



# 地理加权回归分析技术综述

卢宾宾<sup>1,3</sup> 葛咏<sup>2</sup> 秦昆<sup>1</sup> 郑江华<sup>3</sup>

1 武汉大学遥感信息工程学院,湖北 武汉,430079

2 中国科学院地理科学与资源研究所,北京,100101

3 新疆大学资源与环境科学学院,新疆 乌鲁木齐,830046

**摘要:**空间数据关系中的异质性或非平稳性特征是近期空间统计或相关应用领域研究热点之一,而局部空间统计分析技术的提出与发展是其关键环节。地理加权回归分析技术(geographically weighted regression, GWR)通过关于位置的局部加权回归分析模型求解,以随着空间位置不同而变化的参数估计结果,量化反映空间数据关系中的异质性或非平稳性特征。GWR技术已在众多领域内广泛应用,逐渐成为重要的空间关系异质性建模工具之一。针对GWR模型解算、结果解读、模型检验等基础技术环节进行了系统总结,分别分析回顾了其对应的相关研究进展以及应用过程中存在的问题。同时,系统梳理了近年来GWR技术的主要拓展与延伸,重点阐述了其在采用灵活的距离度量选择、参数的多尺度估计以及时空数据建模方面的GWR技术扩展研究。此外,还简要介绍了现有的主要GWR技术软件工具,以期读者和用户提供相对全面的GWR技术信息参考与知识总结。

**关键词:**空间异质性;空间非平稳性;地理加权建模技术;空间统计;空间分析

**中图分类号:**P208

**文献标志码:**A

早期空间统计分析技术多从全局假设的角度出发,认为空间变量关系是固定的,不随空间位置的变化而改变<sup>[1]</sup>。这个前提假设明显违背了现实地理世界空间关系的异质性或非平稳性规律<sup>[2]</sup>。因此,区别于传统研究“单一普适”的空间关系统计分析方法,如何对空间异质性进行精确局部描述的空间分析方法越来越多地受到重视<sup>[3]</sup>,如随机系数模型<sup>[4]</sup>、空间变参数回归分析模型<sup>[5]</sup>和局部加权回归分析模型<sup>[6]</sup>。

1996年,Brunsdon等提出地理加权回归分析(geographically weighted regression, GWR)<sup>[7]</sup>技术,提供了直观、实用的空间关系异质性和多相性分析手段<sup>[8]</sup>,已发展成为局部空间统计分析的重要方法之一。本文以“Geographically Weighted Regression”和“地理加权回归分析”为主题关键词,通过Web of Science (WOS)核心合集(<https://apps.webofknowledge.com/>,检索时间:2019-07-19)和中国知网(China National Knowledge Infrastructure, CNKI) (<http://www.cnki.net>,检索时间:2019-07-21)对已发表的GWR技术相关文章进行搜索,分别得到了1 795

篇和391篇文章记录。针对上述文献检索记录,本文通过R函数包Bibliometrix<sup>[9]</sup>中提供的文献分析工具进行了分析。图1显示了每年发表的GWR技术相关文章数量,英文(来自WOS)和中文(来自CNKI)论文数量均呈现了明显的逐年上升趋势,尤其自2015年之后论文数量大幅提升。

通过对两个记录中论文关键词进行统计分析,分别制作了关键词词云图(图2)。如图2(a)所示的WOS英文论文,学者们利用GWR技术对环境(如高频关键词“PM2.5”“PM10”“AIR QUALITY”)、健康(如高频关键词“MALARIA”“HEALTH”)、安全(如高频关键词“FIRE RISK”“FOREST FIRES”)、大数据分析(如高频词汇“BIG DATA”)、遥感(如高频词汇“REMOTE SENSING”“LiDAR”)等领域开展了深入和广泛的应用。值得注意的是,GWR技术也较多地应用于中国相关问题的研究(如高频词汇“CHINA”“BEIJING”)。如图2(b)所示,CNKI所记录的中文论文主要集中于GWR技术的应用层面,如房地产市场(如高频词汇“房价”)、生态(如高频词汇“生物量”“耕地”)等领域。此外,本文分别统

收稿日期:2019-09-17

项目资助:国家自然科学基金(41725006,41871287,U1833201)。

第一作者:卢宾宾,博士,讲师,主要从事空间统计、地理加权回归分析、地理加权建模技术等方面的研究。binbinlu@whu.edu.cn



性特征,尤其随着空间距离增大,其关联作用程度衰减的规律。GWR技术将地理学第一定律深度融入局部空间统计方法,通过对独立抽样的分析点分别进行回归分析模型解算,得到与空间位置一一对应的空间回归系数,以随着空间位置不同而变化的参数估计量化表征空间关系异质性特征。基础GWR模型一般可表达如下:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^m \beta_k(u_i, v_i) x_{ik} + \epsilon_i \quad (1)$$

式中,  $y_i$  为在位置  $i$  处的因变量值;  $x_{ik}$  ( $k=1, 2, \dots, m$ ) 为位置  $i$  处的自变量值;  $(u_i, v_i)$  为回归分析点  $i$  的坐标;  $\beta_0(u_i, v_i)$  为截距项;  $\beta_k(u_i, v_i)$  ( $k=1, 2, \dots, m$ ) 为回归分析系数。

### 1.1 GWR模型解算

针对GWR模型(1),对每个回归分析点  $i$  分别采用加权线性最小二乘方法进行模型求解:

$$\hat{\beta}(u_i, v_i) = (\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{y} \quad (2)$$

式中,  $\mathbf{X}$  为自变量抽样矩阵,其第一列取值为1,以估计截距项  $\beta_0(u_i, v_i)$ ;  $\mathbf{y}$  为因变量抽样值列向量;  $\hat{\beta}(u_i, v_i) = (\beta_0(u_i, v_i) \beta_1(u_i, v_i) \dots \beta_m(u_i, v_i))^T$  为回归分析点  $(u_i, v_i)$  处的回归分析系数向量;  $\mathbf{W}(u_i, v_i)$  为对角矩阵,其对角线元素值为每个数据点到回归分析点  $(u_i, v_i)$  的空间权重值:

$$\mathbf{W}_i = \begin{bmatrix} w_{i1} & 0 & \dots & 0 \\ 0 & w_{i2} & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & w_{in} \end{bmatrix} \quad (3)$$

式中,对角线值  $w_{ij}$  ( $j=1, 2, \dots, n$ ) 表示第  $j$  个数据点到回归分析点  $i$  的权重值。根据地理学第一定律, GWR模型计算权重的基本原则为:距离越近,赋予的权重值越高;反之,权重值越低<sup>[13]</sup>。因此,可通过任意值域为  $[0, 1]$ 、关于空间距离的单调减函数实现权重计算,称之为核函数(Kernel function)<sup>[14-15]</sup>。针对GWR模型求解,常用核函数包括高斯(Gaussian)函数(式(4))、指数(Exponential)函数(式(5))、盒状(Box-car)函数(式(6))、二次(Bi-square)函数(式(7))和立方体(Tri-cube)函数(式(8))。

$$w_{ij} = e^{-\frac{(d_{ij}/b)^2}{2}} \quad (4)$$

$$w_{ij} = \exp\left(-\frac{|d_{ij}|}{b}\right) \quad (5)$$

$$w_{ij} = \begin{cases} 1, & d_{ij} \leq b \\ 0, & \text{其他} \end{cases} \quad (6)$$

$$w_{ij} = \begin{cases} \left(1 - (d_{ij}/b)^2\right)^2, & d_{ij} \leq b \\ 0, & \text{其他} \end{cases} \quad (7)$$

$$w_{ij} = \begin{cases} \left(1 - (d_{ij}/b)^3\right)^3, & d_{ij} \leq b \\ 0, & \text{其他} \end{cases} \quad (8)$$

式(4)~(8)中,  $d_{ij}$  表示位置  $i$  与位置  $j$  之间的空间距离或邻近度量;  $b$  为带宽值。

根据核函数的值域分布特征,可将其分为连续型(如 Gaussian 函数、Exponential 函数)和截断型(Box-car 函数、Bi-square 函数和 Tri-cube 函数)两种。截断型核函数明确定义了GWR模型解算过程中所考量的局部范围(结合带宽值),相对更有利于模型结果的解读,但易造成  $\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X}$  矩阵奇异,尤其当自变量样本值存在较多的“0”值时;而连续型核函数的特点则恰好相反。在GWR模型实际应用中,最常用的两种核函数为 Gaussian 函数和 Bi-square 函数。建议读者尝试不同类型的核函数,权衡利弊,以选择特定模型和数据条件下的最为合适的核函数。

带宽是GWR模型权重计算的重要控制参数,可分为固定型和可变型带宽,如图3所示。通过定义一个固定的距离阈值  $b$ ,是最为直接的带宽定义方法,但当数据点分布疏密程度不均匀时,则可能造成参与局部模型解算的有效样本数不足(如图3(a)示例中,右侧回归分析点模型求解时有效参与的样本点仅为2个);为了避免固定型带宽的这个缺点,可通过定义最近邻域个数  $N$ ,将回归分析点与第  $N$  个最近邻域之间的距离作为对应模型解算的带宽值,因此其在每个回归分析点对应的带宽值可能是不同的,即可变型带宽。带宽大小直接决定了权重随距离增加而衰减的速率,带宽越大,权重衰减越快,反之越慢。而针对截断型核函数,距离回归分析点超过带宽的数据点对应权重值均为0,带宽的大小则直接决定了GWR模型解算过程中围绕每个回归分析点的有效数据点范围。同时,带宽过小可能导致局部模型过拟合,参数估计空间变化剧烈;带宽过大则可能导致GWR模型趋于全局化,参数估计趋于平滑。因此,选择合适的带宽是GWR模型解算的必要程序。针对特定GWR模型,可通过交叉验证(cross validation, CV)<sup>[16]</sup>或赤池信息量准则(Akaike information criterion, AIC)<sup>[17]</sup>对带



宽值进行优选,表达式如下:

$$CV(b) = \sum_{i=1}^n [y_i - \hat{y}_{\neq i}(b)]^2 \quad (9)$$

$$AIC_c(b) = 2n \ln \hat{\sigma} + n \ln 2\pi + n \left\{ \frac{n + \text{tr}(\mathbf{S})}{n - 2 - \text{tr}(\mathbf{S})} \right\} \quad (10)$$

式中,  $\hat{y}_{\neq i}(b)$  为在数据点  $i$  处, 将其本身排除后进行模型求解所得到的因变量预测值;  $\hat{\sigma}$  为模型标准差估计;  $\text{tr}(\mathbf{S})$  为帽子矩阵  $\mathbf{S}$  的迹;  $AIC_c$  表示校正 AIC 值 (corrected AIC,  $AIC_c$ )<sup>[13]</sup>。通过最小化 CV 或  $AIC_c$  值, 选取对应的“最优”带宽值。一般来说,  $AIC_c$  值相对于 CV 优化程度较好, 但计算复杂度也更高, 读者可酌情选择。

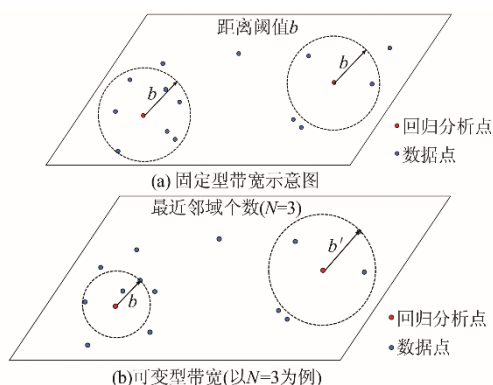


图3 不同类型带宽作用示意图

Fig.3 Schematic Diagram of Different Types of Bandwidth Definitions

## 1.2 GWR 模型结果解读

在带宽优选的基础上即可实现 GWR 模型求解。而 GWR 模型结果的解读需要关注模型诊断信息和结果可视化分析两个环节。GWR 模型诊断信息主要包括: ①残差平方和 (residual sum of squares, RSS), 反映模型预测精度; ②  $R^2$  或其调整版本 (Adjusted  $R^2$ ), 反映模型拟合优度 (goodness of fit, GoF); ③  $AIC_c$ : 综合反映 GWR 模型结果的 GoF 和模型复杂度。一般情况下, 可将 GWR 模型的诊断统计量与对应一般线性回归分析结果进行横向对比, 以观察 GWR 模型是否体现显著改进以及改进的程度。注意  $AIC_c$  值是针对特定建模过程的相对统计量, 即针对同一套数据和同一个因变量, 对应的  $AIC_c$  值才具有可比性。当  $AIC_c$  值变化大于 3 时, 可认为模型结果之间具有显著不同<sup>[13]</sup>。

GWR 模型是关于位置的解算, 也决定了其最典型的特点: 结果可便捷地进行地图可视化。

如图 2 所示, “SPATIALLY VARYING RELATIONSHIPS”“影响因子”等高频词汇的出现表示 GWR 技术作为解释性工具频繁应用, 而利用地图对 GWR 模型结果的可视化解读是此类研究的关键, 直观展示空间关系异质性特征<sup>[3,18]</sup>, 对其空间变化结合地理含义进行准确解读。在目前 GWR 技术的应用过程中, 多存在“重解算、轻解读”的问题, 值得读者注意。

## 1.3 GWR 模型检验

作为一个特点鲜明的技术, GWR 模型被广泛应用, 但相对于一般线性回归分析而言, 它是一种复杂度更高的技术。在建模过程中, 采用 GWR 技术的一个重要假设就是“空间数据关系中存在显著的空间异质性或非常平稳性特征”。因此, 采用 GWR 技术对空间关系进行建模时, 需要进行必要的模型关系异质性特征检验。

早期 Brunsdon 等<sup>[7]</sup>提出了蒙特卡洛模拟方法, 对 GWR 模型中的每一个自变量检验其是否具有显著的空间异质性特征。之后 Leung 等<sup>[19-20]</sup>提出了一系列的  $F$  检验方法, 分别提出了对模型整体和单个自变量进行检验的  $F$  统计量。因此, 在正式采用 GWR 技术对模型求解之前, 有必要利用上述检验方法确认其空间异质性特征。

类似于一般线性回归分析, GWR 技术针对每一个回归分析点的局部模型求解也会输出对应的  $t$  检验结果, 以判断对应参数估计的非零显著性特征。 $t$  检验结果可与 GWR 模型参数估计进行综合体现, 如屏蔽掉不显著的参数估计值部分<sup>[21]</sup>, 以更加精确地呈现结果。

## 1.4 GWR 技术问题分析

GWR 技术的发展过程中, 也因为一些问题而饱受诟病, 如多重共线性问题<sup>[22-23]</sup>。多重共线性是回归分析技术中的常见问题, 由于自变量之间存在较强相关关系, 导致模型求解结果失真甚至非唯一解现象。而针对 GWR 模型, 由于仅有部分样本有效参与到每个局部模型的求解, 从而存在更高的多重共线性风险, 即使某些变量在全局意义上线性无关, 也可能存在较强的局部共线性<sup>[23]</sup>。Wheeler<sup>[24-25]</sup>提出了地理加权 Lasso 回归 (geographically weighted Lasso, GWL) 和地理加权岭回归分析 (geographically weighted ridge regression, GWRR) 方法, Gollini 等<sup>[26]</sup>探讨了岭参数局部补偿地理加权回归分析技术, 一定程度上缓解了 GWR 模型中潜在的共线性风险。Fotheringham 和 Oshan<sup>[27]</sup>利用一系列模拟实验尝试证明

GWR技术对共线性风险是稳健的。而笔者认为多重共线性风险仍然是GWR技术的桎梏,除了以上针对性的技术扩展,全面细致的数据分析和变量选择是规避这种风险的有效途径。

GWR技术所面临的另一问题是统计推断问题,围绕其相关的检验和推论能否成立一直存在较大的争议。如空间关系异质性检验的结果可能受到变量错误、多重共线性等方面的影响,甚至出现假象误导<sup>[28-29]</sup>。而前述GWR模型的 $t$ 检验结果,因为带宽选择和模型参数估计的过程中样本不断被复用,也存在典型的多重依赖假设检验问题,针对这个问题,学者们也提出了多个 $t$ 检验值纠正算法<sup>[30-31]</sup>,值得研究者进一步探索与尝试。

## 2 GWR技术扩展

随着研究与应用的不断深入,GWR技术也不断得到扩展,以改善其求解精度、增强模型的适用性。本节将简要回顾GWR技术的主要扩展,并展望其未来发展方向。

### 2.1 基于灵活距离度量选择的GWR技术

在GWR模型的传统解算过程中,人们往往将距离默认为欧氏距离度量,即直线距离。但是现实世界不仅仅是一个简单的几何容器,而是人们通过对复杂空间现象和关系的感知而建立的综合认知空间<sup>[32]</sup>。Lu等<sup>[33]</sup>尝试将非欧氏距离度量应用于GWR模型解算,并指出非欧氏距离度量能够提升模型求解精度的同时,更易于发现地理空间中非规则分布而造成的空间关系异质性特征,如采用道路网络距离和旅行时间发现河流在城市房价分布中发挥的天然屏障效果。此项研究说明任意合理定义的距离度量都可用于GWR模型求解,进而构成基于灵活距离度量选择的GWR建模技术,在此框架下,未来值得探索更加多样的距离度量,如空间语义、定性距离等。

而针对距离度量未知的模型场景,Lu等<sup>[34]</sup>提出了Minkowski方法实现距离度量自动优选,采用Minkowski距离函数逼近潜在的“最优”距离度量。这个方法解决了针对未知GWR模型的度量优选问题,但算法复杂度较高,且一般选取的距离度量难以直观体现,不利于模型结果解读。因此,在GWR技术作为预测工具时可用性较强,而在关系解释方面的作用有待于进一步探索。

### 2.2 多尺度GWR技术

基础GWR模型往往采用单一核函数和带宽

计算权重,这也导致所有参数估计的空间变化呈现相同的尺度特征。因此,即使多元空间数据关系对应不同的变化尺度,基础GWR却忽略了这种差异并以空间关系的“最佳平均”尺度反映所有参数估计的空间变化。近几年,出现了GWR技术的一个重要分支——多尺度GWR技术(multiscale GWR)<sup>[35-36]</sup>,反映多元GWR模型中变量估计的尺度差异性。

Brunsdon等<sup>[37]</sup>提出了混合GWR模型,将模型参数分为全局和局部两种尺度特征,利用后向迭代算法对模型进行估计,覃文忠等<sup>[38]</sup>和玄海燕等<sup>[39]</sup>、聂磊等<sup>[40]</sup>也曾对混合GWR模型的求解方法进行讨论。Mei等<sup>[41-42]</sup>、Harris等<sup>[43]</sup>利用Bootstrap方法对混合GWR模型参数估计的(全局或局部)尺度选择进行检验。

混合GWR技术虽能够对GWR模型参数估计尺度进行差异化呈现,参与局部估计的参数对应同样的尺度特征,区分仍然相对单一,缺乏对细节差异的精准呈现。Yang等<sup>[44-45]</sup>开始尝试对多元GWR模型中不同参数分别采用对应的不同带宽对其进行求解,以准确呈现不同参数估计对应的异质性尺度特征。Brunsdon等<sup>[37]</sup>对此项技术进行综合描述,将其命名为多尺度GWR。同年,Leong和Yue<sup>[46]</sup>提出了类似的条件GWR技术。这些技术均采用与参数一一对应的差异化带宽值对GWR模型进行解算。

综合在GWR模型中应用灵活的距离度量,Lu等<sup>[47]</sup>提出了采用各异的距离度量和优选带宽的方法对多元GWR模型不同参数进行解算,即距离-变量对应的地理加权回归分析(geographically weighted regression with parameter-specific distance metrics, PSDM GWR)。同时,Lu等<sup>[48]</sup>指出每个距离-变量对应的最优带宽值具有明显的不变特征,即针对模型中每一个特定的变量-度量组合对应的最优带宽值相对稳定不变,与其他变量的组合情况无关,也说明了GWR模型中的每个参数均对应着一个“最优”的表征尺度。

在数据建模场景日趋多样的背景下,空间数据尺度日趋复杂<sup>[49]</sup>,作为GWR技术的最新扩展,多尺度GWR技术更具鲁棒性和普适性<sup>[50]</sup>,围绕其相关的研究也在不断拓展与深入,如空间推论<sup>[51]</sup>、时间维扩展<sup>[52]</sup>等。经历更多的实践应用与技术探索,多尺度GWR技术能够更加准确和全面地实现模型估计,进而成为多元GWR模型的默认技术入口,在空间数据关系异质性建模方面

将发挥更大的作用。

### 2.3 时空地理加权回归分析

Huang 等<sup>[53]</sup>将时间维融入 GWR 技术,提出了时空地理加权回归分析技术(geographically and temporally weighted regression, GTWR),其模型表达式如下:

$$y_i = \beta_0(u_i, v_i, t_i) + \sum_{k=1}^n \beta_k(u_i, v_i, t_i) x_{ik} + \epsilon_i \quad (11)$$

从式(11)可看出,GTWR 模型估计是关于时空回归分析点 $(u_i, v_i, t_i)$ 的解算,即:

$$\hat{\beta}(u_i, v_i, t_i) = (X^T W(u_i, v_i, t_i) X)^{-1} X^T W(u_i, v_i, t_i) y \quad (12)$$

其中针对权重矩阵  $W(u_i, v_i, t_i)$  的计算, Huang 等<sup>[53]</sup>提出了时空椭圆坐标系,以计算时空距离:

$$d_{ij}^{ST} = \sqrt{\lambda[(u_i - u_j)^2 + (v_i - v_j)^2] + \mu(t_i - t_j)^2} \quad (13)$$

其中,  $\lambda$  和  $\mu$  为时间和空间距离的比例调整系数。这种距离定义可操作性较强,能够与基础 GWR 技术无缝结合,但调整系数相对难以确定,时空距离一定程度上较为抽象,会为结果解读带来一定困难。Fotheringham 等<sup>[54]</sup>提出了另一版本的 GTWR 技术,在假设时间维和空间维相互独立的基础上,分别计算时间权重和空间权重,通过式(14)计算时空权重:

$$w_{ij}^t = K_S(d_{s_{ij}}, b_S) \times K_T(d_{t_{ij}}, b_T) \quad (14)$$

其中,  $K_S$ 、 $K_T$  分别表示空间权重、时间权重核函数;  $d_{s_{ij}}$ 、 $d_{t_{ij}}$  分别表示空间距离和时间距离;  $b_S$ 、 $b_T$  分别为对应的空间维带宽和时间维带宽。此算法保留了时间和空间距离的原始值特征,利于 GTWR 结果解读,但算法复杂度高,需要对两种带宽分别进行优选。

经过 GTWR 技术的不断演化, Wu 等<sup>[55]</sup>提出了时空地理加权自回归模型(geographically and temporally weighted autoregressive model); Du 等<sup>[56]</sup>综合考量时间维度的周期性特征,提出了周期性时空地理加权回归分析技术(geographically and circle-temporally weighted regression); 赵阳阳等<sup>[57]</sup>提出了一种基于局部多项式求解的 GTWR 分析方法; Liu 等<sup>[58]</sup>提出了时空地理加权相关逻辑回归分析技术(geographically weighted temporally correlated logistic regression)。

随着时空数据场景的不断涌现,GTWR 技术应用范围日趋扩大,但它也面临较大的挑战:一

方面时空数据一般体量较大,对模型解算能力和效率造成非常大的困难,已成为当前 GTWR 技术重要瓶颈之一;另一方面,GTWR 模型是关于时间和空间维度的解算,参数估计结果的有效展示与分析异常困难。

### 2.4 其他扩展

除了以上扩展, GWR 技术在其他方面也在不断演化和改进,如考量异方差特征的 GWR 方法(heteroskedastic GWR)<sup>[59]</sup>, 弱化异常值影响的鲁棒性地理加权回归分析(robust GWR)<sup>[60]</sup>, 将参数估计扩展为空间坐标的线性函数的基于局部线性估计的 GWR 技术(local linear estimation-based GWR)<sup>[61]</sup>, 以 GWR 技术作为预测工具(GWR as a predictor)<sup>[62]</sup>, 非高斯建模框架下的广义地理加权回归分析(如 geographically weighted poisson regression<sup>[63]</sup>和 geographically weighted logistic regression<sup>[64]</sup>), 地理加权序数回归分析(geographically weighted ordinal regression)<sup>[65]</sup>, 用于海量数据 GWR 模型求解的高性能解决方案(如 grid-based GWR<sup>[66]</sup>、fastGWR<sup>[67]</sup>和 scalable GWR<sup>[68]</sup>)等。多样的 GWR 技术扩展提供了适应异方差、异常值、共线性、空间自相关<sup>[69]</sup>等不同应用场景的技术,但需要在充分了解数据状况的基础上选取对应的技术,但篇幅有限,本文不再一一详述。

## 3 GWR 软件工具

随着 GWR 技术的不断发展,涌现了一系列 GWR 软件工具。Charlton 等<sup>[70]</sup>采用 FORTRAN 语言开发了最早的 GWR3.0 软件,支持 Windows 操作系统,用户界面友好,在计算效率方面表现优异。但因为收费原因,用户受众相对有限。之后, Nakaya 等<sup>[71]</sup>采用 C++ 语言开发了全新版本的 GWR 软件,沿用版本号称之为 GWR4.0。两个软件在功能和界面上均有很大区别,后者除了包含基础 GWR 技术之外,重点支持混合 GWR 模型和广义 GWR 模型的解算,而且因其是免费软件,用户量大幅增加。Oshan 等<sup>[72]</sup>采用 Python 语言开发了 MGWR 软件,包含了与 GWR4.0 类似的功能,并将原有的混合 GWR 模型解算功能扩展为支持多尺度 GWR 模型解算,它同时对应了 Python 语言函数包 PySAL 的一个模块和一个 GUI 封装版本。此外,在 ESRI 公司推出的 ArcGIS9.0 以及之后的软件版本中([http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm? Topic=](http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?Topic=)



Name=Interpreting\_GWR\_results),在空间统计工具箱中集成了独立的GWR工具,能够实现基础GWR模型求解。借助ArcGIS的强大功能,它在结果可视化方面优势明显,但缺少必要的模型诊断信息,功能更新远远落后于GWR技术的进化与扩展。

近些年来,以R软件平台为基础,开发了多个GWR函数工具包。Bivand和Yu开发了spgwr(<http://cran.r-project.org/web/packages/spgwr/index.html>),包含了基础GWR和广义GWR模型的求解,是最早关于GWR技术的R函数包。Lu等<sup>[26,73]</sup>开发了函数包GWmodel,基本囊括了本文所提及的基础GWR技术及其扩展模块。相对于其他软件工具,GWmodel具有以下典型的特点与优势:

1)技术覆盖全面:除了GWR技术及其相关扩展模块外,还集成了地理加权汇总统计(geographically weighted summary statistics)、地理加权主成分分析(geographically weighted principal components analysis)和地理加权判别分析(geographically weighted discriminant analysis)技术模块。

2)模型参数选项丰富:函数包提供了5种不同的核函数和灵活的距离度量控制参数进行权重计算;

3)运行效率高:函数包中所有的核心函数均采用了C++语言和Rcpp<sup>[74]</sup>嵌套开发,大大提升了函数运行效率。

此外,Wheeler<sup>[75]</sup>开发了函数包gwrr,提供了地理加权岭回归分析工具以应对GWR模型中共线性问题。Murakami等<sup>[68]</sup>开发了函数包scgwr,提供了GWR模型的高性能求解工具Scalable GWR,该工具也被包含在函数包GWmodel中。

## 4 结 语

在过去的15 a中,GWR技术不断发展与演化,较好地解决了空间数据关系异质性或非平稳性特征的量化估计与表征问题,其在各个领域内的广泛应用及技术的持续发展也说明它已成为重要的空间计量建模工具之一。本文回顾了GWR模型解算、结果解读、模型检验等基础技术环节,并对其间存在的问题进行了剖析与总结,强调了GWR技术使用过程中可能存在的误区与注意事项。同时,本文梳理了GWR技术在距离度量、多尺度估计、时空综合等方面的扩展与延

伸,介绍了现有的GWR软件工具,以期为读者和用户提供全面的信息参考与知识总结。

GWR技术的算法与理论研究虽日趋丰富与完备,但仍存在一些问题待于深入探讨。本文对这些问题进行了总结与展望:

1)GWR技术中存在的多重共线性问题和统计推论问题,需要读者在充分理解数据的基础上合理选择模型变量,而针对问题本身的分析与研究也是未来的研究点之一;

2)新的GWR技术扩展不断出现,如多尺度GWR技术,而其模型诊断、统计推断、计算效率等问题也是未来研究的热点方向之一;

3)随着数据量逐渐变大,尤其是时空大数据体量激增,GWR及其扩展GTWR等技术都面临着计算效率瓶颈问题,而针对于此问题的高性能解决方案也是未来有待于突破的方向之一。相信随着GWR技术研究不断深入及其应用场景的持续拓展,更多的问题会被发现进而被解决,形成较好的良性互动,不断推动GWR技术体系的发展与成熟。

## 参 考 文 献

- [1] Fotheringham A S, Brunsdon C. Local Forms of Spatial Analysis[J]. *Geographical Analysis*, 1999, 31(4): 340-358
- [2] Goodchild M F. The Validity and Usefulness of Laws in Geographic Information Science and Geography[J]. *Annals of the Association of American Geographers*, 2004, 94(2): 300-303
- [3] Páez A. Local Analysis of Spatial Relationships: A Comparison of GWR and the Expansion Method [C]. The 5th International Conference on Computational Science and Its Applications, Singapore, 2005
- [4] Swamy P A V B, Conway R K, Leblanc M R. The Stochastic Coefficients Approach to Econometric Modeling, Part 1: A Critique of Fixed Coefficients Models[R]. US:Board of Governors of the Federal Reserve System, 1988
- [5] Gamerman D, Moreira A R B, Rue H. Space-Varying Regression Models: Specifications and Simulation [J]. *Computational Statistics & Data Analysis*, 2003, 42(3): 513-533
- [6] Cleveland W S. Robust Locally Weighted Regression and Smoothing Scatterplots[J]. *Journal of the American Statistical Association*, 1979, 74(368): 829-836
- [7] Brunsdon C, Fotheringham A S, Charlton M E. Geographically Weighted Regression: A Method for

- Exploring Spatial Nonstationarity[J]. *Geographical Analysis*, 1996, 28(4): 281-298
- [8] Páez A, Wheeler D. Geographically Weighted Regression[M]// Kitchin R, Thrift N. *International Encyclopedia of Human Geography*. Oxford: Elsevier, 2009: 407-414
- [9] Aria M, Cuccurullo C. Bibliometrix: An R-Tool for Comprehensive Science Mapping Analysis[J]. *Journal of Informetrics*, 2017, 11(4): 959-975
- [10] Xuan Haiyan, Li Qi, Zhang Yunhu. Spatial Characteristics Analysis of Total Population in Various Cities Based on Geographically Weighted Regression[J]. *Journal of Biomathematic*, 2016, 31(2): 223-228(玄海燕, 李琪, 张运虎. 基于地理加权回归的我国各市人口总数的空间特征分析[J]. *生物数学学报*, 2016, 31(2): 223-228)
- [11] Zhao Yangyang, Liu Jiping, Zhang Fuhao, et al. An Approach of Characteristics Variable Selection of Geographically Weighted Regression Based on Greedy Algorithm[J]. *Science of Surveying and Mapping*, 2016, 41(7): 41-46(赵阳阳, 刘纪平, 张福浩, 等. 贪心算法的地理加权回归特征变量选择方法[J]. *测绘科学*, 2016, 41(7): 41-46)
- [12] Tobler W R. A Computer Movie Simulating Urban Growth in the Detroit Region[J]. *Economic Geography*, 1970, 46(2): 234-240
- [13] Fotheringham A S, Brunsdon C, Charlton M. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships[M]. Chichester: Wiley, 2002
- [14] Cho S H, Lambert D M, Chen Z. Geographically Weighted Regression Bandwidth Selection and Spatial Autocorrelation: An Empirical Example Using Chinese Agriculture Data[J]. *Applied Economics Letters*, 2010, 17(8): 767-772
- [15] Cameron A C, Trivedi P K. *Microeconometrics: Methods and Applications*[M]. New York: Cambridge University Press, 2005
- [16] Farber S, Páez A. A Systematic Investigation of Cross-Validation in GWR Model Estimation: Empirical Analysis and Monte Carlo Simulations[J]. *Journal of Geographical Systems*, 2007, 9(4): 371-396
- [17] Akaike H. Information Theory and an Extension of the Maximum Likelihood Principle[C]. 2nd International Symposium on Information Theory. Tsahkadzor, Armenia, 1973
- [18] Mennis J. Mapping the Results of Geographically Weighted Regression[J]. *The Cartographic Journal*, 2006, 43(2): 171-179
- [19] Leung Y, Mei C L, Zhang W X. Testing for Spatial Autocorrelation Among the Residuals of the Geographically Weighted Regression[J]. *Environment and Planning A*, 2000, 32(5): 871-890
- [20] Leung Y, Mei C L, Zhang W X. Statistical Tests for Spatial Nonstationarity Based on the Geographically Weighted Regression Model[J]. *Environment and Planning A*, 2000, 32(1): 9-32
- [21] Comber A, Chi K, Huy M Q, et al. Distance Metric Choice Can Both Reduce and Induce Collinearity in Geographically Weighted Regression[J]. *Environment and Planning B: Urban Analytics and City Science*, 2018, DOI:10.1177/2399808318784017
- [22] Griffith D A. Spatial-Filtering-Based Contributions to a Critique of Geographically Weighted Regression (GWR)[J]. *Environment and Planning A*, 2008, 40(11): 2 751-2 769
- [23] Wheeler D, Tiefelsdorf M. Multicollinearity and Correlation Among Local Regression Coefficients in Geographically Weighted Regression[J]. *Journal of Geographical Systems*, 2005, 7(2): 161-187
- [24] Wheeler D C. Simultaneous Coefficient Penalization and Model Selection in Geographically Weighted Regression: The Geographically Weighted Lasso[J]. *Environment and Planning A*, 2009, 41(3): 722-742
- [25] Wheeler D C. Diagnostic Tools and a Remedial Method for Collinearity in Geographically Weighted Regression[J]. *Environment and Planning A*, 2007, 39(10): 2 464-2 481
- [26] Gollini I, Lu B, Charlton M, et al. GWmodel: An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models[J]. *Journal of Statistical Software*, 2015, 63(17): 1-50
- [27] Fotheringham A S, Oshan T M. Geographically Weighted Regression and Multicollinearity: Dispelling the Myth[J]. *Journal of Geographical Systems*, 2016, 18(4): 303-329
- [28] Jetz W, Rahbek C, Lichstein J W. Local and Global Approaches to Spatial Data Analysis in Ecology[J]. *Global Ecology and Biogeography*, 2005, 14(1): 97-98
- [29] Wheeler D C, Páez A. Geographically Weighted Regression[M]. Fischer M M, Getis A. *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*. Berlin, Heidelberg: Springer-Verlag, 2010: 461-486
- [30] Da Silva A R, Fotheringham A S. The Multiple Testing Issue in Geographically Weighted Regression[J]. *Geographical Analysis*, 2016, 48(3): 233-247
- [31] Byrne G, Charlton M, Fotheringham S. Multiple



- Dependent Hypothesis Tests in Geographically Weighted Regression [C]. The 10th International Conference on Geocomputation, Sydney, Australia, 2009
- [32] Kitchin R. Space II[M]//Kitchin R, Thrift N. International Encyclopedia of Human Geography. Oxford: Elsevier, 2009: 268-275
- [33] Lu B, Charlton M, Harris P, et al. Geographically Weighted Regression with a Non-Euclidean Distance Metric: A Case Study Using Hedonic House Price Data[J]. *International Journal of Geographical Information Science*, 2014, 28(4): 660-681
- [34] Lu B, Charlton M, Brunsdon C, et al. The Minkowski Approach for Choosing the Distance Metric in Geographically Weighted Regression[J]. *International Journal of Geographical Information Science*, 2016, 30(2): 351-368
- [35] Fotheringham A S, Yang W, Kang W. Multiscale Geographically Weighted Regression (MGWR)[J]. *Annals of the American Association of Geographers*, 2017, 107(6): 1 247-1 265
- [36] Lu B, Brunsdon C, Charlton M, et al. A Response to 'A Comment on Geographically Weighted Regression with Parameter-Specific Distance Metrics' [J]. *International Journal of Geographical Information Science*, 2019, 33(7): 1 300-1 312
- [37] Brunsdon C, Fotheringham A S, Charlton M. Some Notes on Parametric Significance Tests for Geographically Weighted Regression[J]. *Journal of Regional Science*, 1999, 39(3): 497-524
- [38] Qin Wenzhong, Wang Jianmei, Liu Miaolong. Algorithm for Mixed Geographically Weighted Regression Mode[J]. *Geomatics and Information Science of Wuhan University*, 2007, 32(2): 115-119(覃文忠, 王建梅, 刘妙龙. 混合地理加权回归模型算法研究[J]. 武汉大学学报·信息科学版, 2007, 32(2): 115-119)
- [39] Xuan Haiyan, Liu Shuqun, Luo Shuanghua. Two Kinds of Estimation of Mixed Geographically Weighted Regression Mode[J]. *Journal of Lanzhou University of Technology*, 2007, 33(3): 142-144(玄海燕, 刘树群, 罗双华. 混合地理加权回归模型的两类估计[J]. 兰州理工大学学报, 2007, 33(3): 142-144)
- [40] Nie Lei, Shu Hong, Liu Yan. Interpolation of Monthly Average Temperature by Using (Mixed) Geographically Weighted Regression Kriging in the Complex Terrain Region[J]. *Geomatics and Information Science of Wuhan University*, 2018, 43(10): 1 553-1 559(聂磊, 舒红, 刘艳. 复杂地形地区月平均气温(混合)地理加权回归克里格插值[J]. 武汉大学学报·信息科学版, 2018, 43(10): 1 553-1 559)
- [41] Mei C L, Xu M, Wang N. A Bootstrap Test for Constant Coefficients in Geographically Weighted Regression Models[J]. *International Journal of Geographical Information Science*, 2016, 30(8): 1 622-1 643
- [42] Mei C L, Wang N, Zhang W X. Testing the Importance of the Explanatory Variables in a Mixed Geographically Weighted Regression Model[J]. *Environment and Planning A*, 2006, 38(3): 587-598
- [43] Harris P, Brunsdon C, Lu B, et al. Introducing Bootstrap Methods to Investigate Coefficient Non-stationarity in Spatial Regression Models[J]. *Spatial Statistics*, 2017, 21: 241-261
- [44] Yang W, Fotheringham A S, Harris P. An Extension of Geographically Weighted Regression with Flexible Bandwidths [C]. GISRUk, Lancaster, UK, 2012
- [45] Yang W. An Extension of Geographically Weighted Regression with Flexible Bandwidths [D]. UK: Centre for GeoInformatics, University of St Andrews, 2014
- [46] Leong Y Y, Yue J C. A Modification to Geographically Weighted Regression[J]. *International Journal of Health Geographics*, 2017, 16(1): 11
- [47] Lu B, Brunsdon C, Charlton M, et al. Geographically Weighted Regression with Parameter-Specific Distance Metrics[J]. *International Journal of Geographical Information Science*, 2017, 31(5): 982-998
- [48] Lu B, Yang W, Ge Y, et al. Improvements to the Calibration of a Geographically Weighted Regression with Parameter-Specific Distance Metrics and Bandwidths [J]. *Computers, Environment and Urban Systems*, 2018, 71: 41-57
- [49] Ge Y, Jin Y, Stein A, et al. Principles and Methods of Scaling Geospatial Earth Science Data[J]. *Earth-Science Reviews*, 2019, 197: 102 897
- [50] Wolf L, Oshan T, Fotheringham A. Single and Multiscale Models of Process Spatial Heterogeneity [J]. *Geographical Analysis*, 2018, 50(3): 223-246
- [51] Yu H, Fotheringham A S, Li Z, et al. Inference in Multiscale Geographically Weighted Regression[J]. *Geographical Analysis*, 2020, 52(1): 87-106
- [52] Wu C, Ren F, Hu W, et al. Multiscale Geographically and Temporally Weighted Regression: Exploring the Spatiotemporal Determinants of Housing Prices [J]. *International Journal of Geographical Information Science*, 2019, 33(3): 489-511

- [53] Huang B, Wu B, Barry M. Geographically and Temporally Weighted Regression for Modeling Spatio-Temporal Variation in House Prices[J]. *International Journal of Geographical Information Science*, 2010, 24(3): 383-401
- [54] Fotheringham A S, Crespo R, Yao J. Geographical and Temporal Weighted Regression (GTWR) [J]. *Geographical Analysis*, 2015, 47(4): 431-452
- [55] Wu B, Li R, Huang B. A Geographically and Temporally Weighted Autoregressive Model with Application to Housing Prices [J]. *International Journal of Geographical Information Science*, 2014, 28(5): 1 186-1 204
- [56] Du Z, Wu S, Zhang F, et al. Extending Geographically and Temporally Weighted Regression to Account for Both Spatiotemporal Heterogeneity and Seasonal Variations in Coastal Seas[J]. *Ecological Informatics*, 2018, 43: 185-199
- [57] Zhao Yangyang, Zhang Xiaolu, Zhang Fuhao, et al. A Local Polynomial Geographically and Temporally Weight Regression [J]. *Acta Geodaetica et Cartographica Sinica*, 2018, 47(5): 663-671(赵阳阳, 张小璐, 张福浩, 等. 一种局部多项式时空地理加权回归方法[J]. *测绘学报*, 2018, 47(5): 663-671)
- [58] Liu Y, Lam K, Wu J, et al. Geographically Weighted Temporally Correlated Logistic Regression Model [J]. *Scientific Reports*, 2018, 8(1): 1 417
- [59] Lesage J P. A Family of Geographically Weighted Regression Models [M]//Anselin L, Florax R J, Rey S J. *Advances in Spatial Econometrics*. Berlin, Heidelberg: Springer, 2002
- [60] Harris P, Fotheringham A S, Juggins S. Robust Geographically Weighted Regression: A Technique for Quantifying Spatial Relationships Between Freshwater Acidification Critical Loads and Catchment Attributes[J]. *Annals of the Association of American Geographers*, 2010, 100(2): 286-306
- [61] Wang N, Mei C L, Yan X D. Local Linear Estimation of Spatially Varying Coefficient Models: An Improvement on the Geographically Weighted Regression Technique[J]. *Environment and Planning A*, 2008, 40(4): 986-1 005
- [62] Harris P, Juggins S. Estimating Freshwater Acidification Critical Load Exceedance Data for Great Britain Using Space-Varying Relationship Models [J]. *Mathematical Geosciences*, 2011, 43 (3): 265-292
- [63] Nakaya T, Fotheringham A S, Brunsdon C, et al. Geographically Weighted Poisson Regression for Disease Association Mapping[J]. *Statistics in Medicine*, 2005, 24(17): 2 695-2 717
- [64] Atkinson P M, German S E, Sear D A, et al. Exploring the Relations Between Riverbank Erosion and Geomorphological Controls Using Geographically Weighted Logistic Regression [J]. *Geographical Analysis*, 2003, 35(1): 58-82
- [65] Harris R, Singleton A, Grose D, et al. Grid-Enabling Geographically Weighted Regression: A Case Study of Participation in Higher Education in England [J]. *Transactions in GIS*, 2010, 14(1): 43-61
- [66] Dong G, Nakaya T, Brunsdon C. Geographically Weighted Regression Models for Ordinal Categorical Response Variables: An Application to Geo-referenced Life Satisfaction Data [J]. *Computers, Environment and Urban Systems*, 2018, 70: 35-42
- [67] Li Z, Fotheringham A S, Li W, et al. Fast Geographically Weighted Regression (FastGWR): A Scalable Algorithm to Investigate Spatial Process Heterogeneity in Millions of Observations [J]. *International Journal of Geographical Information Science*, 2019, 33(1): 155-175
- [68] Murakami D, Tsutsumida N, Yoshida T, et al. Scalable GWR: A Linear-Time Algorithm for Large-Scale Geographically Weighted Regression with Polynomial Kernels [OL]. <https://arxiv.org/abs/1905.00266>, 2019
- [69] Harris P. A Simulation Study on Specifying a Regression Model for Spatial Data: Choosing Between Autocorrelation and Heterogeneity Effects [J]. *Geographical Analysis*, 2019, 51(2): 151-181
- [70] Charlton M, Fotheringham A, Brunsdon C. GWR 3: Software for Geographically Weighted Regression [OL]. <http://www.uvm.edu/rsenr/gradgis/GWR3Manual.pdf>, 2003
- [71] Nakaya T, Charlton M, Fotheringham S, et al. How to Use SGWRWIN (GWR4.0) [R]. Ireland: National Centre for Geocomputation, 2009
- [72] Oshan M T, Li Z, Kang W, et al. MGWR: A Python Implementation of Multiscale Geographically Weighted Regression for Investigating Process Spatial Heterogeneity and Scale [J]. *ISPRS International Journal of Geo-Information*, 2019, 8(6): 1-31
- [73] Lu B, Harris P, Charlton M, et al. The GWmodel R Package: Further Topics for Exploring Spatial Heterogeneity Using Geographically Weighted Models [J]. *Geo - Spatial Information Science*, 2014, 17 (2): 85-101
- [74] Eddelbuettel D, Francois R. Rcpp: Seamless R and C++ Integration [J]. *Journal of Statistical Software*, 2011, 1(8): 1-19

- [75] Wheeler D. GWRR: Fits Geographically Weighted Regression Models with Diagnostic Tools; R Package Version 0.2-1[OL]. <https://CRAN.R-project.org/package=gwrr>, 2013

## A Review on Geographically Weighted Regression

LU Binbin<sup>1,3</sup> GE Yong<sup>2</sup> QIN Kun<sup>1</sup> ZHENG Jianghua<sup>3</sup>

<sup>1</sup> School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China

<sup>2</sup> Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

<sup>3</sup> College of Resources and Environment Sciences, Xinjiang University, Urumqi 830046, China

**Abstract:** Spatial heterogeneity or non-stationarity in data relationships is one of the hot topics in spatial statistics or relative application fields, while the development of local techniques forms an essential part for the relative studies. Geographically weighted regression (GWR) provides spatially varying coefficient estimates via location-specific weighted regression model calibrations, to explore spatial heterogeneities or non-stationarities, quantitatively. It has been widely used in a number of fields, and become one of the most important tools for exploring spatial heterogeneities in data relationships. We summarized the GWR basics in model calibration, result interpretation, model diagnostics, reviewed its research progress and problems in its applications, respectively. Meanwhile, we sorted out the important extensions of the basic GWR technique, particularly in applying flexible distance metric choices in GWR model calibration, multiscale parameter estimates and spatiotemporal data modeling. In addition, we also introduced the main GWR tools or software accordingly to provide the users or readers comprehensive reference and knowledge on the GWR technique.

**Key words:** spatial heterogeneity; spatial non-stationarity; geographically weighted models; spatial statistics; spatial analysis

**First author:** LU Binbin, PhD, lecturer, specializes in spatial statistics, geographically weighted regression, geographically weighted models. E-mail: binbinlu@whu.edu.cn

**Foundation support:** The National Natural Science Foundation of China(41725006, 41871287, U1833201).

**引文格式:** LU Binbin, GE Yong, QIN Kun, et al. A Review on Geographically Weighted Regression[J]. Geomatics and Information Science of Wuhan University, 2020, 45(9):1356-1366. DOI:10.13203/j.whugis20190346(卢宾宾, 葛咏, 秦昆, 等. 地理加权回归分析技术综述[J]. 武汉大学学报·信息科学版, 2020, 45(9):1356-1366. DOI:10.13203/j.whugis20190346)