

Accounting for and Predicting the Influence of Spatial Autocorrelation in Water Quality Modeling

Lorrayne Miralha¹ and Daehyun Kim²

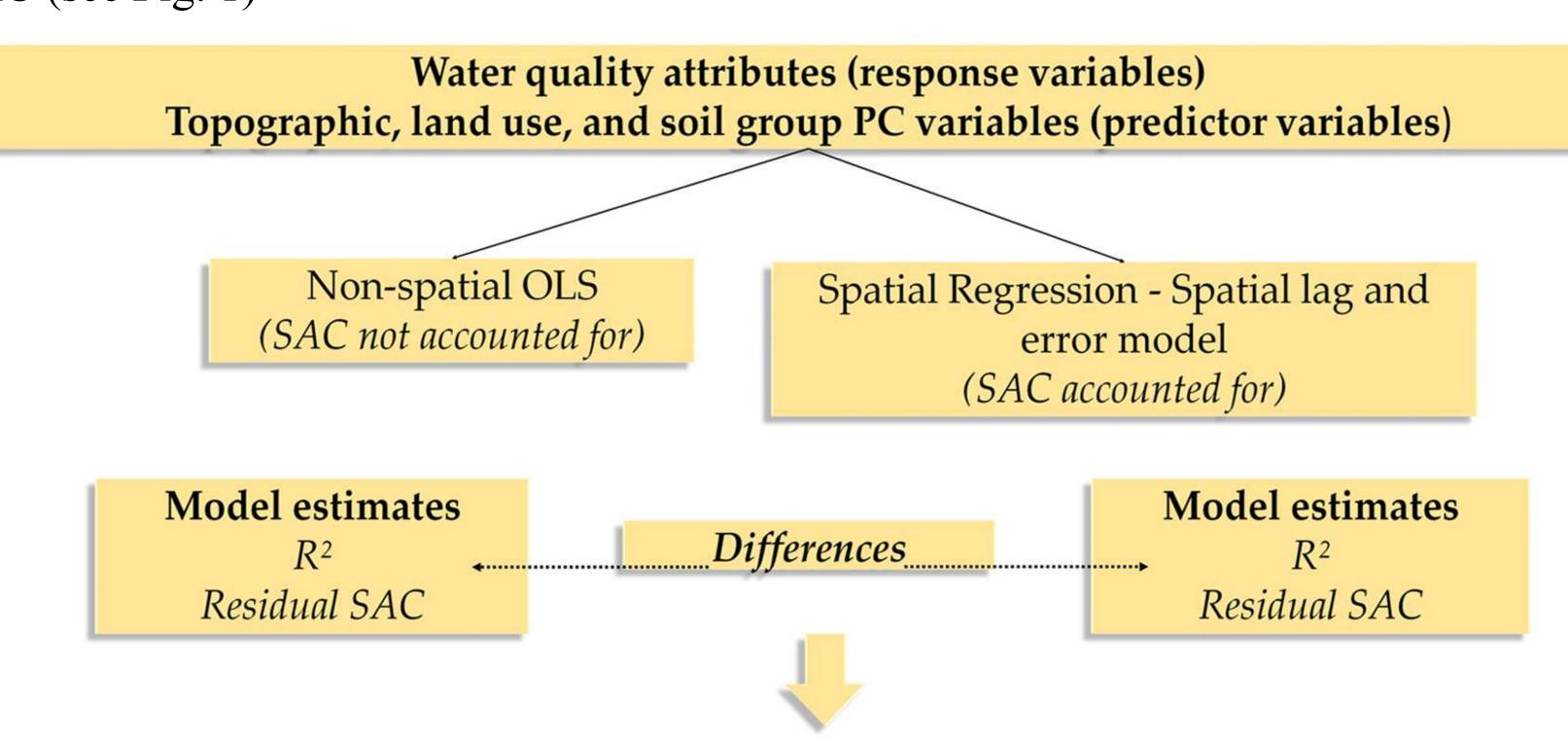
¹ Department of Geography, University of Kentucky, Lexington, KY 40508, USA ² Department of Geography, Seoul National University, Seoul 08826, South Korea



1. Introduction

Several studies in the hydrology field have reported differences in outcomes between models in which spatial autocorrelation (SAC) is accounted and those in which SAC is not [1-3]. However, the capacity to predict the magnitude of such differences is still ambiguous.

We hypothesize that SAC possessed by a response variable (i.e., a water quality variable) influences the outcomes of spatial modeling. We test if water quality variables with a higher amount of SAC would exhibit greater improvement in model outcomes than those with a lower amount of SAC (see Fig. 1)



Are these differences in model outcomes related to the degree of inherent SAC (Moran's I) possessed by the response variable?

Figure 1. Conceptualization of the main ideas of the study (PC, principal components; OLS, ordinary least squares; SAC, spatial autocorrelation).

2. Materials and Methods

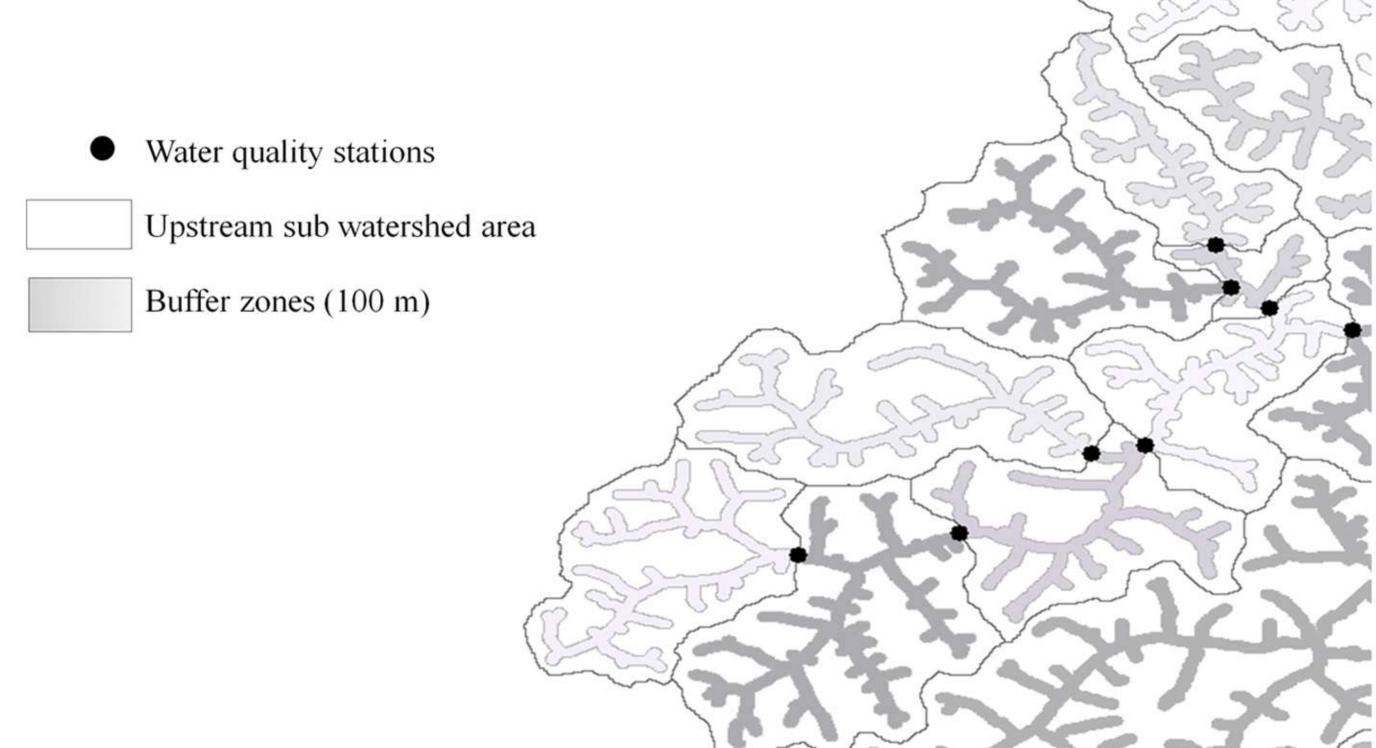
Study areas and model variables

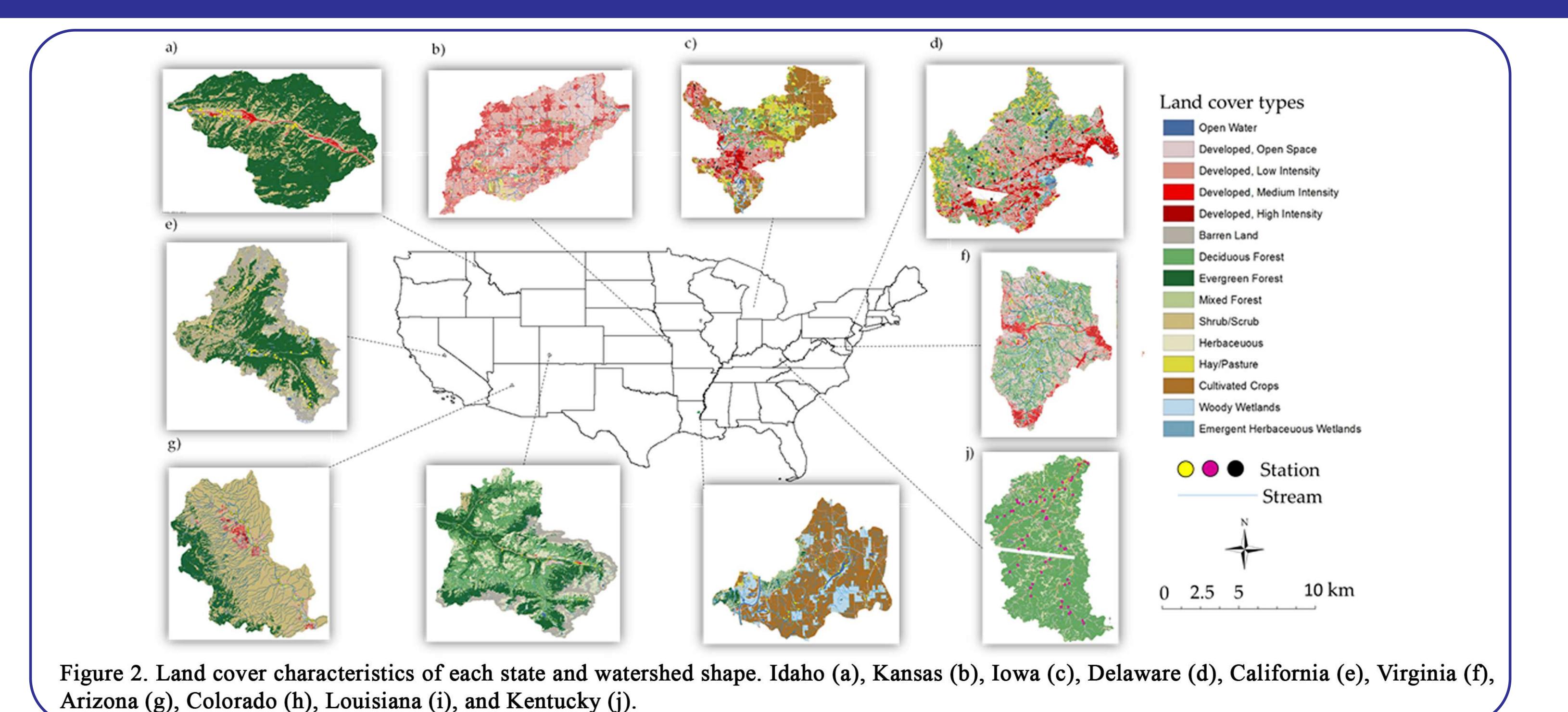
- Ten divergent watersheds located each in one state of the Unites States. The watershed areas varied from 150 km² to 764 km² (Fig.2).
- 29 to 54 sampling sites (water stations) were selected from each watershed. From these, we collected water quality data (e.g., pH, temperature, specifc conductance, calcium, etc). The number and type of variables vary across watershed.
- Dependent variables were the water quality variables and the predictors of the models were principal component groups (PC) summarized in the table below:

Variable	PC group	Derived Variable	Original Data
Dependent	-	-	Physical water quality variables
Independent	Topographic	Mean elevation Elevation standard deviation Mean slope Slope standard deviation	National Elevation dataset (10 m) Digital Elevation Model
Independent	Land use	% Agriculture % Forest % Urban % Wetland	National Land Cover dataset (30 m) Pasture, cultivated crops. Deciduous forest, evergreen forest, mixed forest. Low-, medium-, high-intensity urbanized areas, open space. Woody wetland, emergent herbaceous wetland
Independent	Soil	% A, B, C, D, A/D, B/D, C/D	Soil Survey Geographic (SSURGO) database

Upstream area and buffer zones

We delineated upstream sub-watershed boundaries and generated buffer zones of 100 meters total, as illustrated:





Moran's I

We quantified the inherent degree of SAC for each water quality parameter using Moran's I function (Equation 1).

$$I = \frac{n}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \frac{\sum_{i=1}^{n} \sum_{j}^{n} W_{ij} (X_i - \bar{X}) (X_j - \bar{X})}{\sum_{i=1}^{n} \sum_{j}^{n} W_{ij}},$$
(1)

Regression Models

- Non-spatial OLS (Ordinary least squares) Equation 2
- Spatial lag Model Equation 3
- Spatial Error Model Equation 4

$$Yi = \beta o + \beta 1X1 + \dots + \beta iXi + \varepsilon i,$$

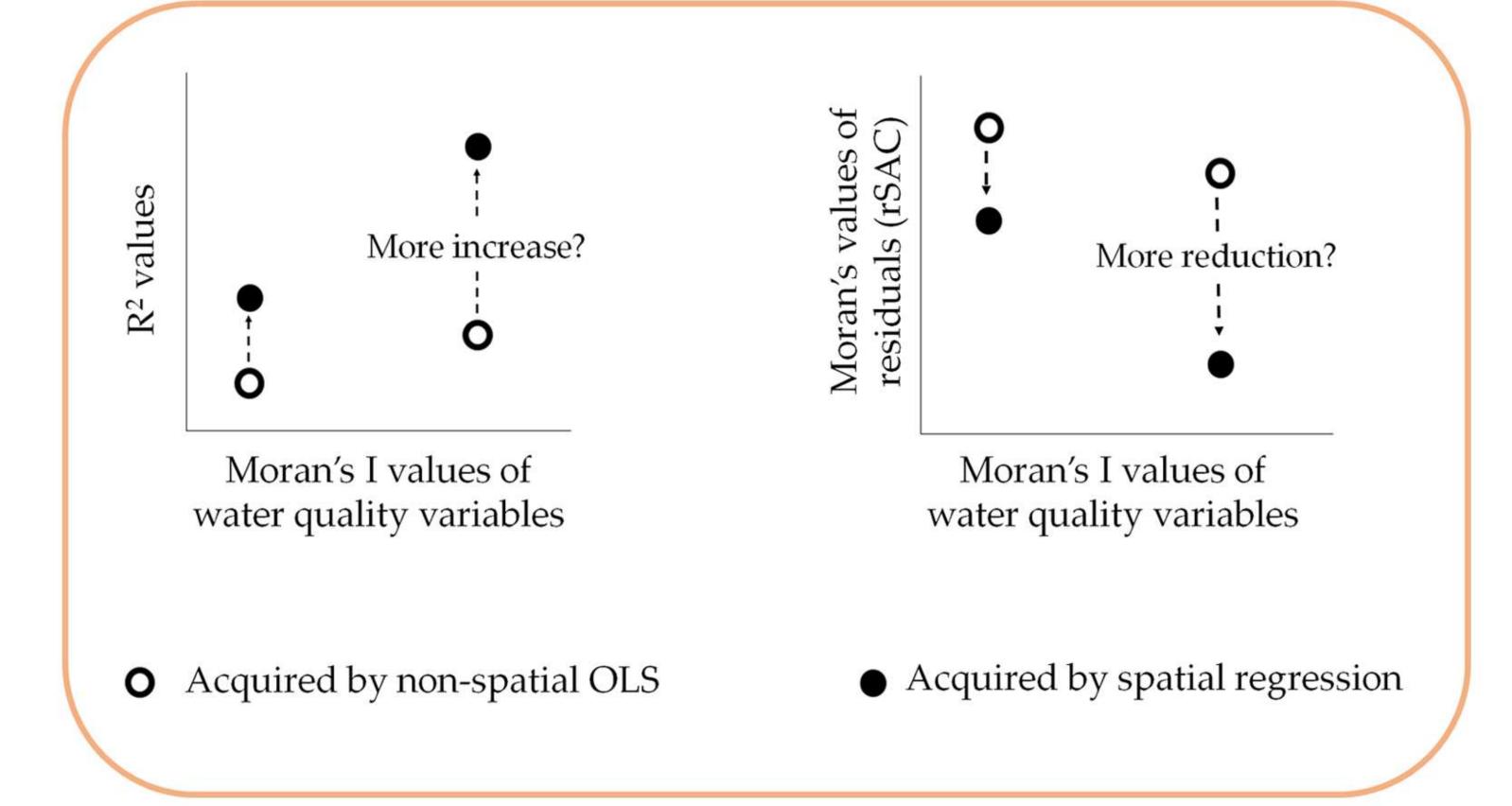
$$Yi = Xi\beta i + \rho WYj + \varepsilon,$$

$$Yi = Xi\beta i + \varepsilon = \lambda W \varepsilon + \varepsilon,$$

$$(2)$$

$$(3)$$

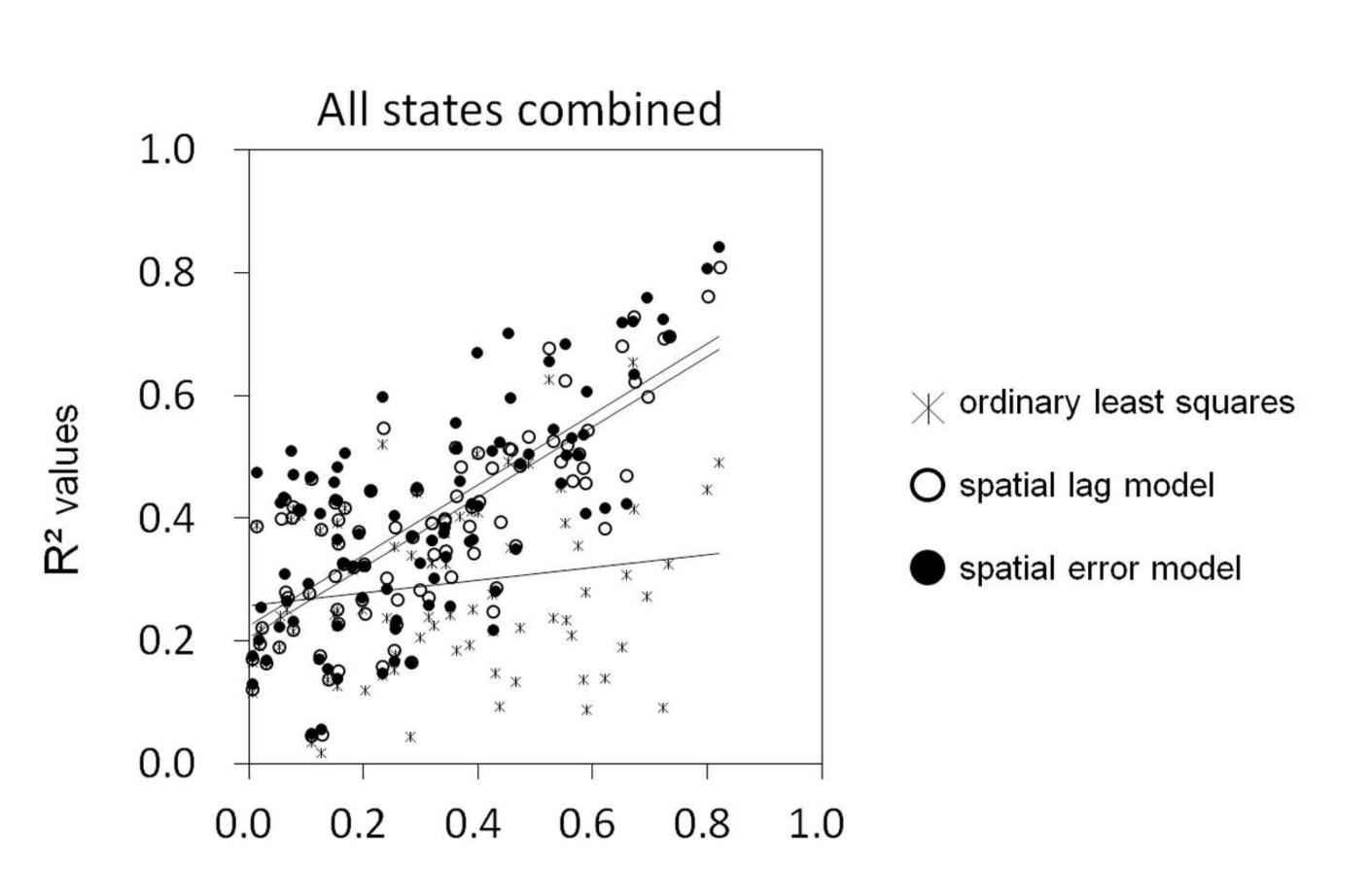
Model Comparison



3. Results

Changes in R²

- Strongly autocorrelated water quality parameters over space exhibited greater increases in R² values after spatial regression compared to weakly autocorrelated variables (Fig. 3).

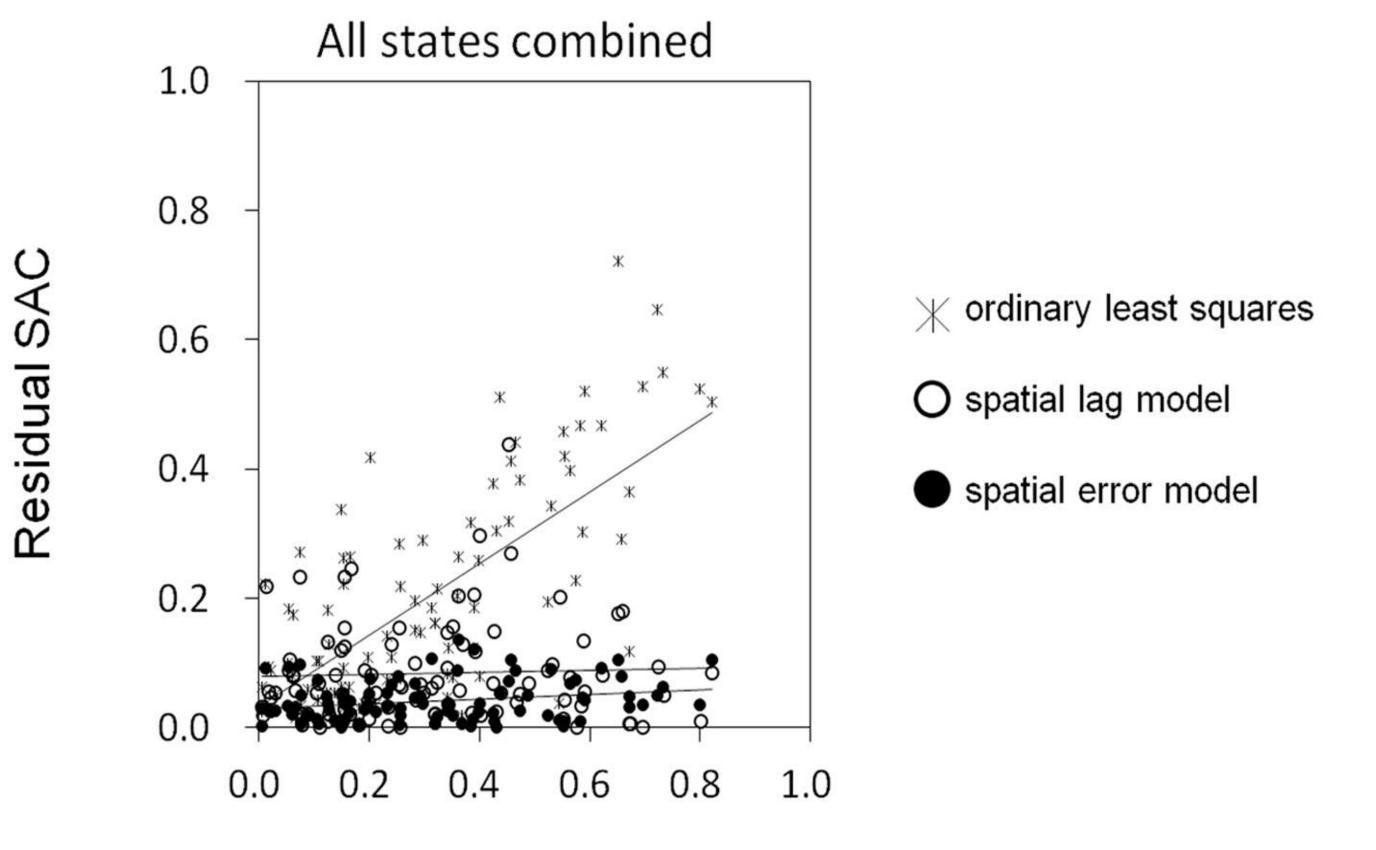


Moran's / values of water quality variables

Figure 3. The hypothesis predicting increases in R² as a function of the degree of SAC is supported.

Changes in Residual SAC (rSAC)

-The larger the Moran's I values possessed by water quality variables, the greater the reduction in residual SAC after running spatial models.(Fig. 4).



Moran's / values of water quality variables

Figure 4. The hypothesis predicting greater decreases in rSAC, proportional to the degree of SAC in water quality variables, is supported.

Summary of findings

- We found that the magnitude of model improvement (i.e., increases in R² and decreases in rSAC), after both spatial lag and error modeling, is significantly and linearly a function of the SAC inherently possessed by water quality variables (i.e., response variables) (Fig. 5).

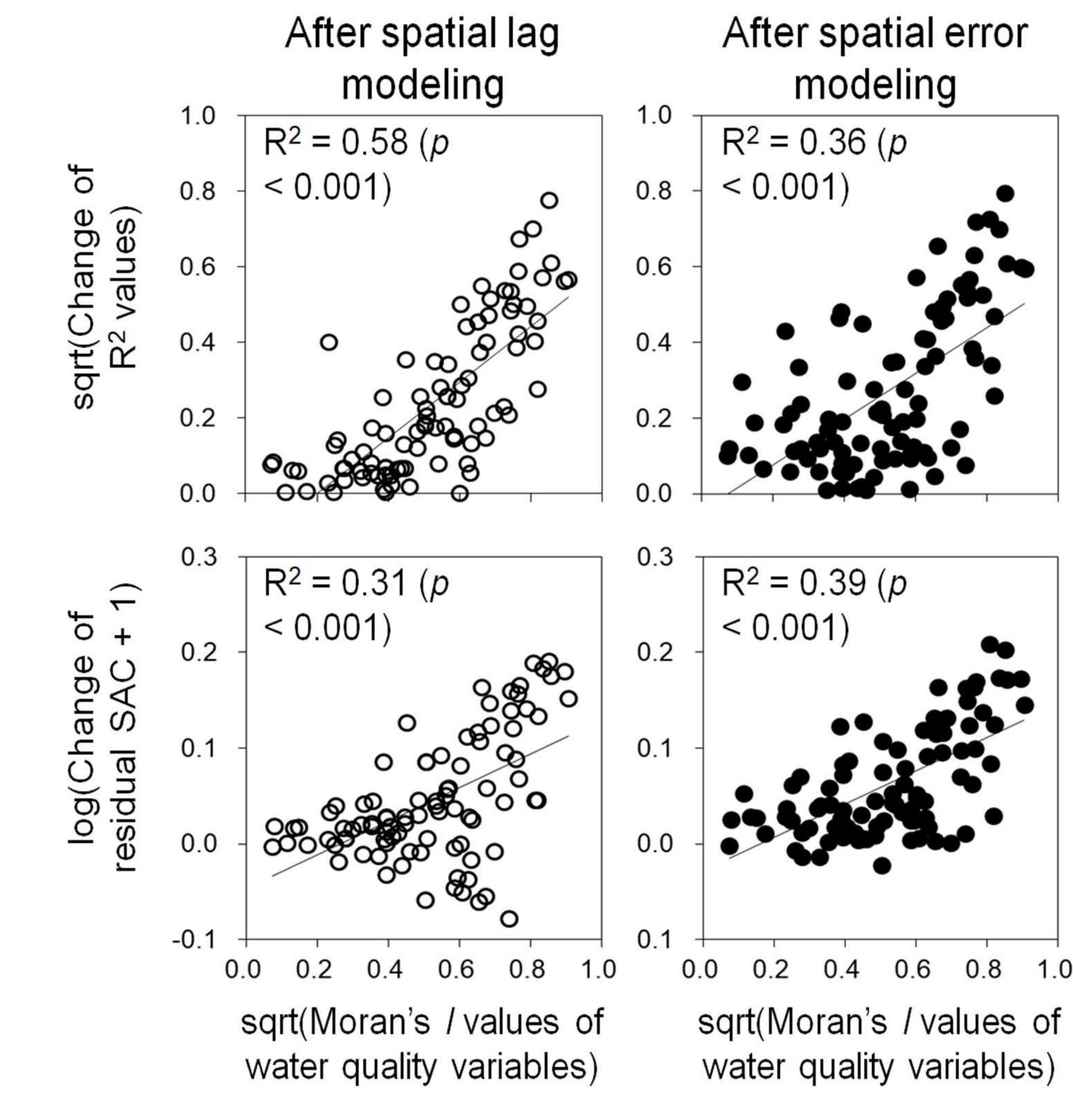


Figure 5. Linear regression models demonstrating that the magnitude of improvement of model performance after spatial lag and error modeling is significantly and linearly explained by the SAC inherently possessed by water quality variables

4. Discussions

- The need of—and, potentially, the benefit from—accounting for SAC in water quality modeling increases as the degree of SAC increases.
- This study reveals a consistent and linear relationship between the SAC of water quality variables and changes in the model parameters (R² and rSAC).

This study can serve as a useful screening technique where modelers run Moran's *I* to predict the spatial pattern of independent variables using a spatially explicit method.

- Analyzing SAC in water quality modeling can lead to a better understanding of the spatial organization extent of water quality variables, and be beneficial to the process of watershed management.

References

1. Pratt, B.; Chang, H. Effects of land cover, topography, and built structure on seasonal water quality at 269 multiple spatial scales. Journal of hazardous materials 2012, 209, 48-58. DOI: 10.1016/j.jhazmat.2011.12.068

2. Kim, D.; Hirmas, D. R.; McEwan, R. W.; Mueller, T. G.; Park, S. J.; Šamonil, P.; Thompson, J. A.; Wendroth, 292 O. Predicting the Influence of Multi-Scale Spatial Autocorrelation on Soil–Landform Modeling. Soil Science 293 Society of America Journal 2016, 80 (2), 409. DOI:10.2136/sssaj2015.10.0370

3. Chang, H. Spatial analysis of water quality trends in the Han River basin, South Korea. Water research 2008, 288 42 (13), 3285-3304. DOI: 10.1016/j.watres.2008.04.006



