

DOI: 10.13671/j.hjkxxb.2015.0013

陈强, 朱慧敏, 何溶, 等. 2015. 基于地理加权回归模型评估土地利用对地表水质的影响[J]. 环境科学学报, 35(5): 1571-1580

Chen Q, Zhu H M, He R, et al. 2015. Evaluating the impacts of land use on surface water quality using geographically weighted regression[J]. Acta Scientiae Circumstantiae, 35(5): 1571-1580

基于地理加权回归模型评估土地利用对地表水质的影响

陈强¹, 朱慧敏¹, 何溶¹, Randy A. Dahlgren², 张明华^{1,2,*}, 梅琨^{1,*}

1. 温州医科大学浙南水科学研究院, 浙江省流域水环境与健康风险重点实验室, 温州 325035

2. 加州大学戴维斯分校农业与环境科学学院, 陆地、大气与水资源系, 美国戴维斯 CA 95616

收稿日期: 2014-11-21 修回日期: 2015-01-10 录用日期: 2015-01-10

摘要: 针对传统线性回归模型大多忽视空间数据局部变化特征这一缺陷, 引入地理加权回归模型(GWR)用于评估土地利用对地表水质的影响, 分析了不同子流域内两者关系出现空间变化的规律并阐释了原因. 同时, 对比了 GWR 模型与普通最小二乘模型(OLS)的校正 R^2 、Akaike 信息准则(AIC_c)及残差的空间自相关指数(Moran's I), 验证了 GWR 模型在预测精度和处理空间自相关过程中是否优于 OLS 模型. 结果表明, 同一土地利用类型对水质的影响随空间位置的改变而发生方向或大小的变化. 以温瑞塘河流域总氮(TN)与农用地的关系为例, 从 GWR 模型局部回归系数的方向分析, 两者关系表现为农村正、城区负的现象, 从大小分析, 旧城区 TN 与农用地回归系数的绝对值高于其它区域; 在溶解氧(DO)与人口密度所构建的 GWR 模型中, 两者关系在整个研究区域内均表现为负值, 与 OLS 结果吻合, 从回归系数的大小分析, 人口密度对 DO 的作用在郊区及农村更为显著. 针对此类关系出现空间变化的原因分析表明, 相邻子流域土地利用百分比的改变及水体主要污染源的不同, 是导致土地利用对水质作用发生变化的根本因素. 最后, 对比所构建的 80 个 GWR 与 OLS 模型校正 R^2 、AIC_c 指标, 验证了 GWR 作为一种局部统计模型, 其预测精度优于 OLS 等传统全局模型且更能反映实际空间特征.

关键词: 地理加权回归; 土地利用; 水质; OLS; 温瑞塘河

文章编号: 0253-2468(2015)05-1571-10 中图分类号: X32 文献标识码: A

Evaluating the impacts of land use on surface water quality using geographically weighted regression

CHEN Qiang¹, ZHU Huimin¹, HE Rong¹, Randy A. Dahlgren², ZHANG Minghua^{1,2,*}, MEI Kun^{1,*}

1. Southern Zhejiang Water Research Institute (iWATER), Wenzhou Medical University, Key Laboratory of Watershed Science and Health of Zhejiang Province, Wenzhou 325035

2. Department of Land, Air and Water Resources, College of Agricultural and Environmental Sciences, University of California Davis, CA 95616, USA

Received 21 November 2014; received in revised form 10 January 2015; accepted 10 January 2015

Abstract: Most traditional linear regression models ignore local variations of spatial data. In this study, a new technique called geographically weighted regression model (GWR) was introduced to evaluate the impacts of land use on surface water. The reason for the spatial variations of relationships between land use and water quality were explored. Meanwhile, the adjusted R^2 , Akaike information criterion (AIC_c) and spatial autocorrelation index (Moran's I) of residuals were compared with ordinary least squares model (OLS) to verify if the GWR model is better than OLS in the prediction accuracy and the capacity of conducting spatial autocorrelation. The results showed that impact of the same types of land use on water quality changes in direction or size along with the variation of spatial position. For example, the relationships between TN and agricultural land in Wen-Rui Tang River showed a positive correlation in countryside and negative correlation in urban area in GWR models. The absolute values of regression coefficients in old downtown area were higher than other places. In the GWR model of dissolved oxygen (DO) and population density, the relationships were negative in the whole study area, which was consistent with the OLS results, and the effect of population density on DO is greater in suburban and rural areas. The reason for these spatial changes over different sub-watersheds indicated that the changes of land use percentage and the varied main pollution sources are fundamental factors. Furthermore, the adjusted R^2 and AIC_c values from the 80 established models confirmed that as a local statistical model, GWR had better prediction

基金项目: 浙江省科技厅重大专项(No. 2008C03009); 温州市地方专项(No. KJXH1347)

Supported by the Zhejiang Province Technology Office Major Projects (No. 2008C03009) and the Wenzhou Local Major Projects (No. KJXH1347)

作者简介: 陈强(1990—), 男, E-mail: chenqiang0107@126.com; * 通讯作者(责任作者), E-mail: kunjuly@wmu.edu.cn, mhzhang@ucdavis.edu

Biography: CHEN Qiang(1990—), male, E-mail: chenqiang0107@126.com; * **Corresponding author**, E-mail: kunjuly@wmu.edu.cn, mhzhang@ucdavis.edu

accuracy than OLS model and could better reflected the actual spatial characteristics.

Keywords: geographically weighted regression; land use; water quality; OLS; Wen-Rui Tang River

1 引言 (Introduction)

相对于点源污染,非点源污染由于其自身的隐蔽性和难以量化性,备受研究人员的关注并得到了广泛研究(Dowd *et al.*, 2008; Zhang *et al.*, 2011; 白凤娇等, 2012; 刘亚群, 2014). 而作为坡面径流、壤中流等输送污染物的重要载体,土地利用则成为研究非点源污染与地表水质关系的关键对象.

纵观土地利用对地表水质影响的定量分析研究,从分析方法上看,应用最为广泛的是基于统计学的方法,包括相关分析、聚类分析、主成分分析、线性回归模型、指数模型等.例如,官宝红等(2008)、孟祥巍等(2014)分别利用相关分析、普通线性回归模型评价了不同土地利用类型对水质影响的显著性;曹芳芳等(2013)通过 Spearman 相关分析、指数模型定量计算出土地利用结构与水质指标的关系,并利用层次聚类法,研究了流域内水质指标的相似性分布;Ahearn 等(2005)则采用优化的线型混合效应模型,讨论了不同水文年及上下游地区土地利用、人口密度对水质的影响;Tran 等(2010)在分析河流不同邻近度下土地利用类型对水质的影响时,用到了主成分分析、相关分析及断点回归等技术;Wang 等(2013)运用聚类分析及逐步线性回归模型,探讨了不同人为干扰梯度下,土地利用与水质的关系.该类研究结果普遍表明,与人类活动关系密切的土地利用类型,如农用地、城镇建设用地、商业用地、工矿用地等,通常与河流水质污染物(如总氮、总磷、氨氮等)浓度呈正相关关系,而受人类活动影响较小的林地、草地则多与水体污染物浓度呈负相关关系.

以上传统统计方法如线性回归模型、指数模型的应用均基于同一个假设前提:即在整个研究区域内,水质指标与土地利用类型的关系是保持稳定的,不随空间位置改变而发生变化.这是一种全局统计的方法,回归方程在很大程度上表现为现有关系混合的平均值.而实际中,随着研究区域的不同,同一土地利用类型对水质指标的影响也在发生变化.例如,王瑶等(2012)在关于北京市土地利用结构与水质关系的研究中发现,农用地与高锰酸盐指数、氨氮浓度呈正相关;而在孙金华等(2011)关于滇池流域的研究中发现,农用地与 TN、氨氮呈负相关.实

际上,水质与土地利用往往表现出局部变化特征,即使在同一流域的不同位置,其相应社会经济特征、主要土地利用方式也会不同,导致影响该局部地区水质的第一因素发生改变,同一种土地利用类型对水质的影响有时甚至会呈现相反的关系.此时若依然采用全局模型,将掩盖可能具有意义的局部特征. Williams 等(2005)的研究结果表明,在美国 Massachusetts 东部,农用地与水质指标 Ca、Na、Mg 等表现为显著正相关;而针对同一地区, Tu 等(2008)运用普通最小二乘模型的研究结果却发现,农用地与这 3 种指标显著负相关.可见,使用全局统计模型所得结果有时并不可靠.

地理加权回归(Geographically Weighted Regression, GWR)模型(Fotheringham *et al.*, 1996; 2002)作为一种局部统计方法,被 Tu 等(2008)首次引入到评估土地利用与水质关系的研究中,并得到了优于传统线性回归模型(如 Ordinary Least Squares, OLS)的结果.该模型通过将数据的空间位置嵌入到回归参数之中,以探索自变量与因变量在空间范围内的变化关系,有效地解决了空间非平稳性(Fotheringham *et al.*, 2002)的问题;同时,针对传统统计方法中难以处理的空间数据自相关性(Brown *et al.*, 2012), GWR 模型也进行了考虑.作为一门新兴技术, GWR 模型已被应用到多个领域,如疾病健康研究(Nakaya *et al.*, 2005)、犯罪频发地区调查(Cahilland *et al.*, 2007)、森林损失评估(Pineda *et al.*, 2010)、地下水受土地利用影响评价(Javi *et al.*, 2014)等.

然而,在以往应用 GWR 模型分析土地利用对地表水质影响的研究中(Tu *et al.*, 2008; Tu, 2013; Pratt *et al.*, 2012),侧重点往往是对比 GWR 模型相对其它模型的优势及定量描述 GWR 结果,却很少有对 GWR 回归结果出现空间变化的具体原因进行阐述.为此,本研究基于 GWR 模型,采用子流域分析的方式,同时结合水质污染空间分布特征来探讨不同土地利用类型对地表水质的影响,主要尝试解决以下问题:①在不同子流域或测站点,土地利用类型对水质影响的空间分布特征如何,在何处正相关,何处负相关? 何处影响大,何处影响小? 并对此现象进行解释;②对比 GWR 模型与传统线性回归模型的性能,并检验 GWR 模型在处理空间自相

关中是否具备优势。

2 研究区域(Study area)

本文选取温瑞塘河温州段为研究区域(120°27′~120°46′E, 27°51′~28°02′N),流域面积 371.1 km²(图 1)。温瑞塘河位于浙江省东南沿海瓯江和飞云江之间的温瑞平原地区,河道北起温州市鹿城区,南至瑞安市城关东门白岩桥,河网总长 1178 km,主河道长 33.9 km,水面面积 22 km²,流域总面积近 740 km²,河网密集。温瑞塘河流域属亚热带季风气候,多年平均降雨量为 1818 mm,其中,

70%的降水集中在 4—9 月的丰水期。塘河年径流量 9.1 亿 m³,对温州市农田灌溉、水产养殖、城市排洪及居民、工业用水有重要作用。

根据河网分布和流域地形特征,将研究区域划分为 7 个流域盆地。其中,盆地 I 被山地环绕,土地利用类型以灌木林、果林主导的林地及农用地为主;II、III、IV、V 区北部是温州市主城区,主要用地为城市建设用地等,南部为三垟湿地,湿地陆域面积以人工栽培果木和城镇建设用地为主;VI、VII 区主要为林地和农田,此处河网多经帆游(丽岙镇)进入瑞安市境内,向南汇流至飞云江并最终注入东海。

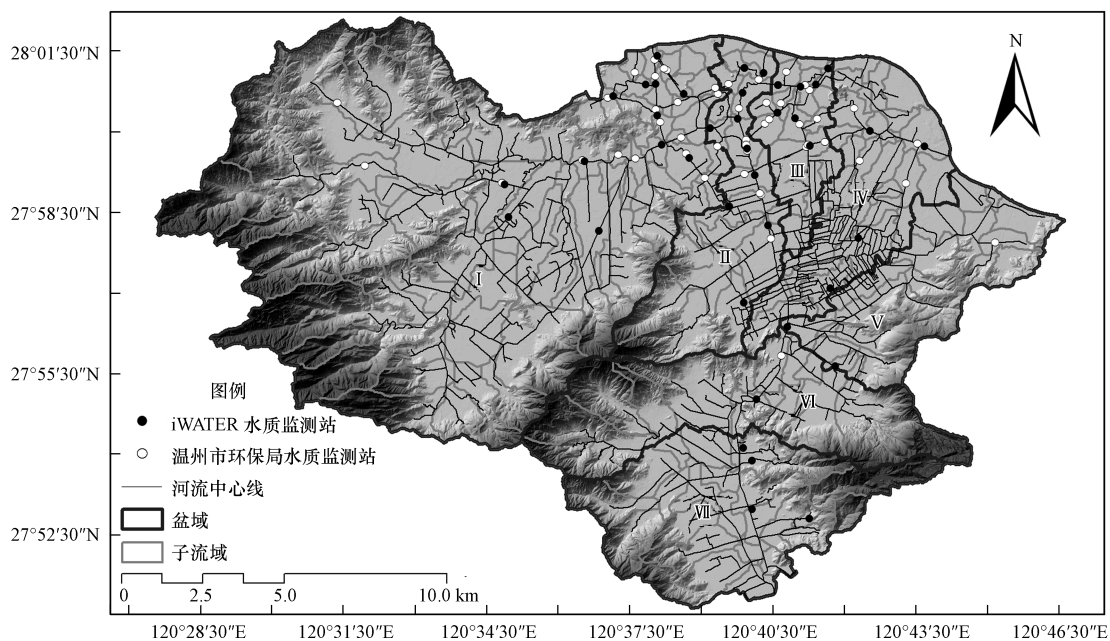


图 1 温瑞塘河流域温州段水质监测点布设及盆域划分

Fig.1 Location of monitoring sites and basin area in Wen-Rui Tang River

截至 2013 年底,温州市常住人口达 919.7 万,10 年内增长了 18.7%,国民经济三次产业结构比为 2.9:50.3:46.8。根据温州市水利局《2012 年温州市水资源公报》统计,2012 年温州平原河网水质达到地表水环境质量 III~IV 类标准的河长为 31.8 km,占评价总河长的 16%,其余断面水质皆为劣 V 类。Mei 等(2014)通过对该地区 11 年水质时空变化特征的分析发现,氮污染是温瑞塘河流域最主要的污染问题,温瑞塘河水污染现状不容乐观。

3 数据采集和方法(Data collection and methods)

3.1 水质数据

研究共选取 70 个水环境监测断面数据进行分析,分析时段为 2008—2010 年。其中,30 个断面由温

州市环保局所设,采样间隔为 4~8 周;其余 40 个断面由本研究中心所设,每月进行 1 次采样,用以对数据进行补充。入选水质指标包括 pH、溶解氧(DO)、总氮(TN)、氨氮(NH₄⁺-N)及电导率(EC)。所有采集水样均按照国家《地表水环境质量标准》(GB3838—2002)和《水和废水监测分析方法》进行测定与分析。

分别统计每个监测断面不同水质指标实测结果的年均值,并针对每种水质指标进行 Kolmogorov-Smirnov 检验,为保证后期模型分析过程的准确性,采用自然对数变换的方法对数据进行处理。

3.2 土地利用及子流域划分

以分辨率为 0.5 m 的温州市航空影像为数据源,将解译后的 96 种详细土地利用类型重分类为 7

种一级类别:农用地、绿地(园地、林地、草地及城市绿化带)、商服用地(商务金融用地、行政办公用地、文化娱乐用地、医疗卫生用地及市政公用设施用地等)、交通用地、工矿用地、住宅用地和水域(图2)。由于流域内未利用地如裸岩石砾地、迹地等其斑块数量极少且均处于山地,为统计方便统一合并到住宅用地。

以分辨率为5 m的温州市DEM为数据源,利用ArcGIS 10.2软件下的水文分析模块计算流域栅格的流向与流量,并提取出流量大于10000(栅格Value值,无量纲)的栅格得到河网,通过盆域分析确定出7个出水口不同的流域盆地(图1中I~VII);利用Arc Hydro Tools生成河网支流所在的集水多边形并与盆地要素叠加,消除碎屑多边形后,以流域盆地边界为基准,对面积小于10万m²且相

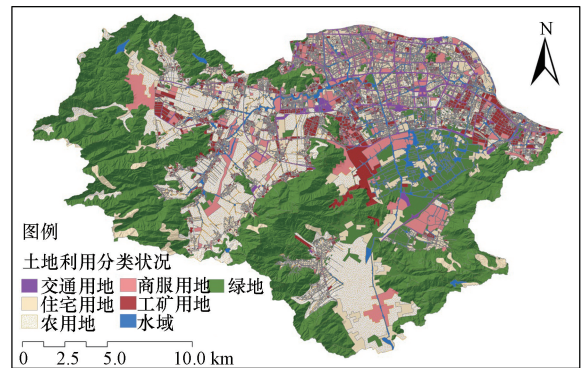


图2 研究区域内7种土地利用类型空间分布

Fig.2 Spatial distribution of the 7 types of land use in study area

互毗邻的集水多边形进行合并,得到201个子流域.最后,利用ArcGIS统计分析工具计算得到每个子流域内不同土地利用类型的面积及百分比(图3).

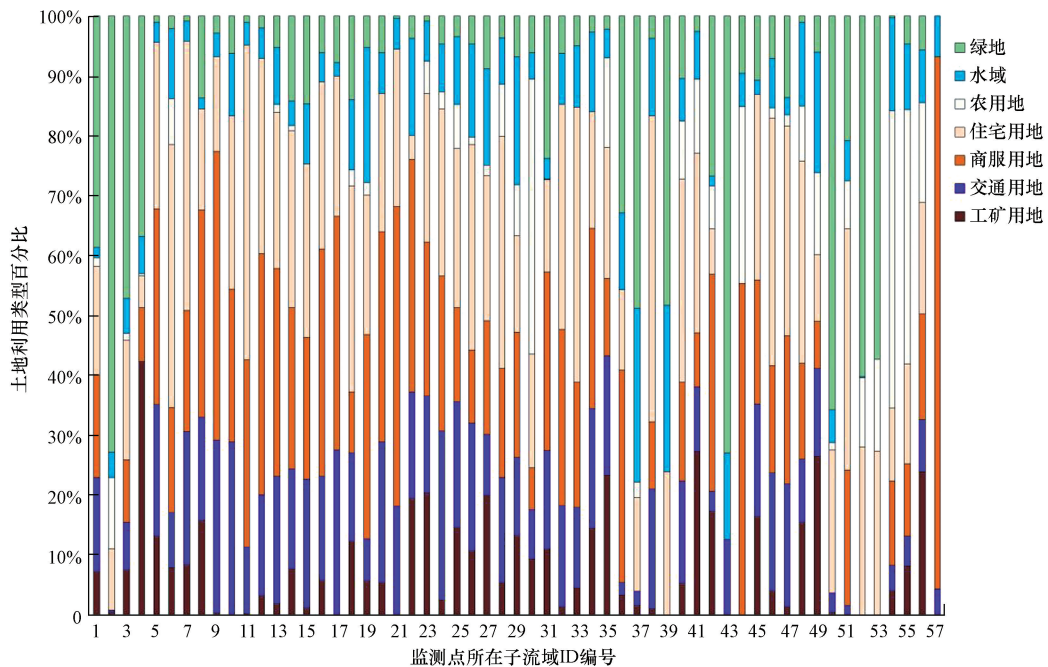


图3 监测站所在的57个子流域土地利用结构比较

Fig.3 Comparison of land use structure in the 57 sub-watershed where monitoring sites are located

3.3 人口密度

原始数据为温州市每个行政区划(街道为单位)下的2010年第六次全国人口普查数据,为了得到每个测站点的人口密度,采用如下方法:①查找每个行政区划的质心位置,赋予其对应区划的人口数;②采用Kernel算法(牟乃夏等,2012),由质心点内插出人口密度栅格图;③提取各测站所对应栅格的Value值,即得到该点人口密度。

3.4 模型构建方法

传统回归模型如普通线性回归模型假定回归参数在整个研究区域保持稳定一致,与样本的空间位置无关,回归参数的估计通常采用最小二乘估计方法求解极值得到.作为一种全局统计,该方法往往忽略了不同地区回归参数的变异性.地理加权回归模型作为普通线性回归模型的扩展,将数据的空间位置嵌入到回归参数之中(Fotheringham *et al.*, 2002),通过求得局部参数评估自变量与因变量关

系在空间尺度上的变动,是一种局部统计模型.其基本形式为:

$$y_j = \beta_0(u_j, v_j) + \sum_{i=1}^p \beta_i(u_j, v_j) \chi_{ij} + \varepsilon_j \quad (1)$$

式中, (u_j, v_j) 为第 j 个采样点的坐标, $\beta_i(u_j, v_j)$ 为自变量 χ_i 在 j 点处的局部回归系数, $\beta_0(u_j, v_j)$ 及 ε_j 分别为模型在 j 点处的回归常数和误差项. 针对回归参数 $\beta_i(u_j, v_j)$ 的估计, 可通过式 (2) 达到最小来估计 (Brunsdon *et al.*, 1998; 覃文忠, 2007).

$$\sum_{k=1}^n w_{jk} \left(y_k - \beta_0(u_j, v_j) - \sum_{i=1}^p \beta_i(u_j, v_j) \chi_{ij} \right)^2 \quad (2)$$

式中, w_{jk} 为回归点 j 与其它已知观测点 k 之间的距离衰减函数, 其基本假设为距离 j 点越近的观测点对求解局部回归参数的重要性越大, 越远的观测点重要性越小. w_{jk} 作为空间权重矩阵, 是地理加权回归模型的核心, 计算方法包括距离阈值法、Gauss 函数法及 bi-square 函数法等 (覃文忠, 2007), 其中, Gauss 空间权函数因其普适性而得到广泛应用, 本研究即采用此方法, 函数形式为:

$$w_{jk} = \exp(-d_{jk}^2/b^2) \quad (3)$$

式中, d_{jk} 为观测点 j, k 之间的距离; b 为描述权重与距离之间函数关系的非负衰减参数, 也称为带宽 (Fotheringham *et al.*, 2002), 当观测点间的距离大于 b 值时, 权重 w_{jk} 迅速趋近于 0, 保证远距离的对回归点处的参数估计影响最小. 关于 b 的设置, ArcGIS10.2 中提供固定带宽和自适应带宽两种形式, 前者查找的是最佳距离, 后者计算的则为最佳邻近点个数.

为避免由于多重共线性而导致 GWR 模型出现设计错误, 本研究选取单独一种土地利用类型作为解释变量, 分别和不同水质指标构建普通最小二乘模型 (OLS) 与 GWR 模型, 两者分别由 SPSS 及 ArcGIS10.2 构建, 共得到 80 个模型. 在带宽的选择上, 与文献 (Tu *et al.*, 2008; Tu, 2013) 所用自适应带宽不同, 本研究通过对比验证, 发现固定带宽法在构建温瑞塘河流域的 GWR 模型中具备更好的优势.

3.5 模型评价

针对 GWR 和 OLS 模型的评价, 本文主要从模型精度和处理空间自相关能力两方面进行分析. 首先, 选取校正 R^2 及 AIC_c (Akaike 信息准则) (Tu *et al.*, 2008) 两个指标对两模型进行对比, 以确定 GWR 模型是否优于普通线性回归模型: R^2 越大表明

自变量对因变量的解释程度越高, AIC_c 作为评价模型精确度及复杂性的指标, 取值越低表明模型预测能力越好, 通常两模型 AIC_c 的差值大于 3, 则可认为具有较低 AIC_c 的模型更佳 (Brown *et al.*, 2012).

其次, 根据水质实测值与回归值的残差分别计算两模型的全局 Moran's I 指数值, 以对比 GWR 和 OLS 模型处理空间自相关的能力. 如果模型较好, 则其偏高估计值和偏低估计值将随机分布; 残差若存在空间聚类, 则表明模型可能丢失了某个关键解释变量. 在 GIS 中, Moran's I 是反应空间数据聚散与否的有效指标, 其计算公式如下:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{i,j} (x_i - \bar{X}) (x_j - \bar{X})}{\sum_{i=1}^n (x_i - \bar{X})^2} \quad (4)$$

$$S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{i,j} \quad (5)$$

$$w_{i,j} = 1/d_{i,j} \quad (6)$$

式中, x_i, x_j 分别为要素 i 和邻近要素 j 的值, $w_{i,j}$ 是两者之间的空间权重, 此处定义为要素 i 和 j 之间距离 d 的倒数; n 是要素的个数, S_0 是所有空间权重的聚合. 全局 Moran's I 从 -1 到 1 进行取值, 在统计显著时, 若 I 为 0 则数据集为随机分布; I 为正, 表明存在聚类趋势的空间自相关, I 为负表示离散趋势, I 值越接近于 0, 表明模型处理空间自相关的能力越强. 本研究中, 校正 R^2 、 AIC_c 及全局 Moran's I 均在 ArcGIS 10.2 中分析并进行显著性检验.

4 结果与讨论 (Results and discussion)

4.1 水质与土地利用及人口密度相关分析

针对研究区域内 70 个监测站, 计算其对应子流域 (部分测站位于同一子流域内) 土地利用与不同水质指标的 Spearman 相关系数 (表 1). 由表 1 可知, 交用地、商服用地及人口密度均与 DO 显著负相关, 前两者又与 TN、 $\text{NH}_4^+\text{-N}$ 显著正相关; 水域则与 TN、 $\text{NH}_4^+\text{-N}$ 显著负相关, 这可能是由于随着水体面积的增加, 其对水质污染物起到了一定的稀释作用; 农用地与 TN、 $\text{NH}_4^+\text{-N}$ 、EC 均表现为负相关, 表明随着农用地的增多河流水质也在提高, 这与很多同类研究结果不同 (王娇等, 2012; 曹芳芳等, 2013). 孙金华等 (2011) 对此现象解释为工矿用地等城市建设用地掩盖了农用地本身对河流的污染贡献. 然而, 具体原因有待进一步说明.

表 1 水质指标与土地利用、人口密度的 Spearman 相关分析

Table 1 Spearman correlations between water quality and land use, population density indicators

	交通用地	住宅用地	农用地	商服用地	工矿用地	水域	绿地	人口密度
DO	-0.428**	-0.061	0.221	-0.363**	-0.086	0.220	0.125	-0.390**
TN	0.312**	0.071	-0.393**	0.288*	-0.012	-0.360**	-0.085	0.039
NH ₄ ⁺ -N	0.342**	0.074	-0.420**	0.280*	-0.036	-0.324**	-0.054	-0.010
pH	0.378**	0.447**	-0.190	0.090	-0.058	0.157	-0.261*	0.023
EC	0.058	0.211	-0.262*	0.015	-0.118	-0.075	0.000	-0.298*

注: * $p < 0.05$, ** $p < 0.01$.

4.2 GWR 与 OLS 模型结果分析

由于 GWR 模型在局部回归系数的计算中对相邻要素赋予了一定权重,因此,在分析系数变化时必须考虑周围子流域的情况.受篇幅所限,此处未全部列出 40 个 GWR 模型结果,仅挑选出具有代表性且显著相关的 TN 与农用地、TN 与交通用地、DO 与人口密度共 3 组模型进行详细讨论,对应 OLS 模型如表 2 所示.

表 2 所选 3 种 OLS 模型结果

Table 2 Results of the 3 selected OLS models

因变量 y	自变量 x	OLS 方程	p
$\ln[\text{TN}]$	农用地比例	$y = 2.428 - 0.016x$	0.016
$\ln[\text{TN}]$	交通用地比例	$y = 1.935 + 0.029x$	0.001
$\ln[\text{DO}]$	人口密度	$y = 1.630 - 0.013x$	<0.001

4.2.1 TN 与农用地关系 根据 TN 浓度空间分布图(图 4),不难发现 TN 污染呈现城市高、郊区低的空间分布特征,城区又以靠北的旧城区污染为重.结合所构建的 GWR 模型结果(图 5),发现农用地对 TN 的影响并非如 OLS 模型(表 2)及 Spearman 结果所示完全表现为负相关,而显示出明显的局部特征.从回归系数的方向分析,在盆地 I 区,农用地与 TN 首先表现为正相关,随着河道进入城区,开始变为负相关,在 IV、V 区北部及 VII 区两者又表现为正相关;从回归系数的大小分析,旧城区农用地与 TN 局部回归系数的绝对值高于其它区域.

究其原因, I 区左侧监测站(瞿溪、郭溪)位处河道上游,水质污染相对较低, TN 浓度均值为 $2.3 \text{ mg} \cdot \text{L}^{-1}$, 周边相邻 10 个子流域内土地利用类型以林地(57.7%)、农用地(20.2%)及住宅用地(6.6%)为主,此处氮含量主要来自畜禽粪便及农药化肥的施用,人为耕作活动对水质起主要影响作用,因而表现出农用地与 TN 正相关的结果.而在 I 区中部位处郊区的河道, TN 浓度均值为 $7.3 \text{ mg} \cdot \text{L}^{-1}$, 周边 20 个子流域内不透水面积(包括工矿、住宅及商服用地

等)占 36.9%,其次为农用地 34.6%和绿地 22.6%;且随着河道步入城区, TN 浓度均值达 $12.1 \text{ mg} \cdot \text{L}^{-1}$, 对应测站周边 60 个子流域内以住宅用地为主,约占 26.1%,其次为商服用地和交通用地,总不透水面积比例为 73.3%,而农用地和绿地仅分别为 6.5%、12.4%,且存在部分子流域内无农用地的现象.这样,随着子流域靠向城区,对应农用地所占比例在逐渐降低,而其 TN 浓度却在增加,因此,回归系数

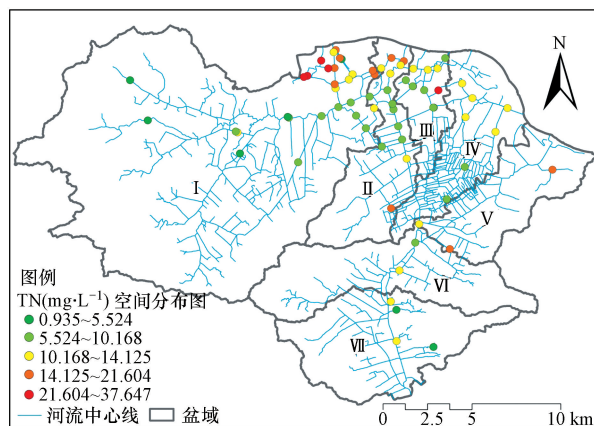


图 4 水质指标 TN 浓度空间分布专题图

Fig.4 Distribution diagram of TN concentration

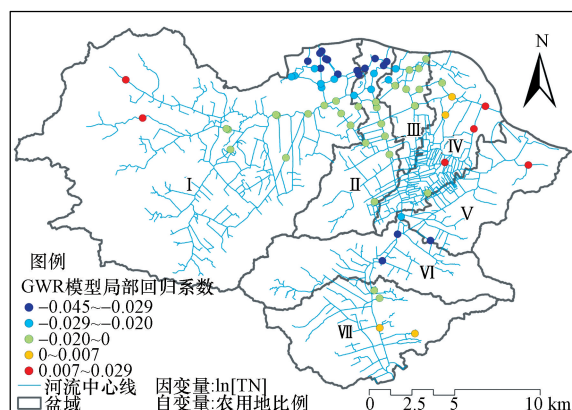


图 5 农用地与 TN 所构建 GWR 模型之局部回归系数分布图

Fig.5 Distribution diagram of local regression parameters for TN and agricultural land based on GWR

表现为负值.这也是为何 OLS 模型及 Spearman 分析结果中农用地与 TN 整体表现负相关的原因.但是,此时的负相关并不能用来反向表征农用地的增加抑制或降低了 TN 的污染,而是因为造成河流氮污染的主要污染源随空间位置发生了改变,城区不透水面积迅速增加、点源污染增多是导致 TN 浓度上升的主要原因;此外,旧城区河道狭窄、流速缓慢且底泥相对较厚, TN 浓度迅速升高,从而使得其与农用地的局部回归系数绝对值更大.

而在 IV 区北部, TN 平均浓度为 $12.5 \text{ mg} \cdot \text{L}^{-1}$, 与城区持平,且进入 V 区有增加趋势,农用地比例与此变化一致,两者呈正相关,但此时快速增加的工矿用地(21.7%)是导致 TN 浓度上升的主要原因. VI 区测站多设于塘河主河道,水质受上游城区影响显著, TN 浓度均值为 $10.8 \text{ mg} \cdot \text{L}^{-1}$, 对应农用地逐渐增加,表现出与 TN 负相关的结果. VII 区河道自北向南流入瑞安市,上游山脚处水质明显优于河道,周边土地利用变化与瞿溪、郭溪所在 I 区类似, TN 受农用地影响显著,两者又表现为正相关关系.

4.2.2 TN 与交通用地关系 在 TN 与交通用地所构建的 OLS 模型(表 2)中,两者回归系数为正值,表明交通用地的增大一定程度上加剧了水质的恶化.而在 GWR 模型中(图 6),交通用地对水质的正相关作用表现为城区影响小、郊区及农村影响大的空间变化特征,且东部工业园区(盆地 IV、V 区北部)出现负相关结果.

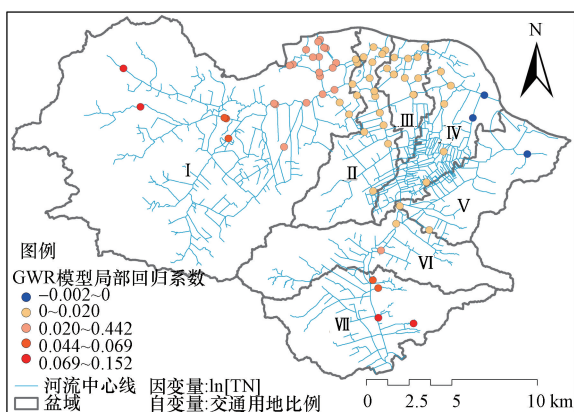


图 6 交通用地与 TN 所构建 GWR 模型之局部回归系数分布图

Fig.6 Distribution diagram of local regression parameters for TN and transportation land based on GWR

结合 TN 浓度空间分布专题图(图 4)分析可知,随着河道由 I 区自西向东进入 II、III、IV 区,并于 VII 区自北向南流出, TN 浓度及交通用地密度均呈

现先增加后减少的变化趋势,因而回归系数为正.而 IV、V 区东北部,交通用地比例由城区的 16.2%降至 11.9%,工矿用地则由 11.3%增至 21.7%,因而在比例变化上交通用地与 TN 回归系数呈现为负相关,但此处导致 TN 浓度变化的决定因素并非交通用地而是工矿用地的改变.可见,此时所构建的 GWR 模型丢失了关键解释变量.如何同时引入多个有效解释变量并同时避免多重共线问题的产生,这是本文今后所需开展的方向.

4.2.3 DO 与人口密度关系 从局部回归系数的方向分析,由 DO 与人口密度所构建的 GWR 模型结果(图 7)在整个研究区域内均表现为负值,与 OLS 结果(表 2)吻合;从回归系数的大小分析,在 GWR 模型中人口密度对 DO 的影响在郊区及农村表现的更为明显.这是因为从农村、郊区再到城市,对应测站平均人口密度由 12.8 、 $60.8 \text{ 人} \cdot \text{hm}^{-2}$ 增至 $170.9 \text{ 人} \cdot \text{hm}^{-2}$,区域间人口密度基数在增加而增涨幅度却在逐步降低,城市区域内人口密度则趋于稳定,因而在局部回归系数上表现为郊区人口密度与 DO 的回归系数数值更大.这与实际情况也是符合的,因为在城区,影响水质的因素复杂而多样,人口因素相对其它指标如城市不透水面积、污染点源等对水质的影响则表现的较弱;而在郊区及农村,水质主要受农用地、住宅用地等人为活动影响,而此两者则与人口密度息息相关.

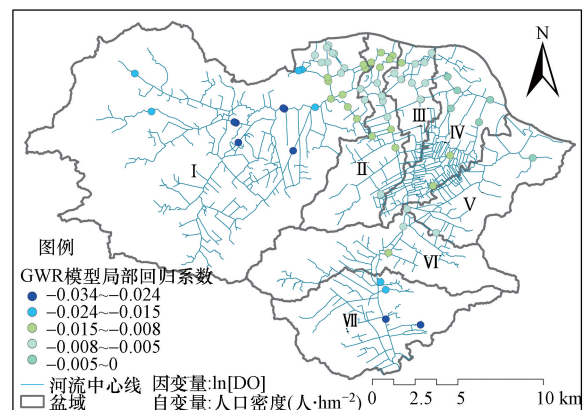


图 7 人口密度与 DO 所构建 GWR 模型之局部回归系数分布图

Fig.7 Distribution diagram of local regression parameters for DO and population density based on GWR

4.3 GWR 与 OLS 模型性能评价

查阅两模型的校正系数 R^2 、AIC_c 及全局 Moran's I 计算结果(表 3)可知,相对 OLS 模型,所有 GWR

表3 GWR与OLS两模型结果对比(校正R²、AIC_c及残差Moran's I)

Table 3 Comparison of adjusted R square, AIC_c and residuals' Moran's I from GWR and OLS models

参数	DO				TN				NH ₄ ⁺ -N				pH				EC	
	校正R ²	AIC _c	Moran's I	校正R ²	AIC _c	Moran's I	校正R ²	AIC _c	Moran's I	校正R ²	AIC _c	Moran's I	校正R ²	AIC _c	Moran's I	校正R ²	AIC _c	Moran's I
交通用地比例	OLS	0.243	133.8	0.052	0.13	127.1	0.268	0.136	154	0.178	154	0.178	0.115	-316.2	-0.005	0.013	85.4	0.140
	GWR	0.334	128.6	-0.034	0.321	113.9	0.188	0.436	128.7	0.124	128.7	0.124	0.216	-322.1	-0.043	0.159	76.2	0.128
住宅用地比例	OLS	-0.002	153.4	0.123	-0.012	137.7	0.138	-0.010	164.9	0.087	164.9	0.087	0.072	-312.9	0.048	0.023	84.6	0.085
	GWR	0.365	130.4	-0.029	0.378	112.5	0.153	0.468	128.9	0.085	128.9	0.085	0.232	-322.3	-0.045	0.172	75.7	0.100
农用地比例	OLS	0.103	145.7	0.056	0.069	131.8	0.155	0.067	159.3	0.092	159.3	0.092	0.034	-310.1	0.074	0.032	84.0	0.119
	GWR	0.330	133.3	-0.024	0.319	118.0	0.126	0.418	134.4	0.071	134.4	0.071	0.194	-319.1	-0.023	0.158	78.9	0.127
商服用地比例	OLS	0.048	149.9	0.090	0.080	131.1	0.183	0.071	159	0.111	159	0.111	-0.006	-307.2	0.089	-0.010	87.0	0.104
	GWR	0.350	133.0	-0.017	0.352	116.3	0.159	0.447	132.6	0.083	132.6	0.083	0.201	-318.9	-0.024	0.217	79.0	0.060
工矿用地比例	OLS	0.007	152.8	0.160	-0.007	137.3	0.133	-0.008	164.8	0.089	164.8	0.089	-0.013	-306.7	0.109	-0.015	87.3	0.094
	GWR	0.318	135.2	0.003	0.289	121.7	0.138	0.395	137.8	0.075	137.8	0.075	0.205	-320.3	-0.028	0.195	79.9	0.095
绿地比例	OLS	0.057	149.2	0.101	0.045	133.6	0.170	0.056	160.1	0.115	160.1	0.115	0.086	-314.0	0.019	-0.002	86.4	0.108
	GWR	0.379	129.1	0.002	0.376	112.9	0.170	0.477	128.0	0.115	128.0	0.115	0.187	-319.2	-0.034	0.123	79.7	0.109
水域比例	OLS	0.041	150.4	0.136	0.02	135.4	0.130	0.001	164.1	0.082	164.1	0.082	0.022	-309.2	0.130	-0.014	87.3	0.093
	GWR	0.496	113.6	-0.017	0.56	87.7	0.174	0.614	105.9	0.080	105.9	0.080	0.249	-325.1	-0.018	0.311	68.5	0.107
人口密度	OLS	0.243	133.8	0.024	0.099	129.6	0.189	0.108	156.2	0.124	156.2	0.124	0.077	-313.3	0.043	-0.015	87.3	0.097
	GWR	0.335	131.8	0.029	0.322	115.2	0.172	0.423	131.5	0.115	131.5	0.115	0.265	-322.2	-0.064	0.352	63.0	0.061

注:黑斜体表示所建OLS模型不具有显著性(p>0.05);黑体表示所求Moran's I指数不具有显著性(p>0.05)

模型的校正 R^2 均得到了较大的提高.以交通用地与不同水质指标所构建的模型为例,其 OLS 模型校正 R^2 取值范围为 0.013~0.243,对应 GWR 模型中取值为 0.159~0.436,平均提升幅度为 323.2%.而在以氨氮为因变量的所有模型中,其与水域面积百分比所构建的 GWR 模型校正 R^2 达 0.614,不仅远大于对应 OLS 模型结果,也高于与其它土地利用类型的 R^2 值,线性模拟能力最好.即使 OLS 模型中与土地利用类型回归系数均较小的 pH 指标,其对应的 GWR 模型的校正 R^2 也得到了提升.

作为衡量模型精确度和复杂性的重要标准, AIC_c 值越低表明模型预测结果与实际结果越接近.由表 3 不难发现,除去 DO 与人口密度所构建的模型外,其余所有 GWR 模型的 AIC_c 值都低于对应 OLS 模型结果,且差值均大于 3.在 TN 与不同土地利用所构建的模型中, AIC_c 均值由 OLS 模型中的 133.0 降低为 GWR 模型中的 112.3,平均降幅为 15.5%.综合上述两种指标的比较,可以判断,GWR 作为一种局部统计模型,其模型预测精度优于 OLS 等传统全局模型,这也验证了 Pratt 等(2012)及 Tu (2013)的研究结果.

然而,针对由残差计算的 Moran's I 指数值,对比发现所有 OLS 及 GWR 模型的指数值均不具有统计学上的显著性($p>0.05$),这表明回归残差的空间分布很可能是随机空间过程的结果,两模型均不存在明显的预测过度(高残差聚类)和预测不足(低残差聚类)现象.这一点并未表现出 Tu 等(2008)及 Javi 等(2014)文中所提出的结论,即 GWR 相对 OLS 模型具备更好的处理空间自相关能力.这主要是因为,在评比 GWR 与 OLS 处理空间自相关能力前,一个假设前提就是初始数据存在某种自相关.然而,通过对本研究中不同水质指标进行空间自相关分析(表 4)发现,研究区域内初始水质监测数据本身就不存在显著的空间聚类模式,自然也无从谈起两模型处理自相关能力的优劣.

表 4 不同水质指标空间自相关分析结果

Table 4 Spatial autocorrelation analysis of different water quality indicators

水质指标	Moran's I	p
DO	0.137	0.428
TN	0.140	0.408
NH_4^+-N	0.090	0.571
pH	0.115	0.407
EC	0.095	0.549

5 结论(Conclusions)

1) 土地利用与水质的关系往往随着空间位置的改变表现出局部变化特征.即使在同一研究区域的不同子流域,同一土地利用类型对水质的影响在大小、方向上均可能表现不同.以温瑞塘河流域 TN 与农用地关系为例,两者局部回归参数在方向上表现出农村正、城区负的结果,在大小上又以旧城区回归参数的绝对值较高.

2) 相邻子流域土地利用百分比的改变及水体污染源的不同,是导致土地利用对水质作用发生变化的根本原因.从土地利用角度分析,在农村地区,农用地比例和 TN 浓度变化一致,随着河道步入城区,农用地比例减少而 TN 浓度增加使得两者表现为负相关,在旧城区 TN 浓度增加迅速从而使负相关性更为显著;从污染源角度分析,农村地区 TN 含量受农药化肥、牲畜粪便等人类耕作活动影响密切,郊区城区 TN 含量主要来自皮革工业、城市生活污水等点源污染及大范围的不透水用地等,旧城区 TN 浓度的迅速增加主要受所处河道狭窄、流速缓慢且底泥相对较厚所致.

3) 对比校正 R^2 和 AIC_c 两个指标发现,所构建的 GWR 模型预测精度整体优于 OLS 传统全局模型.根据残差和初始水质数据各自的 Moran's I 结果得知,若研究区域不存在明显空间聚类模式,则评比 GWR 模型与 OLS 模型处理空间自相关的能力并无意义.

责任作者简介:梅琨(1981—),女,理学博士,助理研究员,长期从事 GIS 技术在水环境领域的应用研究.目前已发表论文 20 余篇,其中 SCI 收录论文 5 篇.

张明华,女,教授,博士生导师,加州大学农业与环境学院国际合作交流中国项目主任.研究领域包括水质量模型、面源污染传输与环境评价、湿地生态学等.

参考文献(References):

- Ahearn D S, Sheibley R W, Dahlgren R A, *et al.* 2005. Land use and land cover influence on water quality in the last free-flowing river draining the western Sierra Nevada, California [J]. *Journal of Hydrology*, 313(3/4): 234-247
- 白凤娇,李天宏.2012.基于 GIS 和 L-THIA 模型的深圳市观澜河流域非点源污染负荷变化分析[J].*环境科学*,33(8): 2667-2673
- Brown S, Versace V L, Laurenson L, *et al.* 2012. Assessment of spatiotemporal varying relationships between rainfall, land cover and surface water area using geographically weighted regression [J]. *Environmental Modeling & Assessment*, 17(3): 241-254
- Brunsdon C, Fotheringham A S, Charlton M. 1998. Spatial nonstationarity

- and autoregressive models[J]. *Environment and Planning A*, 30(6): 957-973
- Cahill M, Mulligan G. 2007. Using geographically weighted regression to explore local crime patterns[J]. *Social Science Computer Review*, 25(2): 174-193
- 曹芳芳, 李雪, 王东, 等. 2013. 新安江流域土地利用结构对水质的影响[J]. *环境科学*, 34(7): 2582-2587
- Dowd B M, Press D, Los H M. 2008. Agricultural nonpoint source water pollution policy: The case of California's Central Coast [J]. *Agriculture Ecosystems & Environment*, 128(3): 151-161
- Fotheringham A S, Charlton M, Brunsdon C. 1996. The geography of parameter space: An investigation of spatial non-stationarity [J]. *International Journal of Geographical Information Systems*, 10(5): 605-627
- Fotheringham A S, Brunsdon C, Charlton M. 2002. *Geographically weighted regression: the analysis of spatially varying relationships*. Chichester: Wiley
- 官宝红, 李君, 曾爱斌, 等. 2008. 杭州市城市土地利用对河流水质的影响[J]. *资源科学*, 30(6): 857-863
- Javi S T, Malekmohammadi B, Mokhtari H. 2014. Application of geographically weighted regression model to analysis of spatiotemporal varying relationships between groundwater quantity and land use changes [J]. *Environmental Monitoring and Assessment*, 186(5): 3123-3138
- 刘亚群. 2014. 旭水河贡井段非点源污染负荷估算研究[D]. 成都: 西南交通大学. 1-12
- Mei K, Liao L L, Zhu Y L, *et al.* 2014. Evaluation of spatial-temporal variations and trends in surface water quality across a rural-suburban-urban interface [J]. *Environment Science and Pollution Research*, 21(13): 8036-8051
- 孟祥巍, 许学工. 2014. 平原城市河流面源污染研究范围及方法的选择与效果比较[J]. *生态环境学报*, 23(1): 145-150
- 牟乃夏, 刘文宝, 王海银, 等. 2012. ArcGIS 10 地理信息系统教程: 从初学到精通[M]. 北京: 测绘出版社. 328-330
- Nakaya T, Fotheringham A S, Brunsdon C, *et al.* 2005. Geographically weighted Poisson regression for disease association mapping [J]. *Statistics in Medicine*, 24(17): 2695-2717
- Pineda J N, Bosque S J, Gomez D M, *et al.* 2010. Exploring the driving forces behind deforestation in the state of Mexico (Mexico) using geographically weighted regression [J]. *Applied Geography*, 30(4): 576-591
- Pratt B, Chang H J. 2012. Effects of land cover, topography, and built structure on seasonal water quality at multiple spatial scales [J]. *Journal of Hazardous Materials*, 209: 48-58
- 覃文忠. 2007. 地理加权回归基本理论与应用研究[D]. 上海: 同济大学. 6-24
- 孙金华, 曹晓峰, 黄艺. 2011. 滇池流域土地利用对入湖河流水质的影响[J]. *中国环境科学*, 31(12): 2052-2057
- Tran C P, Bode R W, Smith A J, *et al.* 2010. Land-use proximity as a basis for assessing stream water quality in New York State (USA) [J]. *Ecological Indicators*, 10(3): 727-733
- Tu J, Xia Z G. 2008. Examining spatially varying relationships between land use and water quality using geographically weighted regression I: Model design and evaluation [J]. *Science of the Total Environment*, 407(1): 358-378
- Tu J. 2013. Spatial variations in the relationships between land use and water quality across an urbanization gradient in the watersheds of Northern Georgia, USA [J]. *Environmental Management*, 51(1): 1-17
- 王娇, 马克明, 张育新, 等. 2012. 土地利用类型及其社会经济特征对河流水质的影响[J]. *环境科学学报*, 32(1): 57-65
- Wang R Z, Xu T L, Yu L Z, *et al.* 2013. Effects of land use types on surface water quality across an anthropogenic disturbance gradient in the upper reach of the Hun River, Northeast China [J]. *Environmental Monitoring and Assessment*, 185(5): 4141-4151
- 王瑶, 石爱军. 2012. 北京市土地利用分布与水质响应空间关系研究[J]. *中国环境监测*, 28(3): 58-62
- Williams M, Hopkinson C, Rastetter E, *et al.* 2005. Relationships of land use and stream solute concentrations in the Ipswich River basin, Northeastern Massachusetts [J]. *Water, Air, & Soil Pollution*, 161(1/4): 55-74
- Zhang X D, Huang G H, Nie X H. 2011. Possibilistic stochastic water management model for agricultural nonpoint source pollution [J]. *Journal of Water Resources Planning and Management*, 137(1): 101-112