



Analysis of time-space varying relationship between land use and water quality in a tropical watershed

Moriken Camara¹ · Nor Rohaizah Jamil^{1,2} · Ahmad Fikri Abdullah^{2,3} · Rohasliney Hashim²

Received: 28 May 2020 / Accepted: 22 January 2021 / Published online: 19 February 2021
© Saudi Society for Geosciences 2021

Abstract

Over the years, studies have been conducted to examine not only the relationship between land use and water pollution using ordinary least square (OLS) regression but also their spatially varying relationships using geographical weighted regression (GWR). However, the relationships between land use and water quality indicators may vary not only spatially but also over time. Therefore, this study was conducted to analyse the spatiotemporal variations in land use-water quality relations over time within a tropical watershed. To achieve this goal, land use data for 2006, 2010, and 2015 and the corresponding years of frequently sampled water quality data were utilised. The kriging interpolation technique was applied to estimate the unknown water quality values at 12 unmonitored sampling points using the measured values from the 9 monitored sites that were used for the regression analyses. Both OLS and GWR were applied using four groups of land use and seven water quality variables. The study found better performances of GWR models over OLS models throughout these investigation periods. The GWR results indicated that the ability of land use indicators to explain water quality change is not the same with time and space and neither the same among the different water quality parameters. For examples, in 2015, agricultural land most predicted the change in most water quality variables compared with its prediction proportion in 2010 and 2006, while urban land most predicted the change in most water quality variables in 2010 compared to other years. However, other groups of lands were more positively associated with most of the water pollutants compared to forest, agricultural, and urban areas. Therefore, this study suggests that control and management policies be adjusted to different periods and areas according to the time-space varying sources of pollution and good predictors of water quality. Although the approach used in this study tends to reveal the complex spatiotemporal relationships between land use and water quality over different time periods, it should be noted that this approach is more appropriate for comparative analysis to understand the implication of the land use change pattern into surface water pollution

Keywords Land use · Water quality · GWR · OLS · Spatiotemporal relationships

Introduction

The quantity and quality of water resources have been affected by patterns of accelerated land use around the world. Rapid

urbanisation and industrial and agricultural development have been associated with the deterioration of aquatic and terrestrial ecosystems in the watersheds of developing countries. Numerous studies have shown a significant correlation

Responsible Editor: Amjad Kallel

✉ Nor Rohaizah Jamil
norrohaizah@upm.edu.my

Moriken Camara
mkcamara2015@gmail.com

Ahmad Fikri Abdullah
ahmadfikri@upm.edu.my

Rohasliney Hashim
rohasliney@upm.edu.my

¹ Department of Environment, Faculty of Forestry and Environment, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

² Aquatic Ecosystem and Management Laboratory, International Institute of Aquaculture and Aquatic Sciences (i-AQUAS), Lot 960 Jln Kemang 6, 71050 Port Dickson, Negeri Sembilan, Malaysia

³ Department of Biological and Agricultural Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

between land use planning activities and water quality in various water bodies (Li et al. 2008; Buck et al. 2004; Baker 2003; Tong and Chen 2002). In general, the level of concentration of water pollutants is linked to the higher percentage of land use change resulting from land development in watersheds, while good water quality is mainly found in the natural forest and undeveloped areas. However, these associations are not constant across different zones because the anthropogenic and natural-induced factors of watersheds such as the physical landscape, economic development activities, sources of pollution, and regulation policies vary often with space (Tu 2011a). As such, the relationships between land use and water quality studied by traditional statistical approaches, such as Pearson correlation and OLS regression, which suppose that relationships are stationary over space, would not be reliable when analysing spatial relationships among geographical data. Relationships may often vary spatially because the characteristics of pollutants and their sources within a watershed may not be the same in different locations. Therefore, the relationships between one type of land use and different water pollutants are inconsistent and different types of land use are not correlated with the same water pollution issues (Tu 2011b). To address these problems, in recent years, a simple and more powerful statistical method known as GWR has been developed in order to explore the continuing varying interactions between land use and water quality across space (Brunsdon et al. 2002) and time.

GWR provides a useful technique for exploring the spatially varying associations between land use and water pollutants across different regions. It has been the applied method to analyse the variations in spatial relationships between various phenomena including natural and anthropogenic factors (Mohammadi et al. 2019; Nkeki and Asikhia 2019; Wang et al. 2017; Huang et al. 2015; Yu et al. 2013; TU and XIA 2008). It is likely that when global models such as OLS are used for these analyses, a certain level of relationships may be masked. Monitoring sites are assumed to have a distinct relationship model between land use factors and water pollutants when the results are spatially analysed. By employing GWR to analyse the interactions between the land use and water quality variables, it is expected that each associated land use-water quality variable exhibits a more or less distinct pattern of relationship across study region. In fact, the GWR is able to present regression coefficient values and a spatially varying relationship strength for each associated variable on a map. GWR tries to account for spatial differences by letting the parameters of the regression model to vary in space. Local estimation of the model parameters is attained by weighting all adjacent counts using a function of distance decay, supposing that nearby counts have a greater effect on the regression point than those far away. Like the OLS system, GWR constructs a model to analyse how points may be different (Tu 2011b). Local model parameters are commonly mapped using

visualisation such as GIS, which allows local spatial variation in regression outputs to be studied (Tu 2011b). Subsequently, Global Moran I is performed on the OLS and GWR models' residuals to evaluate the performance of these models with respect to spatial autocorrelation. Moran's I is generally used as an indicator of spatial autocorrelation and has been used in previous studies to test the significance of spatial autocorrelation in regression models to ensure no violation of the randomness assumption and independence of residuals (Buyong 2006; TU and XIA 2008; Clark 2007; Harries 2006).

However, most previous studies examined the relationship between land use and water pollution in terms of total areas rather than densities (Huang et al. 2015; Chen et al. 2016; Erfanian and Alijanpour 2016). For example, they investigated the concentration of pollutants based on the percentage of a land use class in the sub-basin area, but it would also be relevant to evaluate pollutants in relation to land use class within pollution potential zone (contributing area) across the study region. In addition, it will be useful to see how these relationships vary over time, because the major concern could be peak pollution in time, and not regular pollution over time. Thus, more frequent sampling data for three different periods were needed in this study to evaluate the performance and predictive capability of GWR model in examining how water quality and land use variable relations vary with time and space.

Material and methods

Study area

The study area of this research is Selangor River watershed, located in the State of Selangor, Malaysia. It is the main river in the State of Selangor, with an extent area of about 2200 km², nearly a quarter of the total area of the State of Selangor (Chowdhury et al. 2018). The Selangor River streams in a southwest direction and runs through a total distance of 110 km before ending into the Strait of Malacca. The watershed is the largest water source for the States of Kuala Lumpur and Selangor, and about 60% of the water consumed in these States comes from the Selangor River (Sakai et al. 2017). The main tributaries include Sungai Rawang, Sungai Buloh, Sungai Batang Kali, Sungai Serendah, Sungai Kundang, Sungai Kerling, and Sungai Sembah. The watershed is approximately 70 km long by 30 km wide and covers almost 28% of the State of Selangor, where about 406,000 people lived in 2006 (Fulazzaky et al. 2010). Approximately 57% of the watershed is still covered by natural forests, while agricultural activities use 22%, 17% is used for development areas and 4% occupied by water (Kusin et al. 2016). The land use maps of the Selangor River watershed are shown in Fig. 1.

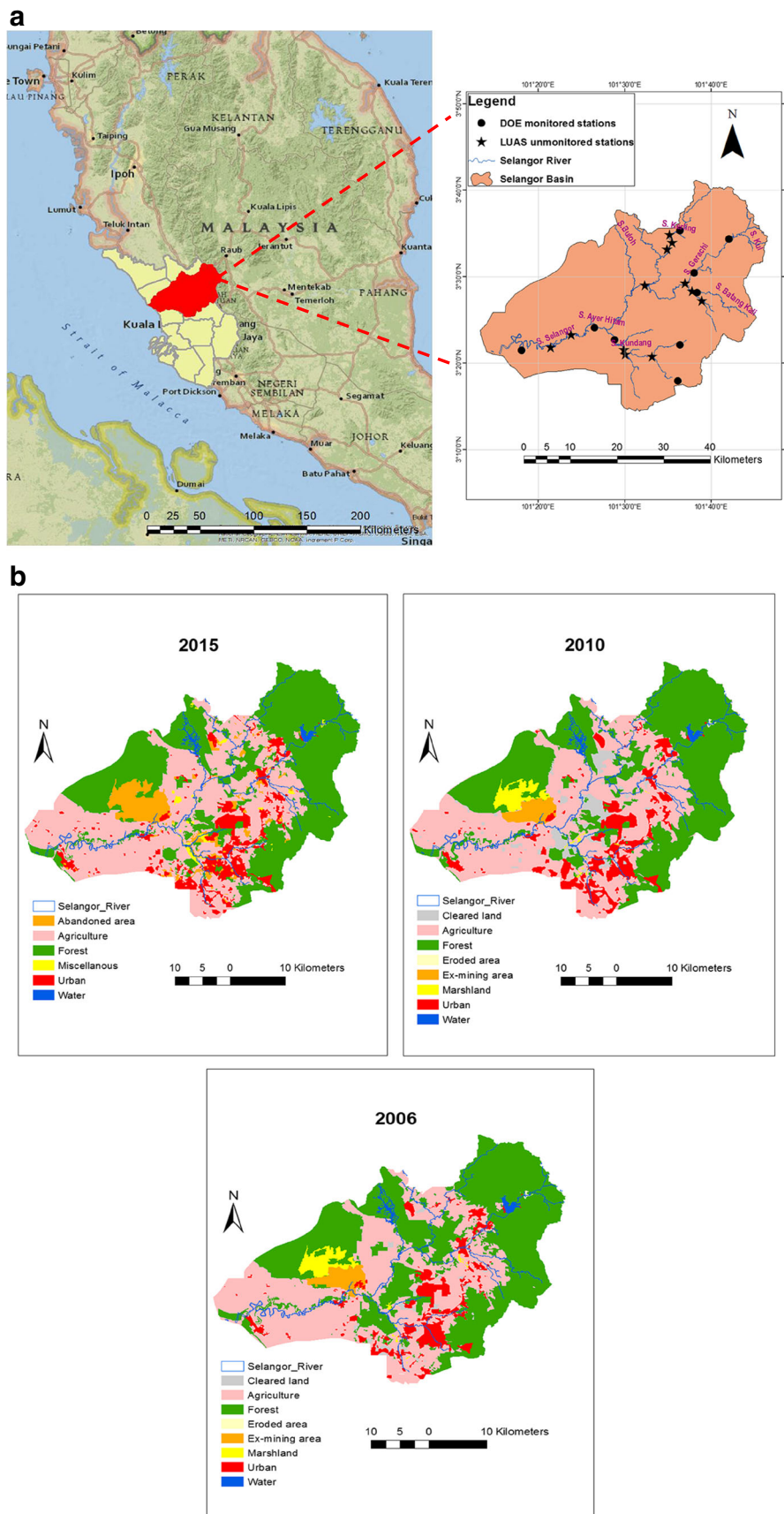


Fig. 1 Location of the study area along with sampling stations (a) and land use maps of the Selangor River Basin (b) from DOA (2019)

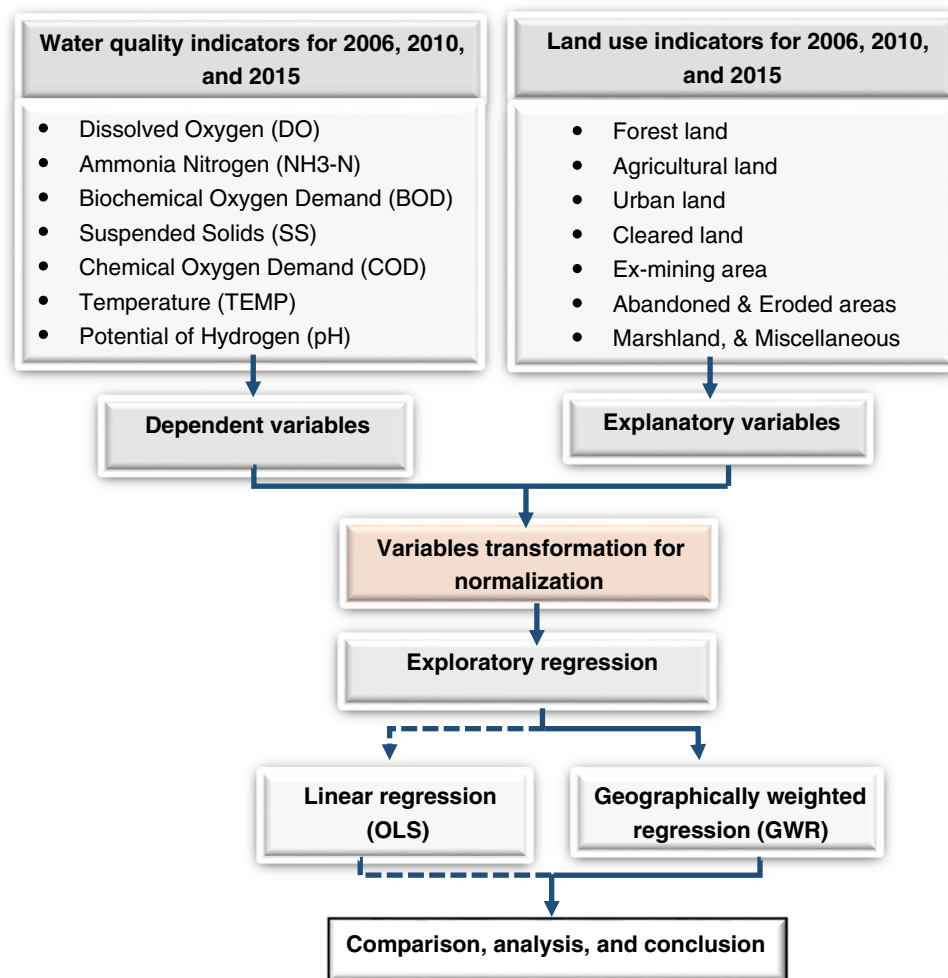
Land use and water quality data

To consistently and adequately analyse the interactions between land use and water quality pollutants over time and space, we used land use data for 2006, 2010, and 2015 and the corresponding years of frequently sampled water quality data, from two different sources. First, the land use maps were obtained from the Department of agriculture (DOA), Malaysia. The interview with the technicians of this department indicated that the land use maps were derived from high resolution satellite imageries (spots 2, 4, and 5). All pre-processing and processing analyses of the satellite imageries, including radiometric and geometric corrections, were performed at DOA in Malaysia. These imageries were spatially registered, corrected, and classified into several land use classes using the field ground control points and appropriate classification algorithm. In this study, the land use attributes were aggregated based on level I from Anderson (1976) Classification Scheme. Second, the water quality data used in this study were acquired from nine monitoring stations as part of the river water quality monitoring programme by the Department of environment (DOE) of

Malaysia in 2006, 2010, and 2015. Since 2000, the DOE regularly (every 2 months) monitors the water quality of these stations within the Selangor River basin. This dataset includes of a set of water quality indicators such as chemical oxygen demand (COD), dissolved oxygen (DO), biochemical oxygen demand (BOD), suspended solids (SS), ammonia nitrogen ($\text{NH}_3\text{-N}$), temperature (TEMP), and potential of hydrogen (pH) for monitoring sites that involve the lower, middle, and upper streams of the basin profile (Fig. 1).

In this study, LUAS comprising 12 points are estimates of unknown values using measured values at 1SR 9 sampled points based on a kriging estimation technique. The GWR and OLS models were conducted using water quality pollutants as dependent variables and land use classes as explanatory variables. To avoid possible multicollinearity among land use indicators, each of the GWR and OLS models used only one land use class as an independent variable to analyse its associations with each quality pollutant (Fig. 2). In this study, there were four classes of land use and seven water quality pollutants for three different years. As a result, 84 OLS models and 84 GWR models were built in this study. Given that land units located far from the river cannot generate pollution potential for bodies of

Fig. 2 Methodology flowchart



surface water (Sivertun and Prange 2003), pollutants generated at a distance greater than 1000 m cannot reach the river or influence the water quality of the river (Sivertun and Prange 2003; Do et al. 2011; Alilou et al. 2018). Therefore, in this study, a 1000-m buffer radius from each monitoring station is used to estimate the land use area and its relationship to water quality at each monitoring site. Through this step, a linkage between each water quality variable and land use indicators for all sampling points was established. Apart from the three main land use categories such as forest, agricultural, and urban lands, the exploratory regression indicated that the other land use classes do not individually contribute much to the prediction of water quality in the study area. Therefore, these land use classes have been grouped into one category (known as others) for further modelling.

Statistical analysis and modelling methods

Prior to modelling the spatiotemporally varying interactions between land use and water quality pollutants, the normality test for all land use classes and water quality variables was conducted using the Shapiro-Wilk test and quantile-quantile

Table 1 Standardised water quality values used in OLS and GWR analyses for 2015

Station	DO	BOD	COD	SS	pH	NH3-N	TEMP
1SR01	0.000	0.875	1.000	0.876	0.000	0.472	1.000
1SR03	0.990	0.000	0.011	0.262	0.780	0.017	0.274
1SR04	0.976	0.042	0.258	0.083	0.783	0.111	0.333
1SR05	0.933	0.167	0.596	0.000	0.870	0.023	0.325
1SR06	0.994	0.125	0.348	0.415	0.824	0.000	0.000
1SR07	0.938	0.000	0.000	0.049	1.000	0.193	0.206
1SR08	0.816	0.208	0.607	0.086	0.822	0.406	0.469
1SR09	0.182	0.500	0.337	1.000	0.715	1.000	0.733
1SR10	0.264	1.000	0.596	0.894	0.681	0.540	0.696
LUAS1	0.012	0.861	0.695	0.935	0.293	0.561	0.889
LUAS2	0.099	0.805	0.626	0.898	0.483	0.618	0.785
LUAS3	0.289	0.482	0.435	0.942	0.791	0.725	0.699
LUAS4	0.284	0.471	0.430	0.929	0.805	0.713	0.685
LUAS5	0.548	0.307	0.351	0.486	0.875	0.546	0.537
LUAS6	0.716	0.398	0.416	0.515	0.762	0.359	0.380
LUAS7	0.935	0.132	0.316	0.262	0.784	0.128	0.291
LUAS8	1.000	0.075	0.291	0.259	0.782	0.116	0.283
LUAS9	0.902	0.175	0.336	0.223	0.786	0.139	0.317
LUAS10	0.934	0.285	0.406	0.382	0.800	0.106	0.120
LUAS11	0.964	0.228	0.383	0.383	0.816	0.071	0.062
LUAS12	0.979	0.238	0.388	0.467	0.817	0.069	0.002

LUASs are estimates of unknown values using measured values at 1SR sampled points based on kriging estimation technique

Table 2 Standardised water quality values used in OLS and GWR analyses for 2010

Station	DO	BOD	COD	SS	pH	NH3_N	TEMP
1SR01	0.194	0.727	0.410	0.946	0.674	0.200	0.689
1SR03	0.947	0.545	0.429	0.019	0.028	0.371	0.170
1SR04	0.875	0.000	0.000	0.099	0.628	0.237	0.179
1SR05	0.743	1.000	0.632	0.084	0.000	0.000	0.294
1SR06	0.952	0.636	0.397	0.014	0.730	1.000	0.000
1SR07	0.830	0.727	0.313	0.000	1.000	0.020	0.285
1SR08	0.786	0.182	0.200	0.019	0.605	0.241	0.475
1SR09	0.035	1.000	1.000	0.679	0.116	0.823	1.000
1SR10	0.194	0.727	0.410	0.946	0.674	0.200	0.689
LUAS1	0.124	0.652	0.461	0.987	0.516	0.309	0.723
LUAS2	0.186	0.652	0.461	1.000	0.516	0.309	0.723
LUAS3	0.015	0.616	0.421	0.598	0.495	0.343	0.901
LUAS4	0.000	0.616	0.421	0.576	0.495	0.343	0.892
LUAS5	0.313	0.616	0.421	0.300	0.495	0.343	0.684
LUAS6	0.709	0.616	0.421	0.403	0.495	0.343	0.361
LUAS7	0.899	0.616	0.421	0.064	0.495	0.343	0.213
LUAS8	1.000	0.616	0.421	0.002	0.495	0.343	0.217
LUAS9	0.856	0.616	0.421	0.108	0.495	0.343	0.204
LUAS10	0.943	0.616	0.421	0.154	0.495	0.343	0.067
LUAS11	0.946	0.602	0.435	0.112	0.432	0.384	0.040
LUAS12	0.984	0.602	0.435	0.093	0.432	0.384	0.021

LUASs are estimates of unknown values using measured values at 1SR sampled points based on kriging estimation technique

(QQ) plots. Most land use indicators and water quality pollutants were not normally distributed. Therefore, the square root and natural log transformation techniques were performed for transforming them to meet the normality distribution requirements for further analyses. Table 1, Table 2, and Table 3 show the standardised water quality values used in OLS and GWR analyses for 2015, 2010, and 2006, respectively. Prior to these regression analyses, a Pearson correlation was applied to observe the overall association between land use indicators and water quality pollutants. In this study, the spatiotemporal relationships between land use and water quality were examined by applying both the OLS and GWR modelling methods. However, OLS is a globally homogeneous model and is unable to reveal the heterogeneity and non-stationarity of spatial relationships between geographic data and assumes a consistent relationship between variables in a given spatial region. Thus, we relied on GWR in this study to examine the spatiotemporally varying interactions between land use and water quality across the study area (Eqs. (1) and (2)).

In order to understand the evolution of the DOE water quality index (DOE-WQI) in relation to land use, a graphical model was developed. The local WQI mainly used in Malaysia emanated from an opinion polling formula of a

Table 3 Standardised water quality values used in OLS and GWR analyses for 2006

Station	DO	BOD	COD	SS	pH	NH3_N	TEMP
1SR01	0.150	0.125	0.817	0.577	0.000	0.142	0.565
1SR03	1.000	0.104	0.337	0.690	0.658	0.013	0.039
1SR04	0.841	0.042	0.221	0.116	0.566	0.000	0.274
1SR05	0.794	0.000	0.000	0.000	0.470	0.135	0.689
1SR06	0.967	0.021	0.087	0.049	0.634	0.005	0.000
1SR07	0.927	0.083	0.125	0.471	1.000	0.048	0.465
1SR08	0.843	0.146	0.279	0.233	0.754	0.079	0.477
1SR09	0.000	1.000	0.990	1.000	0.581	1.000	1.000
1SR10	0.364	0.292	1.000	0.472	0.512	0.269	0.785
LUAS1	0.114	0.289	0.963	0.605	0.222	0.259	0.631
LUAS2	0.134	0.323	0.996	0.606	0.348	0.259	0.697
LUAS3	0.296	0.362	0.917	0.519	0.659	0.188	0.850
LUAS4	0.300	0.344	0.909	0.522	0.676	0.188	0.815
LUAS5	0.559	0.258	0.571	0.489	0.784	0.188	0.623
LUAS6	0.684	0.214	0.534	0.400	0.618	0.188	0.413
LUAS7	0.900	0.145	0.320	0.337	0.633	0.188	0.107
LUAS8	0.940	0.159	0.305	0.340	0.650	0.188	0.145
LUAS9	0.877	0.146	0.322	0.334	0.623	0.188	0.198
LUAS10	0.886	0.146	0.224	0.303	0.618	0.188	0.182
LUAS11	0.907	0.145	0.163	0.286	0.628	0.205	0.128
LUAS12	0.910	0.145	0.139	0.279	0.628	0.205	0.103

LUASs are estimates of unknown values using measured values at 1SR sampled points based on kriging estimation technique

panel of experts consulted on the choice of parameters and the weighting of each parameter (Gazzaz et al. 2012). The six parameters chosen for the WQI are DO, COD, BOD, SS, pH, and NH3-N. The calculations are done on the sub-indices rather than the parameters themselves. From the computed WQI, a river can be categorised into a number of classes, each indicating the valuable uses to which that river can be used. This classification is based on allowable limits of designated pollution parameters. For this reason, the DOE has defined the values of the water quality variables (WQVs) and WQI indicators that determine each water quality class (DOE 2007). Details on the WQI calculation procedures are provided in the [supplementary materials](#) of this study.

Rather than global, GWR model lets local parameters to estimate the location of the sample and the model is expressed as follows:

$$y_j = \beta_o(u_j, v_j) + \sum_{i=1}^p \beta_i(u_j, v_j)x_{ij} + \varepsilon_j \quad (1)$$

In which y_j denotes the dependent variable, u_j and v_j are considered the coordinates for each location j , $\beta_o(u_j, v_j)$

considered the intercept of location j , and $\beta_i(u_j, v_j)$ is the estimation of local parameter for variable x_i at location j .

The GWR is executed by weighting all observed counts about a sampling point using a function of distance decay, by supposing that observed counts nearer to the location of the sampling point have a greater effect on the estimates of local parameter for the location (TU and XIA 2008). The weighting function can be defined using the following exponential decay formula:

$$w_{ij} = \exp(-d_{ij}^2)/b^2 \quad (2)$$

In which w_{ij} is the observation weight j for observation i , d_{ij} is considered the distance between the two observations (i and j), and b is known as the kernel bandwidth.

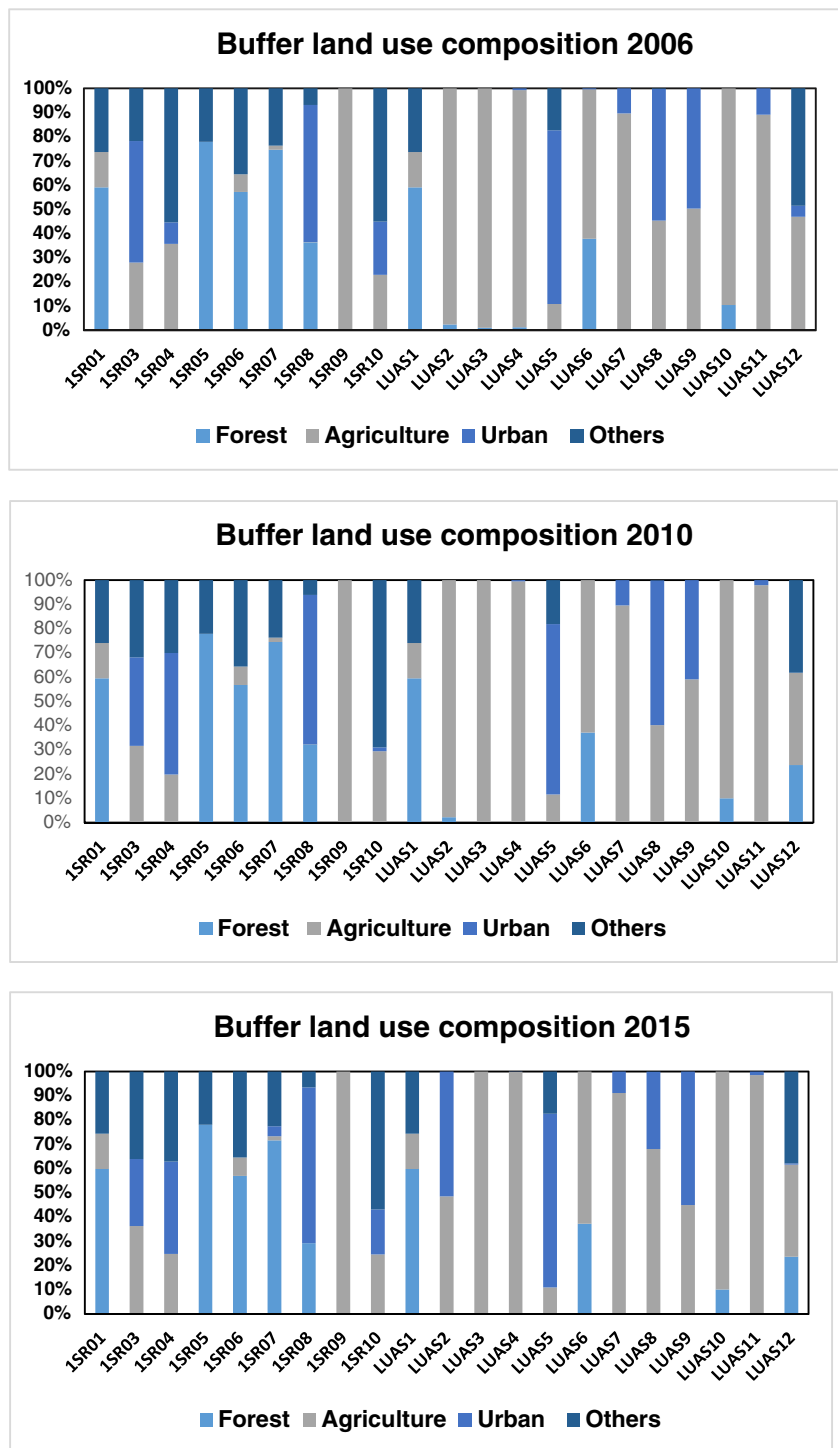
When the distance is superior to the kernel bandwidth, the weight quickly moves towards zero. The bandwidth of either fixed or adaptive kernel can be selected for GWR. However, GWR model generally works better with a large number of sampling sites, but with a small number of monitoring sites in a delimited space from the sampling points (pollution potential zone), and by choosing a fixed kernel which has a constant bandwidth across space, the GWR also provides desirable results.

In this study, the comparison between OLS and GWR models was done on the basis of adjusted R^2 and the AICc values of both models. Higher R^2 indicates that the explanatory variable may explain further variance in the dependent variable. While a lower AICc value means a closeness of the model to reality, a lower AICc value indicates better model performance. If the difference in AICc between two models is less than 3, they are considered equivalent in terms of explanatory power (Fotheringham and Chris Brunsdon 2003). When the difference between AICc_{OLS} and AICc_{GWR} is greater than 3, one model shows a statistically significant improvement over the other model (Yu et al. 2013).

For comparing the ability between OLS and GWR models in dealing with spatial autocorrelation, global Moran's I statistics was computed for the residuals of each of the OLS and GWR models. As it is a regularly used indicator to assess spatial autocorrelation, Moran's I value varies from -1 to 1 . Where a value of 1 indicates a perfect positive spatial autocorrelation, while a value of -1 means a perfect negative spatial autocorrelation, and a value of 0 designates a perfect spatial randomness. If significant autocorrelation is found in the residuals, the results of the regression analysis are not reliable (Ishizawa and Stevens 2007).

Afterwards, the local R^2 values resulted from the GWR models were used for mapping in order to give an ideal visualisation of the spatiotemporal variations in the relationships between land use and water quality at 1SR 9 regular monitoring stations, and the capabilities of the land use indicators to explain variations in water quality. All analyses and mappings were done using ArcGIS 10.2, SPSS, and MS Excel.

Fig. 3 Land use composition in the pollution zones



Results

Land use composition in the pollution zones

In this study, a 1-km buffer zone from each sampling station is used to estimate the land use area and its relation to water quality at each monitoring site in 2006, 2010, and 2015.

Figure 3 shows the composition of land uses in the buffer zones which include forest, agricultural, urban, and other land groups. It is not surprising that agricultural and forest lands are the most dominant in the contributory buffer zones, as these lands occupied about 80% of the total area of the river basin. However, the urban area, which is significant in only a few contributing areas, is expected to increase more rapidly over

Table 4 The results of Pearson correlations between land use and water quality variables in 2015, 2010, and 2006

Year	Variables	DO (mg/L)	BOD (mg/L)	COD (mg/L)	SS (mg/L)	pH	NH ₃ -N (mg/L)	TEMP
2015	Forest	0.377	- 0.415	- 0.264	- 0.600	0.480	- 0.233	- 0.318
	Agriculture	- 0.541	0.495	0.087	0.712	- 0.593	0.334	0.489
	Urban	0.286	0.270	0.112	- 0.457	0.186	0.037	- 0.064
	Others	- 0.523	0.567	0.164	0.582	0.043	0.643	0.421
2010	Forest	0.277	0.210	- 0.081	- 0.362	0.486	- 0.304	- 0.228
	Agriculture	- 0.030	- 0.565	- 0.472	0.262	0.030	0.110	0.006
	Urban	0.379	0.784	- 0.622	- 0.336	0.031	0.221	- 0.325
	Others	- 0.746	0.552	0.805	0.534	- 0.403	0.366	0.773
2006	Forest	0.128	- 0.128	- 0.427	- 0.262	0.306	- 0.239	0.118
	Agriculture	- 0.187	- 0.042	0.367	0.471	- 0.470	0.123	- 0.255
	Urban	0.265	0.165	- 0.229	- 0.355	0.222	- 0.234	- 0.101
	Others	- 0.260	0.076	0.488	0.038	- 0.070	0.081	0.377

Values in bold are different from 0 with a significance level $\alpha = 0.05$

the next decades, while other land use areas which include cleared land, ex-mining areas, marshes, and abandoned and eroded areas changed relatively little during these periods.

Pearson correlations between land use and water quality variables

Table 4 shows the results of Pearson correlation analysis between land use and water quality variables. From the results, it can be seen that in 2015, forest land is negatively correlated with all the water quality variables except DO and pH which are positively correlated with it. However, agricultural and urban lands are positively correlated with most variables, and agricultural area is significantly correlated with SS. With the exception of DO, all other water quality variables in 2015 are positively correlated with other land groups that include abandoned mining areas and miscellaneous land. Based on this result, it is evident that in 2015, apart from forest area, all other land use activities including agricultural and urban areas were important non-point sources of water pollution in the Selangor River. In addition, the result of correlation analysis between land use and water quality pollutants in 2010 indicate that forest land is positively correlated with three water pollutants which are DO, BOD, and pH, and negatively correlated with other pollutants. However, agricultural activities, urban area, and other land groups exhibit positive relationship with most of the pollutants, with urban land showing significantly positive correlation with BOD, and other land groups indicating significant positive correlation with COD and TEMP and significantly negative correlation with DO. Moreover, in 2006, while forest, agricultural, and urban lands are negatively correlated with most of the water pollutants, other land groups which include cleared land and ex-mining areas, however, are positively correlated with all the pollutants except DO and pH. These land groups could be considered the

major non-point source of the river pollution in this year. In general, the level of concentration of water pollutants at a particular site is related to the change of forest or natural areas to urban or agricultural land, while good water quality is mainly found in natural forest and undeveloped areas. However, it should be noted that various qualitative and management criteria may also influence the concentration of pollution at sites where the river receives pollutant loads from poultry farms, municipal wastewater and industrial wastewater (point sources). This may alter the trends of some parameters in time and space, regardless of the land use type or percentage in that area. Therefore, efforts to control both pollutant sources and pollution processes are needed.

Evaluation of performances of OLS versus GWR models

The performance of the OLS and GWR models was compared primarily based on the average adjusted R^2 values (ranging from 0 to 1) of all the land use indicators and the result is given in Table 5. A considerable improvement in GWR R^2 over the OLS R^2 for each water quality variable and the land use indicators is observed during these investigation periods. The R^2 values from GWR range from 0.65 to 0.07, 0.50 to 0.04, and 0.25 to 0.03 for 2015, 2010, and 2006, respectively. The R^2 values from OLS range from 0.44 to 0.04, 0.25 to 0.01, and 0.10 to 0.01 for 2015, 2010, and 2006, respectively. Thus, the R^2 values from GWR are all considerably higher than the corresponding R^2 values from OLS. In addition, the result also indicates that the two models most predict the variation in DO for all the three periods compared with the other pollutants, while the models explain the least the variation in NH₃-N for these periods.

In addition to evaluating the performance of the OLS and GWR models by considering higher values of global R^2 (adjusted R^2), to identify the statistical significance of one model

Table 5 Comparison of average coefficient of determination (adjusted R^2) between GWR and OLS

Dependent variables	Predictor variables	GWR (R^2)			OLS (R^2)		
		2015	2010	2006	2015	2010	2006
DO (mg/L)	Forest, agriculture, urban, others	0.65	0.39	0.23	0.44	0.19	0.10
BOD (mg/L)	Forest, agriculture, urban, others	0.34	0.26	0.05	0.20	0.25	0.03
COD (mg/L)	Forest, agriculture, urban, others	0.07	0.25	0.15	0.04	0.25	0.07
SS (mg/L)	Forest, agriculture, urban, others	0.35	0.50	0.11	0.26	0.23	0.04
pH	Forest, agriculture, urban, others	0.42	0.04	0.25	0.13	0.04	0.06
NH3-N (mg/L)	Forest, agriculture, urban, others	0.15	0.04	0.03	0.09	0.01	0.01
TEMP (mg/L)	Forest, agriculture, urban, others	0.44	0.27	0.09	0.22	0.16	0.06

improvement over the other, we used another measure of goodness-of-fit which is the AICc values presented in Table 6. In this case, models with smaller AICc values are better than models with higher values. When the difference in AICc values between two models is greater than 3, one model shows a statistically significant improvement over the other model (Yu et al. 2013).

As shown in Table 6, significant perfections in the fit of the GWR over OLS models are found for the forest models with DO, BOD, COD, and TEMP in 2015; DO, TEMP, and SS in 2010; and TEMP in 2006. For agriculture, the GWR models present significant improvements compared with their corresponding OLS models for DO and pH in 2015; SS and TEMP in 2010; and DO, SS, and pH in 2006. For urban, the improvement of GWR models is significant for DO, SS, and TEMP in 2015, and only for TEMP in 2006. The improvement of GWR models over OLS models for others is significant for DO, BOD, and pH in 2015; for DO, SS, and TEMP in 2010; and only for

BOD and COD in 2006. Overall, 27 of the 84 GWR models in total show significant perfections over their corresponding OLS models. The remaining 57 models are considered to be insignificant improvements over the OLS models, although in few cases the OLS models show improvements slightly greater than their corresponding GWR models, but none of these improvements is greater than 3 in values of difference.

Comparison of spatial autocorrelations of OLS and GWR residuals

The results of the Moran’s I statistics based on the residuals of the GWR and OLS models for 2015, 2010, and 2006 are presented in Table 7. Both positively and negatively spatial autocorrelations are identified for all the OLS models. Significantly positive spatial autocorrelations are also identified in the models. In Table 7, bold number indicates values with significant autocorrelation at $p \leq 0.05$; bold and italic number indicates significant

Table 6 Results of statistical test showing the improvement in the model fit of GWR over OLS

Water quality	Model	Forest			Agriculture			Urban			Others		
		2015	2010	2006	2015	2010	2006	2015	2010	2006	2015	2010	2006
DO (mg/L)	OLS	33.90	34.56	35.13	34.17	34.27	35.96	34.51	33.88	34.62	32.40	27.95	34.65
	GWR	26.98	31.04	35.94	30.66	35.27	30.97	31.08	33.35	34.14	26.97	21.54	34.08
BOD (mg/L)	OLS	35.58	34.88	33.13	32.75	31.82	35.27	34.60	26.71	35.03	34.60	32.01	35.23
	GWR	31.95	34.89	35.14	32.34	30.83	33.28	33.86	26.52	33.04	31.44	32.03	29.24
COD (mg/L)	OLS	34.63	35.22	33.47	35.21	33.02	33.98	33.17	30.87	34.80	35.03	25.89	32.83
	GWR	29.63	33.23	33.48	34.70	32.03	31.99	35.17	28.88	34.09	35.04	25.91	27.22
SS (mg/L)	OLS	31.26	34.02	35.64	28.91	34.64	33.02	35.17	34.20	34.07	31.56	32.25	35.27
	GWR	31.12	30.39	34.05	28.22	30.61	29.04	30.76	32.59	34.07	31.46	29.84	35.27
pH	OLS	32.92	31.85	34.39	31.38	35.27	35.04	34.97	35.27	34.83	35.26	33.69	35.24
	GWR	32.84	32.86	34.36	27.89	35.29	31.53	34.48	35.28	32.83	29.21	33.69	34.92
NH3-N (mg/L)	OLS	32.78	34.41	34.75	34.22	35.17	33.14	35.27	34.83	34.77	30.48	33.99	35.22
	GWR	34.78	34.42	34.27	34.13	33.19	35.16	35.20	33.84	34.79	30.48	34.00	33.23
TEMP (mg/L)	OLS	34.32	34.80	35.15	32.82	35.28	34.68	35.24	34.28	35.19	33.53	28.09	33.90
	GWR	28.85	27.92	28.16	32.20	30.29	34.39	31.49	34.17	30.20	32.31	24.80	33.21

Bold numbers indicate a significant perfection of GWR over OLS, with $AIC_{OLS} - AIC_{GWR} > 3$

Table 7 Comparison of Moran's I of the residuals from the OLS and GWR models

Year	Variables		DO	BOD	COD	SS	pH	NH3-N	TEMP
2015	Forest	OLS	<i>0.5953</i>	<i>0.5402</i>	-0.1590	<i>0.6144</i>	0.0930	<i>0.5532</i>	0.3822
		GWR	<i>0.5720</i>	0.4724	-0.1595	0.5765	0.0623	<i>0.5527</i>	0.3074
	Agriculture	OLS	0.3990	0.2760	-0.1390	0.3084	-0.0951	0.4469	0.2109
		GWR	0.3980	0.2283	-0.1749	0.3080	-0.1165	0.4463	0.1672
	Urban	OLS	0.4120	0.3852	0.0034	0.2305	0.0626	0.4343	0.3161
		GWR	0.2212	0.2798	0.0032	0.1362	0.0234	0.3924	0.1283
Others	OLS	0.2471	0.3852	-0.1285	0.0344	0.1467	0.2426	0.1542	
	GWR	-0.0930	0.2023	-0.1290	-0.0359	0.0177	0.2417	0.0031	
2010	Forest	OLS	<i>0.6000</i>	-0.0406	0.1645	<i>0.6396</i>	-0.2371	-0.3034	<i>0.5820</i>
		GWR	0.3991	-0.0409	-0.1647	0.4730	-0.2376	-0.3034	0.4222
	Agriculture	OLS	<i>0.4716</i>	0.0707	-0.1838	<i>0.5259</i>	-0.2318	-0.3430	<i>0.4487</i>
		GWR	<i>0.4711</i>	0.0701	-0.1841	<i>0.5172</i>	-0.2323	-0.3430	<i>0.4480</i>
	Urban	OLS	0.2429	-0.2545	-0.3708	0.3254	-0.2230	-0.3089	0.2590
		GWR	0.1572	-0.2545	-0.3708	0.2705	-0.2234	-0.3091	0.2115
Others	OLS	0.3756	-0.1295	-0.2813	0.4013	0.0430	-0.2580	0.2781	
	GWR	0.1590	-0.1295	-0.2814	0.2564	0.0421	-0.2581	0.0944	
2006	Forest	OLS	0.2788	0.1092	0.3872	0.0257	0.0394	0.0950	0.3070
		GWR	0.2782	0.1088	0.3867	0.0251	0.0279	0.0945	0.3063
	Agriculture	OLS	0.3287	0.0746	0.4963	0.3230	-0.0024	0.0099	0.3053
		GWR	0.3282	0.0741	0.4957	0.3225	0.0062	0.0094	0.3045
	Urban	OLS	0.2716	0.1079	0.3965	0.1570	0.0557	0.0507	0.2432
		GWR	0.1519	0.1074	0.2713	0.1563	0.0558	0.0503	0.2425
Others	OLS	0.0845	0.0392	0.0669	0.0050	0.0971	-0.0164	0.0647	
	GWR	-0.0350	0.0388	-0.0594	0.0043	0.0458	-0.0167	0.0642	

Bold number indicates values with significant autocorrelation at $p \leq 0.05$; bold and italic number indicates significant autocorrelation at $p \leq 0.01$

autocorrelation at $p \leq 0.01$. Based on these results, in 2015, significant autocorrelations are found in 11 of the 28 OLS models, while they are found in 9 models for 2010, and only in 5 models for 2006. However, most OLS models show no spatial autocorrelation which indicates that most of these models are suitable for examining the relationships between land use and water quality variables. In addition, the results of the Moran's I statistics derived from the residuals of the GWR models are also presented in Table 7. From these results, both positively and negatively spatial autocorrelations are found for all the GWR models. As shown in Table 7, in 2015, significant autocorrelations are identified in 6 of the 28 GWR models, while they are only found in 4 models for 2010, and in 5 models for 2006. This result indicates that GWR models increase the reliability of the relationships by minimising the spatial autocorrelations in the model residuals.

Temporal variations in relationship between forest land and water quality pollutants

Figure 4 shows the result of temporal variations in local R^2 values for forest land in relation to each water quality variable. This result indicates the capacity of forest land to explain the variance of each water quality variable at each monitoring site according to its proportion of annual prediction. In Fig. 4, we can observe that in 2015, forest land most predicts the variation in DO,

BOD, SS, and TEMP compared to its prediction proportion in 2010 and 2006. However, for pH and NH3-N, the 2010 proportion of prediction is the highest, while for the COD, the highest proportion is recorded in 2006, followed by the prediction of 2015, and no notable prediction in 2010 for this variable.

Temporal variations in interactions between agricultural land and water quality pollutants

The result of temporal variations in local R^2 values for agricultural land in relation with each water quality pollutant is shown in Fig. 4. This result indicates the ability of agricultural land to explain the variance in each water quality variable at each monitoring station according to its proportion of annual prediction. Figure 4 shows that in 2015, agricultural land most predicts the change in most water quality variables compared with its prediction proportion in 2010 and 2006. However, it is only for BOD and COD that the proportion of prediction for 2010 is dominant compared to other years.

Temporal variations in relationship between urban land and water quality pollutants

The result of temporal variations of local R^2 values for urban land in relation to each indicator of water quality is shown in

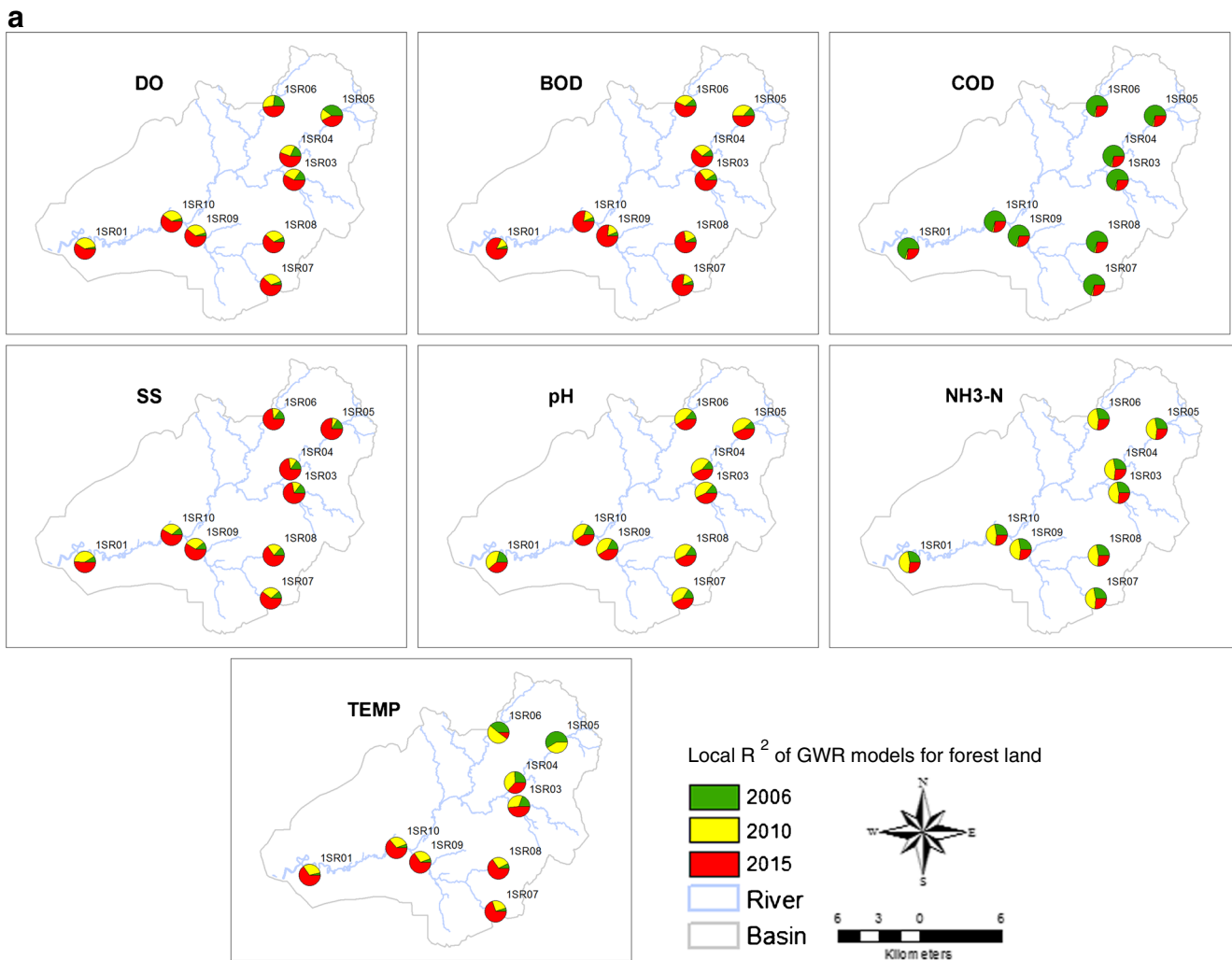


Fig. 4 Local R^2 of GWR models for forest (a), agricultural (b), urban (c), and other lands (d)

Fig. 4. This result indicates the capacity of urban land to explain the variance of each water quality variable at each monitoring site based on its annual prediction proportion. Figure 4 indicates that in 2010, the proportion of annual prediction of urban land was higher in DO, BOD, COD, and TEMP compared with its prediction in 2015 and 2006 for these variables. However, the proportion of prediction from urban land is higher in 2006 for pH and NH3-N, while for SS, its prediction for 2015 is dominant over other periods.

Temporal variations in relationship between other lands and water quality pollutants

Figure 4 shows the result of temporal variations in local R^2 values for other lands in relation to each water quality variable; based on this result, the ability of other lands to explain the variance in each water quality variable at each monitoring site based on its yearly prediction proportion is observable. Figure 4 indicates that other lands have higher prediction proportion for most water quality variables in 2010 compared to

other periods. However, only for BOD and NH3-N, the prediction proportion from these lands is higher in 2015 compared with 2010 and 2006 predictions.

Spatial relationship between land use and water quality variables in 2015

Figure 5 shows the result of spatial variations in local R^2 values between land use and water quality variables at the monitoring sites in 2015. This result indicates that the association between land use and water quality variables varies spatially from one monitoring station to another; thus, other land groups which include abandoned mining areas and miscellaneous land can explain most of the spatial variations in most of the water quality indicators, including DO, BOD, NH3-N, and TEM. However, forest land can explain the most the variation in COD while agricultural land explains the most the variation in SS and pH parameters in 2015. This result is consistent with the result obtained from the Pearson correlation analysis between land use and water quality in 2015 (Table 4).

b

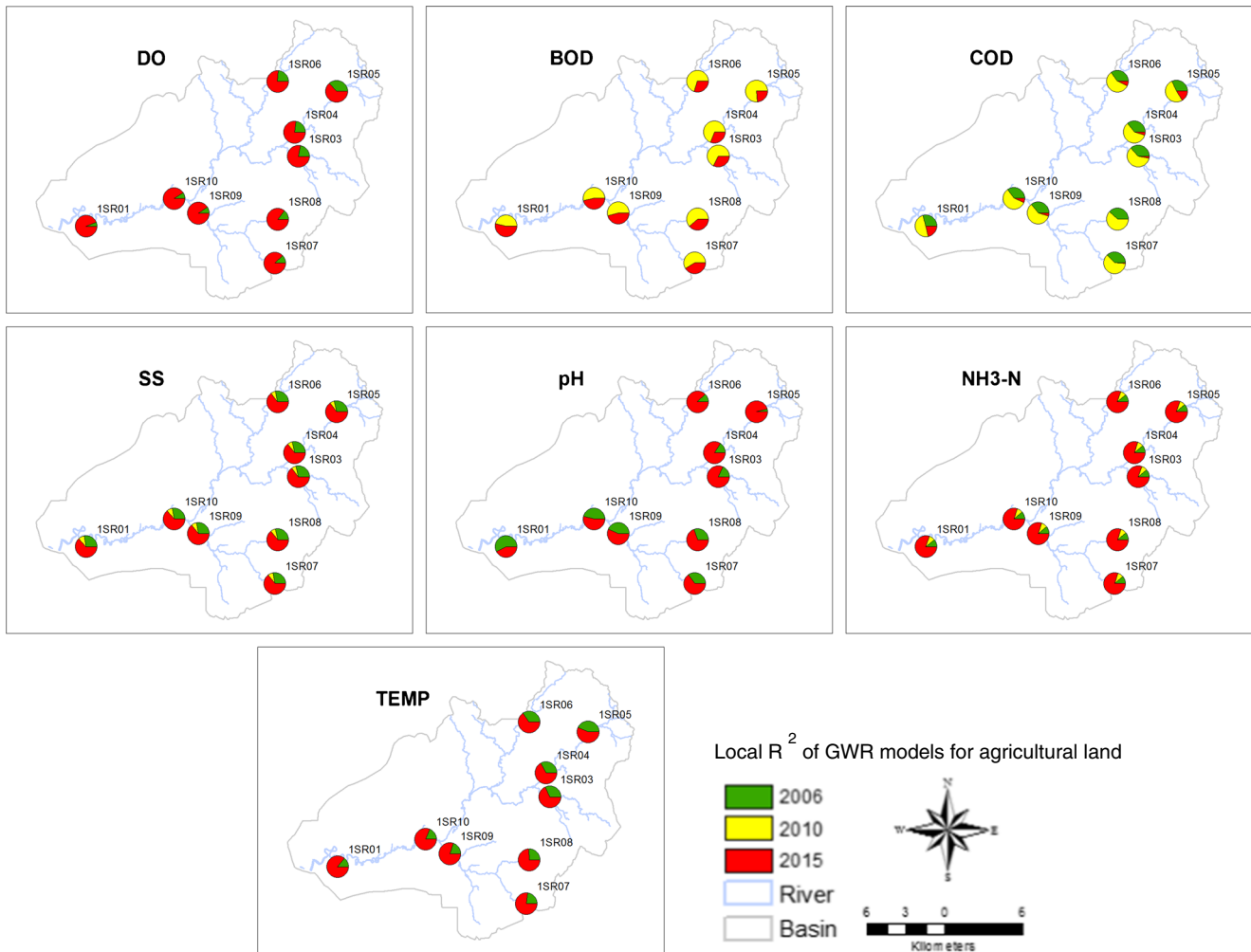


Fig. 4 (continued)

Spatial relationship between land use and water quality variables in 2010

Figure 5 shows the result of spatial variations in local R^2 values between land use and water quality variables at the monitoring sites in 2010. This result indicates that the association between land use and water quality variables varies spatially from one monitoring station to another; therefore, other land groups including cleared land, abandoned mining areas, and eroded land may account for most of the spatial variation in most of the water quality indicators in this year. While urban land appears to be the most important predictor for BOD and a good predictor for COD, forest land indicates the greatest association only with pH. And agricultural land has considerable association with BOD and COD. This result also reflects the result of the correlation analysis between land use and water quality in 2010 (Table 4).

Spatial relationship between land use and water quality variables in 2006

Figure 5 shows the result of spatial variations in local R^2 values between land use and water quality variables at the monitoring sites in 2006. This result indicates that the association between land use and water quality variables is not the same at all the monitoring sites. Figure 5 indicates that agricultural land can account for most of the SS and pH variations in 2006, while urban and forest lands are the greatest predictors of BOD and NH₃-N this year. However, other lands predict more variation in DO, COD, and TEMP. This result is also in line with the result of Pearson correlation analysis between land use and water quality in 2006 presented in Table 4. However, unlike Pearson's correlation and OLS regression, the result of local R^2 from GWR models allows us to observe how the relationship between variables varies in space between sampling sites.

C

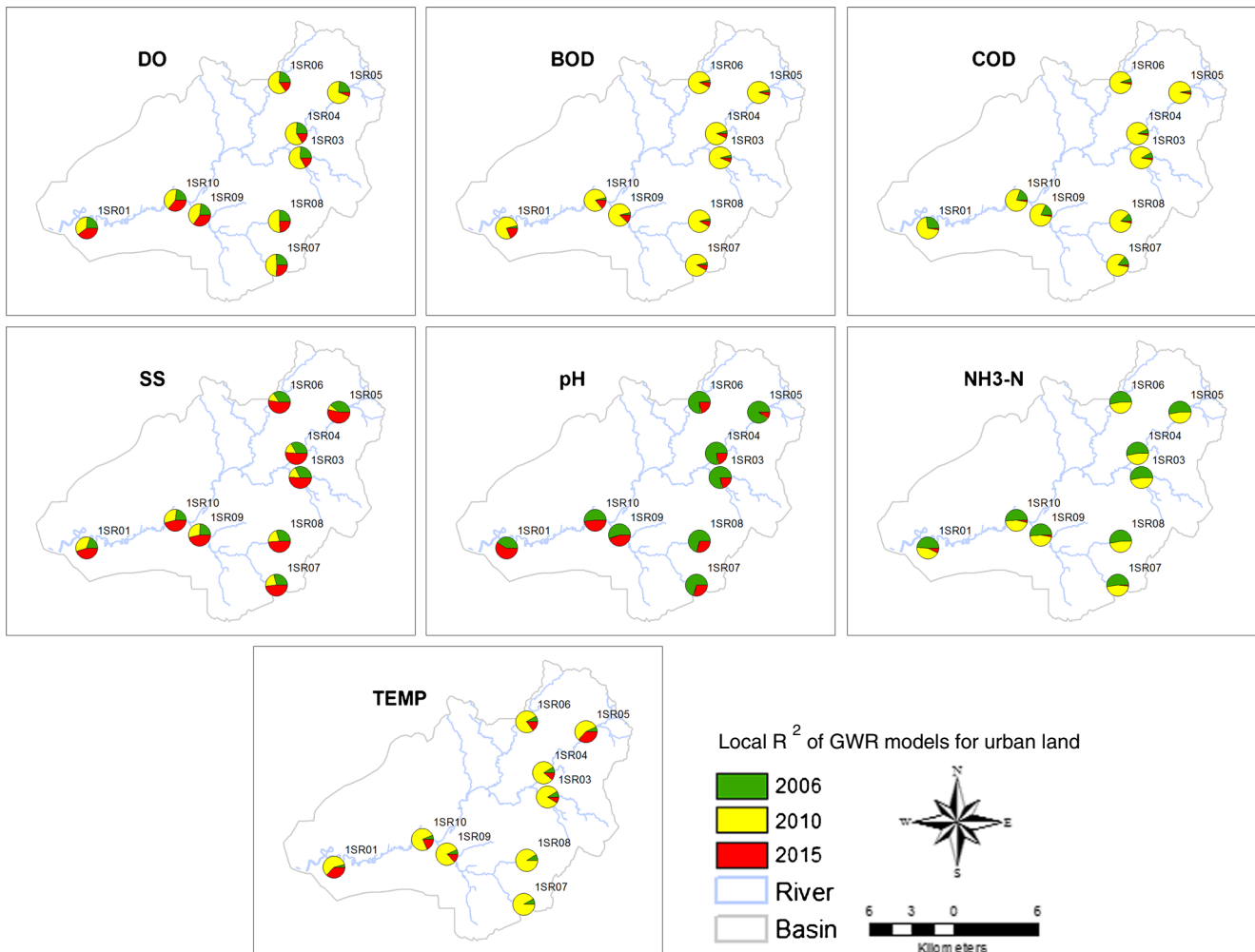


Fig. 4 (continued)

Temporal variations in relationship between land uses and local water quality index

This research employed the local WQI to evaluate the state of water quality in relation to land use indicators for the three different periods. In this process, the water quality data was converted into usable information that reflects the level of water quality degradation in the river (Fig. 6). The water quality status expressed in terms of WQI indicates that the river water is generally of good quality and can therefore be used directly for recreational activities with body contact, but conventional treatment is required for other uses such as domestic supply. However, the river water is of average quality at SR01, 1SR09, and 1SR10 for all 3-year periods, indicating the level of water quality degradation requiring extensive treatment. These monitoring stations are located downstream of the watershed where anthropogenic activities are dominant and they are classified in class III by the DOE-water quality index (WQI). In addition, Fig. 6 clearly shows that

predominantly agricultural areas are associated with good water quality status, while areas of agricultural, urban, and other land use activities locate sampling stations that have generally a moderate water quality status in the river basin.

Discussion

Relationship between land use and water quality indicators

All the land use indicators have more or less interactions with all the water quality variables at certain monitoring sites or even throughout the study area. These interactions varied between different variables of water quality and land use. These variables also showed an important spatial non-stationarity. The interactions for the same pair of land use and water quality variables at different monitoring sites varied not only in response to different time periods but also according to different

d

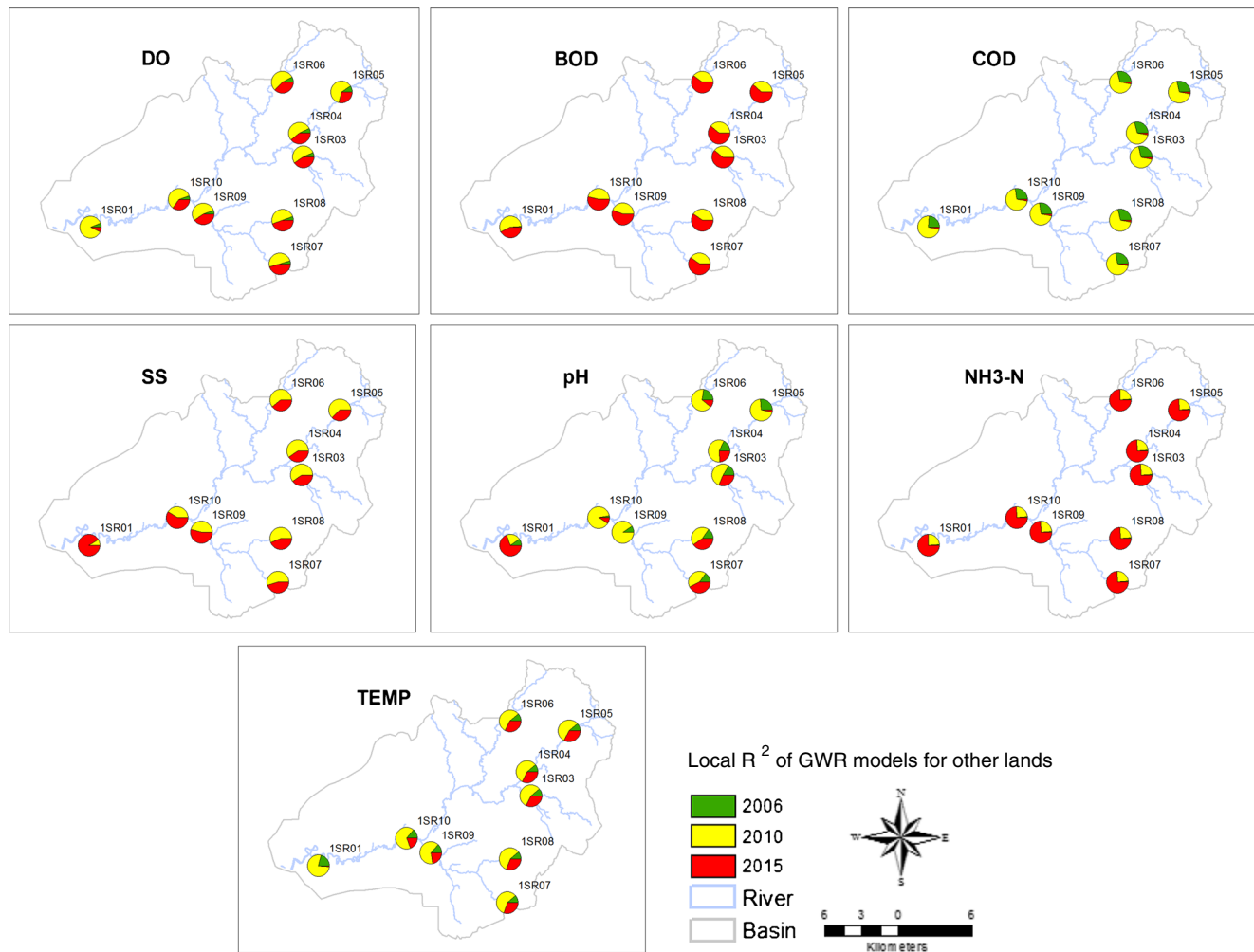


Fig. 4 (continued)

levels of land use indicators in the contributing areas. Large spatial differences in interactions were observed between a low and a high percentage of each land use indicator at different monitoring sites. While the Pearson's correlation and OLS results primarily indicated the global associations between land use indicators and water quality pollutants, GWR results enabled to observe the spatial non-stationarity of these associations, as also demonstrated by the previous studies (Mohammadi et al. 2019; Tu 2011; TU and XIA 2008; Yu et al. 2013; Chen et al. 2016; Huang et al. 2015). However, unlike these previous studies, the present study also demonstrated considerable temporal differences in the interactions between the same pair of land use indicators and water quality pollutants at different monitoring sites over different periods. It was relevant to see how these interactions vary over time, as the main concern could be the peak of pollution over time, not regular pollution over time. Thus, both the temporal irregularity and the spatial non-stationarity between land use and water quality indicators at different sites were covered in this study.

Temporally varying relationship between land use and water quality pollutants

While all of the land use indicators used in this study were relatively important predictors of water quality, it is important to note that the predictive ability of the same pair of land use indicators varies from period to period, which could be due to the fact that the percentage of each category of land use was not the same in different periods. The Selangor River watershed has undergone various land use changes in recent decades, with a considerable conversion of forest land to either agricultural, urban, or other development areas (Fig. 1). Although the forest area is affected by various anthropogenic activities, however, the vegetative landscape of the river basin is still covered by about 57% of natural forests (Kusin et al. 2016). This is an important area for maintaining water quality in good condition, as forest land showed a negative correlation with most water quality pollutants in all the investigation periods. However, forest became a more important predictor of most pollutants

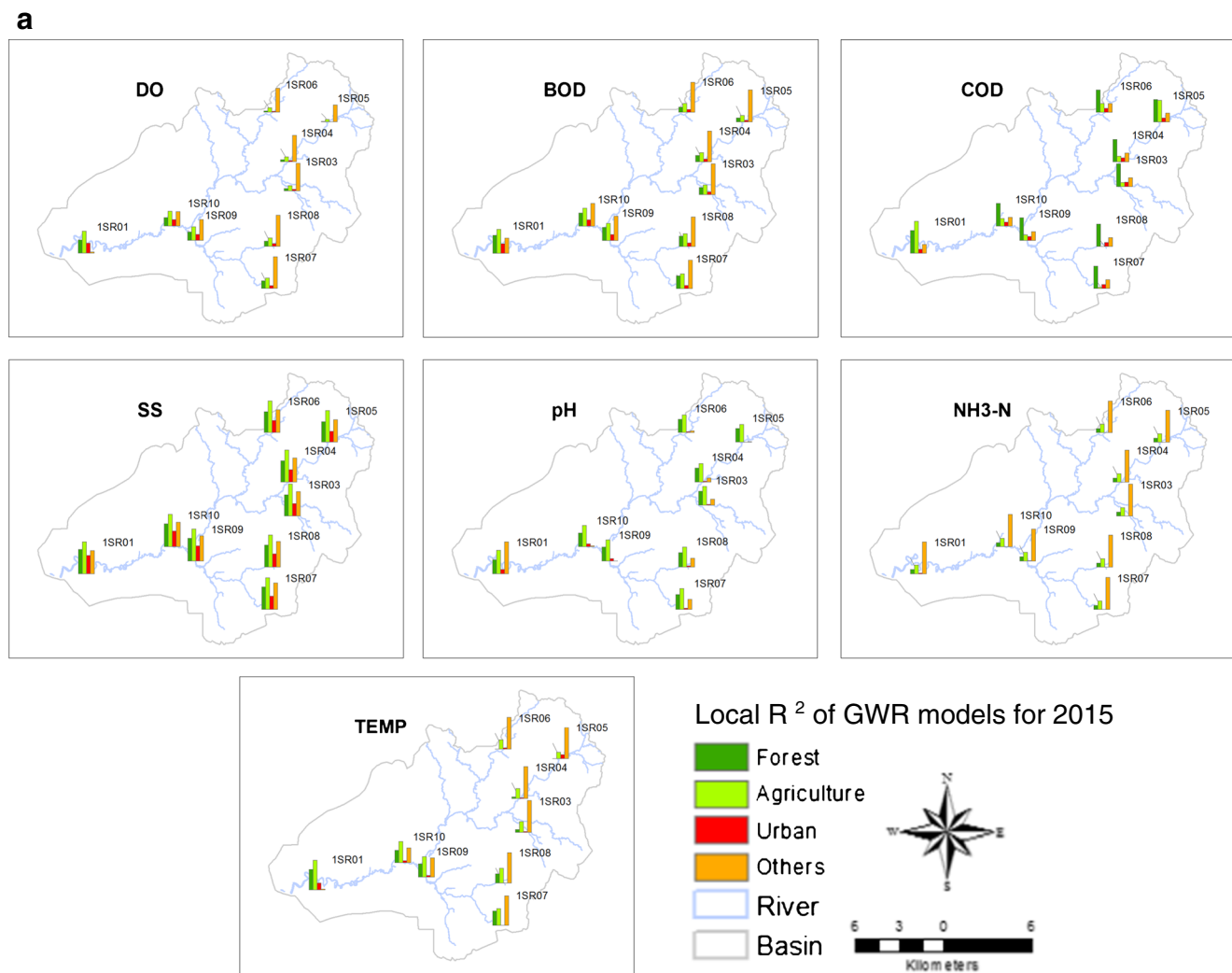


Fig. 5 Local R^2 of GWR models for land use indicators in 2015 (a), 2010 (b), and 2006 (c)

in 2015 compared to 2010 and 2006 where the forest area was correlated with fewer pollutants. Likewise, agricultural land most predicted the change in most water quality variables in 2015, while urban areas and other land groups including cleared land, abandoned mining areas, and eroded land were the most important predictors of water quality in 2010. Thus, the variation in prediction capability of an explanatory variable is consistently the result its level of correlation with the dependent variable under observation. As such, only GWR models have the predictive ability to examine how these interactions vary across the study area and over time. Traditional statistical methods such as the OLS and Pearson correlation techniques are global and do not reveal local relationships.

Spatially varying relationship between land use and water quality pollutants

Forest is commonly associated with good water quality in various studies on watersheds worldwide (Bahar et al. 2008;

Huang et al. 2014; Camara et al. 2019; Nainar et al. 2017). Similarly, in this study, negative associations were found between forest land and most water quality pollutants. The GWR results showed that these associations were not constant among the different sampling sites. Although natural forest covers the largest area of the total study area, however, in most contributing areas, local estimates indicate that forest fragmentation is not very useful in improving the quality of water in the study area, since the forest area has a higher explanatory power only for COD in 2015, pH in 2010, and BOD and NH₃-N in 2006. Thus, the ability of forest land to explain the change in water quality not only varied with time and space but also differed among different water quality pollutants.

Agricultural land is generally considered to be an important source of non-point pollution for surface water quality, significantly positive association is often found between agricultural land and increased pollution of water quality due to various farming activities such as fertiliser application, crop production, and livestock (Camara et al. 2019; Stutter et al. 2007;

b

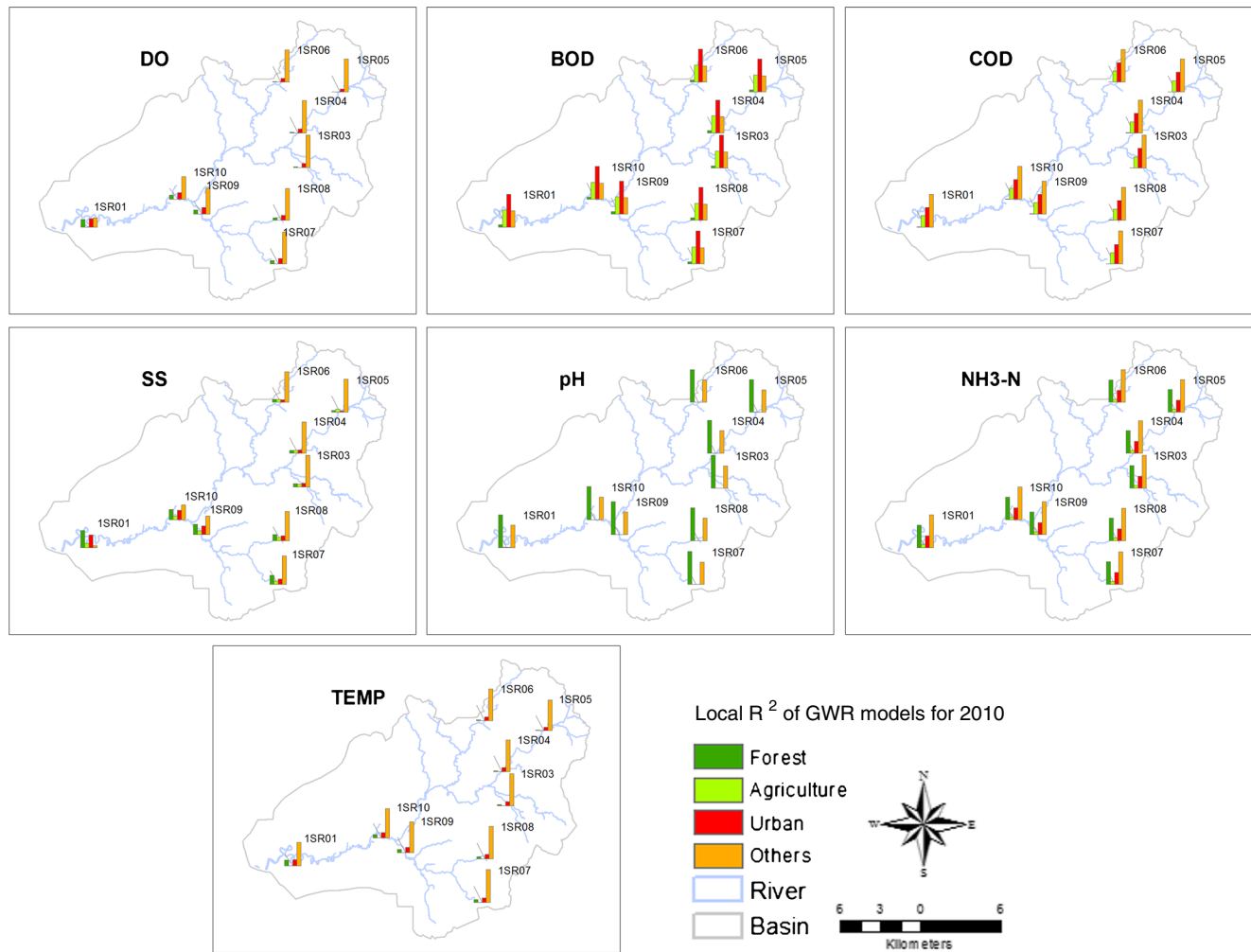


Fig. 5 (continued)

Ariffin et al. 2016). Although agricultural land represented the lowest percentage of land use in the contributing areas at 1SR sites, it could nonetheless make a great contribution to water pollution in the river watershed by having positive correlation with several water quality parameters and significantly positive association with SS in 2015. It is not surprising that with its lowest percentage, agricultural land had a high predictive power for the parameters SS and pH in 2015, BOD and COD in 2010, and SS and pH in 2006, because studies indicated that agricultural land usually represent less than 10% in most watersheds, but continues to contribute considerably to water pollution (Tu 2011a, b; Mehaffey et al. 2005). In this study, the ability of agricultural land to explain the variation in water quality not only varied over time and space but also differed among the different water quality pollutants (Fig. 5).

Urban land, including residential and recreational lands, facilities, and public service areas, is generally related to

water quality deterioration resulting from various anthropogenic activities, such as food waste, construction residues, municipal sewage, and wastewaters from treatment plants, which more correlate with higher concentration of water quality pollutants in watersheds, particularly in developing countries (Azyana et al. 2012; Tong and Chen 2002; Nurhidayu et al. 2016; Camara et al. 2019). Urban areas also influence surface runoff and erosion patterns and alter hydrological processes (Camara et al. 2019). However, in the present study, urban land exhibited positive relationship with most of the water quality pollutants, with significantly positive correlation with BOD in 2010. Tong and Chen 2002 also found significantly positive relationships between urban land and many water quality parameters including BOD in the State of Ohio watersheds, USA. In this study, the ability of urban land to explain water quality change was revealed by GWR (Fig. 5), the results also indicated that this explanatory ability was not

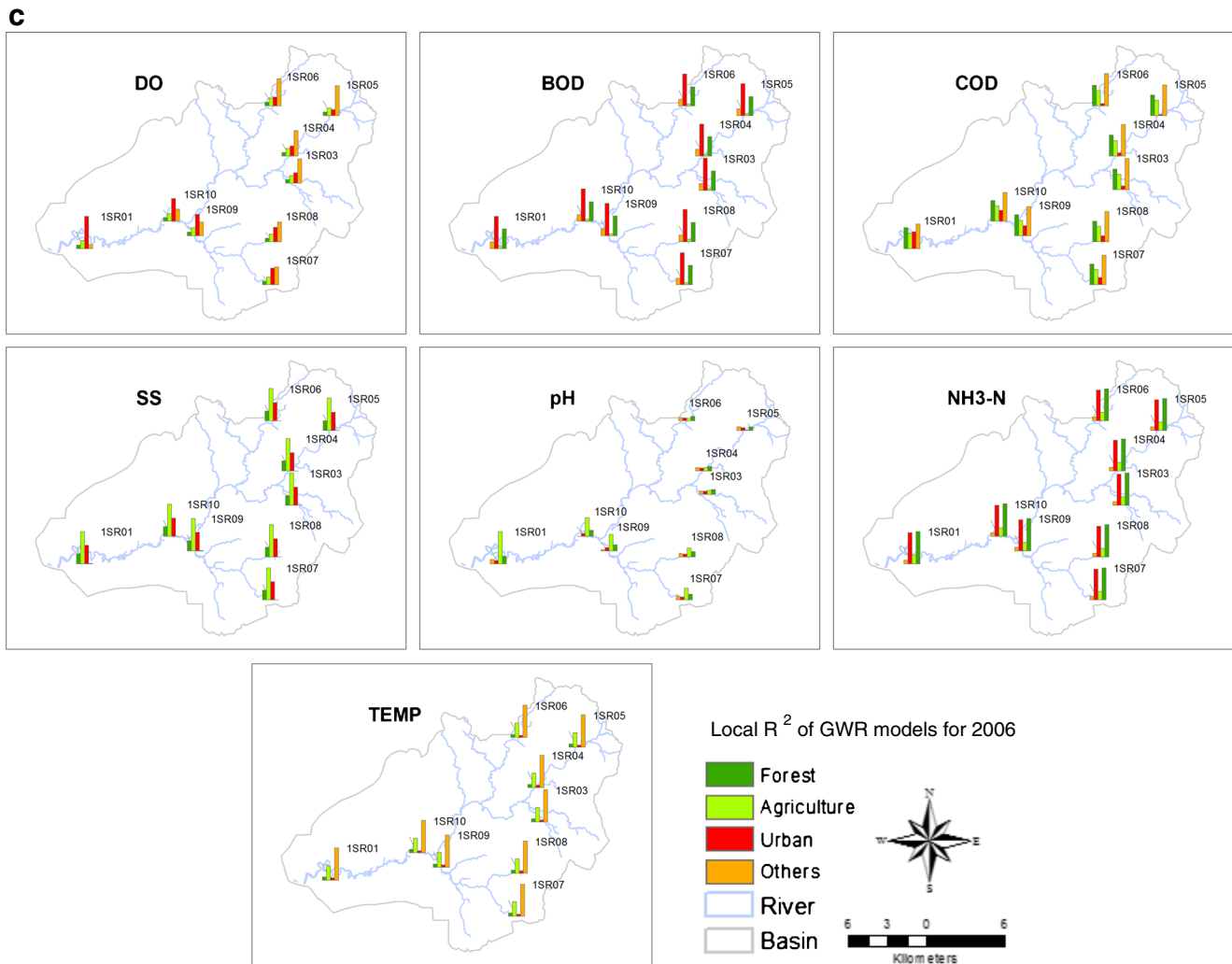


Fig. 5 (continued)

the same over time and space and neither the same among the different water quality parameters. Thus, urban land appeared to be the most important predictor of BOD and a good predictor of COD in 2010, and a greater predictor of BOD and NH₃-N in 2006.

Another important group of land use in this study included cleared land, abandoned mining area, eroded land, and miscellaneous area. This land use group showed positive relationships with most of the water quality variables throughout these investigation periods. The results of the Pearson correlation and OLS, as well as the GWR results, indicated that this group of lands was more positively associated with most of the water pollutants in the study area compared to other land use indicators such as forest, agricultural, and urban areas. Mining activities were a major source of pollution in many parts of the world, and several studies have addressed this issue (Kusin et al. 2018; Gomez-Gonzalez et al. 2016; Diami et al. 2016). Cleared and eroded lands could also make a significant contribution to water pollution through continuous accumulation

of sediments in nearby river systems. However, the results of GWR in this study indicated that these lands could explain most of the spatial variations of most of the water quality parameters, including DO, BOD, NH₃-N, and TEM in 2015 and 2010. Thus, this explanatory power was different not only over time and space but also among the different water quality parameters.

Conclusion

This study analysed the influence of four predictors of water quality, including forest and agricultural, urban, and other lands, on seven water quality parameters in the Selangor River watershed for three different periods. By performing the analysis within the potential pollution zones, the results of the Pearson correlation and OLS, as well as the GWR results, showed consistency regarding interactions between land use and water quality indicators.

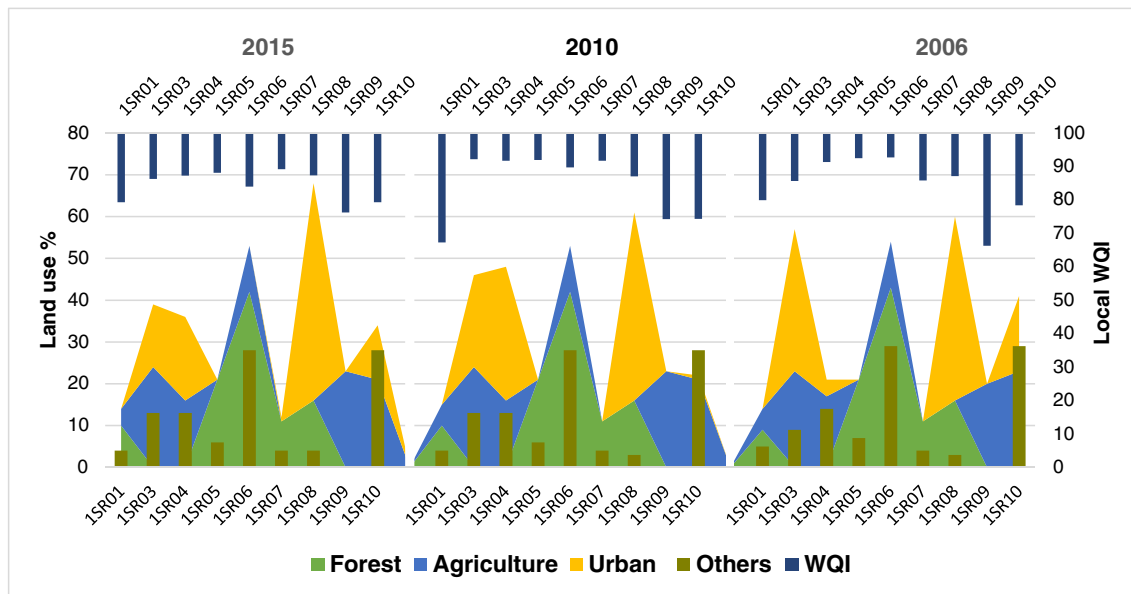


Fig. 6 Temporal evolution of WQI in relation to land use changes over the monitoring stations.

The performance of the OLS and GWR models was compared using the average adjusted R^2 values of all the land use indicators; the results showed considerable improvement in GWR R^2 over the OLS R^2 for each water quality variable for all the three different periods. The performance of the GWR models over OLS models was also demonstrated by measuring the model goodness-of-fit based on AICc values. In this regard, there are many significant improvements in the fit of the GWR models compared to the OLS models. In addition, the capacity of the two models in dealing with spatial autocorrelation was compared; the results of Global Moran's I indicated that the GWR models increase the reliability of relationships by minimising spatial autocorrelations in the model residuals.

To examine the temporal variations in relationship between each land use and water quality pollutants, the local R^2 values from the GWR models were mapped at 1SR 9 regular monitoring stations for clear visualisation of their varying temporal relationship across the river basin. The proportion of predictive ability of the same land use type varied not only with time and space but also among the water quality parameters. In addition, to compare spatial relationships between land use and water quality indicators for the different years, the local R^2 values from the GWR models were equally mapped. The results indicated that the ability of land use indicators to explain water quality change was not the same in time and space and neither the same among the different water quality parameters. GWR can be a useful tool for water resource managers and multi-decision-makers to discover the causes of local pollution on a temporal and spatial scale, to improve understanding of the state of local pollution and to adopt appropriate land use planning and management policies to ensure the sustainability of the local watershed. However, this study

recommends that future research considers more land use and water quality variables, as well as more sampling sites. The study also suggests that future studies take into account topography, precipitation, surface flow, and the nature of pollutants to identify areas of pollution or contributing area, as these factors, in addition to distance, may also alter the influence level of pollution to rivers. Future studies should also identify the best measures to prevent a serious threat to the availability of water resources in the basin, which could promote the sustainable management of water resources within the framework of the State's access to sustainable development, which involves the achievement of environmental, economic, and social sustainability measures.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12517-021-06596-4>.

Acknowledgements The authors thank the Departments of Environment and Agriculture of Malaysia for providing the water quality and land use data of the Selangor River basin used in this study. We thank the Department of Environmental Sciences of the Faculty of Environmental Studies of Universiti Putra Malaysia (UPM) for financial support to this research.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Alilou H, Moghaddam Nia A, Keshtkar H, Han D, Bray M (2018) A cost-effective and efficient framework to determine water quality

- monitoring network locations. *Sci Total Environ* 624:283–293. <https://doi.org/10.1016/j.scitotenv.2017.12.121>
- Anderson, J. (1976). A land use and land cover classification system for use with remote sensor data. https://books.google.com/books?hl=en&lr=&id=dE-ToP4UpSIC&oi=fnd&pg=PA1&dq=A+Land+Use+and+Land+Cover+Classification+System+for+Use+with+Remote+Sensor+Data.&ots=sXlq_Y0gcC&sig=5W0NmUMEnh-PdMUN0PAIM_IHIF4. Accessed 14 October 2018
- Ariffin SM, Zawawi MAM, Man HC (2016) Evaluation of groundwater pollution risk (GPR) from agricultural activities using DRASTIC model and GIS. *IOP Conf Ser Earth Environ Sci* 37:012078. <https://doi.org/10.1088/1755-1315/37/1/012078>
- Azyana YN, Norulaini NA, Jannah N (2012) Land use and catchment size/scale on the water quality deterioration of Kinta River, Perak, Malaysia. *Malaysian J Sci* 31(2):121–131. <https://doi.org/10.22452/mjs.vol31no2.4>
- Bahar MM, Ohmori H, Yamamuro M (2008) Relationship between river water quality and land use in a small river basin running through the urbanizing area of Central Japan. *Limnology* 9(1):19–26. <https://doi.org/10.1007/s10201-007-0227-z>
- Baker A (2003) Land use and water quality. *Hydrol Process* 17(12):2499–2501. <https://doi.org/10.1002/hyp.5140>
- Brunsdon C, Fotheringham AS, Charlton M (2002) Geographically weighted summary statistics—a framework for localised exploratory data analysis. *Comput Environ Urban Syst* 26(6):501–524. [https://doi.org/10.1016/S0198-9715\(01\)00009-6](https://doi.org/10.1016/S0198-9715(01)00009-6)
- Buck O, Niyogi DK, Townsend CR (2004) Scale-dependence of land use effects on water quality of streams in agricultural catchments. *Environ Pollut* 130(2):287–299. <https://doi.org/10.1016/J.ENVPOL.2003.10.018>
- Buyong, T. (2006). Spatial statistics for geographic information science. <http://eprints.utm.my/id/eprint/30002/>. Accessed 25 August 2018
- Camara M, Jamil NR, Abdullah A, Bin F (2019) Impact of land uses on water quality in Malaysia: a review. *Ecol Process* 8(1):10. <https://doi.org/10.1186/s13717-019-0164-x>
- Chen Q, Mei K, Dahlgren RA, Wang T, Gong J, Zhang M (2016) Impacts of land use and population density on seasonal surface water quality using a modified geographically weighted regression. *Sci Total Environ* 572:450–466. <https://doi.org/10.1016/j.scitotenv.2016.08.052>
- Chowdhury MDS u, Othman F, Jaafar WZW, N. C. M. & M. I (2018) Assessment of pollution and improvement measure of water quality parameters using scenarios modeling for Sungai Selangor Basin. *Sains Malays* 47(3):457–469. <https://doi.org/10.17576/jsm-2018-4703-05>
- Clark SD (2007) Estimating local car ownership models. *J Transp Geogr* 15(3):184–197. <https://doi.org/10.1016/j.jtrangeo.2006.02.014>
- Department of Agriculture Malaysia (DOA). (2019). Land use maps of Selangor River basin. <http://www.doa.gov.my/>. Accessed 25 October 2020
- Department of Environment. (2007). Malaysia Environmental Quality Report 2006. <https://environment.com.my/wp-content/uploads/2016/05/River.pdf>. Accessed 23 August 2018
- Diami SM, Kusin FM, Madzin Z (2016) Potential ecological and human health risks of heavy metals in surface soils associated with iron ore mining in Pahang, Malaysia. *Environ Sci Pollut Res* 23(20):21086–21097. <https://doi.org/10.1007/s11356-016-7314-9>
- Do HT, Lo S-L, Chiueh P-T, Thi LAP, Shang W-T (2011) Optimal design of river nutrient monitoring points based on an export coefficient model. *J Hydrol* 406(1–2):129–135. <https://doi.org/10.1016/j.jhydrol.2011.06.012>
- Erfanian M, Alijanpour A (2016) Modeling the effects of land use on water quality parameters using OLS and GWR multivariate regression methods in Fars Province watersheds. *J Environ Stud*:353–373 <https://www.researchgate.net/publication/317669391>. Accessed 30 October 2019
- Fotheringham AS, Chris Brunsdon MC (2003) *Geographically weighted regression: the analysis of spatially varying relationships*, vol 86. John Wiley & Sons., pp 554–556. https://doi.org/10.1111/j.0002-9092.2004.600_2.x
- Fulazzaky MA, Seong TW, Masirin MIM (2010) Assessment of water quality status for the Selangor River in Malaysia. *Water Air Soil Pollut* 205(1–4):63–77. <https://doi.org/10.1007/s11270-009-0056-2>
- Gazzaz NM, Yusoff MK, Aris AZ, Juahir H, Ramli MF (2012) Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. *Mar Pollut Bull* 64(11):2409–2420. <https://doi.org/10.1016/j.marpolbul.2012.08.005>
- Gomez-Gonzalez MA, Voegelin A, Garcia-Guinea J, Bolea E, Laborda F, Garrido F (2016) Colloidal mobilization of arsenic from mining-affected soils by surface runoff. *Chemosphere* 144:1123–1131. <https://doi.org/10.1016/j.chemosphere.2015.09.090>
- Harries K (2006) Extreme spatial variations in crime density in Baltimore County, MD. *Geoforum* 37(3):404–416. <https://doi.org/10.1016/j.geoforum.2005.09.004>
- Huang J, Huang Y, Zhang Z (2014) Coupled effects of natural and anthropogenic controls on seasonal and spatial variations of river water quality during baseflow in a coastal watershed of Southeast China. *PLoS ONE* 9(3):e91528. <https://doi.org/10.1371/journal.pone.0091528>
- Huang J, Huang Y, Pontius RG, Zhang Z (2015) Geographically weighted regression to measure spatial variations in correlations between water pollution versus land use in a coastal watershed. *Ocean Coast Manag* 103:14–24. <https://doi.org/10.1016/j.ocecoaman.2014.10.007>
- Ishizawa H, Stevens G (2007) Non-English language neighborhoods in Chicago, Illinois: 2000. *Soc Sci Res* 36(3):1042–1064. <https://doi.org/10.1016/j.ssresearch.2006.06.005>
- Kusin FM, Muhammad SN, Zahar MSM, Madzin Z (2016) Integrated River Basin Management: incorporating the use of abandoned mining pool and implication on water quality status. *Desalin Water Treat* 57(60):29126–29136. <https://doi.org/10.1080/19443994.2016.1168132>
- Kusin FM, Azani NNM, Hasan SNMS, Sulong NA (2018) Distribution of heavy metals and metalloid in surface sediments of heavily-mined area for bauxite ore in Pengerang, Malaysia and associated risk assessment. *Catena* 165:454–464. <https://doi.org/10.1016/j.catena.2018.02.029>
- Li S, Gu S, Liu W, Han H, Zhang Q (2008) Water quality in relation to land use and land cover in the upper Han River Basin, China. *CATENA* 75(2):216–222. <https://doi.org/10.1016/J.CATENA.2008.06.005>
- Mehaffey MH, Nash MS, Wade TG, Ebert DW, Jones KB, Rager A (2005) Linking land cover and water quality in New York City's water supply watersheds. *Environ Monit Assess* 107(1–3):29–44. <https://doi.org/10.1007/s10661-005-2018-5>
- Mohammadi M, Khaledi Darvishan A, Bahramifar N (2019) Spatial distribution and source identification of heavy metals (As, Cr, Cu and Ni) at sub-watershed scale using geographically weighted regression. *Int Soil Water Conserv Res* 7:308–315. <https://doi.org/10.1016/j.iswcr.2019.01.005>
- Nainar A, Bidin K, Walsh RPD, Ewers RM, Reynolds G (2017) Effects of different land-use on suspended sediment dynamics in Sabah (Malaysian Borneo)-a view at the event and annual timescales. *11(1):79–84*. <https://doi.org/10.3178/hrl.11.79>
- Nkeki FN, Asikhia MO (2019) Geographically weighted logistic regression approach to explore the spatial variability in travel behaviour and built environment interactions: accounting simultaneously for demographic and socioeconomic characteristics. *Appl Geogr* 108:47–63. <https://doi.org/10.1016/j.apgeog.2019.05.008>
- Nurhidayu, S., Faizalhakim, M., & Shafuan, A. (2016). Long-term sediment pattern of the Selangor River Basin, Malaysia impacted by

- land-use and climate changes. <https://www.researchgate.net/publication/304139457>. Accessed 30 September 2018
- Sakai N, Alsaad Z, Thuong NT, Shiota K, Yoneda M, Ali Mohd M (2017) Source profiling of arsenic and heavy metals in the Selangor River basin and their maternal and cord blood levels in Selangor State, Malaysia. *Chemosphere* 184:857–865. <https://doi.org/10.1016/j.chemosphere.2017.06.070>
- Sivertun Å, Prange L (2003) Non-point source critical area analysis in the Gisselö watershed using GIS. *Environ Model Softw* 18(10):887–898. [https://doi.org/10.1016/S1364-8152\(03\)00107-5](https://doi.org/10.1016/S1364-8152(03)00107-5)
- Stutter MI, Langan SJ, Demars BOL (2007) River sediments provide a link between catchment pressures and ecological status in a mixed land use Scottish River system. *Water Res* 41(12):2803–2815. <https://doi.org/10.1016/j.watres.2007.03.006>
- Tong STY, Chen W (2002) Modeling the relationship between land use and surface water quality. *J Environ Manag* 66(4):377–393. <https://doi.org/10.1006/JEMA.2002.0593>
- Tu J (2011a) Spatial and temporal relationships between water quality and land use in northern Georgia, USA. *J Integr Environ Sci* 8(3): 151–170. <https://doi.org/10.1080/1943815X.2011.577076>
- Tu J (2011b) Spatially varying relationships between land use and water quality across an urbanization gradient explored by geographically weighted regression. *Appl Geogr* 31(1):376–392. <https://doi.org/10.1016/j.apgeog.2010.08.001>
- Tu J, Xia Z (2008) Examining spatially varying relationships between land use and water quality using geographically weighted regression I: model design and evaluation. *Sci Total Environ* 407(1):358–378. <https://doi.org/10.1016/j.scitotenv.2008.09.031>
- Wang C, Du S, Wen J, Zhang M, Gu H, Shi Y, Xu H (2017) Analyzing explanatory factors of urban pluvial floods in Shanghai using geographically weighted regression. *Stoch Env Res Risk A* 31(7):1777–1790. <https://doi.org/10.1007/s00477-016-1242-6>
- Yu D, Shi P, Liu Y, Xun B (2013) Detecting land use-water quality relationships from the viewpoint of ecological restoration in an urban area. *Ecol Eng* 53:205–216. <https://doi.org/10.1016/j.ecoleng.2012.12.045>