Kriging-based Interpolation of Dissolved Oxygen in Chesapeake Bay

Rebecca Murphy

*EA Science, Engineering and Technology, Inc.*

(*formerly Johns Hopkins University*)

CBP Tidal Monitoring and Analysis Workgroup Dissolved Oxygen Seminar

May 7, 2013
Guiding Questions/Outline

- What method(s) do you use for assessing DO?
- Why did you choose that methodology?
- What have you found drives DO patterns in the Bay?
- What lessons have you learned that we should consider as we move forward?
Outline

- What method(s) do you use for assessing DO?
  - Why did you choose that methodology?
    - For analysis of the relationship between hypoxic volume and nitrogen loads in Chesapeake Bay
- What have you found drives DO patterns in the Bay?
- What lessons have you learned that we should consider as we move forward?
Hypoxia: Shift in relation to nutrient loads?

Data Sets
CBP data
USGS River Monitoring
Historic Chesapeake Bay Institute data
Model outputs, etc...

New Tools/Methods


CBEO website: http://cbeo.communitymodeling.org
Hypoxia: Shift in relation to nutrient loads?

Hypotheses investigated by CBEO team

- Is the shift an artifact of the interpolation method used?
- Has there been a change in the stratification of the bay?
- Has there been an increase in other nutrient loads besides TN from the Susquehanna River?
- Could enhanced nutrient recycling from the sediment account for the difference?

Outline

1) Main-stem method
- What method(s) do you use for assessing DO?
  - DO analysis: Hypoxic volume
  - Spatial interpolation: Kriging
  - Extent:
    - Main-stem only
    - Full-bay
- Why did you choose that methodology?
- What have you found drives DO patterns in the Bay?
- What lessons have you learned that we should consider as we move forward?

2) Full-bay method
Spatial Interpolation: Different methods

General: Weighted average prediction

\[ \hat{Y}(s_0) = \sum_{i=1}^{n} w(s_i)Y(s_i) \]

- \( s_0 \) = location (x,y) to predict at
- \( s_i \) = different x,y location with measured value
- \( \hat{Y}(s_0) \) = value predicted at location \( s_0 \)
- \( Y(s_i) \) = value measured at location \( s_i \)
- \( w(s_i) \) = weight assigned to measured value at \( s_i \)

Inverse distance weighting (IDW) squared predictor weights

\[ w(s_i) = \frac{1/d_i^2}{\sum_{j=1}^{n}(1/d_j^2)} \]

d_i = distance between \( s_o \) and location \( s_i \)

Kriging predictor weights

\[ w(s_i) \] is dependent on the spatial relationship between the observed data, as quantified with a variogram
Method 1: Main Channel

Main stem DO

Murphy, Kemp, and Ball. 2011. 
Estuaries and Coasts 34:1293-1309.
Method 2: Full-bay method

Challenge for interpolation in estuaries: Shape

Solution 1: Segmentation

Solution 2: Water distance
Method 2: Full-bay method – surface only

Use all data below fall-line in Bay

Compute “water distance” between all samples and points on a prediction grid

Transform matrix of water distances into Euclidean distances with multi-dimensional scaling

Use transformed distances to compute variograms and perform kriging

Water distances computed by Eric Perlman

CBP data

[Image of map with salinity data and variograms]
Method 2: Full Bay

Compare to two other methods

Perform comparisons for:

- Seven parameters in surface layer
- Multiple datasets in 2007 and 2008

a) Segmented IDW

b) Kriging with Euclidean distance
Surface interpolation comparisons

July 2007 Surface Salinity

(a) IDW using Euclidean Distance in Segments
(b) Euclidean-Distance-Based Kriging in Entire Bay
(c) Water-Distance-Based Kriging in Entire Bay

Legend
Salinity estimates (ppt)

○ Salinity observations (same scale)
**Method 2: Full Bay**

**Cross validation of surface interpolation results**

- Hold out each sample, one at a time, and predict at the sample location
- Examine on plots, compute average errors

![Segmented IDW](image1)

- RMSE = 2.3
- MAE = 1.2
- ME = 0.35

![Kriging: Water Distance](image2)

- RMSE = 1.1
- MAE = 0.72
- ME = 0.036
Water-distance-based method: Layers

- Apply water-distance-based interpolation to data at every 1-m of depth
- To improve the vertical estimates -- use an indicator covariate as to whether each depth is above or below the pycnocline

\[
DO(s) = \beta_0 + \beta_0 P(s) + \varepsilon(s)
\]

- \(P(s)\) = pycnocline indicator
- \(\beta_i\) = fitted coefficient
- \(\varepsilon(s)\) = spatially varying random error
Water-distance-based method: 3-D output
Overall: Water-distance-based method

- Good in shallow water
- Captures lateral mainstem patterns
- Some problems at deeper depths
- Can be used to compute hypoxic volume for the full bay
Outline

• What method(s) do you use for assessing DO?
• Why did you choose that methodology?
• What have you found drives DO patterns in the Bay?
  • Averaging period impacts perceived trend
  • Stratification and nitrogen loads
• What lessons have you learned that we should consider as we move forward?
Main channel DO examples: June and July 1986 and 2005

Hypoxia trends in relation to N-loading

- Early July hypoxia shows increase since 1980
- Inter-annual variations blur long-term trends; clarify with running means
- Late July hypoxia has actually declined since 1980s
- N-loading increased until mid-1980s, then declined gradually into 2000s
- Hypoxia and late July N-load significantly correlated (p<0.001)

From: Murphy et al. 2011.
Decadal average hypoxic volume vs. TN

From: Murphy et al. 2011.
Entire summer hypoxic volume trends

From: Murphy et al. 2011.
Stratification long-term patterns

Average Monthly Main Channel Stratification Strength

Results Part 1: Main-channel hypoxia

From: Murphy et al. 2011.
Outline

- What method(s) do you use for assessing DO?
- Why did you choose that methodology?
- What have you found drives DO patterns in the Bay?
  - Main channel hypoxia drives full-bay long-term trends
- What lessons have you learned that we should consider as we move forward?

Full-bay method
Re-evaluation of hypoxic volume trends

Mainstem Hypoxic Volume (DO<1 mg/L)

(a) Early July

(b) Late July

Water distance

(a) Early July mainstem hypoxic volumes

(b) Late July mainstem hypoxic volumes
Results Part 2: Water distance
Conclusions

• What have you found drives DO patterns in the Bay?
  • A trend in early summer stratification appears to have lead to an increase in early summer hypoxic volume.
  • The different early and late summer hypoxic volume trends observed are not artifacts of the main-channel-only interpolations.
  • More specific analysis could be done in tributaries.

• What lessons have you learned that we should consider as we move forward?
  • Timing is important when analyzing hypoxia.
  • Average monthly or summer hypoxia estimates may result in lost information.
Acknowledgements

- JHU Advisor: Bill Ball
- Committee members: Grace Brush, Frank Curriero, Dominic Di Toro, Michael Kemp
- “Distance” sources: Eric Perlman and Andrew Miglino
- CBEO team: Randal Burns, Damian Brady, Jen Bosch, Jeremy Testa, Maureen Brooks
- Data sources: CBP, CBI, USGS, NOAA, MDNR, VECOS
- NSF Grants no. 0618986 and 0854329
- EA Science, Engineering, Technology
Extras
Chesapeake Bay

Bay bathymetry

- Bay is on average 6.5 m deep, but a much deeper channel follows the center of the Bay
- Most of the year, the Bay is stratified with fresher water on top, saltier water on bottom

Bathymetry data from NOAA
**Hypoxic volume: Temporal extent**

**Hypoxic volume** = volume of water with dissolved oxygen (DO) below certain critical values

Interpolation method used for management

1. Split bay into segments and 1-m depth layers

2. Interpolate samples in or near each segment using IDW
   \[
   \hat{Y}(s_0) = \sum_{i=1}^{n} w(s_i)Y(s_i) \quad w(s_i) = \frac{1/d_i^2}{\sum_{j=1}^{n}(1/d_j^2)}
   \]

4. Combine interpolation output by segment and layer

Images from Vol36 User’s Guide (Bahner 2006)
Variogram

Classical Variogram Estimator

\[ 2\hat{\gamma}(h) = \frac{1}{|N(h)|} \sum_{N(h)} (y(s_i) - y(s_j))^2 \]

- \( N(h) \) contain the pairs of locations \((s_i, s_j)\) that are a distance \(|h|\) apart, \( h = s_i - s_j \).
- \(|N(h)|\) is the number of such pairs.
Some notes on kriging

Variogram
1. Examine an empirical variogram of observed data \( Y_2 \) as a function of the Euclidean distance between points

\[
2\hat{\gamma}(h) = \frac{1}{N(h)} \sum_{N(h)} [Y(s_i) - Y(s_j)]^2
\]

\( s_{ij} \) are locations
\( N(h) \) is number of pairs distance \( h \) apart

2. Fit a valid variogram model

Estimate unknown values \( Y_1 \) given \( Y_2 \)
3. Use variogram model to fill in covariance matrix:

\[
\Sigma = \begin{bmatrix}
\Sigma_{11} & \Sigma_{12} \\
\Sigma_{21} & \Sigma_{22}
\end{bmatrix}
\]

4. Use the conditional distribution of the unobserved data given the observed to generate estimates for \( Y_1 \)

\[
E(Y_1 \mid Y_2) = \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (Y_2 - \mu_2)
\]

\( \mu_2 \) is mean of observed data, \( \mu_1 \) is mean of unknown values (predicted using stationarity assumption)

Optional: Include covariates to account for some spatial variability in \( \mu \) (universal kriging)
Hypoxia/Stratification
Calculation of stratification strength

Results Part 1: Main-channel hypoxia

1. Calculate Brunt Väisälä Frequency

\[ BVF = \frac{g \frac{\partial \rho}{\partial z}}{\rho} \]

2. Interpolate maximum Brunt Väisälä Frequency (BVF)

Average stratification strength = 0.014 s\(^{-2}\)
Hypoxic volume vs. TN

<table>
<thead>
<tr>
<th>Regression Model</th>
<th>HypoxicVol = β₀ + β₁(TN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early July (DO&lt;1 mg/L)</td>
<td>R²=0.12, p-value=0.04</td>
</tr>
<tr>
<td>Late July (DO&lt;1 mg/L)</td>
<td>R²=0.28, p-value&lt;0.001</td>
</tr>
</tbody>
</table>
Hypoxic volume vs. TN residuals

Regression Model

<table>
<thead>
<tr>
<th>Regression Model</th>
<th>HypoxicVol=$\beta_0+\beta_1$(TN)</th>
<th>Residuals=$\beta_0+\beta_1$(year)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early July (DO&lt;1 mg/L)</td>
<td>$R^2=0.12$, p-value=0.04</td>
<td>$R^2=0.25$, p-value=0.001</td>
</tr>
<tr>
<td>Late July (DO&lt;1 mg/L)</td>
<td>$R^2=0.28$, p-value&lt;0.001</td>
<td>$R^2=0.07$, p-value=0.1</td>
</tr>
</tbody>
</table>
Hypoxia Duration

- Number of days that bottom 5 m of water are hypoxic
- In mid-Bay, the hypoxia duration is significantly correlated with Jan-May TN loads

From: Murphy et al. 2011.
Regression for hypoxic volume (DO<1 mg/L)

\[
\text{HypoxicVol}_{\text{EarlyJuly}} = \beta_0 + 1.2 \left( \text{Stratification}_{\text{June}} \right) + 1.8 \left( \text{Stratification}_{\text{EarlyJuly}} \right) + 0.5 \left( \text{VolBelowPyc}_{\text{EarlyJuly}} \right) + \epsilon
\]

\[ R^2 = 0.77, \ p < 0.001 \]

*June stratification is only variable that explains the temporal increase*

From 1985-2009

From: Murphy *et al.* 2011.
Water distance
Water distance

- Distance between locations through water
- Computed between 11,534 grid and sample locations

Water distances computed by Eric Perlman (was at JHU CS)
1. Identify depth of pycnocline

\[ BVF = \frac{g}{\rho} \frac{\partial \rho}{\partial z} \]

2. Interpolate throughout bay

3. Use pycnocline depth indicator (P) as a covariate to interpolate DO at location s

\[ DO(s) = \beta_0 + \beta_0 P(s) + \varepsilon(s) \]

4. “Stack” interpolated DO layers to generate 3-D estimate

5.5 m

6.5 m

7.5 m

Max BVF
Interpolation Results

July 2007 Salinity in James River

(a) IDW: Segments
(b) Euclidean-Distance-Based Kriging: Entire Bay
(c) Water-Distance-Based Kriging: Entire Bay

Legend
Salinity estimates (ppt)
- 0.0 - 2.7
- 2.8 - 5.4
- 5.5 - 8.1
- 8.2 - 10.8
- 10.9 - 13.5
- 13.6 - 16.1
- 16.2 - 18.8
- 18.9 - 21.5
- 21.6 - 24.2
- 24.3 - 26.9

Observations (same scale)

Interpolation Standard Errors (SE)

(d) IDW: Segments
(e) Euclidean-Distance-Based Kriging: Entire Bay
(f) Water-Distance-Based Kriging: Entire Bay

Legend
Standard error (ppt)
- 0.0 - 0.2
- 0.3 - 0.4
- 0.5 - 0.6
- 0.7 - 0.8
- 0.9 - 1.0
- 1.1 - 1.2
- 1.3 - 1.4
- 1.5 - 1.6
- 1.7 - 1.8
- 1.9 - 2.0
- 2.1 - 2.2
- 2.3 - 2.4
- 2.5 - 2.6
- 2.7 - 2.8
- 2.9 - 8.1

Observation locations

km
0 10 20

No Uncertainty Information
## Cross Validation Summary

RMSE for July 2007 (and percent of time method has lowest RMSE for all 2007/2008 sets)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>IDW in segments</th>
<th>Kriging with Euclidean distance</th>
<th>Kriging with water distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water temp</td>
<td>0.79 (12%)</td>
<td>0.76 (8%)</td>
<td>0.74 (80%)</td>
</tr>
<tr>
<td>Salinity</td>
<td>2.3 (0%)</td>
<td>1.5 (0%)</td>
<td>1.1 (100%)</td>
</tr>
<tr>
<td>DO</td>
<td>1.5 (0%)</td>
<td><strong>1.2</strong> (8%)</td>
<td>1.2 (92%)</td>
</tr>
<tr>
<td>NH₄</td>
<td>0.023 (16%)</td>
<td><strong>0.022</strong> (20%)</td>
<td>0.023 (64%)</td>
</tr>
<tr>
<td>NO₂ + NO₃</td>
<td>0.18 (12%)</td>
<td>0.25 (16%)</td>
<td>0.17 (72%)</td>
</tr>
<tr>
<td>PO₄</td>
<td>0.0079 (4%)</td>
<td><strong>0.0073</strong> (12%)</td>
<td>0.0075 (84%)</td>
</tr>
<tr>
<td>Chl-a</td>
<td><strong>15</strong> (8%)</td>
<td>15 (32%)</td>
<td>15 (60%)</td>
</tr>
</tbody>
</table>
Validation Test

- Identify continuous monitoring stations from Maryland Department of Natural Resources and Virginia Estuarine and Coastal Observing System
- Average continuous observations 1 hour before and after the time of closest CBP sample
- Compare validation observation to interpolation output

Surface Salinity July 2007: Difference between water distance kriging and other methods and validation locations

Validation: http://www3.vims.edu/vecos/
http://mddnr.chesapeakebay.net/newmontech/contmon/index.cfm
Validation Results

**Root Mean Squared Errors of Estimates**

<table>
<thead>
<tr>
<th></th>
<th>IDW in segments</th>
<th>Kriging with Euclidean distance</th>
<th>Kriging with water distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water temp</td>
<td>1.0</td>
<td>0.88</td>
<td><strong>0.84</strong>*</td>
</tr>
<tr>
<td>Salinity</td>
<td>1.4</td>
<td>1.2</td>
<td><strong>0.92</strong></td>
</tr>
<tr>
<td>DO</td>
<td>2.0</td>
<td>1.9</td>
<td><strong>1.8</strong></td>
</tr>
<tr>
<td>Chl-a</td>
<td>29</td>
<td>24</td>
<td><strong>21</strong></td>
</tr>
</tbody>
</table>

*Uses a covariate, depth

**Validation sets:** Maryland Department of Natural Resources and Virginia Estuarine and Coastal Observing System continuous monitoring data
Hypoxic volumes computed using water-distance-based method

Results Part 2: Water distance