Geographically Weighted Regression (GWR): A Better Spatial Technique to Model Sediments Quality of Lake Okeechobee

FL-ASPRS Annual Meeting

Yaoyang Yan, PhD, GISP
Senior Geographer

August 23, 2013

Outline

I. Background
II. Objectives
III. Methodologies
IV. Data exploration
V. GWR models for total phosphorus (TP)
VI. Conclusions

I. Background

- Lake Okeechobee is the largest lake in the southeastern US with an area of 1730 km^2 and an average depth of 2.7 m;
- Lake Okeechobee has excessive phosphorus loads:
  - TP accumulated in mud sediments (44% of lake bottom)
  - Increased TP concentration in water column
  - Increased algal blooms
- It is critical to track the nutrient distribution and its changes over-time

II. Objectives

- Develop optimal spatial models for TP in the lake sediments;
- Examine the spatial changes of TP (1988-2006);
- Calculate the amount of nutrients (TP), its spatial-temporal changes, and identify the hotspots.

Outline

I. Background
II. Objectives
III. Methodologies
IV. Data exploration
V. GWR models for total phosphorus (TP)
VI. Conclusions

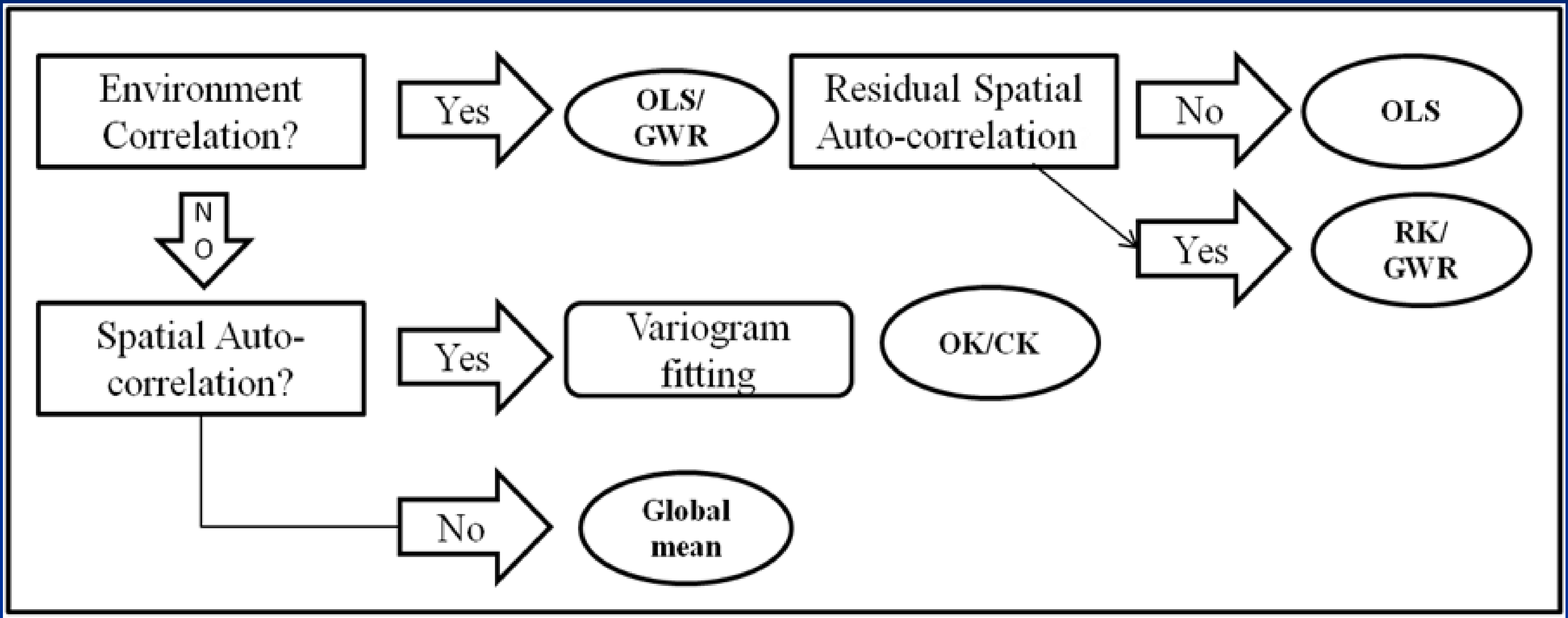
III. Methodologies

- Changes of spatial modeling techniques
- Model selection & model accuracy
- Geographically Weighted Regression (GWR)
  - Weighting functions & bandwidth
  - Optimal bandwidth & nearest neighbors
  - Model performance

Development of Spatial Modeling Techniques

- Great efforts have been made to develop generic and robust spatial interpolation techniques for many years;
- Kriging and its variants (from 1950s -)
- Environmental correlation (Regression) (1990s -)
  - Development of GIS and remote sensing technologies
- Hybrid models
  - Ancillary data is used to improve spatial prediction;
  - Good for non-stationary data;
  - Examples:
    - Regression-Kriging (RK) (late 1990s-)
    - Spatial regression (GWR) (2000s-)

Geostatistical Model Selection Procedure



(OLS: ordinary least square, OK: ordinary kriging; CK: Co-kriging; RK: regression-kriging and GWR: geographically weighted regression)

Model Accuracy and Comparison

- Use objective criteria to evaluate model errors, robustness and reliability
- Mean error:

$$ME = \frac{1}{l}\sum_{i=1}^{l}[\hat{z}(s_i) + Z^*(s_i)];\ E\{ME\} = 0$$

- Root mean square error (RMSE):

$$RMSE = \sqrt{\frac{1}{l}\sum_{i=1}^{l}[\hat{z}(s_i) - Z^*(s_i)]^2}\ ;\ E\{RMSE\} = \sigma(h = 0)$$

- Scatter plots of observed vs. predicted values

What is GWR?

- GWR analyzes spatially varying relationships between the target variable and ancillary variables;
- Developed by Fotheringham and Brunsdon (2000, 2002)
- Based on the "First Law of Geography": everything is related with everything else, but closer things are more related (Tobler 1970);
- βks vary from location to location, and the interpolation can be expressed as:

$$y_i = \beta_0 (u_i v_i) + \sum_{k=0}^{n} \beta_k (u_i v_i) x_{ik} (u_i v_i) + \varepsilon_i$$

$(u_i v_i)$ is the coordinates of the *i*th observation

$\beta_k (u_i v_i)$ is a realization of the function $\beta_k (u, v)$ at point *i*

GWR vs. OLS

- OLS:
  - Global regression model
  - One equation, calibrated using data from all features
  - Relationships are fixed
- GWR:
  - Local regression model
  - One equation for every feature, calibrated using data from nearby features
  - Relationships are allowed to vary across the study area

GWR Weighting Schemes and Bandwidths

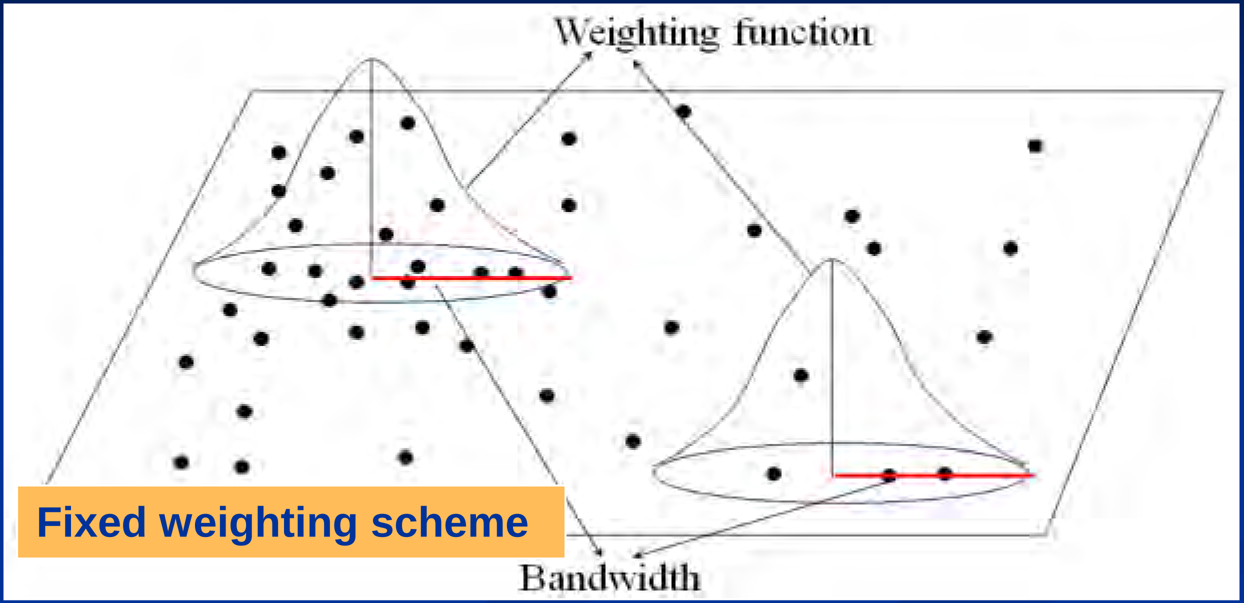
- Model parameters depend on weighting function and bandwidth
- Gaussian scheme:

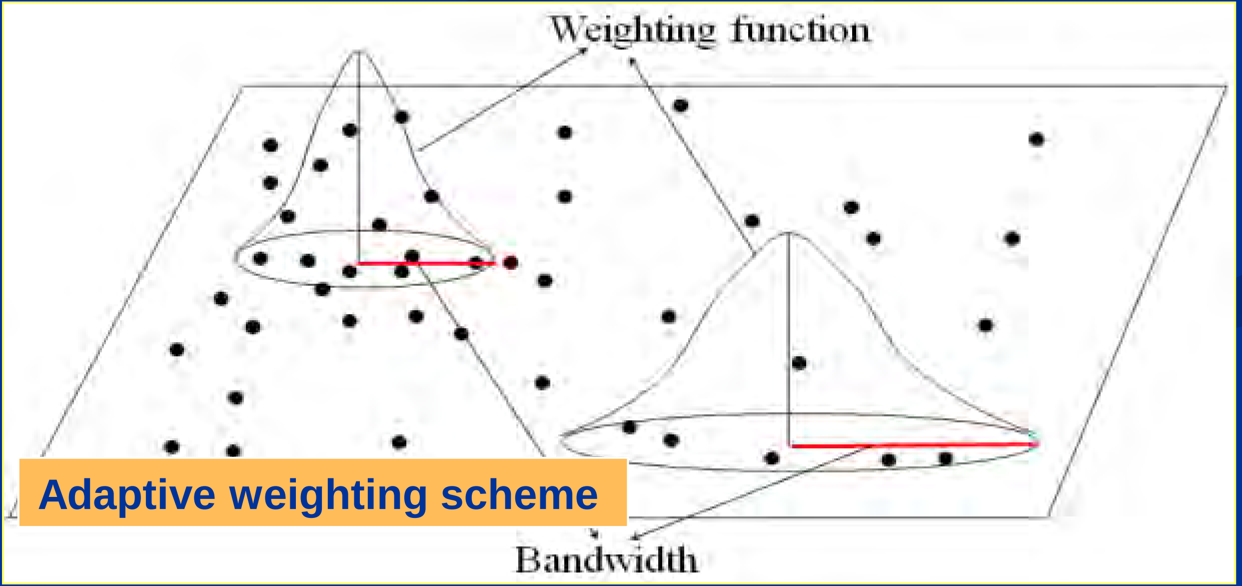
$$w_{ij} = \exp\left[-\frac{\frac{d_{ij}^2}{h^2}}{2}\right]$$

- Bi-square scheme:

$$w_{ij} = \left[1 - \left(\frac{d_{ij}^2}{h^2}\right)\right]^2 \quad \text{if } d_{ij} < h$$

$$= 0 \quad \text{otherwise}$$





Optimal Bandwidth/Neighbor Selection

- An optimal bandwidth (or nearest neighbors) can be selected by satisfying either
  - the least cross-validation (CV) score:

$$\mathrm{CV} = \sum_{i=0}^{n} (\mathrm{y_{i}} - \mathrm{y_{j \neq i}})^2$$

  - or least Akaike Information Criterion (AIC):

$$\mathrm{AIC_c} = 2n\log_e(\hat{\sigma}) + n\log_e(2\pi) + n\{\frac{n+\mathrm{tr}(\mathbf{S})}{n-2-tr(\boldsymbol{S})}\}$$

GWR Model Performance & Diagnosis

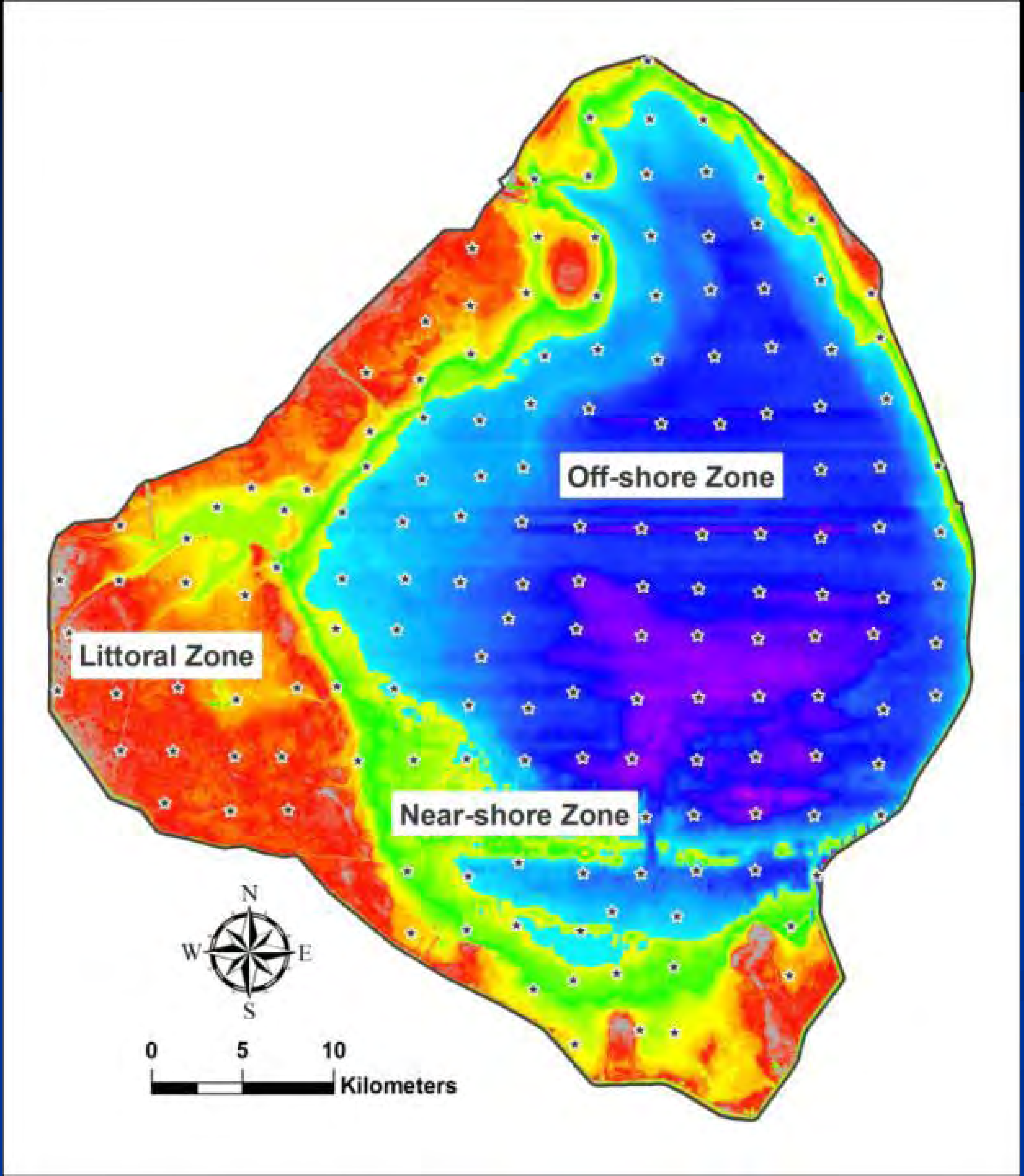
- The corrected AICc is the measure of goodness of fit for GWR models;
- AICc provides a measure of the information distance between the fitted model and the unknown ‘true’ model;
- AICc can be used to compare the global OLS model with a local GWR model;
- Other model performance indicators:
  - Sum of the squared residuals (SSR)
  - Sigma , and
  - R^2

Outline

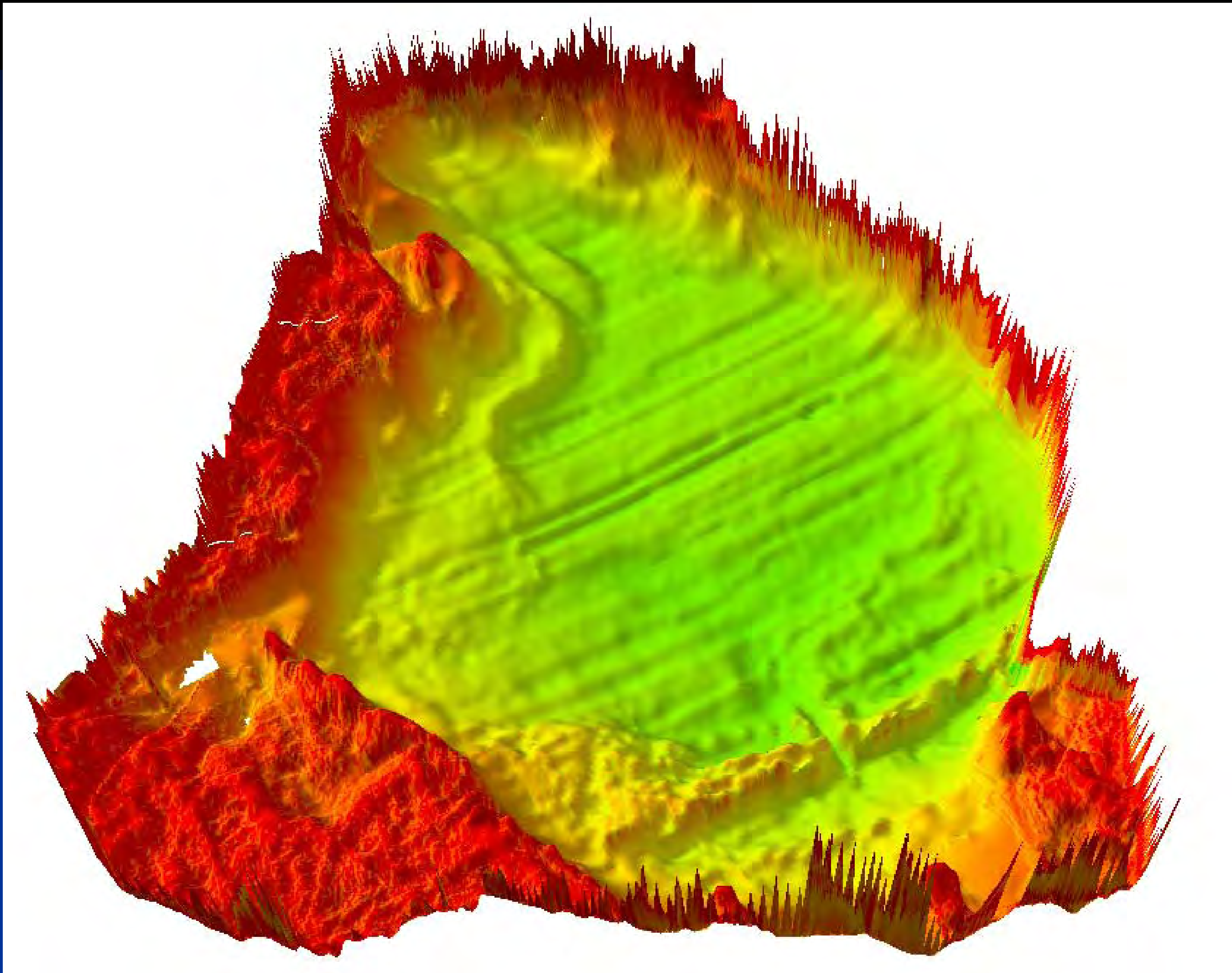
I. Background
II. Objectives
III. Methodologies
IV. Data exploration
V. GWR models for total phosphorus (TP)
VI. Conclusions

IV. Data Exploration

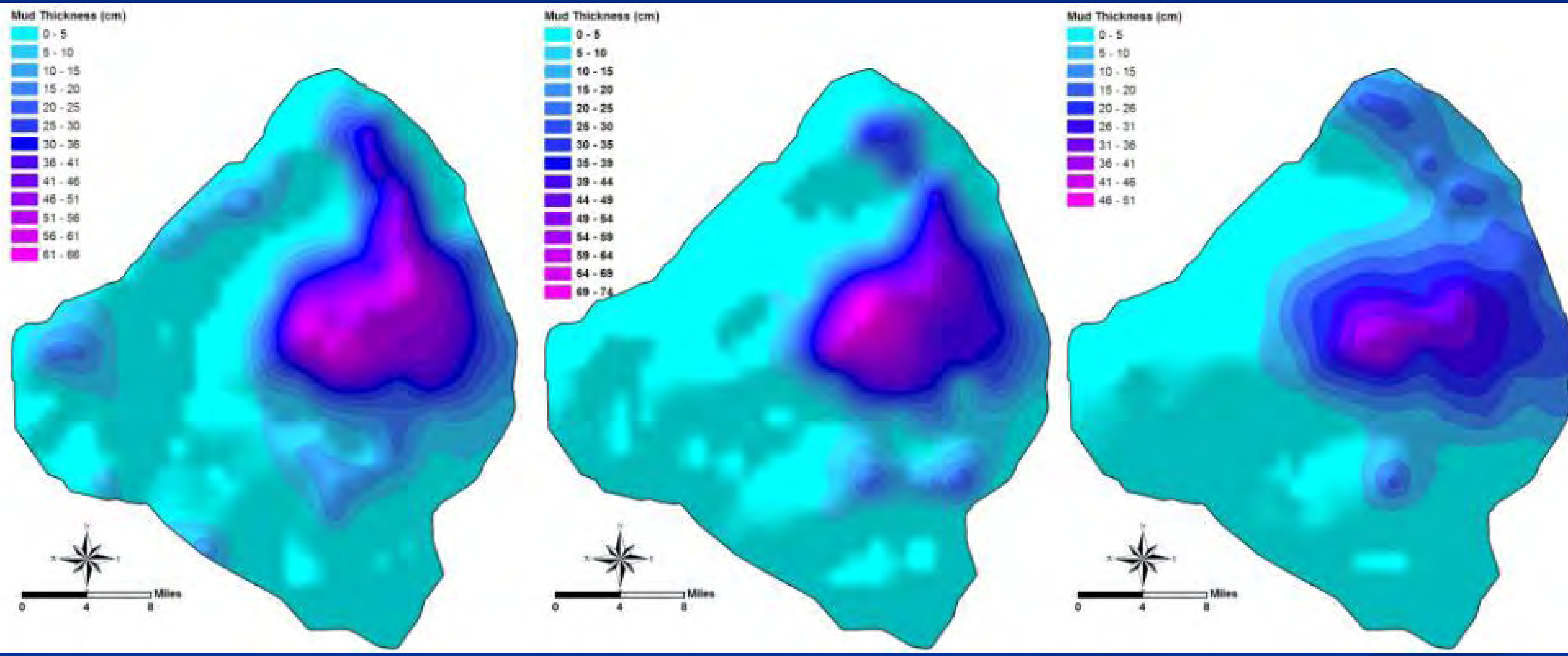
- Sediments Data
  - 170 sediment core samples taken in 1988, 1998 and 2006
  - Mud thickness and nutrients (TP, TN, TC, Fe, Ca, etc)
  - Mud density
- Digital Elevation Model
  - Lake bathymetry data (2008), and LIDAR (2007-2008)
  - Bathymetry & LiDAR data integration
- Data summarization and characteristics
  - Distribution
  - Trends
  - Correlations



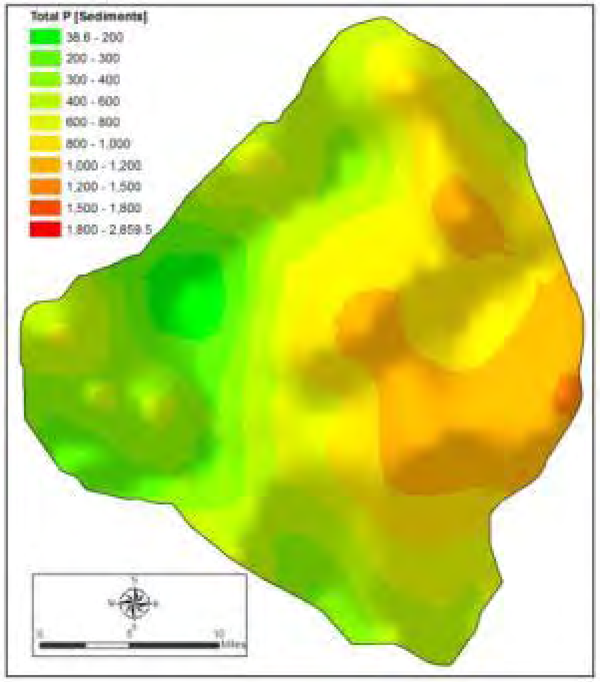
Bathymetry, eco-zones and sample sites Lake Okeechobee

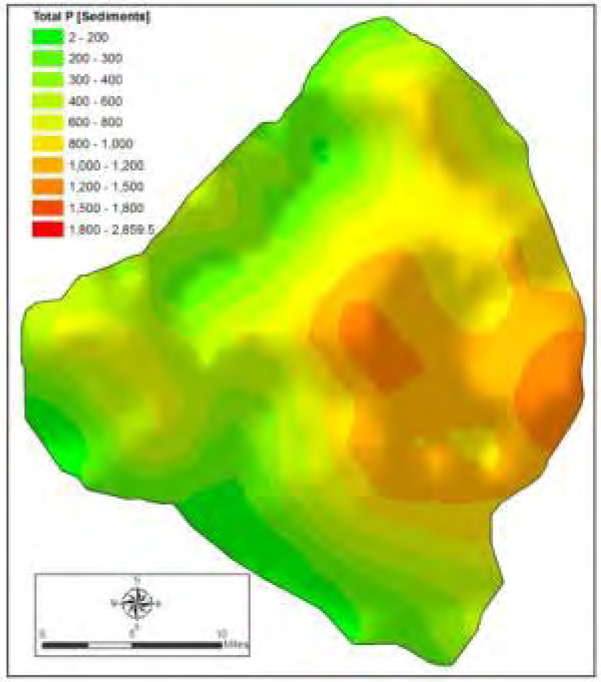


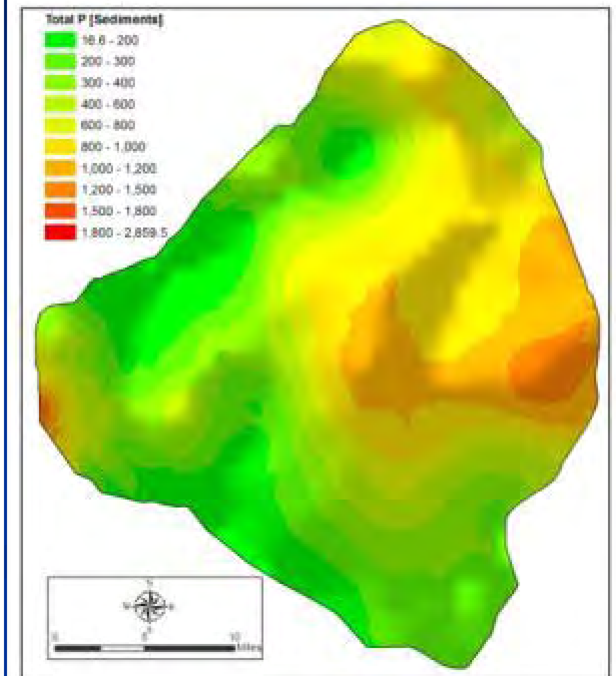
Mud Thickness Variation Using OK



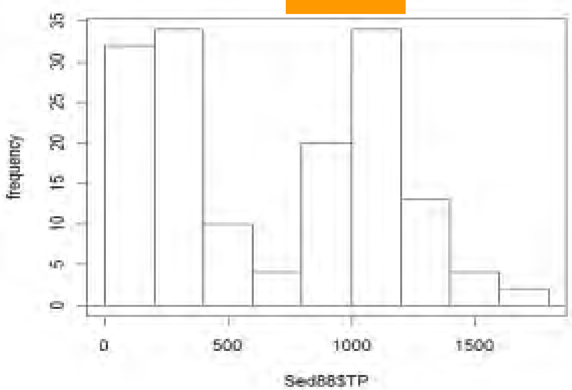
TP Histograms and Maps Using OK



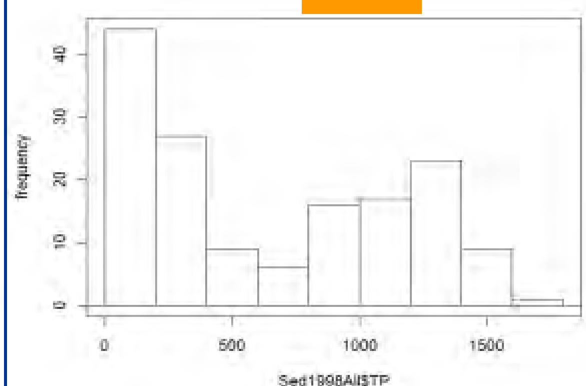




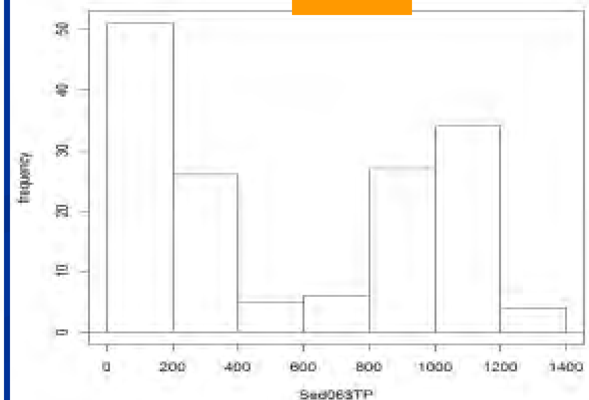
1988



1998

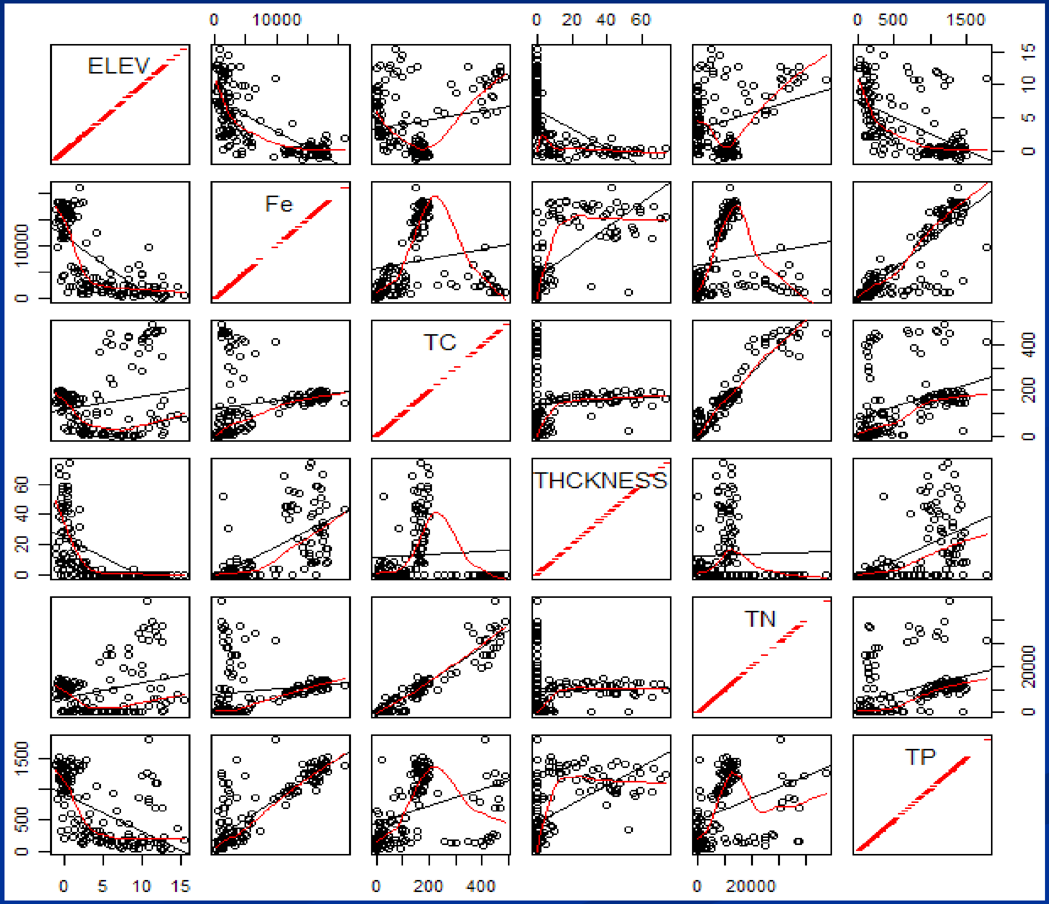


2006



Explore Relationship Using Scatter Plot Matrix

- TP
  - Strong positive correlation with Fe
  - Weak positive correlation with mud thickness and TC
  - Weak negative with site elevation
- All these similar correlations exit in all 3 data sets



1998 data

Outline

I. Background
II. Objectives
III. Methodologies
IV. Data exploration
V. GWR models for total phosphorus (TP)
VI. Conclusions

V. GWR Model Calibration & Validation

- First identify the best OLS regression models using the ArcGIS OLS Tool;
- Then perform GWR analysis using the ArcGIS GWR Tool;
- Compare the fit/performance of the OLS & GWR models using R^2 and AICc;
- Select the best GWR sub models for validation.

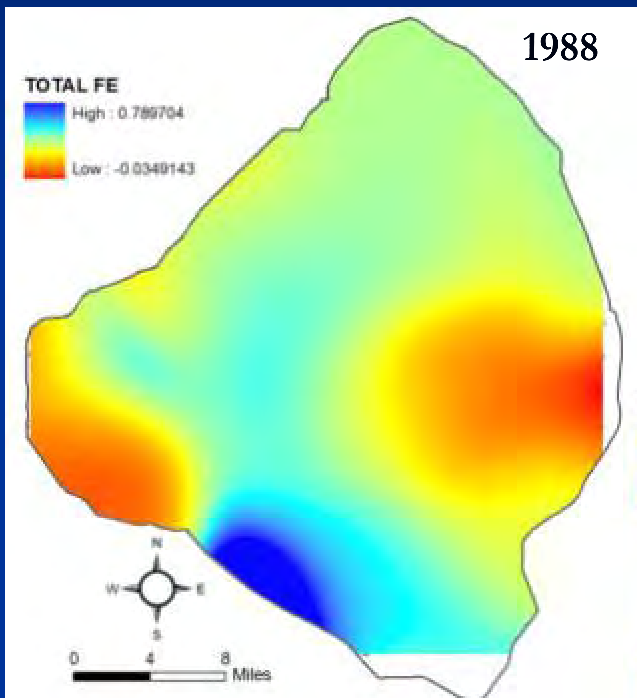
Statistically Significant Regression Models

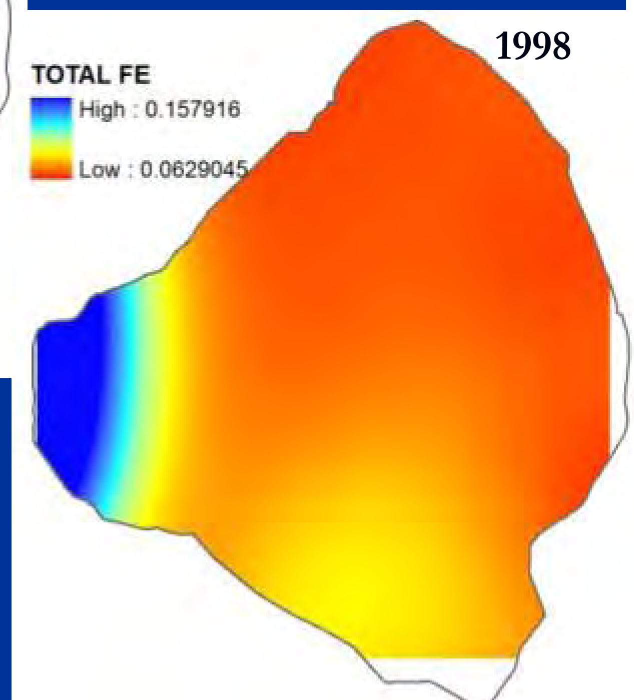
- Best regression models identified using OLS
  - TP vs. Total Fe
  - TP vs. Mud Thickness (Th) and Elevation (Elev)
- Perform GWR model analysis for these models

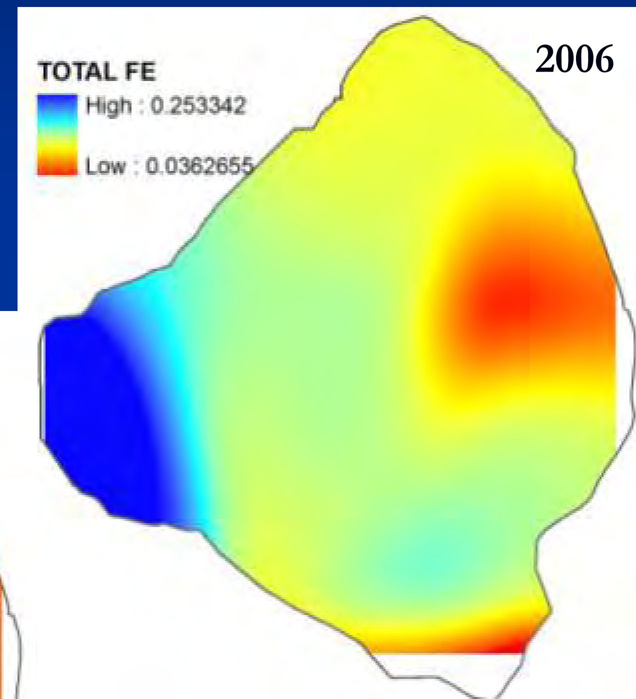
GWR (TP vs. Fe) Model

	GWR Models	Adaptive CV	Adaptive AIC	Fixed CV	Fixed AIC
2006	Neighbours	32	38	**25048.46**	28060.8468
	ResidualSquares	1474546	1559401.91	**1370525**	1479068.08
	Sigma	125.7934	127.1493	**125.8744**	127.5319
	AICc	1442.359	1442.0453	**1445.878**	1445.1828
	R^2	0.9322	0.9283	**0.937**	0.932
	R^2Adjusted	0.9178	0.916	**0.9176**	0.9155
1998	Neighbours	114	46	2583312.66	**38985.12**
	ResidualSquares	5721282	4772089.89	5851677.26	**4754339**
	Sigma	228.3922	219.3559	228.5793	**217.7004**
	AICc	1566.658	1563.9183	1566.1879	**1560.323**
	R^2	0.818	0.8482	0.8138	**0.8487**
	R^2Adjusted	0.8124	0.827	0.8121	**0.8296**
1988	Neighbours	13	38	**21433.62**	24336.93
	ResidualSquares	2256157	4669867.28	**3403058**	3809536.23
	Sigma	189.9278	220.3011	**205.7668**	210.1373
	AICc	1582.818	1567.6434	**1564.545**	1563.3411
	R^2	0.9067	0.8069	**0.8593**	0.8424
	R^2Adjusted	0.8314	0.7732	**0.8021**	0.7936

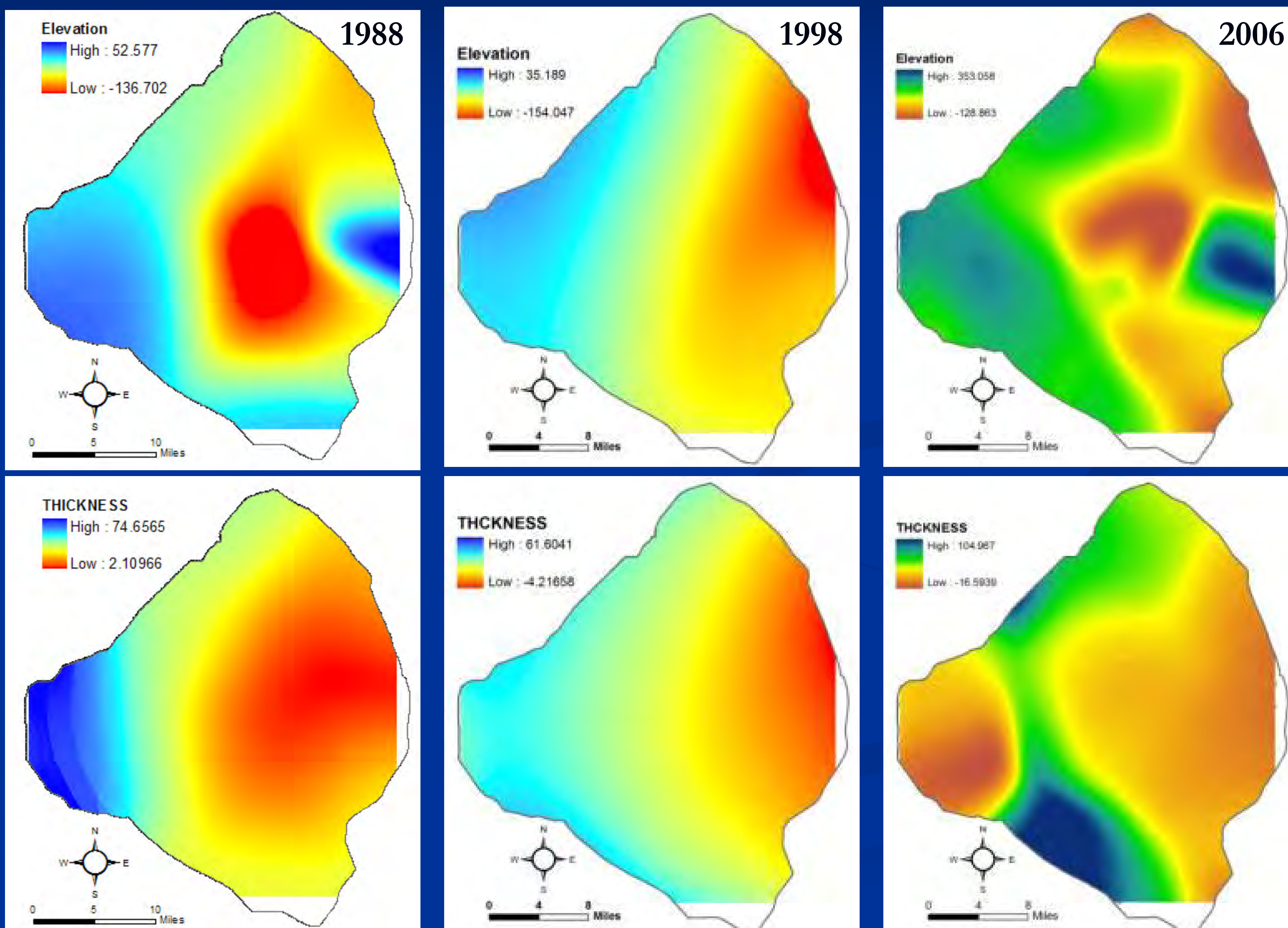
Coefficient Maps of GWR (TP vs. Fe) Model



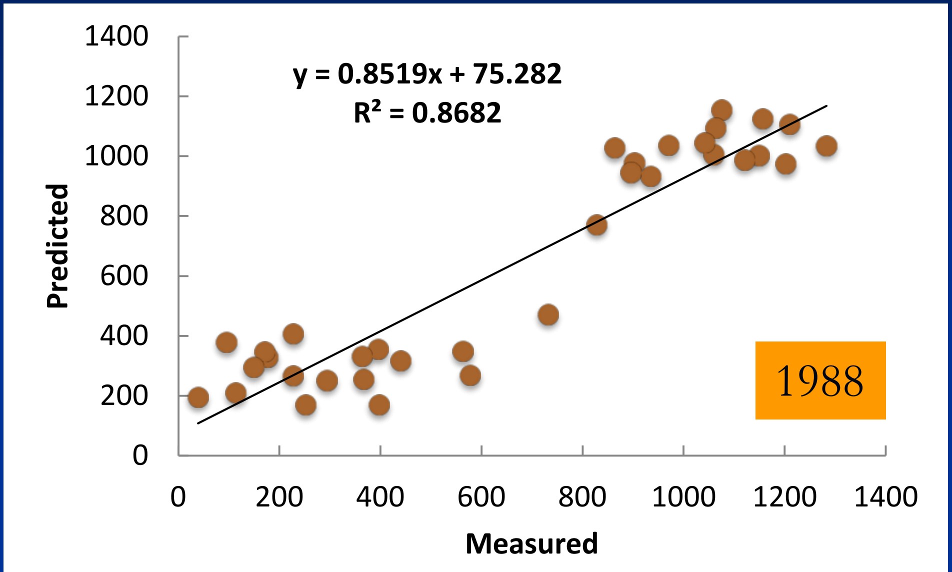




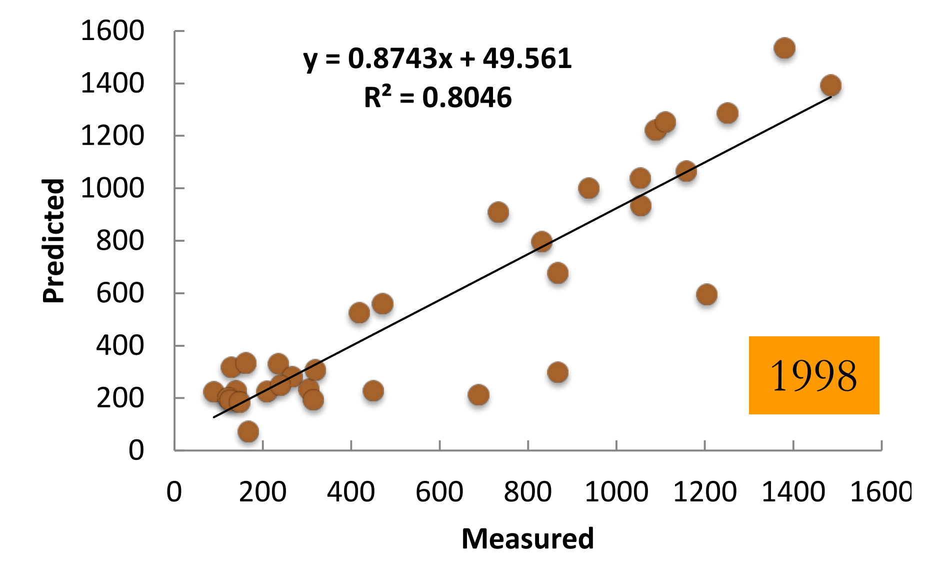
Coefficient Maps of GWR (TP vs. Th & Elev)

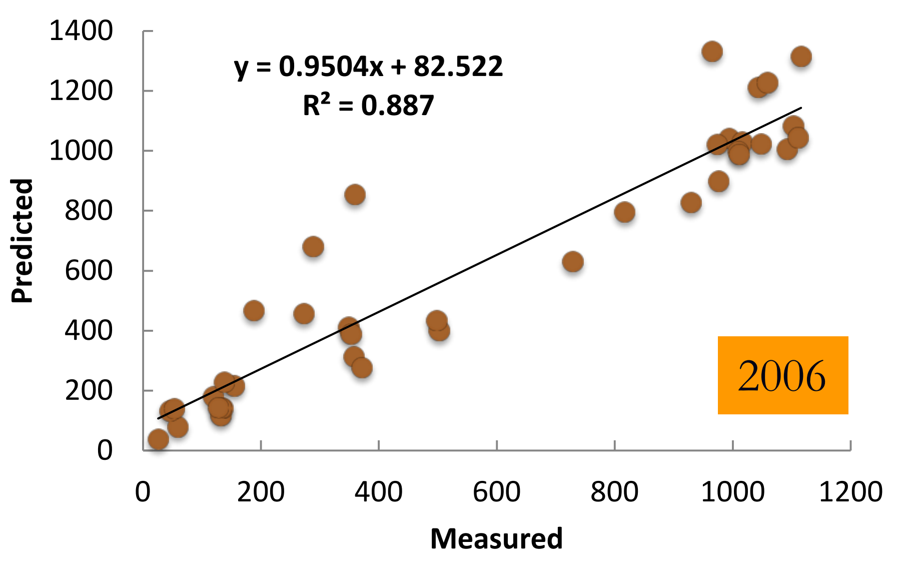


GWR (TP vs. Fe) Validation Results

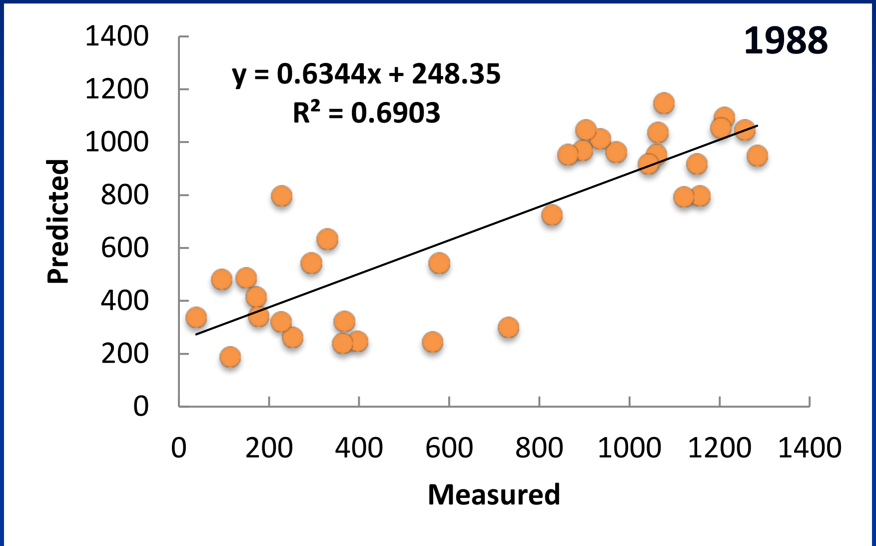


Scatter plots of observed and measured values at validation sites

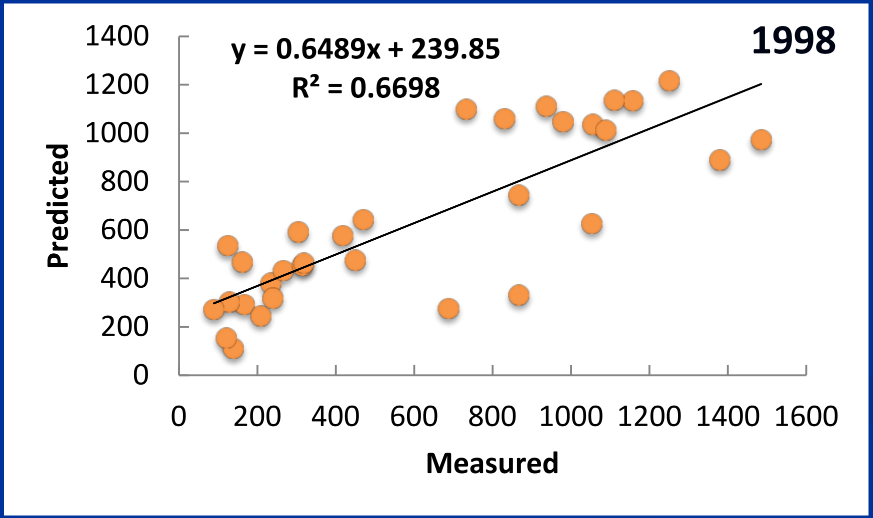


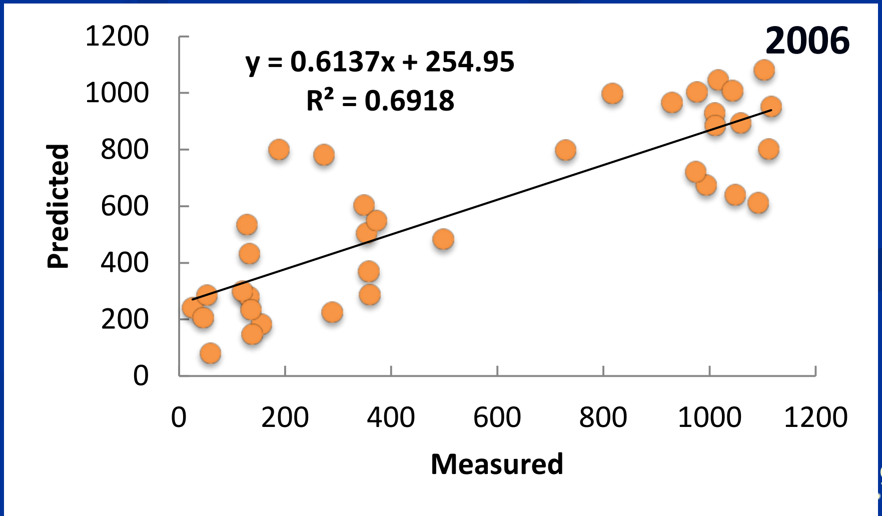


GWR (TP vs. Th & Elev) Validation Result



Scatter plots of observed and measured values at validation sites

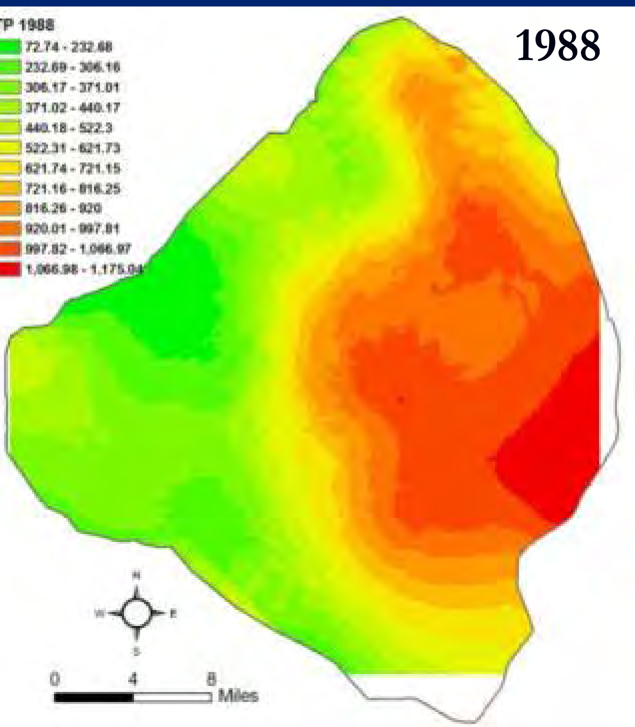


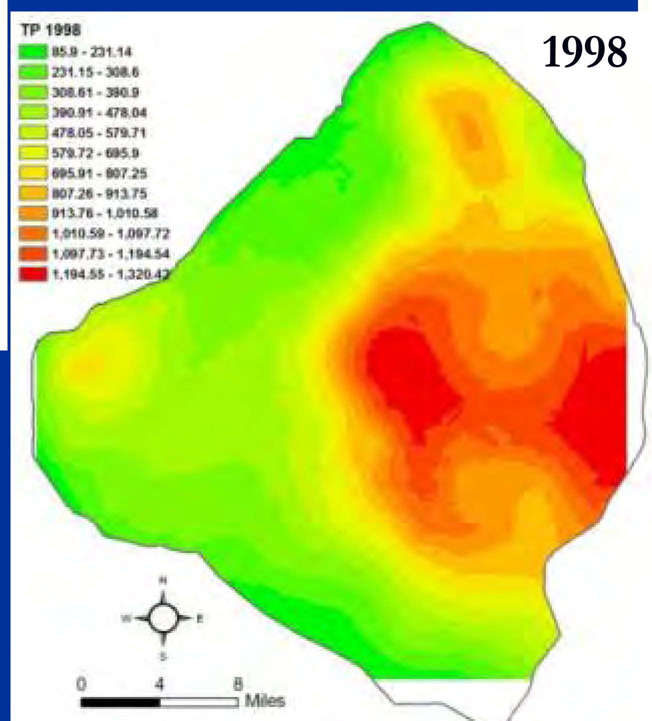


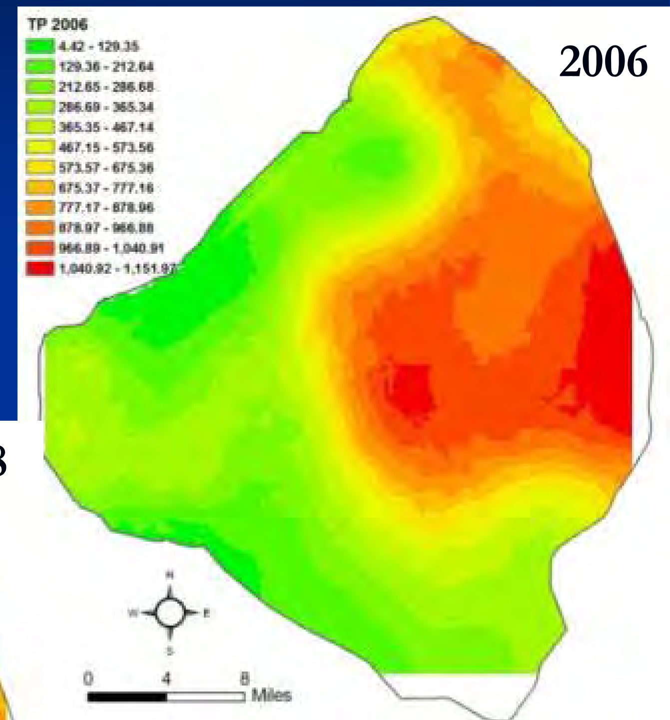
Model Accuracy and Comparison for TP

	Models	OK	OLS (Fe)	GWR (Fe)	OLS (Th, Elev)	GWR (Th, Elev)
2006	Mean Error	-27.4	31.19	-30.19	40.21	20.82
	RMSE	331	169.96	204.61	325.23	307.6
	NRMSE	26.09	13.4	16.13	25.64	24.25
1998	Mean Error	-89.5	136.93	132.81	106.64	100.58
	RMSE	415.1	374.75	359.83	442.98	389.38
	NRMSE	23.18	20.92	20.09	24.73	21.74
1988	Mean Error	-41.7	110.94	109.59	64.15	24.39
	RMSE	402.1	368.36	375.88	371.97	234.34
	NRMSE	0.24	0.22	0.23	0.22	0.14

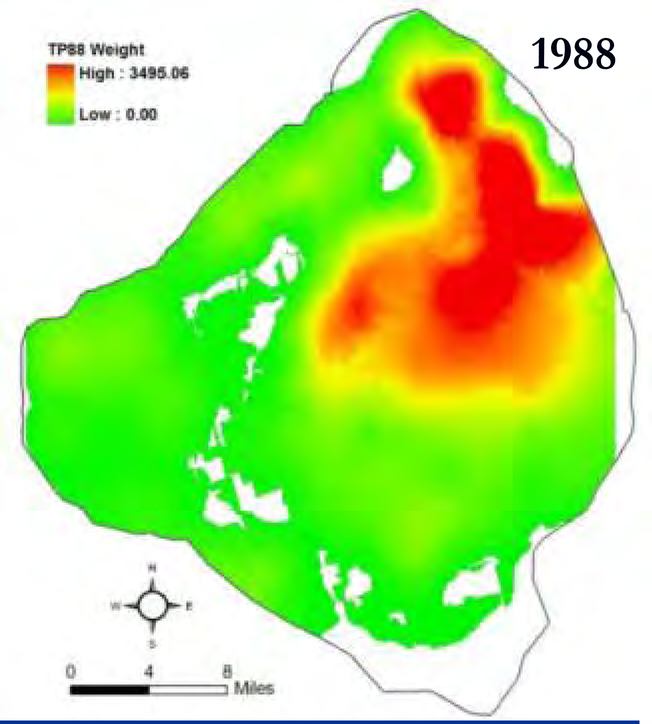
TP concentration Using GWR (TP vs. Fe)

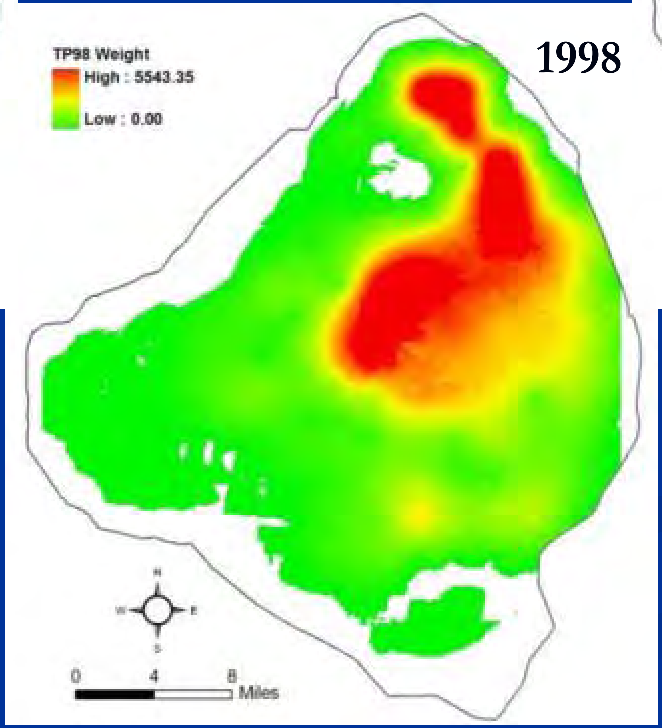


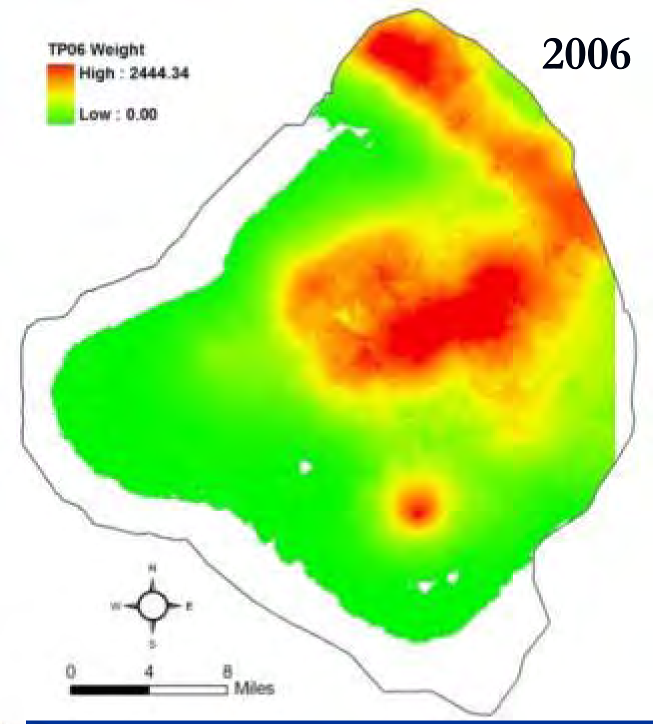




TP Weight Distribution Using GWR (TP vs. Fe)







TP Weights (mt) & Changes

Model	Ordinary Kriging			GWR (TP vs Fe)			GWR (TP vs. Th & Elev)			Average
Year	Weight	Change	%	Weight	Change	%	Weight	Change	%	Weight
1988	42,500			42,000			44,300			42,933
1998	58,900	16,400	39%	60,400	18,400	44%	61,100	16,800	38%	60,133
2006	41,400	-17,500	-30%	41,000	-19,400	-32%	40,100	-21,000	-34%	40,833

Outline

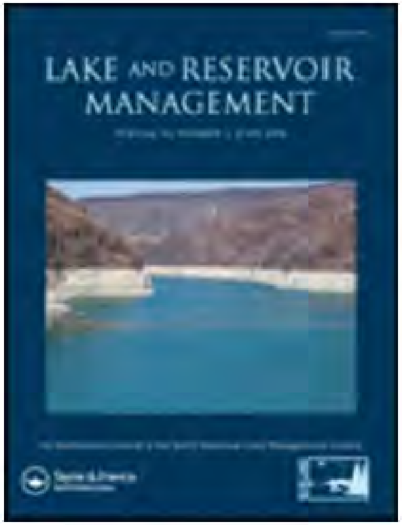
I. Background
II. Objectives
III. Methodologies
IV. Data exploration
V. GWR models for total phosphorus (TP)
VI. VI. Conclusions

Conclusions

- GWR (TP vs. Fe) and GWR (TP vs. Thick & Elevation) models were most accurate based on RMSE;
- GWR models use both spatial auto-correlation and correlation between TP and independent variables, which improves the model performance;
- Ordinary Kriging (OK) models use only spatial autocorrelation data, which were weak in Lake Okeechobee and produced higher RMSE;

Conclusions (cont.)

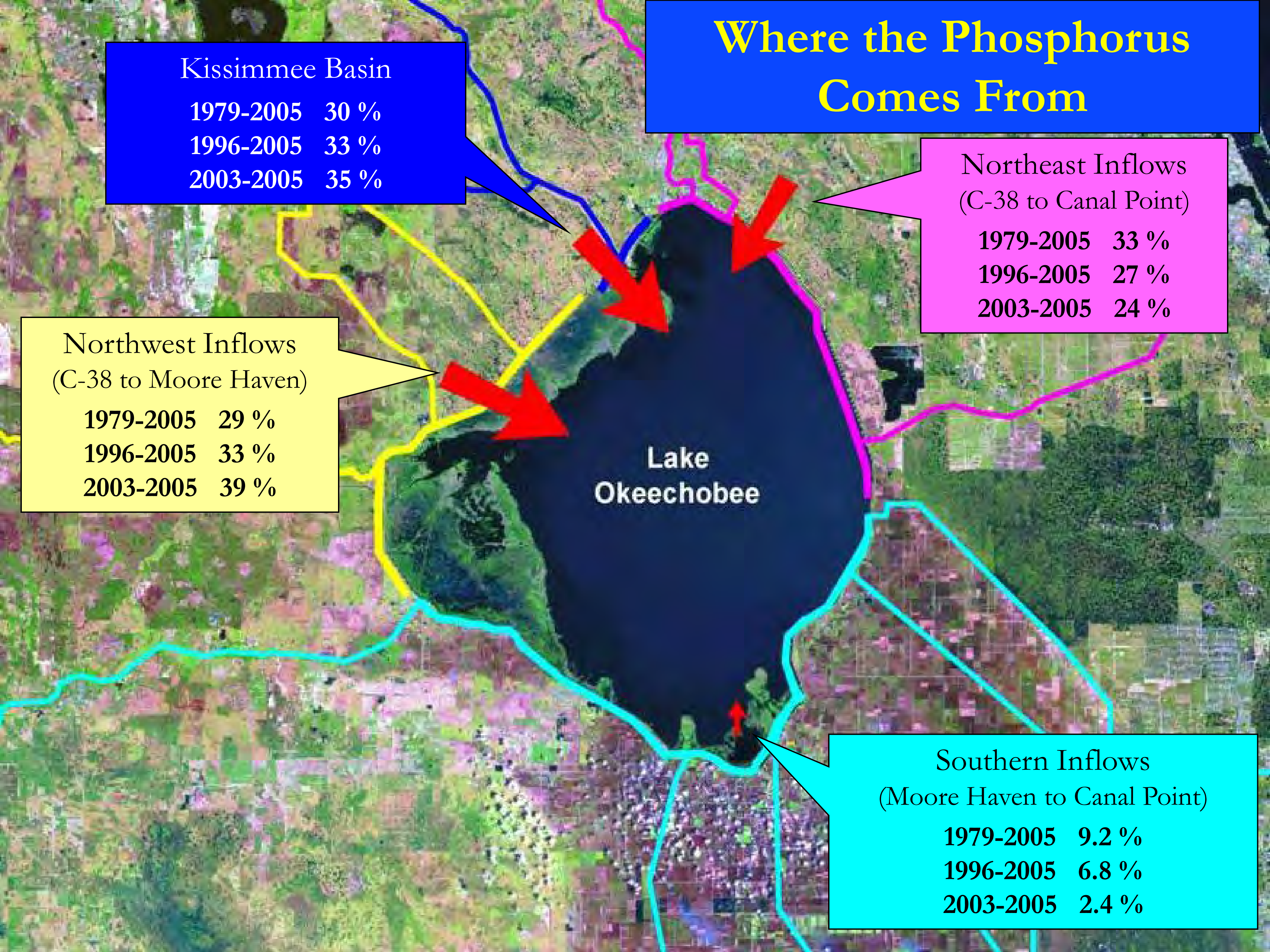
- The two best GWR models were used to calculate TP concentrations and TP mass. The TP mass:
  - increased about 38% - 44% from 1988 to 1998 and
  - decreased about 30% - 34% from 1998 to 2006;
- The TP decline from 1998 to 2006 is likely a result of hurricanes that stirred up the sediments which were then discharged from the lake in 2004 and 2005 hurricane seasons.



Geographically Weighted Regression: A Better Spatial Technique to Model Sediment Quality in Lake Okeechobee

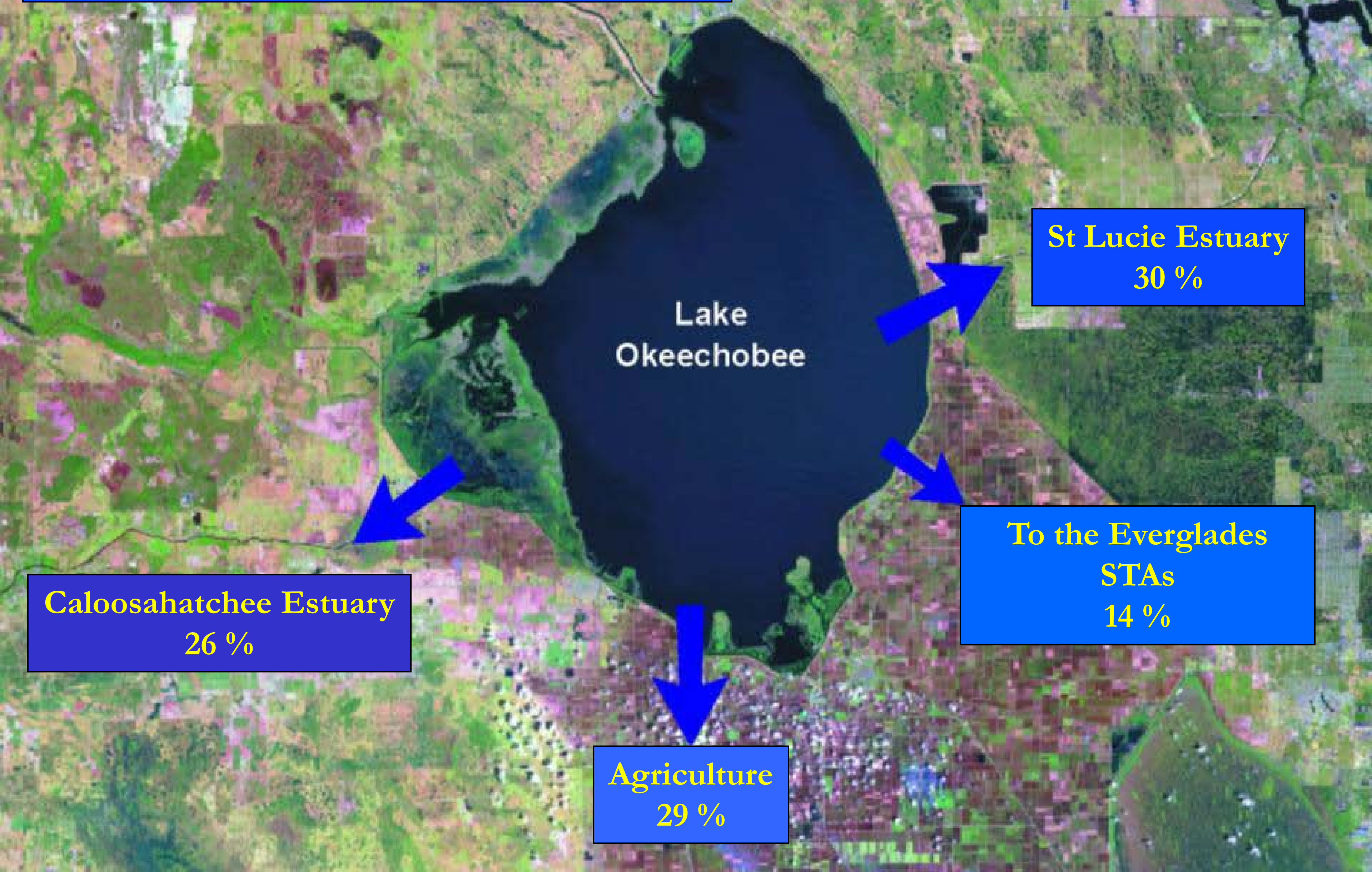
QUESTIONS?

Contact:
Yao Yan, PhD.
yyan@sfwmd.gov
561-6822077


Where the Phosphorus Comes From
Kissimmee Basin
1979-2005 30 %
1996-2005 33 %
2003-2005 35 %
Northeast Inflows
(C-38 to Canal Point)
1979-2005 33 %
1996-2005 27 %
2003-2005 24 %
Northwest Inflows
(C-38 to Moore Haven)
1979-2005 29 %
1996-2005 33 %
2003-2005 39 %
Lake Okeechobee
Southern Inflows
(Moore Haven to Canal Point)
1979-2005 9.2 %
1996-2005 6.8 %
2003-2005 2.4 %

Where the Phosphorus Goes

Based on 1996-2005 data



Major References

- Fotheringham, A. S., C. Brunsdon and M.E. Charlton ME, **Geographically Weighted Regression**: the Analysis of Spatially Varying Relationships, John Wiley & Sons Ltd, West Sussex, 2002.
- Fisher, M. M., K. R. Reddy, and R. T. James. 2001. Long-term changes in the sediment chemistry of a large shallow subtropical lake. Lake and Reservoir Management 17: 217-232.
- Yan, Y. and R. Thomas James, Spatial-Temporal Modeling of Sediments From 1988 to 2006 Lake Okeechobee, Florida. South Florida GIS Expo2008
- Yan, Y., 2011. Development of Methods for Spatial Modeling of Sediments Quality in Lake Okeechobee, Florida. PhD Dissertation, FIU