Evaluating Water Quality using Spatial Interpolation Methods, Pinellas County, Florida, U.S.A.
Cynthia Meyer, Pinellas County Department of Environmental Management

Abstract
In 2003, Pinellas County instituted a stratified random sampling program for ambient water quality monitoring. Pinellas County is a peninsula bounded by Tampa Bay and the Gulf of Mexico with 588 miles of coastline. Previously, water quality for each strata was determined by applying a mean calculation of water quality parameters to the entire area. To visualize the spatial trends of the data, spatial interpolation methods were applied to the chlorophyll and dissolved oxygen data. Investigation determined Inverse Distance Weighting (IDW) as the best interpolation method for the data. The IDW surfaces displayed overall trends and hotspots for extreme high and low values. Comparison with the State Water Quality Scores revealed that only small areas within several of the strata violated the impaired waters rule. The County can use spatial data analysis techniques to develop more effective watershed management plans by targeting the areas in need of remediation.

Introduction
Pinellas County is on the west coast of central Florida. The County, a peninsula, is bounded by Tampa Bay and the Gulf of Mexico with 588 miles of coastline (Figure 1). The County surface waters include coastal and intracoastal areas, lakes and streams, and shallow waters of Tampa Bay.

Figure 1. Location map for Pinellas County, Florida, USA.
In 1989, the County adopted a comprehensive plan that mandated the implementation of an ambient water quality monitoring program. From 1990-2002, Pinellas County Department of Environmental Management (PCDEM) used fixed stations to monitor 45 of the county’s 52 drainage basins, four lakes, and nine receiving water bodies (Moores et al., (1992), Moores et al., (1994), Meyers et al., (2000), Squires et al. (2003), Flock et al., (2004)).

In 2003, a revised monitoring program was implemented to collect samples with an improved geographical extent and produce more statistically defensible results. Developed in conjunction with Janicki Environmental, Inc, the stratified-random design is based on a probabilistic sampling scheme used by the Environmental Protection Agency (EPA) in their Environmental Monitoring Assessment Program (EMAP). The EMAP-based design consists of overlaying a hexagonal grid by strata, and randomly selecting a sample location within each grid cell. This allows the estimation of surface area for water quality conditions within each stratum. The stratified-random design allows for statistical methods to be applied estimating population means and confidence limits for water quality metrics (Janicki, 2003).

The stratified-random design has a temporal and spatial component. Lake Tarpon, Lake Seminole, and the marine waters along the shores of Pinellas County were subdivided into 19 strata (Figure 2). In each calendar year, a total of 36 sample sites were selected for each stratum. At least one of these 36 sites was located in each hexagonal grid in the respective stratum. These 36 sites are called the primary sampling sites. Four sites are assigned to a sample period; there are nine sampling periods in the calendar year. The sites in Tampa Bay were also restricted to depths less than 2 meters defined by the 2- meter mean lower low water isobath in Tampa Bay (Janicki, 2003).

Figure 2. Pinellas County water quality monitoring strata.
At each sampling site, samples are collected for physical and chemical parameters. Parameters measured in situ included temperature, salinity, specific conductance, pH, dissolved oxygen, and Secchi depth. Analyses of grab samples collected from the field included chlorophyll (a, b, c), nutrients (total Kjeldahl nitrogen, ammonia nitrogen, nitrate + nitrite nitrogen, total phosphorus, and dissolved orthophosphorus), total suspended solids, transmissivity, and turbidity (Flock, et al., 2004).

All of the County ambient monitoring data results are periodically uploaded into the United States Environmental Protection Agency's STORET (STOrage AND RETrieval) database. Data can also be downloaded from the Pinellas County Water Atlas (http://www.pinellas.wateratlas.usf.edu/).

The data analysis meets the requirements set forth by the State of Florida under Chapter 62-302, Surface Water Quality Standards, and implemented through Chapter 62-303, Identification of Impaired Surface Waters (IWR), of the Florida Administrative Code (FAC). A table with the surface water quality criteria can be found under Chapter 62-302.530 (FAC) at http://www.dep.state.fl.us/legal/rules/shared/62-302.pdf. Violations of the IWR occur when a certain percentage or number of samples do not meet the State Water Quality Standard for a particular parameter. The analysis for this data focused on calculating a mean criteria score for each water body. The score was categorized into a classification of impairment. The open water strata were labeled good (no IWR violation), fair (additional 25% violations would result in exceedance of IWR), and poor (exceedance of one or more IWR criteria) (Figure 3) (Flock, et al., 2004).

Figure 3. Method of data analysis based on calculating a mean score for the data 2003 and 2004.
The stratified random monitoring design is amendable to the use of spatial interpolation analysis. Several spatial interpolation methods such as Inverse Distance Weighting, Local Polynomial, Radial Basis, and Kriging were investigated to determine the most appropriate analysis (Meyer, 2004). The Inverse Distance Weighting method (IDW) worked best with the limited sample size and random data points. Inverse Distance Weighting (IDW) is a deterministic and exact interpolation method. Exact interpolation means that the surface passes directly through the known points. It requires few parameter decisions and does not assess any prediction errors. IDW is a robust method that assumes a value can be approximated as a weighted average of its neighboring values. In turn, it does not always reproduce the local shape implied by the data. It also tends to produce local extrema at the data points (Mitas, et al. 1999). IDW uses the neighboring points to interpolate the area between the points based on a weighted distance function (Flock, et al., 2004). When additional data are available, Kriging will be the most appropriate method due to its dependence on the use of random functions; data requirements of stationary stochastic process and normal distribution; and interpolation based on statistical methods. This model is stochastic and can be exact or inexact depending on the error associated with the data measures. The flexibility of the model relies on the parameter settings. It also assesses autocorrelation and errors of the prediction (Haining, 2003).

This study focuses on the interpolation of water quality data in select strata on the east coast of Pinellas County in Tampa Bay. Chlorophyll-a and Dissolved Oxygen were evaluated for comparison with the IWR. Water column chlorophyll-a (Chl-a) concentrations measure the quantity or biomass of planktonic algae or phytoplankton. Dissolved oxygen (DO) concentrations measure the amount of available oxygen in the water that enters the aquatic environment from the atmosphere (wind, waves, direct diffusion), plant photosynthesis, and mixing or diffusion from more oxygenated water masses (Flock, et al., 2004). By using spatial interpolation, the areas of concern for exceedances of the IWR can be identified. In addition, spatial extent of the potentially impaired area can be estimated. This information will be critical for the prioritization of Watershed Management Plan development and requirements of the State’s TMDL program.

Methods

Data Collection: The sampling locations are chosen through a series of stratified random selections. The sampling crew collects the water samples within 100 ft of the site’s geographic coordinates. Field sample collections and measurements are carried out according to FDEP Standard Operating Procedures (FDEP, 2004). Physical parameters (temperature, pH, dissolved oxygen, conductivity, and salinity) were measured using Hydrolab® multiprobe units. Surface readings are taken at a depth of 0.2m from the surface. Water samples are collected at 0.2m using a horizontally oriented Alpha bottle water sampler (Nolan et al., 2003). The Pinellas County Utilities Department Laboratory, a National Environmental Laboratory Accreditation Conference (NELAC) certified lab, performs most sample analyses (Flock, et al., 2004).
Data Processing: The dataset including water quality parameters and GPS coordinates was imported into the ESRI ArcMap software. The ESRI Geographic Information System (GIS) software was used for the construction and creation of the interpolation surfaces. The IDWs were processed with the ESRI Geostatistical Analyst extension. The data were examined for normality for each variable, and checked for location and qualifiers before being interpolated. The interpolated surfaces were compared with data collected during the same time period by the Fisheries Independent Monitoring Program (FIM) at Florida Fish and Wildlife Research Institute or the Environmental Protection Commission of Hillsborough County.

The data for each DO and Chl-a value was subset into training and testing sets. Fifteen percent of the sampling sites were used for validation. The IDWs for each DO and Chl-a were calculated using the parameters 5-10 neighbors with a power of 1 and 10-15 neighbors with a power of 1. These surfaces were compared using the Student’s Paired T-test assuming equal variance (alpha=0.05). Additionally, two DO IDWs were calculated from the PCDEM data and the FIM data. These surfaces were statistically compared with the Student’s Paired T- test (alpha=0.05) in Excel. The comparison of the measured, predicted, and error values was tested at alpha = 0.05. The datasets were combined to calculate a more comprehensive and spatially extensive IDW.

The IDWs were then exported to raster layers and extracted using the water quality strata layer as a mask. The scales for the surfaces were set to the same numerical and color classification for each Chl-a and DO surfaces. The ESRI Spatial Analyst Extension was used to further analyze the surfaces. The raster calculator extracted areas that would be considered potentially impaired according to the IWR. The zonal statistics tool was used to estimate the percent of area per strata that could be possibly impaired for each Chl-a and/or DO measurement.

Results

Dissolved Oxygen:

Initially two independent dataset were examined for dissolved oxygen in 2003 and 2004. The Kurtosis (4.10) and the Skewness (-0.07) indicate that the distribution meets the assumption of normality. The FIM data has 2,323 observations in the 2003 and 2004 time span with a mean of 7.02 mg/L (SD 1.88). The PCDEM data has a Kurtosis (4.62) and a Skewness (-0.80) which fulfills the assumption of normality for the distribution. With 533 observations the mean (6.86 mg/L, SD 1.75) is higher than the FDEP single measure standard of 4 mg/L and a daily average standard greater than 5 mg/L.

IDW interpolation models were calculated for each dataset. The model fit the FIM data with prediction errors of mean = -0.015 and RMS = 0.799. The PCDEM data had a slightly higher mean = 0.156 and RMS of 1.731. In comparison, it is apparent that the FIM model, most likely due to the greater amount of observations, had a better fit for the IDW model (Figure 4). However, the PCDEM model also has a fairly good fit.
The FIM and PCDEM datasets were statistically compared using the Student’s T-test for two samples with equal variances. The hypothesized mean difference was zero at alpha = 0.05. There was no significant difference between the observed DO datasets (t-stat = -1.83, p = 0.06, df = 2854). Thus, the datasets were merged together to produce the most comprehensive and spatially extensive IDW surface. The combined data, used for the remainder of the analyses, meets the assumptions of normality (Kurtosis = 4.22, Skewness = -0.18). The mean DO was 6.99 (SD = 1.858).

Prior to calculating the models, the dataset was subset into a training set (90%) and testing set (10%). The training set was used to interpolate the surface and the testing set was used for validation of the model. The IDW models were interpolated using two different parameter settings:

A. minimum neighbors = 10, maximum neighbors = 15, Optimal Power = 1
B. minimum neighbors = 5, maximum neighbors = 10, Optimal Power = 1

Model A and B were compared to determine if the parameter setting caused a significant difference between the models. The hypothesized difference in mean was zero at alpha= 0.05. Using the Student’s T-test for Paired Samples with equal variance, it was determined that there was no significant difference between model A and B for the prediction values (t-stat = 0.778, p = 0.436, df = 285). The Pearson’s Correlation for the prediction values was 0.997 indicating that model A and B are likely to have no significant difference. The error values were compared using the Student’s T-test for Two samples assuming equal variance. There was no significant difference between the models (t-stat = 0.488, p = 0.961, df = 570). To avoid over-generalization of the interpolated surface, Model B was more appropriate for this analysis.

The IDW model B was evaluated for the measured versus predicted values using a Student’s paired T-test. There was no significant difference between the measured values (mean= 6.912) and the predicted values (mean= 6.914) (t-stat = -0.025, p = 0.979, df = 285, alpha = 0.05).

The IDW interpolation is presented (Figure 5) with the locations of the sampling points. The area with the lowest DO values occurs in the northwest portion of bay. In the more southern regions of the bay the DO values increase, likely related to the

![Cross Validation Comparison](image-url)
increased circulation near the mouth of the bay. Areas outside of the Sampling Strata (outlined in red) can not be considered with confidence because of the lack of sampling points.

Figure 5. The IDW model of DO for 2003 and 2004 in Tampa Bay.

The IDW model was converted into a raster grid to calculate the percentage of area in each strata that exceeds the FDEP water quality standards. According to the FDEP (T. Singleton-FDEP, personal communication, 2006), the state standard for DO for Class II and III marine or estuarine waters shall not average less than 5 mg/L in a 24 hr period and shall be never less than 4 mg/L as documented in the Florida Administrative Code in Chapter 62-302.530. The standards can be found at http://www.dep.state.fl.us/legal/rules/shared/62-302.pdf.Using the Spatial Analyst extension, calculations for the potentially impaired areas were based on values where DO <=5 mg/L and DO < 6 mg/L. The areas where DO <= 5 mg/L are potentially impaired; whereas, areas where DO < 6 mg/L could also be at risk for impairment due to the stipulation that the average DO value for a 24 hr period can not be less that 5 mg/L. Since most sampling occurred during the mid-day hours when DO is at the highest concentration, it is possible for the 24-hr average to be lower than the measured value. The northwest strata had the highest percent of potentially impaired areas (Table 1 and Figure 6). The percent of potentially impaired area decreases in the southern strata nearer to the mouth of the bay.
Figure 6. Percent of potentially impaired areas within sampling strata for Dissolved Oxygen concentrations.

<table>
<thead>
<tr>
<th>Strata</th>
<th>DO &lt; 6 mg/L</th>
<th>DO &lt;= 5 mg/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>23.44</td>
<td>5.39</td>
</tr>
<tr>
<td>E2</td>
<td>18.54</td>
<td>1.03</td>
</tr>
<tr>
<td>E3</td>
<td>2.35</td>
<td>0.51</td>
</tr>
<tr>
<td>E4</td>
<td>1.01</td>
<td>0.13</td>
</tr>
<tr>
<td>E5</td>
<td>6.51</td>
<td>0.16</td>
</tr>
<tr>
<td>E6</td>
<td>1.72</td>
<td>0.05</td>
</tr>
<tr>
<td>E7</td>
<td>0.34</td>
<td>0</td>
</tr>
<tr>
<td>RB</td>
<td>20.05</td>
<td>3.71</td>
</tr>
</tbody>
</table>

Table 1. Percent of each strata that is potentially impaired for DO.

Chlorophyll:

The Chl-a data was examined for normality prior to interpolation. The mean (9.48 ug/L, SD 9.0) was lower than the FDEP limit of 11 ug/L. The distribution is not heteroscedastic with a Kurtosis of 55.61 and Skewness of 5.27. The frequency histogram (Figure 7) demonstrates the lack of normality for the 549 observations.

Figure 7. Frequency histogram for the Chl-a data.
The IDW models were interpolated using two different parameter settings:

A. minimum neighbors = 10, maximum neighbors = 15, Optimal Power = 1
B. minimum neighbors = 5, maximum neighbors = 10, Optimal Power = 1

Model A and B were compared to determine if the parameter setting caused a significant difference between the models. The hypothesized difference in mean was zero (alpha = 0.05). Using the Student’s T-test for Paired Samples with equal variance, it was determined that there was no significant difference between model A and B for the prediction values (t-stat = -0.544, p = 0.586, df = 549) nor the error values (t-stat = -0.544, p = 0.586, df = 549). The Pearson’s Correlation for the prediction values was 0.983, and the Error values was 0.995 indicating that model A and B are likely to have no significant difference. Additionally, the error values were compared using a T-test for two-samples assuming equal variance (t-stat = -0.036, p = 0.97, df = 549) showed the models were not significantly different.

Finding no significant difference between the two models, Model A was used for the remainder of the analysis. Due to the variation in the data, Model A generalizes the data more by using a greater number of points for the interpolation. Thus, Model A is more robust for this analysis.

The IDW model A was evaluated for the measured versus predicted values using a Student’s paired T-test. There was no significant difference between the measured values (mean = 9.48) and the predicted values (mean = 9.48) (t-stat = 1.28, p = 0.20, df = 549, alpha = 0.05).

The IDW interpolation is presented (Figure 8) with the locations of the sampling points. The area with the highest Chl-a values occur in the northwest portion of bay. In the more southern regions of the bay the Chl-a values decrease. Most likely the decrease in Chl-a is related to the increased circulation near the mouth of the bay. Areas outside of the Sampling Strata (outlined in red) can not be considered with confidence because of the lack of sampling points.
The IDW model was converted into a raster grid to calculate the percentage of area in each strata that exceeds the FDEP water quality standards. According to the FDEP (T. Singleton-FDEP, personal communication, 2006), there is no numeric State standard for Chlorophyll in marine or freshwater; the State, however, uses the criteria of 11 ug/L for marine and estuarine waters documented in the Florida Administrative Code in Chapter 62-302.530. Using the ESRI Spatial Analyst extension, the potentially impaired areas (Chl-a >= 11 ug/L) were identified. The northwest strata had the highest percent of potentially impaired areas (Table 2, Figure 9). The percent of potentially impaired area decreases in the southern strata nearer to the mouth of the bay.
# Table 2. Percent of each strata that is potentially impaired for Chl-a concentrations &ge; 11 ug/L.

<table>
<thead>
<tr>
<th>Strata</th>
<th>Percent of Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>66.34</td>
</tr>
<tr>
<td>E2</td>
<td>76.63</td>
</tr>
<tr>
<td>E3</td>
<td>63.88</td>
</tr>
<tr>
<td>E4</td>
<td>4.42</td>
</tr>
<tr>
<td>E5</td>
<td>1.97</td>
</tr>
<tr>
<td>E6</td>
<td>0.08</td>
</tr>
<tr>
<td>E7</td>
<td>0.29</td>
</tr>
<tr>
<td>RB</td>
<td>1.17</td>
</tr>
</tbody>
</table>

**Discussion**

The IDW spatial analysis models suggest that the previous analysis of impairment (Figure 3) overestimates the area per sampling strata that is potentially impaired. Through spatial analyses of statistically defensible interpolation models, only 1.10% (DO &le; 5 mg/L) of the Pinellas County estuarine waters in Tampa Bay show potential impairment for the FDEP Dissolved Oxygen standards. Additionally, the analysis determined that 7.14% of the study area may be in risk of potential impairment (DO &lt; 6 mg/L) due to the criteria that the average DO over a 24 hour period can not be less than 5 mg/L. Since the water samples are mostly collected during the mid-day hours, it is feasible that the 24-hr average is less than the sample reading. The models suggest that 31.5% of the study area may violate the IWR for exceeding the Chlorophyll-a requirements (Chl-a &ge; 11 ug/L). Dissolved oxygen concentrations strongly influence the species distribution in the marine and estuarine environment. Oxygen enters the aquatic environment from the atmosphere (wind, waves, direct diffusion), plant photosynthesis, and mixing or diffusion from more oxygenated water masses. Less oxygen can be dissolved in water as water temperature increases. Biological factors such as increased metabolic rates and oxygen uptake rates of aquatic organisms may further reduce dissolved oxygen levels. (Flock, et al., 2004)

In this study, areas of potentially impaired DO tend to coincide with areas known to have mucky sediment. The mucky sediment typically has a lower dissolved oxygen concentration due to the consumption of available oxygen during the decomposition of organic materials. The potentially impaired areas for DO in strata E1 (Figure 3 and Figure 5) occur near the Lake Tarpon Canal outfall, and Alligator Lake discharge. The increase in organic materials from these sources may effect the benthic sediment composition and in turn, the dissolved oxygen concentration. Additional analyses could focus on integrating the habitat information such as bottom type with the dissolved oxygen hotspots to determine the correlation between sediment composition and DO levels.
Increased concentrations of Chlorophyll-a can be caused by phytoplankton blooms usually catalyzed by excessive nutrient loadings. The blooms reduce water clarity and affect the light availability for bottom vegetation such as seagrasses. In turn, the reduction in seagrass growth has an adverse effect on the fauna in the affected area. In this study, the occurrences of extremely high Chl-a readings were in the northwest area (E1). At least two samples of greater than 100 ug/L were collected immediately following a hurricane in Fall 2004. The high readings indicate that the increase in pollutant loads from terrestrial run-off may have a strong effect on catalyzed phytoplankton blooms.

The characteristics of water circulation in Tampa Bay may reflect on the extreme minimum and maximum DO and Chl-a values in the northwest strata. According to previous studies using circulation simulation models (Burwell et al., 2000, Meyers et al., 2006, Weisberg and Zheng, 2006), the residence time for water in the bay can be upwards of 100 days (Weisberg and Zheng, 2006). Three causeways traverse the bay causing a reduction in the water circulation. In addition, the majority of the bay is less than 10 ft deep. Hence, the water circulation concentrates in the deep shipping channels that run through the bay (Burwell et al. (2000), Weisberg and Zheng (2006)). Although the residence times in the deep channels are relatively short, the residence times increase drastically in the shallow waters (Burwell et al., 2000). The causeways cause persistent eddies in the circulation patterns (Weisberg and Zheng (2006), Burwell et al. (2000)). The alteration of the natural circulation patterns may prevent the mixing of the water column. Increasing circulation and mixing may reduce the occurrence of phytoplankton blooms and homogenize the DO concentrations in the water column.

The use of Spatial Interpolation modeling for determining potential impaired waters provide more detailed spatial information, help to develop scientifically sound watershed management plans and improve remediation. By pin pointing the problem areas, Pinellas County can effectively allocate funds for environmental improvement to specific areas resulting in a more efficient improvement for the water body. The water body targeted in the study includes a large portion of the bay that was listed as impaired in 2002 by the FDEP in the 303(d) list for Chlorophyll-a and Dissolved Oxygen. The results of this analysis further improve the FDEP assessment by providing more spatially detailed information about the areas exceeding the state water quality criteria.

The IDW models in this study are limited to the spatial coverage of the data points and the quality of the data. The complex shoreline of Tampa Bay makes the extrapolation of the data less reliable at the edges of the surface. Before any interpolation or analyses, the data must be examined to determine if it meets the assumptions of the spatial model. Diversion from a normal data distribution reduces the reliability of the surface predictions. Also, the location of the data points must adequately cover the target area. In this study, the random stratified sampling design is amendable to the use of spatial interpolation methods. As more data becomes available, more accurate and precise spatial interpolation methods, such as Kriging, can be used for analyzing the water quality in Tampa Bay. In summary, the success of the model interpolation and minimization of prediction errors is deeply based in the characteristics of the dataset and the model parameters.
Conclusion

The spatial interpolation models for dissolved oxygen and chlorophyll-a demonstrate a more comprehensive and spatially extensive representation of the water quality in Tampa Bay. Contrary to the previous evaluation of water quality parameters assessed with a single mean value per strata, the IDW models spatially indicate hotspots of impaired areas. The information provided by this spatial interpolation study should assist the FDEP in the refinement and implementation of the IWR evaluation of Tampa Bay by more accurately identifying areas with potential impairments. Ultimately, the improved analysis will allow the development and implementation best management practices (BMPs) for the protection and remediation of potentially impaired areas.

Acknowledgements

Pinellas County: Mark Flock, Scott Deitche, Kelli Hammer-Levy, Craig Dye, Melissa Harrison, Robert McWilliams, Sue Myers, Kevin O’Donnell, Anamarie Rivera, Andy Squires, Melanie Weed, Geoff Duncan, Joe Torok. Florida Fish and Wildlife Reaseach Institute: Bob Muller, Chris Anderson Fishery Independent Monitoring: Nicole Dunham, Tim MacDonald Janicki Environmental, Inc: David Wade, Tony Janicki Florida Department of Environmental Protection: Tom Singleton Environmental Protection Commission Hillsborough County: Ed Sherwood

References


**Author Information:**

Cynthia Meyer  
Environmental Analyst  
Pinellas County Watershed Management  
300 South Garden Avenue  
Clearwater, FL 33756, USA  
Phone: 727-464-4425  
Email: cmeyer@pinellascounty.org