Modeling the effects of land use on water quality parameters using OLS and GWR multivariate regression methods in Fars Province watersheds

M. Hosseinkhah[1], M. Erfanian[2*], A. Alijanpour[3]

[1] MSc Graduated Student of Watershed Management Engineering, Urmia University, Urmia, Iran

[2] Assistant Professor, Faculty of Natural Resources, Urmia University, Urmia, Iran
*Corresponding Author, **Email: Erfanian.ma@gmail.com**

[3] Associate Professor, Faculty of Natural Resources, Urmia University, Urmia, Iran

Full Paper is available at:

https://jes.ut.ac.ir/article_58738_3ec701a822a32079d8c04951f5d517a5.pdf

Extended Abstract

Introduction

In recent years, several studies around the world have shown that land use has a strong impact on water quality, and significant correlations exist between water quality parameters and land use types. Generally land use types have adverse impacts on water quality, so positive relationships exist between percentages of these land use types and concentrations of water pollutants. In other words, negative relationships are usually found between percentages of un-developed lands (e.g. forest and rangelands) and concentrations of water pollutants (good water quality). In contrast, higher percentages of these developed land use types are related to higher concentrations of water pollutants (worse water quality). The relationships between different land use types and different water quality parameters vary greatly, and a land use type might be positively associated with one water pollutant but negatively related to another. The relationships between water quality and land use obtained in the previous studies are usually analyzed by conventional statistical methods such as Ordinary Least Squares regression (OLS). In recent years, a simple but powerful statistical method named Geographically Weighted Regression (GWR) has been developed to explore the continuously varying relationships over space. Similar to OLS, GWR builds a model to analyze how one dependent variable changes in response to the change in one or more independent variables, and it can calculate a set of local regression results including local parameter estimates. This study applied the GWR technique to explore the spatially varying relationships between land use and water quality in 42 watersheds located in Fars Province, Iran. The main objective of this study is to explore how the relationships between land use and water quality indicators change over selected watersheds.

Materials and Methods

The present study was carried out in Fars Province, Iran, and water quality data from 1971 to 2011 were obtained from the water company authority in Fars. The water quality parameters consist of Ca, Cl, EC, CO_3, HCO_3, K, Mg, Na, P^H, TH, SAR, SO_4 and TDS. Seven land use types, including bare soils, rangelands, fallow, agricultural, Orchards, residential and forest lands, were selected in this study ~~in this study~~. The land use map was created and validated by utilizing the Landsat TM images (12 frames in July 2010) based on the widely-used remote sensing technique known as the maximum likelihood method.

The spatially varying relationships between land use and water quality indicators were analyzed using GWR. Water quality indicators were used as dependent variables, while land use indicators were independent variables. Because high correlations exist among the land use indicators, each GWR model used only one land use indicator to analyze its association with one water quality indicator. There were seven land use indicators and thirteen water quality indicators. Therefore, the relationships for 91 (13 times 7) pairs of water quality and land use indicators were analyzed by building 91 GWR models. GWR analyses were conducted using GWR4 software package. Afterwards, the local parameter estimates, the values of t-test on the local parameter estimates, and the local R^2 values produced by the GWR models were mapped to give a clear visualization of the spatial variations in the relationships between land use and water quality, and the abilities of the land use indicators to explain water quality. All mappings and GIS analyses were performed using the ArcGIS 9.3. The OLS models are like the following:

$$Y = \beta_0 + \sum_{i=1}^{p} \beta_i X + \varepsilon \tag{1}$$

Y represents the dependent variable, β_0 is the intercept, and $\beta_i X$ is the coefficient and the independent variable, ε represents the error term, and p is the number of independent variables. The GWR model differs in that it incorporates the coordinates of each location "i" with a metric coordinates "u" and it is defined as:

$$y_i(u) = \beta_{0i}(u) + \beta_{1i}(u)x_{1i} + \beta_{2i}(u)x_{2i} + ...\beta_{mi}(u)x_{mi} + \varepsilon_i \tag{2}$$

The model GWR is calibrated using an exponential distance decay function:

$$W_{ij} = \exp(\frac{-d_{ij}^2}{b^2}) \tag{3}$$

The weight of site "j" as it effects site "I", W is calculated using the distance (d) between sites "i" and "j" with selecting "b" as the bandwidth. The weight decreases rapidly when the kernel is smaller than the distance. For this study, an adaptive band was used because the density of sample sites varied across the study area. We used the Global Moran's I statistics for the residuals of both OLS and GWR models to test spatial dependence (Autocorrelation). Global models assume that relationships between water quality and explanatory variables are the same across space. This is particularly problematic given the variation in land cover and multiple sources of pollutants. To evaluate two model performances, we utilized the coefficient of determination (R^2) and the corrected Akaike's Information Criterion (AICc). The purpose of comparing GWR with OLS models was to identify whether GWR models have better model performance than the corresponding OLS models. The comparison was performed by comparing the model R^2 and the AICc values from both GWR and OLS models. Higher R^2 means that independent variable can explain more variance in dependent variable. A lower AICc indicates a closer approximation of the model to reality, lower AICc means better model performance. Usually, a decrease of AICc lower than three might be caused by sampling error, but not a result of a real difference in models

Results and Discussion

The global R^2 of GWR with comparison of R^2 of OLS for each pair of dependent variable (water quality indicator) and independent variable (land use indicator) indicate that a dramatic improvement in R^2 of GWR over OLS is observed for every pair of water quality and land use indicators. The R^2 values in GWR in all watersheds were larger than 0.83 and the AICc for all water quality parameters were much smaller than the OLS models. The AICc differences between models for all water quality parameters were larger than 3 and the criteria values for GWR were always less than OLS. The higher values of the global R^2 from GWR than the R^2 from OLS indicate the improvement in model performance of GWR over OLS. However, the statistical significances of the improvements need to be verified with AICc values. The

statistical test results for improvement in model comparisons of the AICc value indicates a closer approximation of the model to reality. Thus, in this study, a GWR model is considered to be significantly improved from its corresponding OLS model if the AICc value of the GWR is at least three lower than that of the OLS and the F-test is significant at p-value < 0.05 level.

The spatial maps of the GWR model parameters, reproduced for the study area in the ArcGIS 9.3 for EC and CL showed that rangeland, fallow lands, orchards and residential in the southeast, bare soil and agricultural lands in the north, and forest lands in the southwest of Fars Province, have the significant increasing impact on EC indicator values. Furthermore, bare soils, rangelands, fallow and forest regions in the west, agricultural areas in the southwest, orchards and residential areas in the south of this Province have decreasing impact on EC values. For the case of chloride (Cl), bare soils, rangelands, agricultural areas and forest in the north, orchards and fallow lands, in the southeast and residential zones in the east show a significant increase impact on this indicator. In addition, bare soils in the northwest, rangelands, fallow lands and forest in central zones, orchards and residential in the south, agricultural lands in the southwest of this province, have not significant increase on the chloride.

Conclusion

This study examined the relationships between seven land use types (%) and thirteen water quality indicators using both OLS and GWR models in 42 watersheds of Fars Province, Iran. Most GWR models show great improvements of model performance over their corresponding OLS models, which is proved by F-test and the comparisons of model R^2 and AICc from both GWR and OLS. Many GWR models also successfully reduce spatial autocorrelations examined by Moran's I statistics. The GWR models improved the reliabilities of the relationships between variables by reducing the spatial autocorrelations in residuals. The visualization of the GWR model local parameter estimates (Beta maps), and local R^2 maps in ArcGIS, highlight the great spatial variations in the impacts of different land use types on different water quality indicators and help identify their spatial patterns.

Key words: Water Quality, GWR, Spatial Autocorrelation