(s_i)$ for the coincident sample is set to 1 and all other $w(s_i)$ are set to zero.

Kriging is a method for performing statistically optimal spatial interpolations (Cressie 1993; Schabenberger and Gotway 2004; Diggle and Ribeiro 2007) that arose out of the field of geostatistics, a subfield of statistics. With kriging, the weights $w(s_i)$ are derived from building a model of the spatial correlation structure of the observations, as described briefly here and more fully in the references listed earlier.

The two types of kriging implemented in this study (OK and UK) can be differentiated most clearly by formulating the kriging method as a general linear regression model among measured values

$$Y(s)=\beta_0+\beta_1 X_1(s)+\varepsilon(s) \tag{3}$$

with s representing a generic spatial location assumed to vary continuously over some domain of interest, $Y(s)$ the outcome of interest measured at s, $X_1(s)$ as a potential covariate indexed by location s (note that it is plausible to have more than one covariate), β_1, as the covariate's associated regression effect, and $\varepsilon(s)$ as the random error term. Eq. (3) is generally referred to as a UK model and allows for the incorporation of auxiliary information $[X_1(s)]$ that is available at all locations into a linear regression with the observations. OK can be represented with the same Eq. (3), but with no covariates (Xs). The OK approach thus contains just the constant term β_0 and the error term $\varepsilon(s)$ accounting for all of the spatial variation in the observations $Y(s)$. The uncertainty in either the OK or UK regression approach is modeled with the random error term $\varepsilon(s)$ assumed to have zero mean, constant variance, and a multivariate correlation structure to accommodate spatial dependence. Information about weights $[w(s_i)]$ for the interpolation is obtained by characterizing the spatial structure of the error term, as described subsequently.

Historic development of kriging approaches that incorporate covariates, such as the $X_1(s)$ in Eq. (3), tended to distinguish when the covariates were just spatial coordinates, as in the original UK development of Matheron (1969), or auxiliary variables also measured or observed spatially throughout the domain (common examples are elevation, bathymetry, or soil type), which has been termed kriging with an external drift (Wackernagel 2003). Because there are no statistical advantages to distinguishing between methods based on whether the covariate used is a spatial coordinate or other information, we proceed by simply referring to Eq. (3) as the UK model. Thus, the term "UK" in this paper represents the inclusion of any auxiliary information as a covariate in the kriging model.

In this study, the WQM-generated concentrations of salinity, water temperature, or DO, which have been generated for all locations in the domain, served as the auxiliary information, or the covariate [i.e., $X_1(s)$], in the UK approach. For example, a UK interpolation of DO observations at one time period and depth can be described in two parts:

1. The observed DO concentrations at n locations were used as the Y values in Eq. (3), and the WQM-generated DO concentrations at the same n locations were used as the X_1 values. We solved for the coefficients β_0 and β_1 and the spatially varying error term, $\varepsilon(s)$, using the UK system of equations (see references and the following text for more details).
2. The fitted β_0 and β_1 coefficients, the spatially varying error, and the WQM-generated concentrations (X_1) at all nonsampled locations were used to estimate the DO ($\hat{Y}$) at all nonsampled locations.

Other combinations of covariates with observations were tested (e.g., modeled salinity as the covariate for observed DO or spatial coordinates as covariates), but did not perform well in comparison. With this formulation for the UK method, the expected value of a parameter (i.e., $\beta_0+\beta_1$[ModeledDO(s)]) varied by location and was a function of the WQM-generated concentration. Deviations from the expected value were accounted for with the residual error term, which also varied spatially. By accounting for a spatially varying expected value with UK, we have lessened the concerns regarding stationarity in a flowing water body.

For both the UK and OK forms of Eq. (3), common distributional assumptions on $\varepsilon(s)$ include normality and lognormality (Cressie 1993; Diggle and Ribeiro 2007). Functions of a specific mathematical type (positive definite) are used to represent the spatial correlation in $\varepsilon(s)$. Under the linear model representation, there is a one-to-one correspondence between the spatial correlation function and the well known variogram function traditionally used in geostatistics for characterizing spatial correlation. When the spatial structure (e.g., the variogram) for Eq. (3) is known, the statistically optimal kriging predictions for the variable Y at unsampled locations can be derived using standard statistical principles. For kriging, optimality criteria are applied in order to give best linear unbiased predictions (BLUPs) that are linear in the data, statistically unbiased, and minimize mean squared prediction error. The equation for the BLUPs can be written in the form of Eq. (1) and contains information from the covariates (the Xs), regression effects (the βs), and most important the correlation structure (Cressie 1993). The minimized mean squared prediction error is then used as a measure of prediction uncertainty. In practice, however, the true spatial structure of the parameter throughout the region of interest is unknown and is estimated from the spatial structure of the observations via an estimated variogram function, described next for this study. There are several methods available and commonly used to estimate variograms based on sampled data [see, for example, Cressie (1993) and Diggle and Ribeiro (2007)].

Kriging is technically and computationally more complex than IDW. Kriging differs from IDW in two important ways. First, for kriging the weighting scheme in Eq. (1) is determined by using the data to characterize the spatial structure of the constituent of interest (i.e., the variogram), whereas for IDW the weighting scheme with distance is fixed via Eq. (2). The other major difference is that a kriging interpolation results in a measure of statistical uncertainty that is not available from IDW.

Computational Details

Computations were performed with the statistical package R (R Development Core Team 2008) and the geoR contributed package (Ribeiro and Diggle 2008). Data were imported into the R computing environment from queries to the CBEO testbed SQL server using the database query functionality in the RODBC contributed package (Ripley and Lapsley 2008). Each query to the CBEO testbed returned: (1) the CBP data collected at all depths in the mainstem during one cruise; (2) the WQM output concentrations temporally averaged over the cruise dates in each model grid cell.

After importing the data and averaged WQM output into R for a cruise and parameter, the data were split into 21 sets corresponding to depth of collection (see Fig. 2). When more than one sample at a station fell within a depth bin, the samples were averaged to obtain a vertical average value at that depth. This only occurred commonly in the top bin (0–1.5 m) and bins below 15.5 m. After the samples were spilt by depth, data sets with 11 or more stations were interpolated. For the UK method, the location of each sample was mapped to a WQM grid cell, as described earlier. For each cruise, samples were also mapped to the grid cell they fell within vertically, with the exception of the occasional bottom samples that were deeper than the deepest grid cell. These were instead matched to the deepest grid cell at the same station location.

The computations involving the IDW interpolations were relatively straight forward. To perform kriging, a variogram model was estimated for each data set. We experimented extensively with individual data sets and variogram model options before settling on this method. Although such case-by-case user interaction is the suggested way to estimate variograms and perform kriging, the large number of interpolations performed for this comparative analysis made the development of an automated method necessary, which is also a problem of more wide spread practical interest. To account for many of the possible variogram shapes encountered in experimentation, each data set was automatically fit to both a spherical and a Matérn variogram model using inferences that were based on restricted maximum likelihood (Diggle and Ribeiro 2007; Cressie 1993). The best fitting model and variogram parameters (sill, range, and nugget) were selected uniquely for each data set by comparing restricted log-likelihood values.

The spherical model was selected as one model option because its form is encountered frequently and it is therefore widely used. The Matérn family of variogram models (Diggle and Ribeiro 2007) was chosen because of its flexibility to represent a wide variety of variogram shapes by changing its smoothness parameter. In this study, options for the smoothness parameter were set at either 0.5 (which is equivalent to an exponential variogram model), 1.0, or 2.5. For both the spherical and Matérn models, the sill and range variogram parameters were fit using a matrix of initial values generated to begin the maximization process. Comparison of results generated with alternative starting points ensured that the final estimates for these parameters were not affected by the initial estimate. The range of initial values tested for the variogram sill was between 0.5 and 1.5 times the variance of the data. Initial values for the variogram range were between 10 and 200% of the maximum interlocation distance. Initial values for the variogram nugget were set to zero. For the UK analysis, residual variance from the regression model [Eq. (3)] was used to determine initial values for the variogram sill. Normality was assumed throughout, as initial experiments with lognormality assumptions did not yield better interpolations. Initial experiments also supported the assumption made throughout the analysis of 2D isotropy (i.e., spatial variation is a function of only distance, not direction) within the evaluated depth layers. The variogram nugget, which was allowed to vary in the parameter fitting, can represent measurement error and/or microscale variation in the observations (Cressie 1993). In this study, we did not have sufficient information to positively confirm whether the nugget represented measurement error. The kriging equations were therefore set to be consistent with the assumption that the nugget error represents microscale variation, which is the more conservative assumption resulting in higher estimated uncertainty. We believe this assumption to be reasonable, in that very close measurements could be quite different due to the complex hydrodynamics of the system, whereas measurement error was controlled carefully with daily instrument calibration.

IDW and kriging (both ordinary and universal) interpolations were generated at the centroids of the WQM grid for each depth. For kriging, the kriged estimation variances were also obtained.

Table 1. Percent of Data Sets That Best Fit Each Possible Variogram Shape

Parameter	Method	Nugget (%)	Linear (%)	Matern 1.0 or 2.5 (%)	Exponential (%)	Spherical (%)
Salinity (n=3,663)	OK	0	69	27	0	3
	UK	7	28	5	10	50
Temperature (n=3,673)	OK	1	26	23	9	41
	UK	2	30	16	10	42
DO (n=3,645)	OK	1	28	15	10	45
	UK	3	25	15	10	48

Performance Evaluation

One method used for evaluating the performance of each interpolation scheme was cross validation. For this technique, one observation was removed from the data set and the interpolation was again performed to generate an estimate at the location of the removed value. This procedure was then repeated for each sample in a data set and for each such cross-validation run, the generated interpolated value, $\hat{Y}(s_i)$, was compared to the true observation, $Y(s_i)$. There are some limitations to using cross validation, especially to gauge performance at unsampled locations (Isaaks and Srivastava 1989). Because the sample locations were spread fairly evenly throughout the Bay (Fig. 1), the cross-validation statistics we employed are expected to be reasonable evaluation tools for comparing relative interpolator performance. In this study, the collection of estimated and observed values was considered in terms of several possible "error metrics" in order to evaluate each interpolator's performance. One such metric is the mean error (ME), which provided a measure of bias

$$\mathrm{ME} = \frac{1}{n}\sum_{i=1}^{n}[\hat{Y}(s_i) - Y(s_i)] \tag{4}$$

The mean absolute error (MAE) provided a measure of the absolute error of the estimates

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{Y}(s_i) - Y(s_i)| \tag{5}$$

The RMS error (RMSE) provided a measure of estimation error that was more sensitive to outliers than the MAE

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[\hat{Y}(s_i) - Y(s_i)]^2} \tag{6}$$

To compare performance across the three different water quality parameters, two scaled error measures were used: the Pearson correlation coefficient and the RMSE divided by the sample standard deviation. An additional measure that also evaluated the kriging statistical uncertainty [i.e., the kriging variance, $\hat{\sigma}^2(s_i)$] was used for both the OK and UK methods

$$\mathrm{StndRMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[(\hat{Y}(s_i) - Y(s_i))/\hat{\sigma}(s_i)]^2} \tag{7}$$

This standardized RMSE should equal one if the kriging variance is representative of the actual interpolation errors (Cressie 1993).

Results

Variograms

The options used to select the most appropriate variogram model to fit each data set allowed for a wide range of possible variogram shapes. The results shown in Table 1 reveal that no one variogram shape provided the best fit to all data sets. Examples of each variogram model, fitted to data using restricted maximum likelihood, are shown as the lines in each graph of Fig. 3. The points in

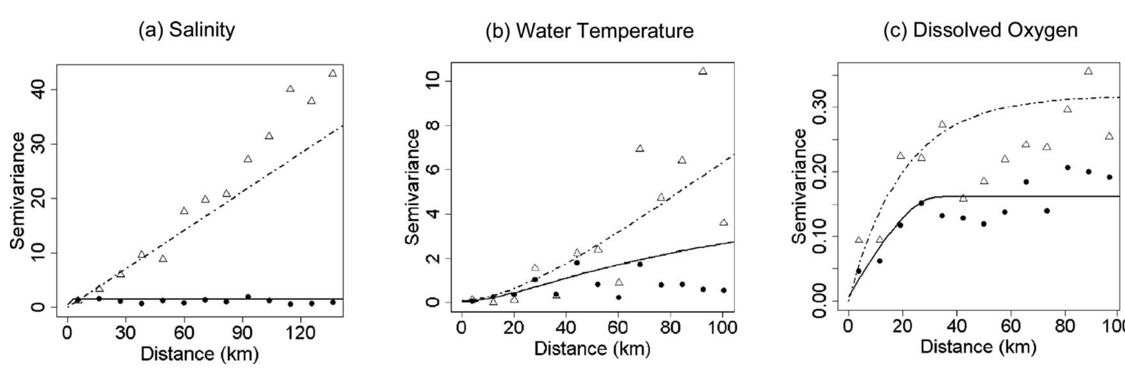

Fig. 3. Variogram estimates (triangles for OK; filled circles for UK) and fitted variogram models (dotted lines for OK; solid lines for UK) for three data sets plotted as the semivariance versus distance: (a) salinity at 5 m from Nov. 16–19, 1992; (b) water temperature at 13 m from June 17–19, 1985; and (c) DO at 0.75 m from March 9–12, 1993

these graphs represent bin-averaged variogram point estimates derived from the classic method of moments variogram estimator (Cressie 1993), and it is important to recognize that the fitted lines were not generated using these points but were instead generated by maximizing the likelihood function for all observed data (in this regard, the points in the graphs are only included to give further visual clues regarding variogram structure). The nugget-only model [Fig. 3(a), solid line] was not one of the two specific model choices (Matérn and spherical) used for this analysis, but was identified in the few cases when results from the Matérn or spherical model fit indicated no spatial correlation by a zero sill and a positive nugget. The linear model [Fig. 3(a), dotted line] was also not one of the specific model options; however, the Matérn and spherical model are each capable of showing an essentially linear fit when both the range and sill parameter values are very large. In this context, we labeled a variogram as linear if the fitted practical range was greater than 1.5 times the maximum distance between samples, meaning that the spatial correlation scale was much longer than the maximum sampling range. When a variogram appears to be linear, the assumption of second-order stationarity may no longer be appropriate. As previously noted, however, our method was not dependent on second-order stationarity, and intrinsic stationarity could still be a reasonable assumption, meaning kriging could still be performed. Nonlinear variograms were also commonly observed. For example, a Matérn variogram model with a smoothness parameter of 1.0 or 2.5 [see Fig. 3(b), both lines] has a slowly increasing slope near the origin, then an almost a linear fit at further distance. Exponential [Matérn with a smoothness of 0.5, Fig. 3(c), dotted line] and spherical models [Fig. 3(c), solid line] both increase at short distances and then level off.

The variogram fitting did reveal some trends. Variograms for the OK method for salinity frequently had a linear shape, whereas UK-based variograms (which incorporated the WQM-generated salinity values as a covariate) more commonly followed the spherical model. For water temperature and DO, the spherical model was most common, indicating that both parameters frequently exhibit spatial correlation that decreases with distance and then levels off at far distances. Comparison of UK and OK among all parameters revealed that the UK method had more data sets with a nugget model and fewer data sets with linear and Matérn models. The majority of the data sets studied were fit to either an exponential or spherical variogram shape. The practical ranges (approximate distance of spatial correlation) and sills (values at which the variograms level off) provided useful information about the distance and degree to which observations were spatially correlated and the degree to which the WQM-generated output (used as the covariate for the UK method) accounted for spatial correlation. For all three water quality constituents analyzed the median sill and range were less for the UK fitted variograms than for the OK fitted variograms. For DO, the median ranges were 146 km and 129 km for the OK and UK sets, respectively. For water temperature the median OK and UK ranges were 174 km and 167 km and for salinity were 2,470 km and 87 km, respectively. The OK salinity median range was the only range larger than the maximum inter-sample distance within the modeling domain (290 km), indicating that the range of spatial correlation for some of the salinity data extended beyond the model domain for these sets. As discussed subsequently, the shorter correlation range observed for the UK method (compared to the OK method) indicates that the WQM output is a useful UK covariate that can help account for some of the large-scale spatial variation in the samples.

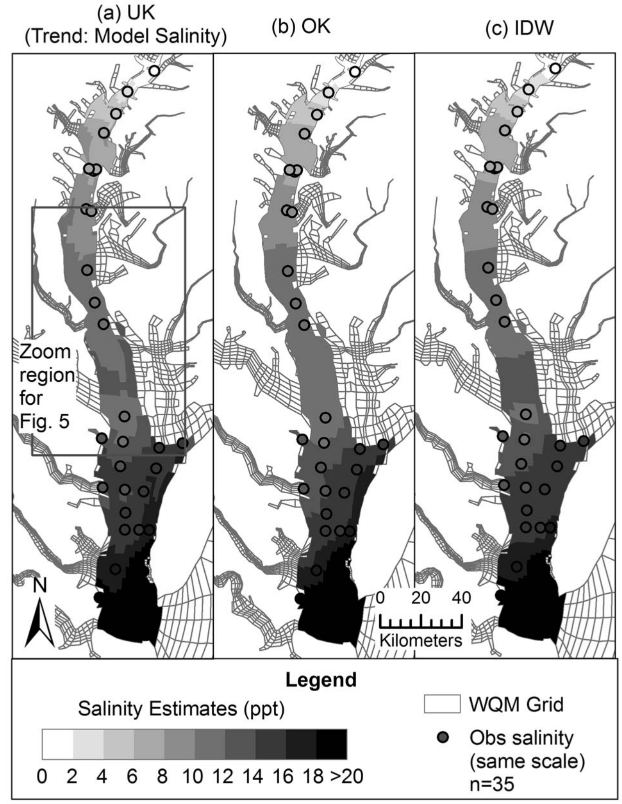

Fig. 4. Estimation of salinity at 2 meters deep from July 12–14, 1993, from (a) UK interpolation with the WQM-generated salinity as a covariate; (b) OK; and (c) IDW

Sample Interpolation Output

Figs. 4–6 are illustrations of selected interpolation results that show some differences between the methods for individual data sets. Fig. 4 is a presentation of salinity interpolation results for samples taken 1.5–2.5 m below the surface, July 12–14, 1993. All three sets of results demonstrate the salinity gradient in the Chesapeake Bay, but there are observable differences among the estimates in regions with few samples. The cross validation error terms for this interpolation are presented in Table 2, and show that the IDW and kriging interpolators did not perform too differently, although one of the kriging methods outperformed IDW for each error measure. The estimation uncertainties that were calculated

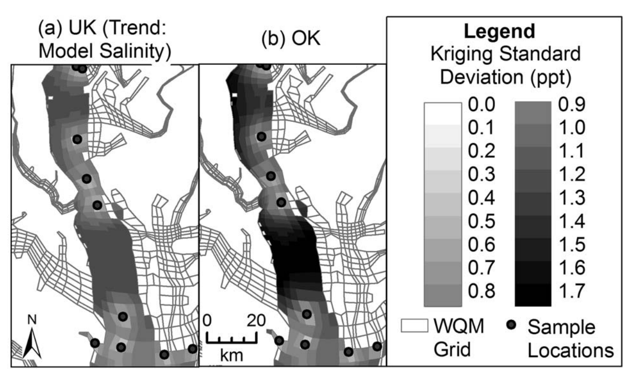

Fig. 5. Kriging standard deviations for interpolation results in selected area of Figs. 4(a and b)

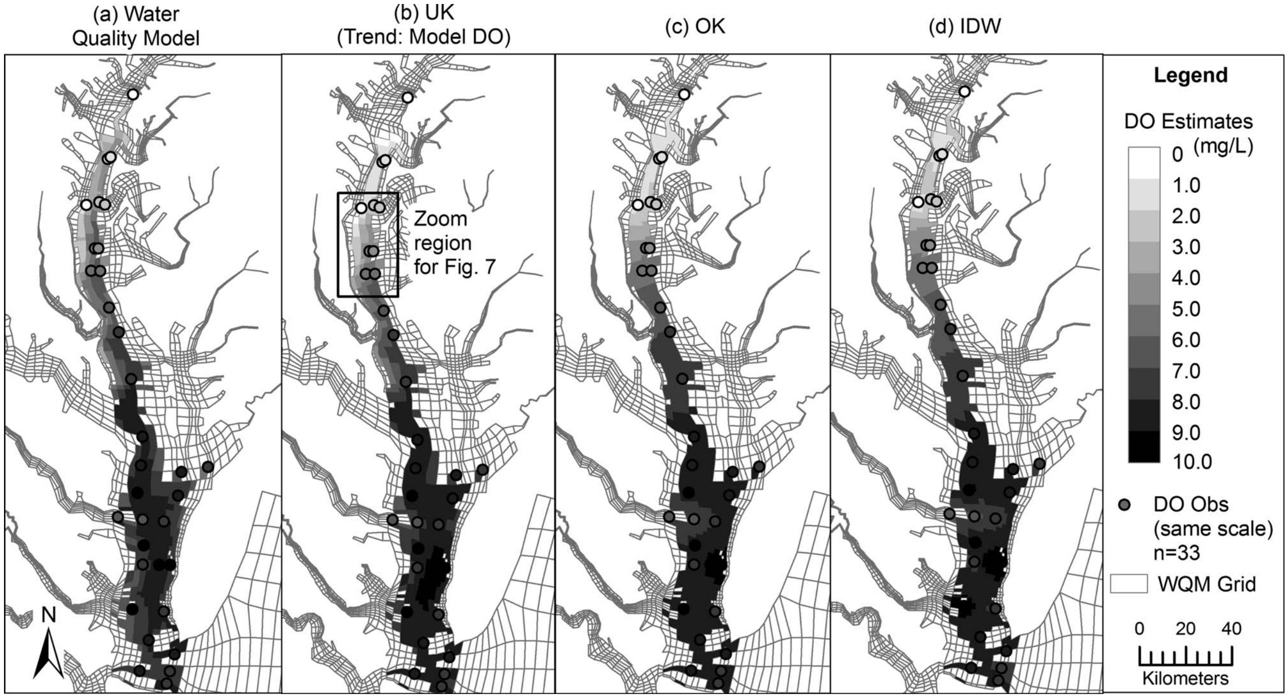

Fig. 6. Estimation of DO at 8 m deep from May 26–28, 1992, from (a) WQM output; (b) UK interpolation with the WQM-generated DO as a covariate; (c) OK method; and (d) IDW

with both of the kriging methods are shown in Fig. 5. Estimated standard deviations from the UK method are slightly smaller than those from the OK method, demonstrating that using the WQM output as a covariate reduced the uncertainty of the estimates. The standardized RMSE [Eq. (7)] results for UK and OK in Table 2 are all close to one and thus indicate that the kriging uncertainty estimates are fairly accurate. Comparison of the UK and OK values (i.e., 1.01 and 0.95, respectively, for salinity results) indicates that the kriging standard deviation estimates from UK are slightly more accurate than those from OK.

Fig. 6 is a set of DO interpolations and WQM results for samples taken 7.5–8.5 m below the surface, May 26–28, 1992. Similar to the results shown in Fig. 4, the cross-validation metrics do not reveal a large difference of performance among the three interpolators (Table 2); however, some potentially important differences do appear in the patterns of the DO interpolations. For example, in the area indicated with the box in Fig. 6(b), on the western side of the Bay, the DO estimates from the WQM and UK method are far less than from the OK and IDW methods. Fig. 7 zooms in on the region and uses a different scale to demonstrate more clearly how the UK method estimated a lateral variation in DO across the Bay that was not observed completely with the samples at this depth. This difference probably arose because the model predicted variations in the lateral hydrodynamics of the Bay at this location. Fig. 8 shows a lateral cross section of the Bay at the CB4.2 stations. In addition to the samples taken at 8 m that were used in the interpolation showed in Fig. 6, samples were also collected by the CBP during the same data collection cruise at other depths. At this particular lateral cross section of the Bay, there was no sample taken at 8 m deep on the western side; however, samples were taken at 7 and 9 m in that spot. These measurements taken at 7 and 9 m were not used in the 8 m-depth interpolation, but their values (3.4 and 1.2 mg/L, respectively) indicate that the lower DO estimate from the UK interpolation along the western side at 8 m (2.6 mg/L) is likely more consistent with the truth than the DO estimate from the OK interpolation

Table 2. Cross-Validation Interpolation Performance for Data Sets in Figs. 4 and 6

Figure and parameter	Measured range	Method	ME [Eq. (4)]	MAE [Eq. (5)]	RMSE [Eq. (6)]	Corr[a]	RMSE/SD	StndRMSE [Eq. (7)]
Fig. 4: Salinity at 2 m deep, July 1993	0–23.4 ppt	IDW	0.12	1.05	1.55	0.97	0.26	
		OK	−0.07	**0.73**	1.26	**0.98**	0.21	0.95
		UK	**0.00**	0.77	**1.17**	**0.98**	**0.20**	**1.01**
		WQM	−1.28	1.35	1.80	**0.98**	0.30	
Fig. 6: DO at 8 m deep, May 1992	0.5–9.9 mg/L	IDW	0.04	0.75	1.00	0.93	0.37	
		OK	0.04	**0.73**	0.95	0.93	0.35	0.99
		UK	**0.01**	**0.72**	**0.91**	**0.94**	**0.34**	**1.00**
		WQM	0.22	1.11	1.45	0.90	0.54	

Note: Bold values indicate the method(s) which performed best using each error metric.

[a]"Corr" = Pearson correlation coefficient calculated from the cross-validation results.

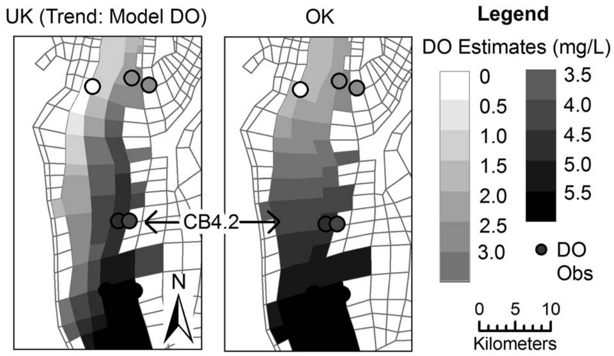

Fig. 7. Zoomed region of Fig. 6 showing differences between UK and OK interpolations

(4.4 mg/L). In this case, a 3D interpolation would have made use of the information from other depths in an effort to estimate more accurately. Within the current context of 2D interpolations, however, this example demonstrates how the mechanistic insights provided by the WQM can be used in conjunction with observations to provide more accurate interpolations.

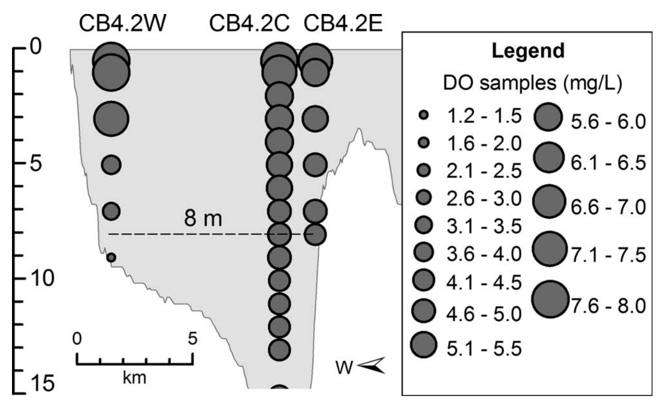

Fig. 8. Lateral transect of the Chesapeake Bay at CB4.2 stations showing DO samples taken at May 26–28, 1992

Patterns across Interpolations

In the 10 years of data collection considered for this study, more than 90,000 data points exist in the mainstem of the Bay for each of three water quality parameters, representing more than 3,640 data sets (i.e., unique depth and cruise combinations) for each parameter. Each of these data sets was interpolated, and Figs. 4–6 only show results from two of these data sets. For a more complete comparison of interpolations, the entire set of results was analyzed to identify major trends in relative interpolator performance. Table 3 shows the average performance indicators from the cross validation tests for all of the methods. For the three interpolated parameters, all performance measures indicate that the kriging methods performed better, on average, than the IDW method. Which kriging method outperformed the other varied with both the water quality parameter and the error metric. Note also that all interpolations performed better than the WQM alone—this was not unexpected, given that the WQM is calibrated to spatially and temporally averaged observations, rather than point values, in an effort to capture major trends in Bay water quality (Cerco and Cole 1993). The WQM is known to overpredict DO and underpredict salinity in the mainstem of the Bay (Cerco and Noel 2005), so these model results are not surprising. In fact, the error statistics shown in Table 3 for the WQM are similar to those presented for this model in an evaluation of incremental improvements to the Chesapeake Bay models by Cerco and Noel (2005).

The interpolation errors presented in Table 3 are averages over all data sets. To explore the results further, the errors were summarized by depth and month as well. Fig. 9 shows the average RMSEs from cross validation by depth for all salinity and DO data sets over the 10-year period. These graphs show how at every depth, on average, the kriging methods performed better than IDW (water temperature is not shown, but relative performance is similar). For salinity and DO, Fig. 9 indicates a decrease in interpolator accuracy in the middepth range of about 5–12 m. The relatively poor performance of salinity and DO interpolations in the middepth region indicates that the existence of a pycnocline in the Chesapeake Bay influences 2D interpolations such as these.

The cross validation errors averaged by month (Fig. 10) also illustrate the extent to which the kriging methods resulted in more accurate time stratified estimates. For water temperature and DO,

Table 3. Cross-Validation Interpolation Performance Averaged over All Data Sets

Parameter	Method	ME [Eq. (4)]	MAE [Eq. (5)]	RMSE [Eq. (6)]	Corr	RMSE/SD	StndRMSE [Eq. (7)]
Salinity ($n=3,663$)	IDW	0.09	1.43	2.04	0.85	0.47	
	OK	0.03	1.09	1.55	0.89	0.36	0.97
	UK	**0.00**	**1.01**	**1.37**	**0.92**	**0.32**	**1.00**
	WQM	−0.58	1.86	2.25	**0.92**	0.56	
Water temperature ($n=3,673$)	IDW	**0.00**	0.50	0.72	0.67	0.67	
	OK	**0.00**	0.45	0.64	0.72	0.60	**1.00**
	UK	0.01	**0.44**	**0.61**	**0.75**	**0.58**	1.01
	WQM	0.09	1.10	1.32	0.47	1.40	
DO ($n=3,645$)	IDW	−0.01	0.68	0.92	0.62	0.71	
	OK	**0.00**	**0.60**	**0.81**	0.66	**0.64**	0.99
	UK	−0.01	0.62	0.83	**0.67**	0.65	**1.00**
	WQM	0.28	1.39	1.69	0.42	1.42	

Note: Bold values indicate the method(s) which performed best using each error metric.

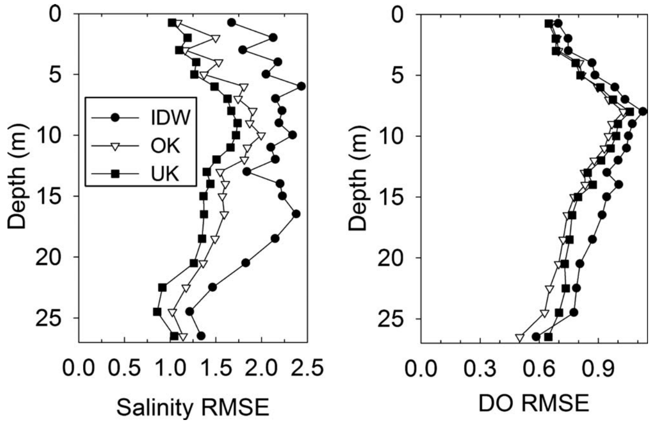

Fig. 9. Average RMSE of all interpolations by depth

all of the interpolators performed better in the winter months than summer. This likely is due to the higher variability of both of these parameters during the summer months. Salinity interpolation performance, which is not shown, was fairly constant throughout the year.

Discussion

The variety of variogram shapes observed in this study is not surprising considering all of the processes that affect these three water quality constituents. For example, the spatial structure of salinity changes with varying influences from tides, seasonally and annually varying freshwater flows, the density gradient caused by these flows, and many other factors (Schubel and Pritchard 1987). The spatial structures of temperature and DO are even more complicated, owing to the addition of many important and complex external and biochemical influences (Smith et al. 1992). Overall, the variability observed in the fitted variogram models demonstrates how important it is to characterize the spatial structure uniquely for each data set. For instance, in practice if a linear-shaped variogram is observed it indicates that the range of spatial correlation is as large or larger than the entire sampling domain. This was the case for many salinity data sets in this study [Fig. 3(a) and Table 1] due to the fact that a decreasing northward salinity gradient in the Bay is very common. In cases like this, it is beneficial to take a covariate into account to reduce the maximum range of spatial correlation to bring it within the sampling domain. In this study we saw a reduction in spatial correlation range by using the WQM-generated salinity as a covariate for observed salinity. Another option for achieving this, and one that is common for the UK method, would be to use the spatial coordinates of the samples as covariates. This latter approach was attempted in preliminary work but found to be generally less fruitful than the more novel WQM-based approach that we have used here.

One of the primary objectives of this study was to use this spatial structure information to perform a comparison of the nonstatistical IDW method and the statistical kriging methods for interpolation of water quality parameters in the Chesapeake Bay. As shown in the results, when considering all data sets over the 10-year period, the kriging methods outperformed IDW for all parameters. The accuracy of each method was compared for each data set using cross validation with various error measures. Using the RMSE, OK outperformed IDW for 86% of the water temperature interpolations, 89% of the DO interpolations, and 97% of the salinity interpolations. The differences in performance were not large in most cases. When OK outperformed IDW, the differences in RMSE on average were only 0.10°C for temperature, 0.12 mg/L for DO, and 0.51 parts per thousand (ppt) for salinity. It is important to keep in mind that we implemented an automation routine, described in the Methods section, for selecting a variogram model for each data set. Automation was necessary due to the large number of data sets, but it is expected that in many cases better variogram fits could be achieved if each variogram was fit manually with interaction between the user and the geostatistics software. Thus, the kriging results here could potentially improve if user interaction were included.

Another important benefit of kriging over IDW is that the former provides a direct measure of uncertainty for the interpolated values, presented as a standard deviation in Fig. 5. Spatial interpolation is frequently a tool used in a scientific study, and not the focus of a study itself. Therefore, having a measure of uncertainty can increase the scientific defensibility of study results and could be used to provide a confidence interval on the estimated volume of water not meeting certain criteria, which is potentially useful information to managers. If interpolation uncertainties are used in these ways, it is important to evaluate their accuracy. In this study, this was done using Eq. (7). The average standardized RMSEs presented in Table 3 for OK and UK are close to 1, indicating that the kriging variances accounted for the variability of the estimates appropriately. For salinity and DO, the UK standardized RMSEs are slightly better than for OK, perhaps because the stationarity assumptions for these data sets were stronger once a trend was accounted for with UK.

One of the other objectives of this study was to evaluate the use of the output from a process-based WQM to inform the kriging of observations. The change in variogram shape for many of the data sets when analyzed with UK compared to OK (Table 1) indicates that the WQM output did account for spatial variation of the samples (as discussed previously). Cross-validation results are less conclusive than for the IDW to OK comparison, as illustrated from the comparisons shown in Table 3. Using RMSE as an error measure, the WQM-informed kriging (UK) outperformed OK for 52% of the water temperature interpolations, 32% of the DO interpolations, and 68% of the salinity interpolations. Apparently, the relative performance of the two kriging methods depends on the water quality parameter. A comparison of the performance of the WQM alone for salinity, temperature, and DO demonstrates the likely reason for this difference. Such a comparison can be made by examining the correlation coefficient or RMSE scaled by the observations' standard deviation in Table 3. Those results

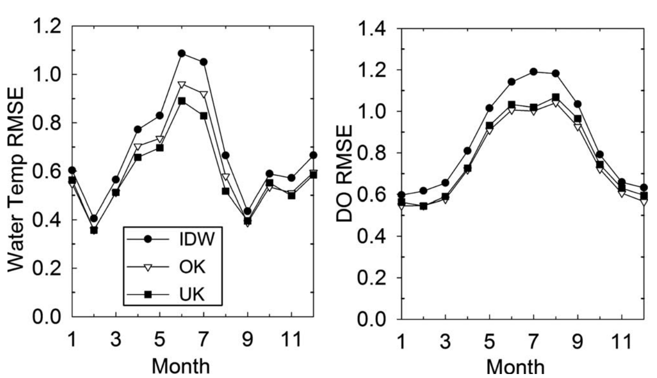

Fig. 10. Average RMSE of all interpolations by month

show that the WQM, by far, performs best for salinity compared to water temperature or DO. This strengthens the intuitive hypothesis that using the WQM-generated concentrations as the covariates in a UK method for interpolating observations will improve the interpolation of the observations most when the WQM itself performs well. The relatively good performance of the WQM for salinity compared to other parameters has been shown by the model developers. Cerco and Noel (2005) explain that modeled salinity has the least error of the water quality parameters modeled because it is calculated solely by physical transport processes and is bounded by the upper and lower ends of oceanic and freshwater salinities.

On the other hand, Figs. 7 and 8 further demonstrate potential other benefits to be derived from using output from the WQM to aid in kriging of observations. Irrespective of some possible inaccuracies and/or biases, the WQM does provide mechanistic simulation of what are believed to be relevant and important hydrodynamic and biochemical processes. For locations where the WQM captures some process or event that the sampling has missed, the UK method will provide the more appropriate interpolation. Of course, conversely, if the WQM is incorrect, it may adversely affect the outcome from the UK interpolation. Thus, the method demonstrated here is a tool to be used with care. An important consideration of using the UK method over OK is that the estimation variances are most often smaller. This will be true as long as the covariate used in the UK approach accounts for at least some of the variation in the data, and can help narrow the uncertainty in interpolation output.

An interesting output to examine related to the application of the UK method is the magnitudes of the fitted β coefficients [from Eq. (3)]. In particular, a truly perfect WQM match to observations would have resulted in $\beta_0=0$, $\beta_1=1.0$, and no residuals at any location. Actual median values of β_0 were 2.2, 5.5, and 4.9 for salinity, water temperature, and DO, respectively. Median β_1 values were 0.90, 0.51, and 0.23 for salinity, water temperature, and DO, respectively. Individual data set results showed a wide range of values around these medians. Overall these results indicate an imperfect fit of WQM results to data, which is consistent with our other findings.

Because a very large number of data sets were analyzed in this study, we have also been able to make certain types of findings that are outside the general topic of method comparison, but could be informative for the application of either interpolation method. For instance, the apparent pycnocline effect on the performance of all three interpolation schemes in the middepth regions (Fig. 9) is further confirmation of the important role played by the pycnocline for making interpolations in the Chesapeake Bay. In this context, Chehata et al. (2007) described a kriging application in the Chesapeake Bay where they estimated the top and bottom depths of the pycnocline layer and split the samples into three separate interpolation sets depending upon where the data point fell relative to the pycnocline. Because the Chesapeake Bay is more stratified in the summer months, the stratification and existence of a pycnocline is also a likely explanation for the relatively poor performance of all three interpolation schemes during those months (Fig. 10). Possible methods for improving these 2D interpolations could include using the position of a sample relative to the pycnocline (i.e., above, within, below) as an "indicator" in the kriging of a parameter [e.g., see Chapter 5 of Cressie (1993)] or using the distance above or below the pycnocline as a covariate in the UK method.

Further work is underway to take still further advantage of available data and tools for Chesapeake Bay interpolation. In particular, studies are being designed to expand these interpolations to three dimensions by taking into account the existence of a pycnocline and the irregular shape and hydrodynamics of the Bay. As discussed in the Methods section, complications arise when kriging in nonconvex regions. This is because to assure validity of most available variogram or spatial correlations functions, the distance metric used is restricted to be Euclidean (Curriero 2007). Using Euclidean distance in nonconvex areas, such as estuaries, is challenging because the straight lines between many pairs of locations go over land boundaries such that relationships among points connected by such lines are fundamentally different than relationships among other sets of points. Thus, other work is being considered to extend the analysis into tributaries by defining new spatial metrics based on "water distance" or, possibly, hydrodynamic travel time, again using dynamic model output as an input toward the understanding of observational data.

Overall, the work described here has been able to take advantage of increased access to the growing quantity of Chesapeake Bay observational data and computer modeling results in order to provide a more comprehensive comparison of some alternative spatial interpolation methods for use in the Bay. The results have shown how deterministic model output can be used to inform interpolations and how interpolation results can be used to provide new insights and information of value to managers. This study made full use of the currently developing CBEO as a compilation of water quality data and WQM output for the Chesapeake Bay. The interpolation methods tested here will be used in further studies with the CBEO-compiled data and are also being integrated into what is intended to become a publicly available "portal" (CBEO Project Team 2008). We view the work presented here as but one small example of how enhanced cyberinfrastructure should increasingly lead to better understanding and management of environmental resources.

Acknowledgments

This study was supported by funding from the National Science Foundation (Grant No. 0618986). We would also like to thank Carl Cerco and Mark Noel of the Army Corps of Engineers for access to and valuable support in the use of the WQM output; EPA Chesapeake Bay Program for access to the observational data; several anonymous reviewers for their comments; and the CBEO team, particularly Randal Burns, Eric Perlman, Dominic Di Toro, Michael Kemp, Damian Brady, Maureen Brooks, and Jeremy Testa for suggestions and aid in this study.

References

Bahner, L. (2006). *User guide for the Chesapeake bay and tidal tributary interpolator*, NOAA Chesapeake Bay Office, Annapolis, Md.

CBEO Project Team. (2008). "Prototype system for multidisciplinary shared cyberinfrastructure: Chesapeake Bay Environmental Observatory." *J. Hydrol. Eng.*, 13(10), 960–970.

Cerco, C. F., and Cole, T. (1993). "3-dimensional eutrophication model of Chesapeake Bay." *J. Environ. Eng.*, 119(6), 1006–1025.

Cerco, C. F., and Meyers, M. (2000). "Tributary refinements to Chesapeake Bay Model." *J. Environ. Eng.*, 126(2), 164–174.

Cerco, C. F., and Noel, M. R. (2004). "The 2002 Chesapeake Bay eutrophication model." *Rep. No. EPA 903-R-04-004*, U.S. Army Corps of Engineers, Waterways Experiment Station, Vicksburg, Miss.

Cerco, C. F., and Noel, M. R. (2005). "Incremental improvements in Chesapeake Bay environmental model package." *J. Environ. Eng.*, 131(5), 745–754.

Chehata, M., Jasinski, D., Monteith, M. C., and Samuels, W. B. (2007). “Mapping three-dimensional water-quality data in the Chesapeake Bay using geostatistics.” *J. Am. Water Resour. Assoc.*, 43(3), 813–828.

Chesapeake Bay Program (CBP). (1993). “Guide to using Chesapeake Bay Program water quality monitoring data.” *Rep. No. CBP/TRS 78/92*, Chesapeake Bay Program, Annapolis, Md.

Chesapeake Bay Program (CBP). (2008a). “Facts & figures.” ⟨http://www.chesapeakebay.net/factsandfigures.aspx⟩ (July 17, 2008).

Chesapeake Bay Program (CBP). (2008b). “CBP water quality database (1984–present).” ⟨http://www.chesapeakebay.net/data_waterquality.aspx⟩ (June 20, 2008).

Cressie, N. A. C. (1989). “The many faces of spatial prediction.” *Geostatistics*, 1, 163–176.

Cressie, N. A. C. (1993). *Statistics for spatial data*, Wiley, New York.

Curriero, F. C. (2007). “On the use of non-Euclidean distance measures in geostatistics.” *Math. Geol.*, 38(8), 907–926.

Day, J. W., Hall, C., Kemp, W. M., and Yanez-Arancibia, A. (1989). *Estuarine ecology*, Wiley, New York.

Diggle, P. J., and Ribeiro, P. J. (2007). *Modeled-based geostatistics*, Springer, New York.

Georgakarakos, S., and Kitsiou, D. (2008). “Mapping abundance distribution of small pelagic species applying hydroacoustics and co-kriging techniques.” *Hydrobiologia*, 612, 155–169.

Goovaerts, P., et al. (2008). “Geostatistical modeling of the spatial distribution of soil dioxins in the vicinity of an incinerator. 1. Theory and application to Midland, Michigan.” *Environ. Sci. Technol.*, 42(10), 3648–3654.

Hagy, J. D., Boynton, W. R., Wood, C. W., and Wood, K. V. (2004). “Hypoxia in the Chesapeake Bay, 1950–2001: Long-term changes in relation to nutrient loading and river flows.” *Estuaries*, 27(4), 634–658.

Isaaks, E. H., and Srivastava, R. M. (1989). *An introduction to applied geostatistics*, Oxford University Press, New York.

Jensen, O. P., Christman, M. P., and Miller, T. J. (2006). “Landscape-based geostatistics: A case study of the distribution of blue crab in Chesapeake Bay.” *Environmetrics*, 17(6), 605–621.

Johnson, B. H., Kim, K. W., Heath, R. E., Hshieh, B. B., and Butler, H. L. (1993). “Validation of a three-dimensional hydrodynamic model of Chesapeake Bay.” *J. Hydraul. Eng.*, 119(1), 2–20.

LoBuglio, J. N., Characklis, G. W., and Serre, M. L. (2007). “Cost-effective water quality assessment through the integration of monitoring data and modeling results.” *Water Resour. Res.*, 43, W03435.

Matheron, G. (1969). “Le krigeage universel.” *Rep. No. 1Cahiers du Centre de Morphologie Mathématique*, l’Ecole des Mines de Paris, Fontainebleau, France.

Mueller, T. G., Pusuluri, N. B., Mathias, K. K., Cornelius, P. L., Barnhisel, R. I., and Shearer, S. A. (2004). “Map quality for ordinary kriging and inverse distance weighted interpolation.” *Soil Sci. Soc. Am. J.*, 68, 2042–2047.

R Development Core Team. (2008). “The R project for statistical computing.” ⟨http://www.r-project.org/⟩ (Jan. 13, 2008).

Ribeiro, P. J., and Diggle, P. J. (2008). “geoR: A package for geostatistical analysis using the R software.” ⟨http://leg.ufpr.br/geoR/⟩ (May 22, 2008).

Ripley, B. D., and Lapsley, M. (2008). “RODBC: ODBC database access.” ⟨http://cran.r-project.org/web/packages/RODBC/index.html⟩ (May 22, 2008).

Rivest, M., Marcotte, D., and Pasquier, P. (2008). “Hydraulic head field estimation using kriging with an external drift: A way to consider conceptual model information.” *J. Hydrol.*, 361, 349–361.

Schabenberger, O., and Gotway, C. A. (2004). *Statistical methods for spatial data analysis*, CRC, Boca Raton, Fla.

Schubel, J. R., and Pritchard, D. W. (1987). “A brief physical description of the Chesapeake Bay.” *Contaminant problems and management of living Chesapeake Bay resources*, S. K. Majumdar, L. W. Hall Jr., and H. M. Austin, eds., Pennsylvania Academy of Science, Phillipsburg, N.J., 1–32.

Shlomi, S., and Michalak, A. M. (2007). “A geostatistical framework for incorporating transport information in estimating the distribution of a groundwater contaminant plume.” *Water Resour. Res.*, 43, W03412.

Simard, Y., Legendre, P., Lavoie, G., and Marcotte, D. (1992). “Mapping, estimating biomass, and optimizing sampling programs for spatially autocorrelated data: Case study of the northern shrimp (Pandalus borealis).” *Can. J. Fish. Aquat. Sci.*, 49, 32–45.

Smith, D. E., Leffler, M., and Mackiernan, G., eds. (1992). *Oxygen dynamics in the Chesapeake Bay: A synthesis of recent research*, Maryland Sea Grant College, College Park, Md.

USEPA. (2008). “Chesapeake Bay Phase 5 community watershed model.” *Rep. No. EPA XXX-X-XX-008*, Chesapeake Bay Program Office, Annapolis, Md., in preparation.

USEPA, and Region III Chesapeake Bay Program Office. (2005). “Chesapeake Bay program analytical segmentation scheme: Revisions, decisions and rationales 1983–2003, 2005 addendum.” *Rep. No. EPA 903-R-05-004*, U.S. Environmental Protection Agency Region III, Chesapeake Bay Program Office, Monitoring and Analysis Subcommittee, and Tidal Monitoring and Analysis Workgroup, Annapolis, Md.

USEPA, and Region III Water Protection Division. (2007). “Ambient water quality criteria for dissolved oxygen, water clarity and chlorophyll a for the Chesapeake Bay and its tidal tributaries 2007 addendum.” *Rep. No. EPA 903-R-07-003*, U.S. Environmental Protection Agency Region III, Chesapeake Bay Program Office, and Region III Water Protection Division, Annapolis, Md.

Wackernagel, H. (2003). *Multivariate geostatistics: An introduction with applications*, Springer, Berlin.

Wang, H. V., and Johnson, B. H. (2000). “Validation and application of the second generation three dimensional hydrodynamic model of Chesapeake Bay.” *Water Qual. Ecosyst. Model.*, 1, 51–90.

Zimmerman, D., Pavlik, C., Ruggles, A., and Armstrong, M. P. (1999). “An experimental comparison of ordinary and universal kriging and inverse distance weighting.” *Math. Geol.*, 31(4), 375–390.