COVER PAGE

Date: 01/31/2012

From : Marc Serre, University of North Carolina at Chapel Hill

To : The New Jersey Department of Environmental Protection

Re: Final project report for NJDEP contract SR08-037 / UNC 5-45861 titled “Version Upgrade, Expansion, and Improvement of Modern Bayesian Spatio-Temporal Geostatistics Code for ArcGIS Users”

Summary deliverable BMEGUI:

As described under contract SR08-037, our project consists of four objectives dedicated to upgrading the BMEGUI software, which are to (1) improve internal error handling, (2) add new soft data types, (3) implement automatic covariance fitting, and (4) implement the river metric. We were successful in finishing the implementation these entire objectives, we completed three cases studies that illustrate the use of the upgraded BMEGUI, version 3.0, and we produced the final report provided below. We are pleased to announce that we have completed the testing, debugging, and documentation of these four objectives, creating tutorials for each completed objectives and delived the BMEGUI software to NJDEP.

Final report:

Version Upgrade, Expansion, and Improvement of Modern Bayesian Spatio-Temporal Geostatistics Code for ArcGIS Users

Abstract

The upgraded BMEGUI, version 3.0, allows users to transparently run the BME programs without having to learn any additional programming language and complexities of tedious mathematical formulations. This upgraded BMEGUI provides an easy-to-use interface with improved internal error handling, with automatic covariance parameter selection, and with expanded types of data types supported, and with the incorporation of a river network distances that substantially improved the previous version of BMEGUI. The input to the upgraded BMEGUI includes environmental monitoring data and river networks that users can obtain from DEP databases. The output will be BME estimated maps showing the spatial distribution of environmental parameters for any time of interest. The upgraded BMEGUI has been delivered to NJDEP with a fully updated user's manual.

1. Objective 1: Internal Error Handling of Input Datasets

1.1 Implementation

The BMEGUI currently provides basic functionality to automatically correct internal inconsistencies in the users' dataset. These inconsistencies are categorized in the following three errors: 1) the same station ID is assigned to different geographic locations, 2) different station IDs are assigned to the same geographic location, and 3) there are conflicting duplicated measurements at a same location and time. The original BMEGUI only displays an error message when any of these errors are detected, but does not provide a reporting of where the errors occur in the data. An example of how the original BMEGUI handles error type 1 (i.e. the same station ID is assigned to different geographic locations) is shown in figure 1(a), which does not report all the stations where this error occurred.

(a) (b)

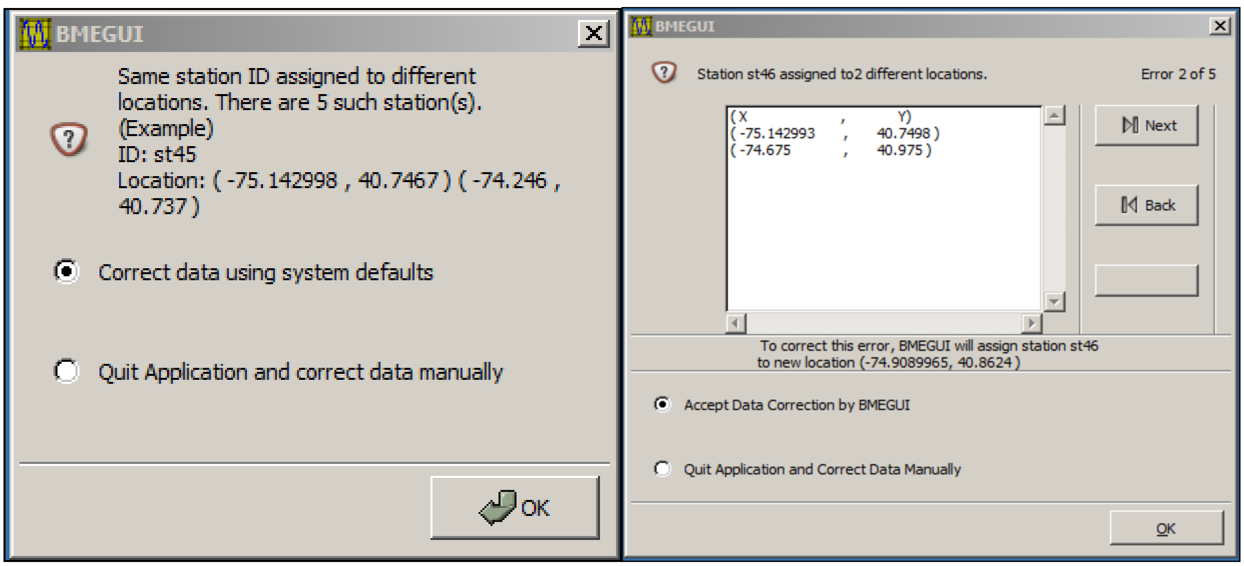


Figure 1(a): Original BMEGUI error handling dialog box. BMEGUI identified 5 monitoring station IDs where theses IDs were assigned to two or more different locations. (b): Newly developed user-friendly error handling dialog box. BMEGUI identified and reported that station "st46" has been assigned to two different locations which are shown in the textbox. To correct this error BMEGUI assigns a new location (-74.90, 40.86) to this station "st46" if the user accepts this data correction offered by BMEGUI.

In order to improve quality control of the data used, it is very important to identify and remove all the stations where error 1 occurred before the BME analysis is carried out. If the user continues to run the BME analysis and doesn't fix internal data errors, then there is an increased risk that the data analysis might be incorrect. To address this issue, an interface window has been developed that facilitates the inspection of the stations where the error occurred in the input dataset, and provides a rudimentary interface to facilitate the correction of these errors. A snap shot of this newly developed user-friendly interface is shown in figure 1(b). This newly added feature tells users about the type of error encountered, the number error instances, and the places where these errors were identified in the dataset. A specific potential solution specific to the error detected is then provided, which the user can accept or reject.

Similarly, a newly developed error handling dialog box also handles the second type of error, where two or more stations IDs are assigned to the same location as shown in figure 2. Figure 2(a) shows the old error handling dialog box and figure 2(b) shows the newly improved error handling dialog box. This new error handling window lists all the stations IDs that are assigned to a same location. For example, four different stations IDs (st200, st201, st202, and st203) are all assigned to the same location (- 74.87, 40.955) as shown in figure 2(b). Additional occurrence of this type of error (where 2 or more station IDs are assigned to a same location) can be seen by pressing the NEXT button on the window.

(a) (b)

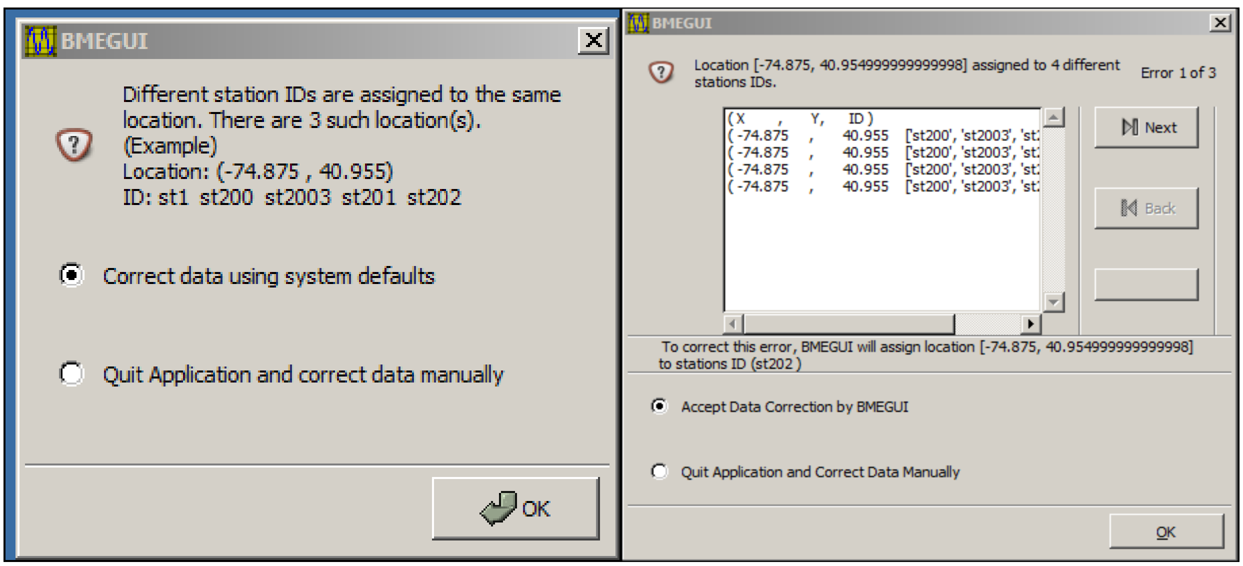


Figure 2: (a) Original BMEGUI's error handling dialog box. BMEGUI identified 3 instances of multiple stations assigned to the same location. (b) Newly developed BMEGUI's error handling dialog box. BMEGUI identified 4stations IDs that are assigned to a same location (-74.875, 40.955). There are three instances of this type of errors. Other instances of this error are viewed by pressing the next button.

1.2 Testing/Simulation

The testing of these newly developed dialog boxes for internal error handling has been done successfully using a small synthetic dataset.

1.3 Quality assurance and quality control

The newly added features for error handling internal data errors have been tested using the small synthetic data file described above. Further tests have been conducted successfully using large data sets for handling internal data errors. QA/QC is ensured as described in Quality Assurance Plan (QAP) along with this report.

1.4 Documentation (tutorial and user manual)

The documentation of these newly added features of error handling has been completed. Details of these newly added features with good graphics have been added in the tutorials and user manual of the upgraded version of BMEGUI.

2. Objective 2: Additional Forms of Soft Data

2.2 Implementation

The old version of the BMEGUI supports only two types of statistical distribution describing the uncertainties associated with the soft data. One is a uniform probability distribution function (PDF) which is defined by two parameters; the lower and upper bounds of the interval. The other is the Gaussian PDF which is also defined by two parameters; the mean and standard deviation. By knowing which data columns in the user's data file contain these two parameters, the BMEGUI is able to automatically construct the PDF for each soft data point. While these two types of PDF are often used in the data analysis of the environmental data, there exist other types of PDF that are useful to consider. This is particularly relevant when considering the emergence of data from multiple sources that were not available to the same extent in the past (increasing sophistication of measurement error models, new secondary information available at low marginal cost, emergence of the availability of remote sensing data, predictions from hydrologic or air quality models, biomarkers of exposure, health records, etc.). As a result, we expanded the types of PDF available in BMEGUI so as to include Truncated Gaussian, and Triangular distributions. Therefore, an interface window has been developed and implemented that accommodates the entry of the Truncated Gaussian, and Triangular types of PDFs. This provides a tool that will improve the user's ability to merge data from multiple data sources. A snap shot of this newly developed user-friendly interface is shown in figure 3. This newly added functionality was implemented by expanding the file format acceptable as input to the BMEGUI, and by adding graphical functionality to display these new types of PDF used to characterize the uncertainty in the user's soft data.

Triangular distribution PDF:

To define the triangular PDF we need three parameters; namely the lower limit, mode, and upper limit of the triangular distribution. This required writing new MATLAB programming functions, which have been implemented and tested in the upgraded version of BMEGUI. If we define the lower limit *a*, the mode *c* and the upper limit *b* as the three parameters of the triangular PDF, then we can write this triangular PDF as:

$$f(x|a,b,c) = \begin{cases} \frac{2(x-a)}{(b-a)(c-a)} & \text{for } a \le x \le c \\ \frac{2(b-x)}{(b-a)(b-c)} & \text{for } c \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

Truncated Gaussian PDF:

The truncated normal distribution is the probability distribution of a normally distributed random variable whose value is either bounded below or above (or both). Thus, at most we need up to four values to define such a PDF. New MATLAB programming functions for the truncated Gaussian PDF have been implemented and tested successfully in BMEGUI.

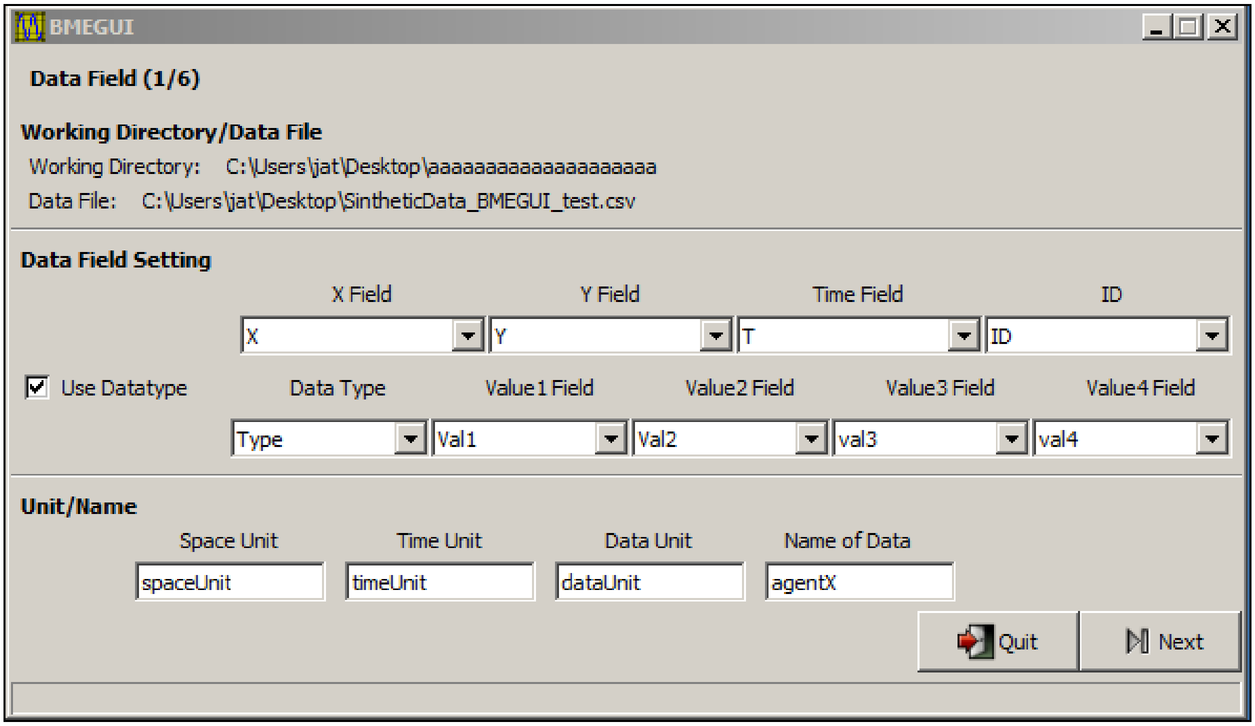


Figure 3: A newly added graphical user interface window developed to incorporate additional forms of soft data types in the upgraded version of BMEGUI.

2.2 Testing /simulation

Implementation and testing of this newly developed user friendly interface for expanded file format has been done successfully.

2.3 Quality assurance and quality control

Further tests have been conducted to assess the robustness and reliability of these newly added types of soft data using a large NJDEP data set on air quality.

2.4 Documentation (tutorial and user manual)

Documentation of this new capability has been completed. Details of this newly added feature with good graphics has been added in the tutorials and in the user manual.

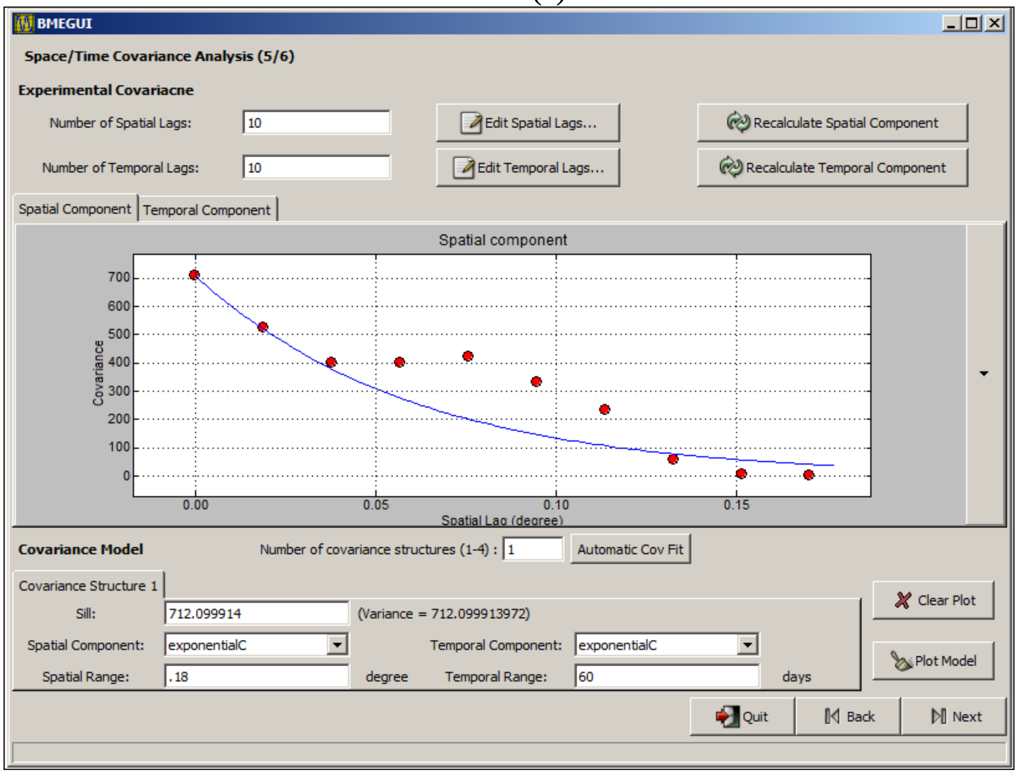
3. Objective 3: Automatic Covariance Parameter Fitting

3.1 Implementation

BMEGUI users need to enter the parameters of a covariance function modeling the spatiotemporal autocorrelation between data points. This covariance function is essential to the subsequent mapping and estimation of environmental variables. The old version of BMEGUI asks users to provide lags at which to calculate experimental covariance values, as well as the parameters of a covariance model that best fit these experimental covariance values. Such parameter selection may be a difficult task for users lacking deep knowledge of geostatistical modeling. Therefore, we have added functionality in the BMEGUI that will provide automatic parameter selection for the covariance modeling. The updated version of BMEGUI now automatically selects optimal values for the covariance lags used to calculate the experimental covariance values, and automatically select covariance model parameters that best fit these experimental covariance values. Hence, the new version of BMEGUI provides a sort of "automatic mapping" procedure. This automatic parameter selection will be of great benefit to a wide range of BMEGUI users without advanced knowledge of spatiotemporal statistical theory. However, advanced BMEGUI users will still be able to override any input parameters of their choice instead of using parameter values suggested by BMEGUI.

We implemented this automatic covariance parameters selection using the dataset on ammonia concentration ($\mu g / L$) described in case study later in this report. Figures 4(a) and (b) are screen shots of the spatial and temporal covariance components fitted in BMEGUI using the automatic covariance parameter selection feature of the new version of BMEGUI

(a)



(b)

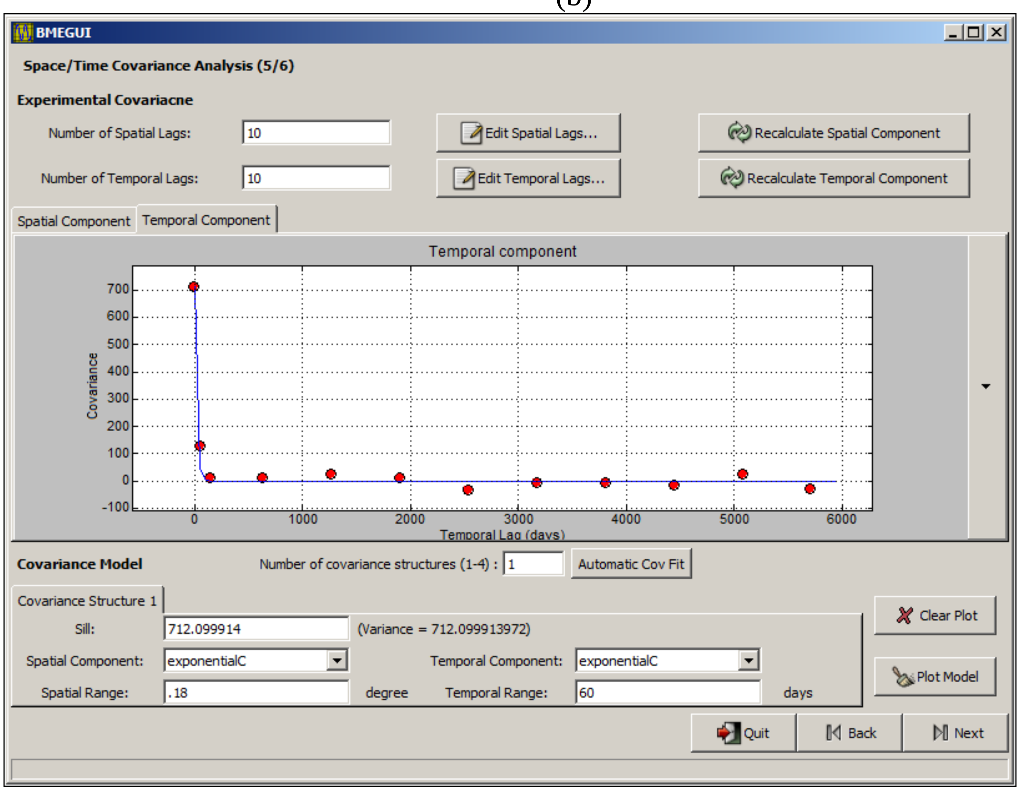


Figure 4: (a) Automatic fitting of the spatial component of the covariance model. (b) Automatic fitting of the temporal component of the covariance model.

3.2 Testing/Simulation

Testing of the newly added feature to automatic select covariance parameters has been performed using ammonia water quality data provided by a NJDEP staff member.

3.3 Quality assurance and quality control

Further testing and debugging have been performed to assess the robustness of this new procedure using both water and air quality data sets. The goodness of fit of the selected theoretical covariance model depends on the data set used. This feature will be of benefit in simple case studies.

3.4 Documentation (tutorial and manual)

A full description this newly added feature has been added in the tutorials and in the user manual.

4. Objective 4: River Metric

4.1 Implementation

Over the first 7 months of this project we have made substantial progress with the *conceptual development* needed for the implementation of a new river metric feature in BMEGI, which has resulted in the publication of a joint paper (Money et al, 2009) between UNC and NJDEP scientists on the use of the river metric in spatiotemporal Geostatistics to map e-coli concentrations along a river basin in New Jersey. This publication fulfills deliverable 2c (joint paper between UNC and NJDEP) described in the attachment E of the contract, and it provides important theoretical concepts that are implemented in the BMEGUI computer program.

We have successfully integrated the river metric in the BMEGUI, and developed a graphical user interface to allow users to select a river network for their analysis. This graphical user interface does not provide tools to manipulate the river network, but it will provide an error message in case the river network is incomplete. The added BMEGUI feature tests for three basic types of river network errors (1) outlet exists outside the river network, (2) outlet exists at location where multiple river stream meets, and (3) broken river network. BMEGUI checks for all these errors and provide a visual snapshot of the error, if the error is detected by BMEGUI. See tutorial for detail.

4.2 Testing/Simulation

The testing of the newly added feature to use a river metric has been performed in BMEGUI using the river network of Money et al, 2009.

4.3 Quality assurance and quality control

Substantial testing and debugging have been performed, however, the testing was limited to the river network data mentioned above. See tutorial for more details.

4.4 Documentation (tutorial and manual)

Documentation of this newly added feature to use river metric has been added to the user's manual. Additionally a tutorial has been developed specifically for this new feature.

References

Money, E., G. Carter, **M.L. Serre**[†] (2009) Modern Space/Time Geostatistics using River Distances: Data Integration of Turbidity and E.coli Measurements to Assess Fecal Contamination Along the Raritan River in New Jersey, *Environmental Science & Technology*, Vol. 43(10), pp. 3736-3742.

5. Case Study1: Comparing BMEGUI and Inverse Distance Weighting (IDW) estimation of surface water ammonia concentration

5.1. Study Area and Data Source:

Surface water quality is monitored by the Water Monitoring and Standards (WM&S) division of the New Jersey Department of Environmental Protection. Ammonia concentrations (µg Nitrogen / L) measured from July 12th, 1989 to Sept 1, 2010 at irregular time intervals were provided by Mr. Mike Kusmiesz, GIS specialist at the Bureau of Marine Water Monitoring at NJDEP, who we are working with on this case study. Figure 5 (a) shows the counties in the State of New Jersey and figure 5(b) shows the location of ammonia concentrations provided by NJDEP. All these observation sites are located near the coastal region of Ocean County.

(a) (b)

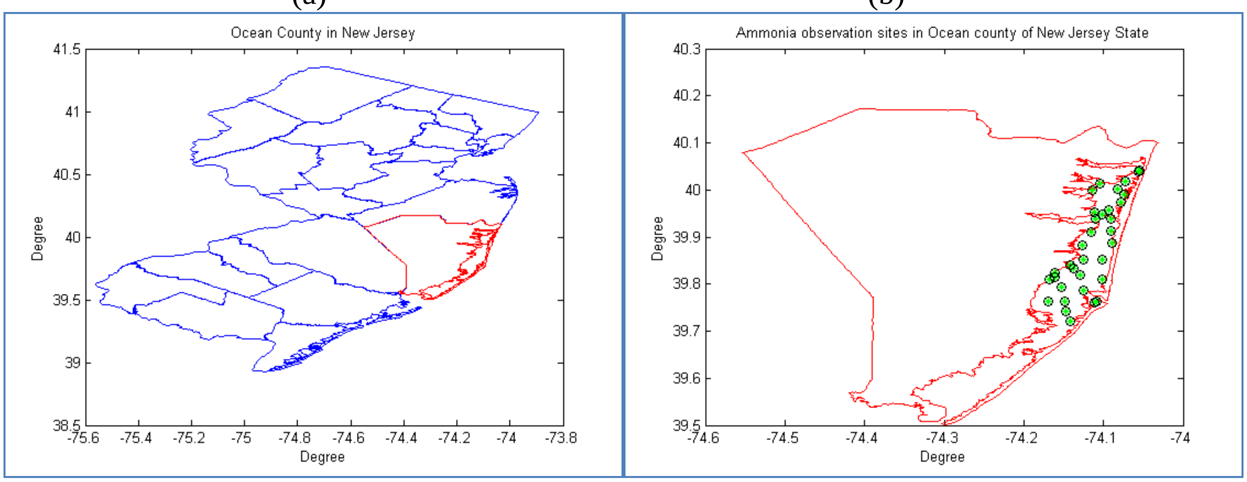


Figure 5: (a) Counties of State of New Jersey and (b) ammonia observation sites in the coastal region of Ocean County in New Jersey

5.2. BMEGUI estimation versus Inverse Distance Weighting interpolation:

Inverse Distance Weighting (IDW) interpolation calculates an interpolated value as the weighted average of neighboring measurements, where the weights are proportional to the inverse of the distance between measurements and interpolation point. On the other hand the BMEGUI estimation is based on the sophisticated Bayesian Maximum Entropy (BME) theory, which provides a sound geostatistical framework that accounts for the space/time autocorrelation of the environmental variable between the measurement points and the estimation point, as well as the uncertainty associated with measured values. As a result we expect BME estimation to provide a better representation of the space/time distribution of environmental variables than that obtained from IDW interpolation. In this case study, we used the BMEGUI to obtain the kriging estimate shown in figure 6a, and we implemented the IWD interpolation in the MATLAB programming platform to obtain the IDW map shown in figure 6b.

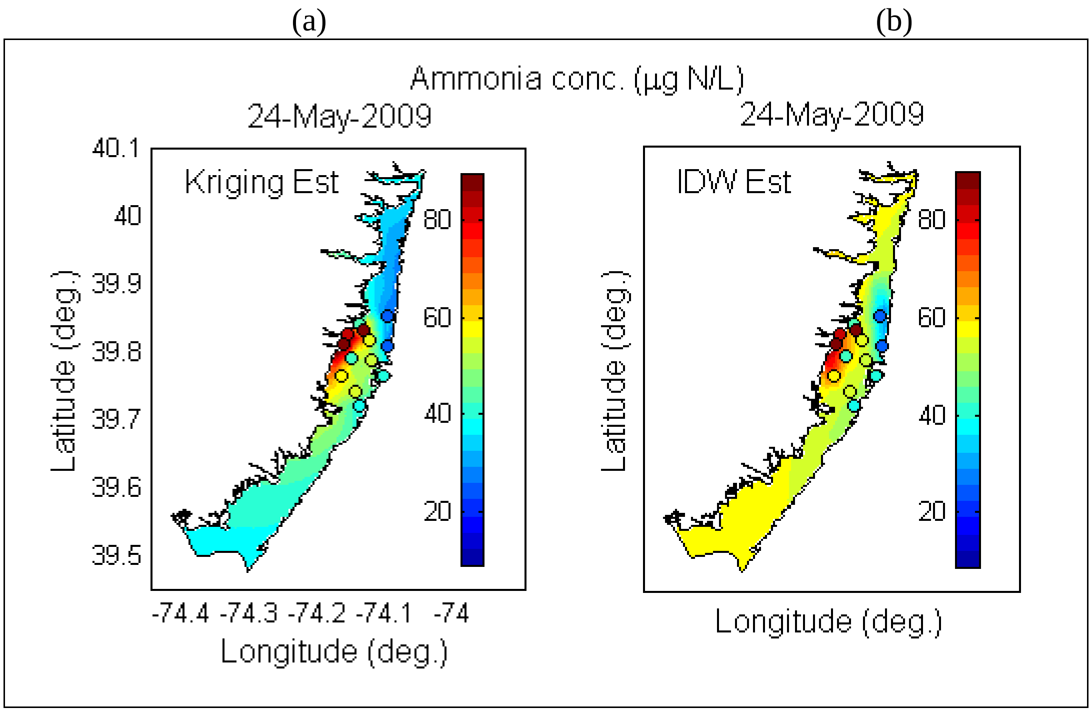


Figure 6: Estimates of surface water quality (ammonia conc. (µg/L)) in Ocean County coastal area of New Jersey on May 24, 2009 using (a) Kriging and (b) the IDW interpolation technique

We can see from figure 6 that each method provides a slightly different water quality map. The objective of this case study is to visualize the relative performance of the IDW and BME mapping methods. The new version of BMEGUI with its added functionality was used to obtain the BMEGUI map shown in figure 6(a), which provides a representation of the spatial distribution of ammonia concentrations on May 24, 2009. The IDW map obtained for the same day is shown in figure 6(b). This map was calculated using only the concentrations measured on that day. A comparison of figure 6(a) and (b) indicates a clear difference in the values estimated by these two methods in the southern area of the study domain, where IDW values are clearly higher than that of BMEGUI. From theory the BMEGUI map is more accurate than the IDW map because BMEGUI accounts for measurements made on days prior and after the estimation day. Hence the accuracy provided by the BMEGUI map leads to a substantial difference in estimation value in the southern area where IDW is only using data available for the estimation day.

There are some fundamental differences in the IDW and BME methods which explain why BME is a very powerful tool. BMEGUI and IDW techniques can be compared as following:

(i) BMEGUI interpolation is based on a modern spatiotemporal geostatistical framework that accounts for space/time autocorrelation and non-Gaussian uncertainties, whereas IDW is a deterministic method that does not account for randomness. As a result IDW only provides an interpolated value, while BMEGUI provides both an estimated value, as well as a corresponding estimation error variance (or equivalently an estimation confidence interval).

(ii) IDW may be obtained as a linear limiting case of BME under the following three simplistic assumptions: (a) there is no measurement error, (b) the autocorrelation between measurements can be neglected, and (c) the autocorrelation between a measurement and the estimation point increases linearly with the inverse of separation distance. However, when these assumptions are violated, BMEGUI will result in a more accurate estimation than IWD

(iii) BMEGUI maps can be estimated on any day, even days when no measurements are taken, on the strength of auto correlated measurements taken on previous or following days. On the other hand IDW maps should only be created for days at which several measurements were taken. As a result BMEGUI is a much better tool to analyze how spatial trends change over time, which is critical for many environmental monitoring efforts.

We estimated the ammonia concentrations using BMEGUI for all dates on which water quality observations have been provided by NJDEP. Figure 7 shows the output from BMEGUI on some of these specific dates.

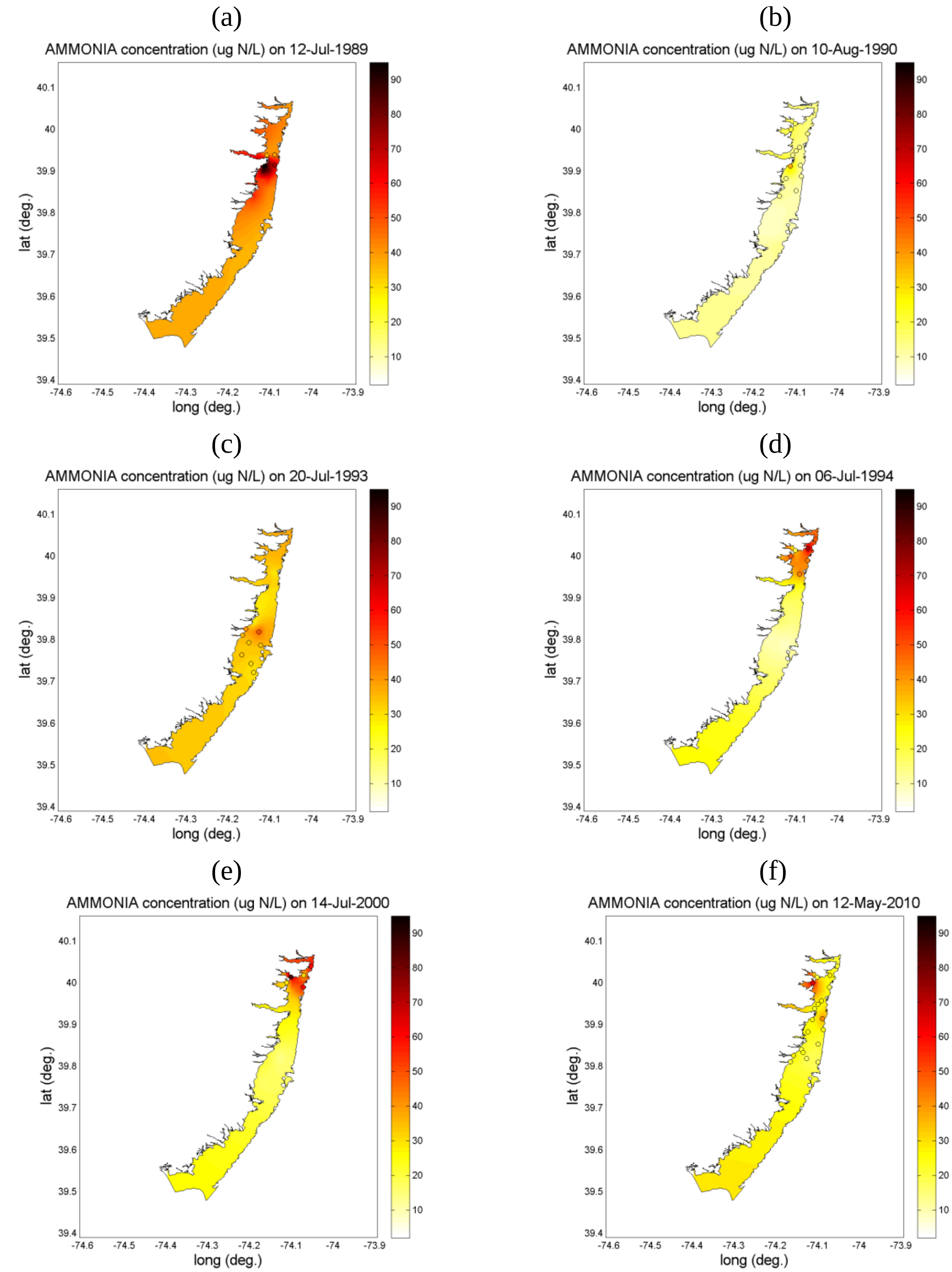


Figure 7: Estimates of surface water quality (ammonia conc. (µg/L)) in the coastal region of Ocean County in New Jersey on July 12th, 1989 (a); August 10th, 1990 (b); July 20th, 1993 (c); July 6th, 1994 (d); July 14th, 2000 (e), and May 12, 2010 (f).

BMEGUI also produces the estimation error variance maps shown in figure 8. We can see from that figure that the estimation error variance is higher in the southern part of the study domain. This makes sense because we do not have any measurement in the southern part of the study domain. On the other hand uncertainty is almost zero at locations where we measurements. These maps may be useful in assessing where new monitoring stations should be located.

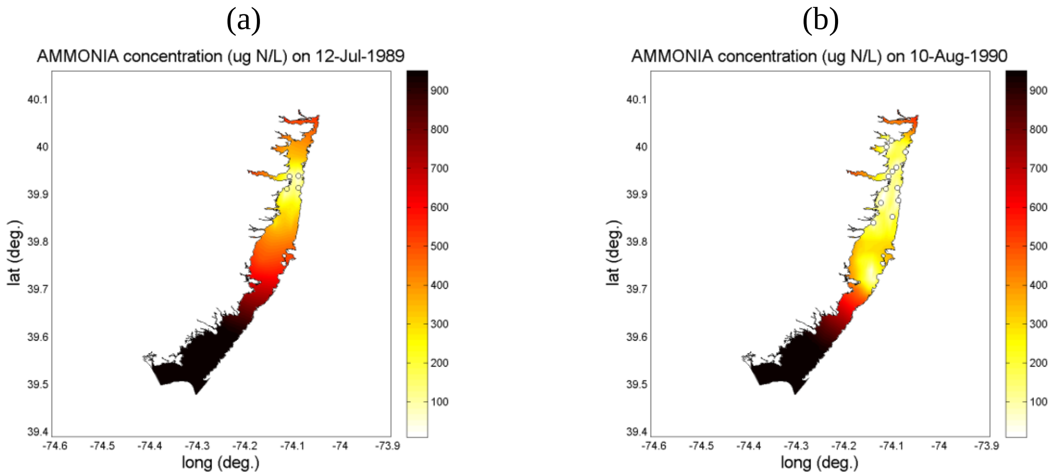


Figure 8: plots of BMEGUI produced uncertainty estimates of ammonia concentration for the days of July 12th, 1989 (a); and August 10th, 1990 (b).

NOTE: Figures for this study were generated in the MATLAB computing platform. Similar figures were generated using BMEGUI as described in the corresponding case study tutorial.

5.3. Cross validation analysis

BMEGUI is a powerful graphical user interface based on a sound geostatistical estimation framework that produces maps that can be substantially different than those obtained with the IDW interpolation technique, as supported by figure 6. The BMEGUI map should be more accurate as it accounts for the space/time autocorrelation in the data. However, the accuracy of these two methods cannot be quantified based only on the visual inspection of these maps.

In order to assess and compare the accuracy of these substantially different maps, we performed a cross validation analysis using the *BMElib* package implemented in the *MATLAB* numerical platform to quantify the estimation error of each technique. The cross validation analysis consists in removing each ammonia value in turn from the data set, and calculating an estimate of that removed value from the remaining data. The difference between the estimated value and the value that was removed (the true value) is the estimation error. Using these estimation errors obtained at 1209 spatio-temporal locations we calculate the cross validation statistics shown on Table 1. These statistics consist in the Mean of the Squares of the estimation Errors (MSE), the Mean of the Absolute Value of estimation Errors (MAE), and the Mean of the estimation Errors (ME). The MSE and MAE are measures of the overall estimation error, as they quantify both the estimation bias and the lack of precision.

Table 1: Estimated errors in IDW and Kriging estimations

Estimation method	IDW	BME
MSE (μg/L)^2	290.5	275.6
MAE (μg/L)	8.446	8.416
ME(μg/L)	0.112	0.012

These statistics reported in Table 1 indicate that the BME method implemented in the BMEGUI is overall better than the IWD method. The ME measures the bias of the estimation. The positive bias reported in Table 1 for both IDW and BME indicate that both methods slightly over predict the true ammonia concentration. The bias is more pronounced for IDW, and it is reduced when using BME. Furthermore the MSE and MAE are consistently lower for BME than for IDW, indicating that BME estimates are more precise and accurate than those of IDW. It can therefore be concluded from the table 1 that the map created using BMEGUI, which implements the BME method, is more accurate and therefore better than the map created using IWD.

We note that in this case study we only used hard data, and in this case the BME method is the same as the space/time kriging method of classical geostatistics. If soft data were included in the analysis, then we would expect that the improvement of BME over IDW would be even larger than those reported in Table 1, because BME can rigorously incorporate soft data while that is not possible with IWD.

6. Case Study2: BMEGUI Estimation of Total Organic Carbon using hard and soft data

6.1. Introduction:

One of the features of the upgraded version of BMEGUI is the extended list of the types of soft data that can be considered, which now includes the triangular and truncated Gaussian PDFs. These added types of PDFs are very useful for the BME spatiotemporal estimation of air quality using data from different sources, as demonstrated in De Nazelle et al. (2010)[(1)]. We therefore completed a second case study to demonstrate how BMEGUI can be used to estimate $PM_{2.5}$ chemical speciation across NJ using the new version of BMEGUI. Chemical speciation of $PM_{2.5}$ provides information about air pollution that is critical to develop policies and regulations that are effective in protecting the public health. The analysis of spatial and temporal patterns of the chemical composition of $PM_{2.5}$ provide critical insights about the contribution of local and urban pollution sources relative to regional background concentrations. For example these insights might help understand which chemical constituents are primary drivers for high $PM_{2.5}$ mass in urban areas.

New Jersey State has 4 monitoring stations that provide $PM_{2.5}$ chemical speciation out of a total of 12 $PM_{2.5}$ monitoring stations. Important components of $PM_{2.5}$ at these stations include sulfate, nitrate, total carbonaceous mass, ammonium, and crustal material. These components have complex spatial-temporal dependency and cross dependency structures. In order to demonstrate how BMEGUI can be used to map these important $PM_{2.5}$ components, we provide here a case study that focuses on Organic Carbon (OC). Using BMEGUI it would be easy to extend the analysis performed here to any the other components of $PM_{2.5}$.

6.2. Materials and Method:

Monitoring Data:

$PM_{2.5}$ speciation stations measure the daily average of both OC and $PM_{2.5}$ concentrations. These stations provide *hard* data on OC average daily concentrations because OC is *directly* measured with a small measurement error. On the other hand regular $PM_{2.5}$ stations only measure the daily average $PM_{2.5}$ concentrations. These stations are more numerous, but they only provide an *indirect* assessment of OC since OC is a fraction of $PM_{2.5}$. We will use these stations to provide *soft* data on OC, i.e. data with associated uncertainty, since these data are not coming from direct measurements of OC.

The State of New Jersey has 4 $PM_{2.5}$ speciation stations, and an additional 8 regular $PM_{2.5}$ stations where only $PM_{2.5}$ is measured. Since at least 10-15 speciation stations measuring OC are needed to conduct a reliable OC space/time geostatistical analysis, we extended our study domain to the rectangular area shown in Figure 9. This rectangular domain consists in New Jersey and surrounding parts of Connecticut, Maryland, Pennsylvania, and New York, so as to include 20 $PM_{2.5}$ speciation stations and 29 regular $PM_{2.5}$ stations. We obtained the $PM_{2.5}$ and OC daily average concentrations observed at these stations from Jan 2007 to Dec 31, 2008. Due to missing values in this time period, 5 of the 20 $PM_{2.5}$ speciation stations did not have any data. On the other hand, all 29 stations where only $PM_{2.5}$ is measured had data in our time period. As a result we ended up with 15 $PM_{2.5}$ speciation stations where both $PM_{2.5}$ and OC were measured, and 29 additional stations where only $PM_{2.5}$ is measured.

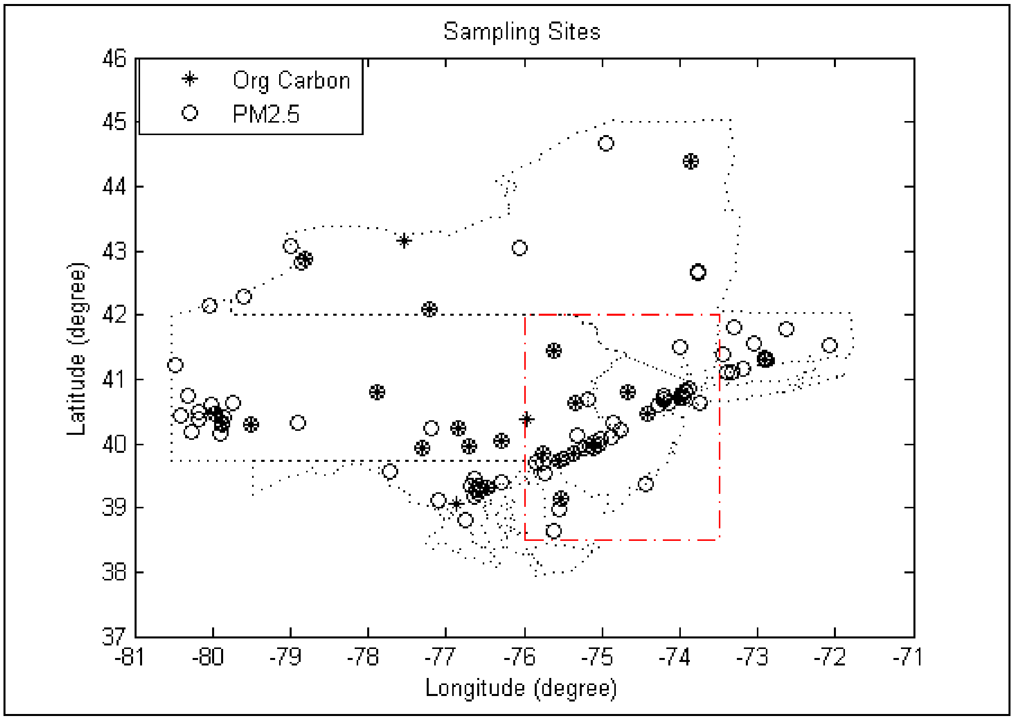


Figure 9: Sampling sites measuring $PM_{2.5}$ and OC.

Mass Fraction Model for organic Carbon:

The procedure used to map OC consists in treating the log of OC daily concentrations observed at the 15 $PM_{2.5}$ speciation stations as hard data for logOC, and in generating a soft datum for logOC for each $PM_{2.5}$ daily concentration measured at the 29 stations where only $PM_{2.5}$ is measured. The mass fraction procedure to generate the logOC soft data from the measured $PM_{2.5}$ concentrations is described in Allshouse et al. (2009) [(2)]. Briefly, logOC is related to $\log PM_{2.5}$ by the following relationship

$$\log OC = \log PM_{2.5} + \log MF,$$

where $MF=OC/PM_{2.5}$ is the Mass Fraction of OC in $PM_{2.5}$. If follows that the mean and variance of OC given a measured value for $PM_{2.5}$ are given by

$E[\log OC|\log PM_{2.5}] = \log PM_{2.5} + E[\log MF]$, and

$var[\log OC|\log PM_{2.5}] = var[\log MF]$

We set the values of E[logMF] and var[logMF] to the mean and variance, respectively, of the MF values observed at the 15 $PM_{2.5}$ speciation stations. Then, for each $\log PM_{2.5}$ value observed at a station where only $PM_{2.5}$ is measured, we calculated the mean and variance of logOC using the above equations, and we entered these values in BMEGUI as the mean and variance of soft data of Gaussian type.

6.3. Results and Discussions:

Using the hard and soft data described above, BMEGUI generated maps of logOC for any day of interest. The map displayed in Figure 10 shows the spatial distribution of OC across New Jersey on Jan 03, 2007. This figure shows the hard and soft data available on that day with circles and squares, respectively. The BME method provides a rigorous mathematical framework that incorporates both hard and soft data to obtain the estimated values shown on the map. This estimation process puts more weights on the hard data, since they represent values of OC that were directly measured at $PM_{2.5}$ speciation stations, and also incorporates (with lesser weights) the information provided by the soft data inferred from $PM_{2.5}$ concentrations observed at stations that do not monitor OC.

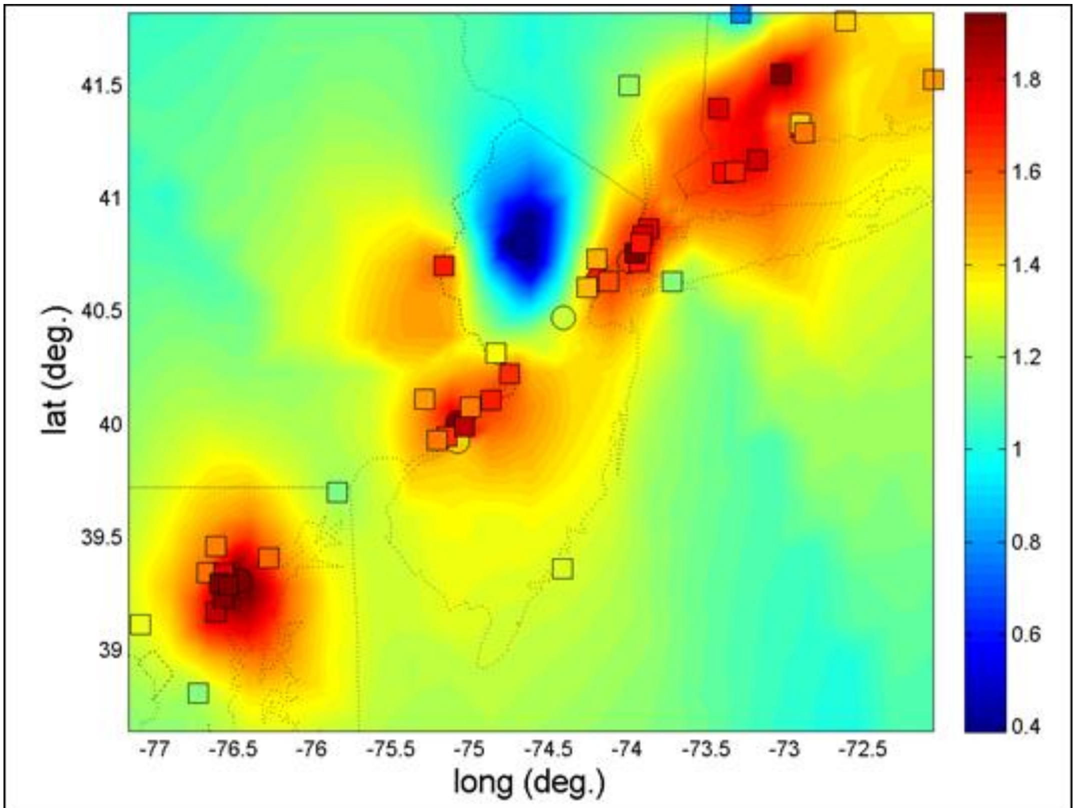


Figure 10: Estimated OC (log μg/m^3) on January 3, 2007. Circles represent observed (hard data) and square represent soft data generated using the mass fraction approach

An important feature of the map of Figure 10 is that it describes the spatial distribution of OC at a finer spatial resolution than a map that would only incorporate the hard OC data obtained at $PM_{2.5}$ speciation stations, or in other words, that would disregard the valuable information provided by the many stations that only monitors $PM_{2.5}$. The estimation error associated with the BME map of Figure 10 is shown in Figure 11. The BME estimation error is zero at the hard data locations, small (but greater than zero) at the soft data locations, and increases with distance away from the hard and soft data points.

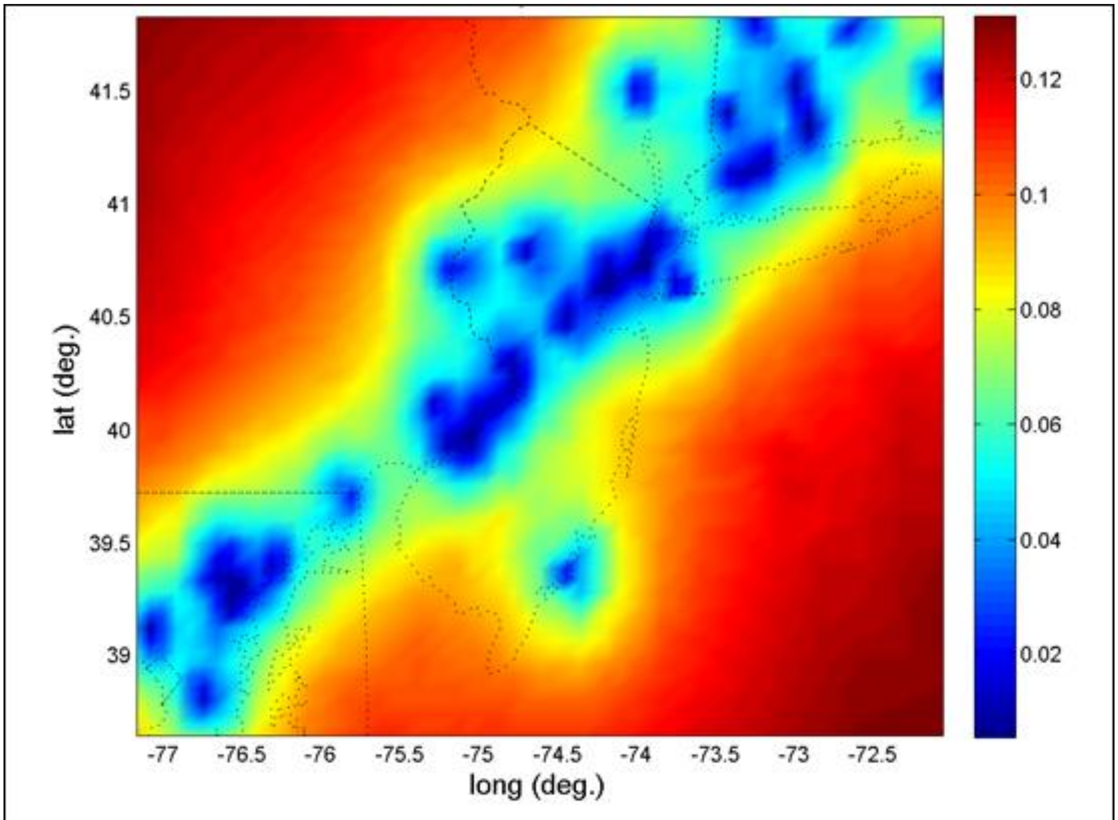


Figure 11: BME error variance of estimated OC (log $\mu g/m^3)^2$ on January 3, 2007.

NOTE: Figures for this study were generated in the MATLAB computing platform. Similar figures were generated using BMEGUI as described in the corresponding case study tutorial.

6.4. Conclusions:

This case study illustrates the power of BMEGUI in mapping air quality data. As shown in this case study, BMEGUI will provide reliable and accurate maps of $PM_{2.5}$ chemical constituents that can be used to better understand its spatial and temporal patterns across New Jersey and help in designing effective policies and regulations to contain air pollution.

6.5. References

De Nazelle[D], A., S. Arunachalam, **M.L. Serre**[†] (2010) Bayesian Maximum Entropy Integration of Ozone Observations and Model Predictions: An Application for Attainment Demonstration in North Carolina, *Environ. Sci. Technol*. Vol. 44, pp. 5707–5713

Allshouse[D], W.B., J.D. Pleil, S.M. Rappaport, **M.L. Serre**[†] (2009) Mass Fraction Spatiotemporal Geostatistics and its Application to Map Atmospheric Polycyclic Aromatic Hydrocarbons after 9/11, *Stochastic Environmental Research and Risk Assessment*. Vol. 23, pp. 1213–1223

7. Case Study3: Water Quality Monitoring Network Design

7.1. Introduction:

There is an ever-increasing demand for environmental monitoring data and other interpretive products derived from those data. However, due to budget constraints it is not possible to collect samples at all locations and times of interests. It is therefore necessary to efficiently design the monitoring network used to collect water quality data. In this work we propose a methodology that can be used to prioritize locations where new water quality monitoring stations should be installed in the coastal area of Ocean County in New Jersey State in order to efficiently expand the existing water quality monitoring network. The monitoring network methodology is based on the BME geostatistical method, which is used to identify locations with high estimation errors. Based on this methodology we identified 20 optimal locations for new monitoring sites in the study area. These new monitoring sites would collectively enhance the reliability of interpolated water quality across the domain of study region.

If the results of this case study are of interest to the NJDEP, then future works could consider the implementation of this proposed monitoring network design methodology in the BMEGUI.

7.2. Monitoring Water Quality:

Water quality monitoring is very important to protect the ecosystem and ultimately the public health. The importance of water quality monitoring has been well accepted among experts in sustainable water resources across the world. However, existing water quality monitoring networks may not be sufficient to provide comprehensive information about current water quality and resources. The U.S. commission on Ocean Policy (2004) emphasized the importance of water quality monitoring, saying that "Ongoing monitoring is essential to assess the health of ocean and coastal ecosystems and detect changes over time. More than any other measure, monitoring provides accountability for management actions. The nation needs a coordinated, comprehensive monitoring network that can provide the information necessary for managers to make informed decisions, adapt their actions as needed, and assure effective stewardship of ocean and coastal resources." (*An Ocean Blueprint for the 21st Century", U.S. Commission on Ocean Policy,* 2004.).

7.3. Study Area and Data Source:

Ocean County is located along the Jersey Shore of New Jersey State. Surface water quality is monitored by the Water Monitoring and Standards (WM&S) division of the New Jersey Department of Environmental Protection. The department collects and provides water quality information from water quality monitoring stations located throughout a extensive water body inlet near the coastal region (Figure 11). Ammonia concentrations (µg Nitrogen / L) measured from July 12th, 1989 to Sept 1, 2010 at irregular time intervals were provided by Mr. Mike Kusmiesz, GIS specialist at the Bureau of Marine Water Monitoring at NJDEP, who we are working with on this case study. Figure 11 shows the monitoring stations where ammonia concentrations were observed in Ocean County of New Jersey State.

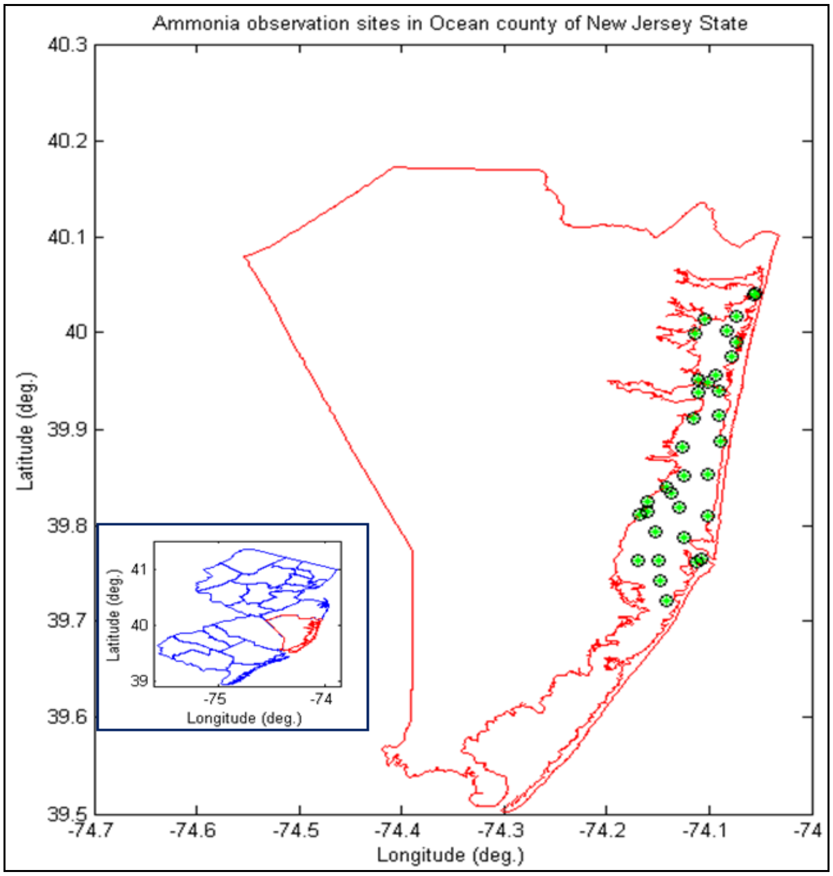


Figure 11: Ammonia monitoring sites in coastal area of Ocean County of New Jersey State

7.4. Network Design:

The proposed water quality network primarily provides optimal sites for additional monitoring stations for reliable and improved water quality information. The optimal monitoring sites in this study are based on a single water pollutant i.e. ammonia concentration in the water. However, the design approach can be extended for multiple water pollutants by incorporating all pollutants into a single number based on their toxicity potential level. For instance, a water quality index can be developed based for all pollutants present in the water and then the monitoring network can be designed for this index. This type of approach has been used successfully for designing air quality monitoring network. Statistical approaches such as principal component analysis (Peterson 1970), cluster analysis (Sabaton, 1976), linear programming (Hougland and Stephens 1976), and optimization based on the inverse of the estimation variance (Husain and Khan 1983) have been used widely for designing air quality monitoring network. However, these statistical design approaches rely substantially on estimates of pollutant concentrations obtained from high quality data on air pollution measurements and model outputs, or a combination of both. The coastal water region in this study poses a unique set of challenges because water pollution is strongly influenced by point and non point sources off its coast. As a result we have developed a monitoring network optimization procedure that mainly focuses on the uncertainty in the geostatistical estimation of water quality, as well as the mean average ammonia concentration over the 20 years observations.

7.5. Interpolation uncertainty:

The kriging method of linear Geostatistics provides the best linear unbiased estimation of water quality across a study domain. Hence the kriging method provides a robust tool to interpolate ammonia concentration in water at unobserved sites. Typically the kriging variance (model uncertainty) is zero at each monitoring station, and increases with distance away from monitoring stations. The distance over which the kriging variance increases away from a monitoring station is described by the spatial covariance function, which is a measure of the autocorrelation of ammonia across space.

The covariance model for ammonia was obtained using the BMEGUI. The spatial component of this covariance model can be mathematically expressed as by the equation below

$$C_X(r,0) = \sigma^2\left[(c_{01})\exp\left(-\frac{3r}{a_{r1}}\right) + (c_{02})\exp\left(-\frac{\sqrt{3}\,r}{a_{r2}}\right)^2\right]$$

where

c_{01}	c_{02}	a_{r1} (degree)	a_{r2} (degree)
285.07	427.61	0.085	0.300

Using this spatial component of the covariance model we calculated the kriging variance across domain of study region.

7.6. Placement of preselected monitoring stations

We calculate the priority index at any location of interest as the multiplication of kriging variance and the average ammonia concentration over the 20 years observations at that location. Thus, the priority index (PI) is higher for locations of higher interpolation uncertainty and higher mean ammonia concentration. The PI is used for the pre selection of 42 sites as follow: We create a regular grid of potential new locations for monitoring sites, and we select the grid point with the highest PI value as an optimal new monitoring site. Adding a station at that new location results in a total of 35 stations (34 existing stations plus 1 new station). We recalculate the kriging variance and PI map for these 35 stations. We then use this updated PI map to find the next grid point with the highest PI value where we locate our second preselected site. This results in a total of 36 stations (34 existing and 2 new stations). We iteratively repeat this procedure until we have 42 preselected sites. These preselected sites are shown in Figure 12 as “candidate” sites.

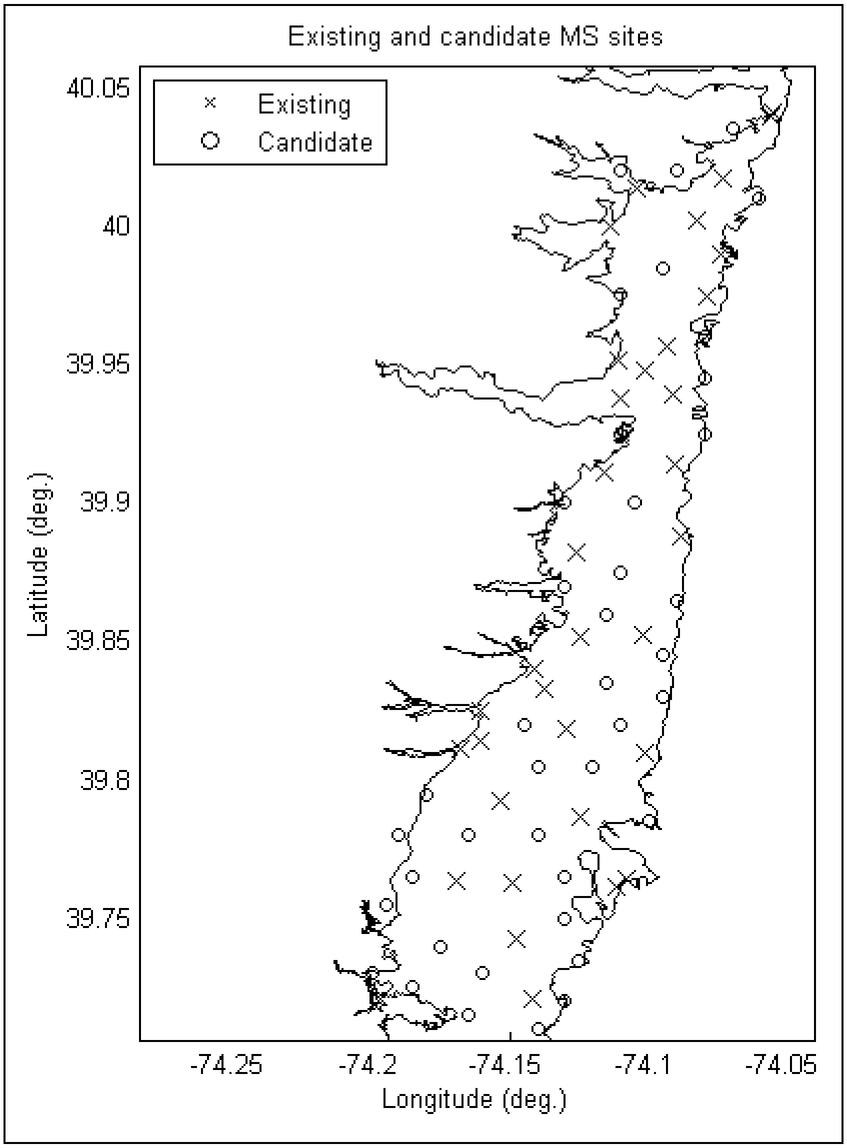


Figure 12: Existing and candidate monitoring sites

Out of the 42 candidate sites we selected 20 recommended sites based on expert judgment. In this work we simply selected recommended sites based on their PI score, but future work should include knowledge from the agency involved, consideration about accessibility to the sampling sites, cost of operation, etc. For each of the recommended site we calculated the percent reduction in aggregate priority index that results from installing a monitoring station at that site. The recommended sites are then ranked from largest to smallest reduction in aggregate PI. Figure 13 shows these final proposed sites and the kriging estimate of ammonia concentration (µg Nitrogen / L) on August 15, 2009.

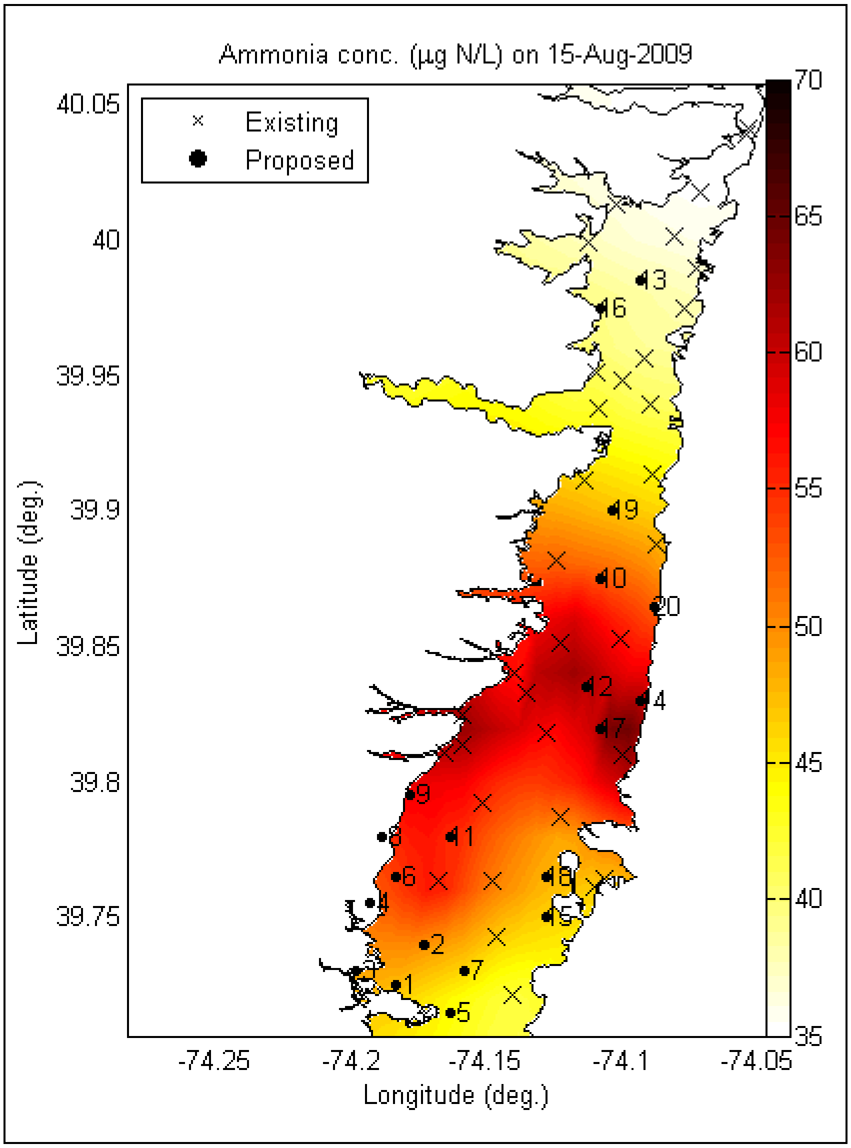


Figure 13: Kriging estimate of ammonia concentration (µg Nitrogen / L) on August 15, 2009

The priority of these proposed monitoring sites were assessed using the total percent reduction in PI value. This percent reduction in PI and priority rank of proposed monitoring sites are tabulated in table 2. The combination of Figure 13 and Table 2 provides NJDEP managers with a powerful tool to select optimal location where to locate new stations in a way that will result in the highest reduction of uncertainty where ammonia concentrations are high.

Table2: Priority rank and percent reduction in PI (over all interpolation uncertainty).

Percent reduction in total uncertainty	Priority Rank	Percent reduction in total uncertainty	Priority Rank
-5.65	1	-1.61	11

-5.49	2	-1.60	12
-4.92	3	-1.58	13
-4.80	4	-1.50	14
-3.80	5	-1.44	15
-3.75	6	-1.42	16
-3.49	7	-1.35	17
-3.06	8	-1.29	18
-2.07	9	-1.22	19
-1.70	10	-1.18	20

7.7. Conclusions:

The proposed water quality monitoring stations primarily provides optimal sites for additional monitoring stations for reliable information on interpolated water quality. First, a total of 42 candidate monitoring sites (Figure 11) were pre-selected based on the PI index. Then out of these candidate sites we selected 20 proposed sites, which we ranked based on percent reduction in overall interpolation uncertainty where ammonia concentration is high. This methodology is applicable to any water body and any combination of water quality parameters, and it is highly flexible as it relies on a priority index that can be adjusted as desired by the relevant agency.

We recommend that future works consider the implementation of a graphical user interface for this methodology, which can be ideally be done in the BMEGUI, and will provide users with a powerful tool to optimize water quality monitoring network operating under limited or reduced budgets. This tool can also be easily extended to choose which stations should be removed or how stations should be relocated in the case of budget cuts. Ultimately this tool may allow monitoring agencies to reduce the cost of their monitoring operation while improving water quality assessment.

7.8. References:

Husain T. and Khan S.M. (1983), Air monitoring network design using Fisher's information measures - a case study. *Atmos. Environ.* **17,** 2591-2598.

Sabaton C. (1976), Etude d’ optimisisation d’ un reseau de surveillance de la pollution atmospherique dans la region Parissienne. In *Atmospheric Pollutio*n, M.M. Benarie, Ed. Amsterdam: Elsevier, pp. 159-171.

Peterson J.T. (1970) Distribution of sulfur dioxide over metropolitan St. Louis, as described by empirical eigenvectors, and its relation to meteorological parameters. *Atmos. Environ.* **4,** 501-518.

Hougland A. and Stephens R.(1976), Sitting of air pollution monitors by analytical techniques. *JAPCA* 26(1), 51-53.

QUALITY ASSURANCE PLAN (QAP)

Quality Assurance Plan (QAP): The following three steps have been adopted to ensure the quality and development of the upgraded version of the BMEGUI software

1. Planning and implementation
2. Testing
3. Corrective action

Planning and Implementation: Many significant improvements on the functionality of BMEGUI have been implemented and successfully executed. These added BMEGUI features will greatly enhance the user's experience when performing a spatiotemporal analysis of environmental variables. A new approach to estimate water quality variables along river network has also been implemented in this upgraded version. We followed exactly the plan for the version upgrade, expansion, and improvement of the BMEGUI as described in the Methods section of contract SR08-037.

Final Testing and Delivery: The final upgraded and improved version of BMEGUI has been tested after each new BMEGUI features are implemented and executed successfully. A comprehensive testing procedure was used to test each feature separately, and then as a complete package, in order to ensure a good product quality. This testing was involved the operation of BMEGUI under normal conditions (i.e. for a typical user with knowledge of this software and the way it should be used) and abnormal conditions (i.e. for a person without any knowledge of this software).

Corrective Action: Any defect or flaw in design and calculation, and any code bug identified in the upgraded version of BMEGUI, have been fixed and tested again up until no further changes were needed. We conclude that the upgraded BMEGUI is bug free under normal operating conditions and is as user friendly.