SoftwareX 21 (2023) 101291

Contents lists available at ScienceDirect

SoftwareX

journal homepage: www.elsevier.com/locate/softx

Original software publication

GWmodelS: A software for geographically weighted models

Binbin Lu [a], Yigong Hu [b], Dongyang Yang [c], Yong Liu [c], Liuqi Liao [a], Zuoyao Yin [a], Tianyang Xia [a], Zheyi Dong [a], Paul Harris [d], Chris Brunsdon [e], Lex Comber [f], Guanpeng Dong [c,g,h,*]

[a] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China
[b] School of Geographical Sciences, University of Bristol, Bristol, UK
[c] Key Research Institute of Yellow River Civilization and Sustainable Development, Henan University, Kaifeng, 475001, China
[d] Net Zero and Resilient Farming, Rothamsted Research, North Wyke, UK
[e] National Centre for Geocomputation, Maynooth University, Maynooth, Ireland
[f] School of Geography, University of Leeds, Leeds, UK
[g] Collaborative Innovation Center on Yellow River Civilization Jointly Built By Henan Province and Ministry of Education, Henan University, Kaifeng, 475001, China
[h] Key Laboratory of Geospatial Technology for the Middle and Lower Yellow River Regions, Ministry of Education, China

ARTICLE INFO

Article history:
Received 18 July 2022
Received in revised form 19 November 2022
Accepted 7 December 2022

Keywords:
Spatial heterogeneity
Spatio-temporal models
Visualization
High-performance
Local techniques

ABSTRACT

Spatial heterogeneity or non-stationarity has become a popular and necessary concern in exploring relationships between variables. In this regard, geographically weighted (GW) models provide a powerful collection of techniques in its quantitative description. We developed a user-friendly, high-performance and systematic software, named ***GWmodelS***, to promote better and broader usages of such models. Apart from a variety of GW models, including GW descriptive statistics, GW regression models, and GW principal components analysis, data management and mapping tools have also been incorporated with well-designed interfaces.

Metadata

Current code version	V1.0.3
Permanent link to code/repository used for this code version	https://github.com/ElsevierSoftwareX/SOFTX-D-22-00202
Permanent link to reproducible capsule	
Legal code licence	GNU
Code versioning system used	GitHub
Software code languages, tools and services used	C/C++, CMake
Compilation requirements, operating environments and dependencies	QT Creator, GDAL, QGIS, Armadillo
If available, link to developer documentation/manual	
Support email for questions	binbinlu@whu.edu.cn

1. Motivation and significance

Spatial heterogeneity or non-stationarity has drawn more and more attentions in data relationships [1]. In line with Tobler's first law of geography [2], a particular branch of spatial statistics, termed geographically weighted (GW) models have evolved to encompass local techniques applicable in situations when data cannot be described well by global models in which relationships between variables are often unrealistically assumed to be spatially invariant. In contrast, a suitably localized calibration method can provide a better description. Typical GW models and techniques include GW regression [3–6] and a number of its extensions [7–18], GW descriptive statistics [19,20], GW principal components analysis [4,21], GW discriminant analysis

* Corresponding author at: Key Research Institute of Yellow River Civilization and Sustainable Development, Henan University, Kaifeng, 475001, China.
E-mail address: gpdong@vip.henu.edu.cn (G. Dong).

https://doi.org/10.1016/j.softx.2022.101291

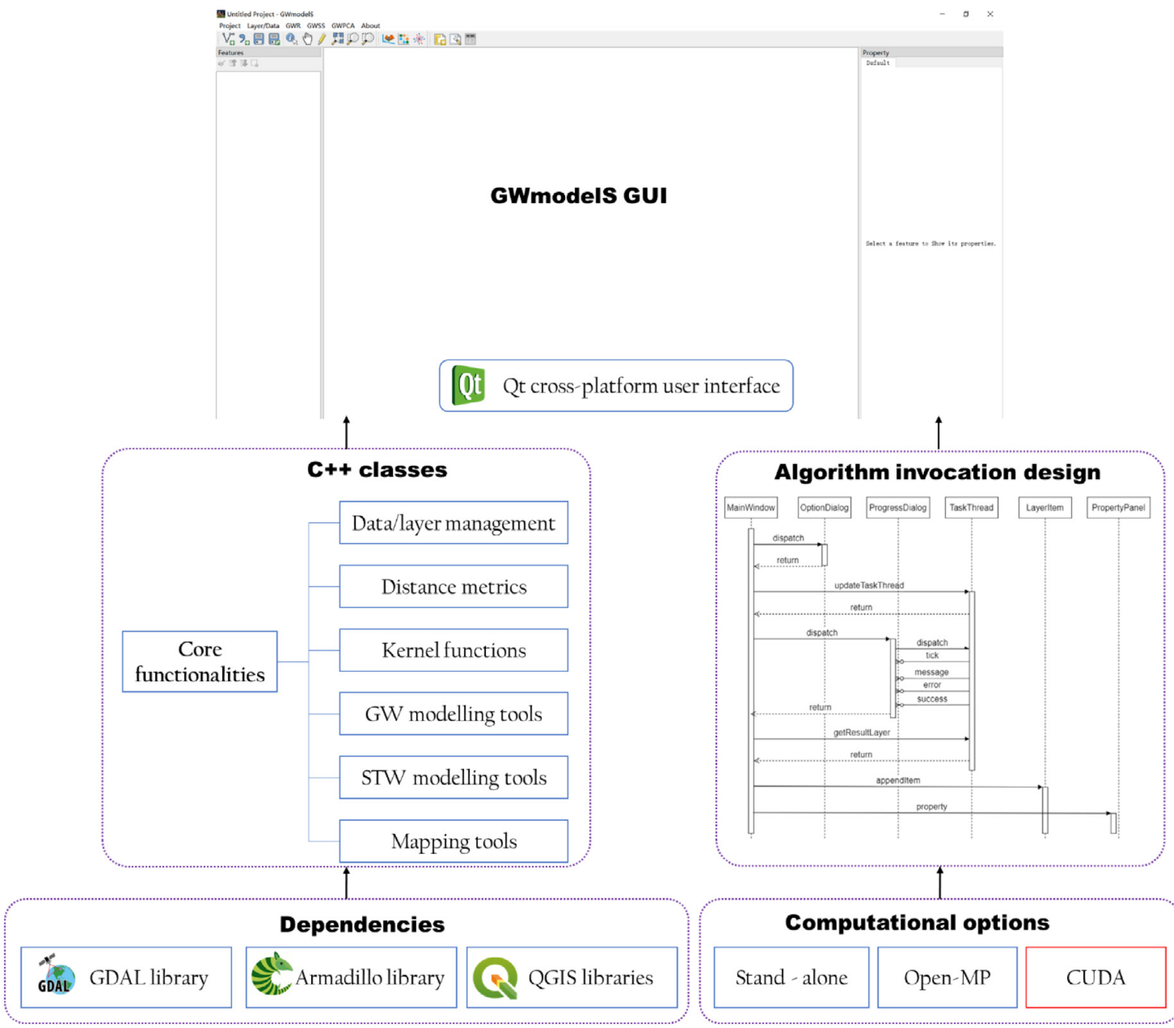

Fig. 1. Software architecture of ***GWmodelS***.

[22,23], GW visualization techniques [24] and GW artificial neural network [25,26]. These GW models form a generic, open, and continually evolving technical framework to explore spatial heterogeneities from a wide range of disciplines in the natural and social sciences.

Many of the listed GW models are incorporated into a range of R [27] packages, including **spgwr** [28], **mgwrsar** [18], **GWLelast** [29], **spMoran** [30], **gwer** [31], **lctools** [32,33], **gwrr** [34], **CARBayes** [35] and **GWmodel** [36,37]. In particular, **GWmodel** contains functions to calibrate and estimate a wide range models or techniques based on geographical weighting schemes. These include GW summary statistics, GW principal components analysis, GW discriminant analysis and various forms of GW regression techniques; some of which are provided in both basic and outlier resistant forms, associated tests and diagnostics, and options for flexible choices of distance metrics [38]. It has been downloaded more than 120,000 times (counted via the R package **cranlogs**) since its first release on CRAN in 2013, and attracting more and more attentions from a wide range of disciplines [39].

The users of **GWmodel**, however might frequently encounter problems when applying with it. Firstly, programming skills with **R** are required, particularly for various scholars from different disciplines. The current version of **GWmodel** manual is, at the time of writing (June 2019), some 85 pages in length and organized alphabetically. Whilst these write-ups conform to the CRAN guidelines, they can be hard to follow. Notably, an R shiny package, namely GeoWeightedModel [40] has been developed to specifically provide a graphical user interface for GW functionalities in **GWmodel**. Secondly, memory and computational limits explicitly exist in R, although high-performance solutions have been expediently developed with multi-core or compute unified device architecture configurations [41]. This natural limitation of **R** tend to lead the efficiency of high-performance solutions to be compromised. Thirdly, data processing and mapping are necessary operations for GW models, but relevant utility tools are not available in **GWmodel**. External packages (e.g. ggplot2 [42]) or tools (e.g. ArcGIS [43]) are usually turned to, which induces even more difficulties in the usages of GW models. As such, a user-friendly, high-performance and integrated software is urgently demanded to implement GW models. We thus aimed to develop a stand-alone software, namely ***GWmodelS***, fully solving the above concerns.

2. Software description

2.1. Software architecture

The development architecture of ***GWmodelS*** is shown in Fig. 1. We adopted functions from the geospatial data abstraction library (GDAL) [44], Armadillo C++ library [45] and QGIS libraries [46] to achieve fundamental functionalities in processing geospatial data, computing linear algebra and geovisualization. With these

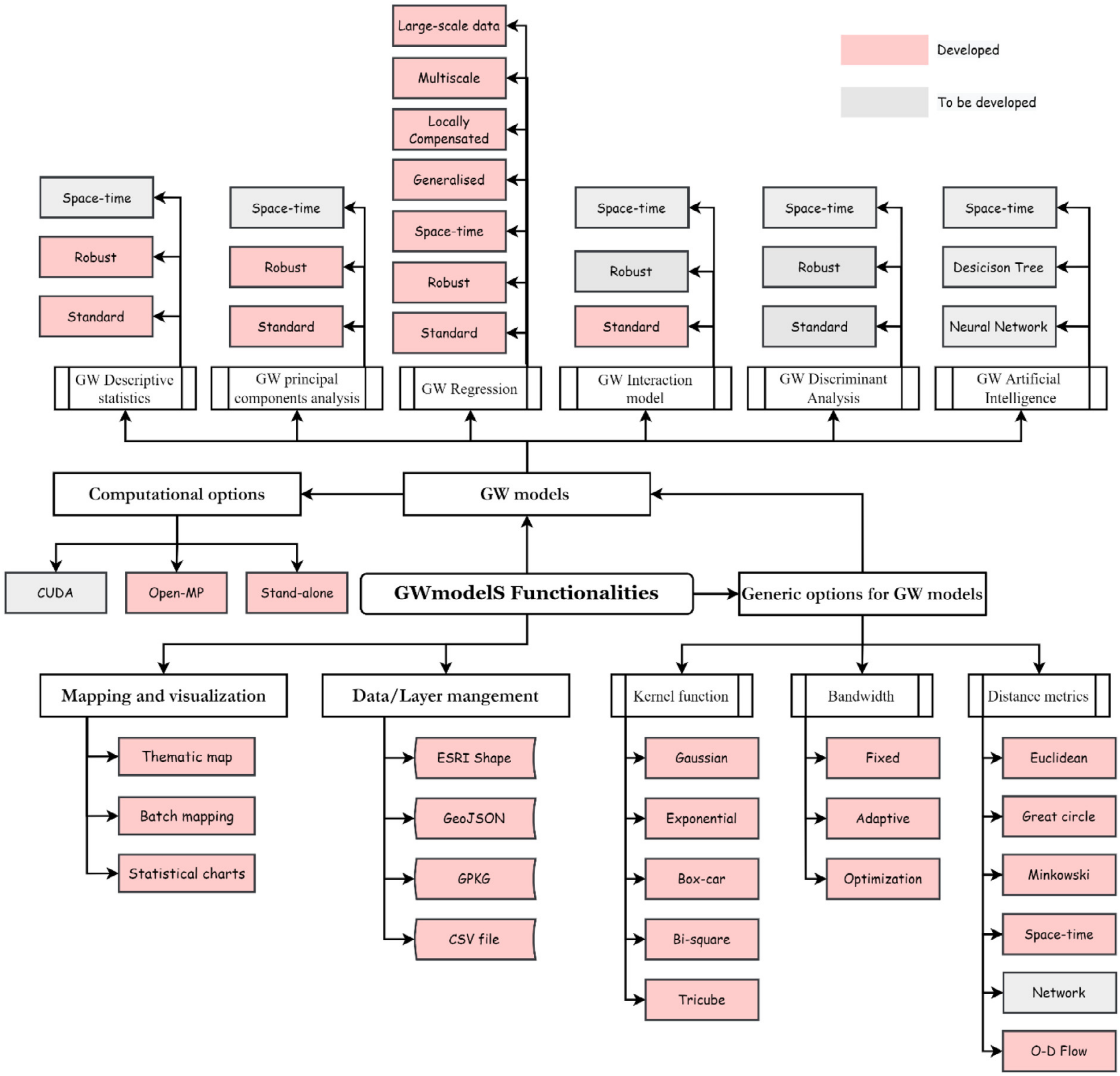

Fig. 2. An ontology of functionalities (developed or to be developed) in ***GWmodelS***.

dependencies, the core functionalities in data/layer management, generic computations of distance metrics and spatial weights (i.e. kernel functions), GW modelling, spatially and temporally weighted (STW) models and mapping tools were programmed as C++ classes. Finally, we incorporated all these modules and algorithms via the cross-platform framework Qt to produce a friendly graphic user interface (GUI) of ***GWmodelS***. Moreover, the algorithm invocations were canonically designed to facilitate its usage and further extensions. In the current release of ***GWmodelS*** (Version 1.0.3), we provide two options, stand-alone or open multi-processing (Open-MP) to compute most of the GW models, and the compute unified device architecture (CUDA) solution to accelerate the computations with graphics processing unit (GPU) devices is also under development.

2.2. Software functionalities

In Fig. 2, we present an ontology diagram of the functionalities developed or to be developed in ***GWmodelS***. The key features of this software are listed as follows:

(1) Data/layer management: geospatial data of different formats, including ESRI shape file, GeoJOSN and GPKG could be imported into *GWmodelS*, while much more choices could be provided on request by means of GDAL. Moreover, the table file in the CSV format could be also used for interchange.
(2) Generic options for GW models: In ***GWmodelS***, we provided a substantial collection of options to calculate geographical weights, one of the core components for calibrating any GW model. In details, a number of distance metrics, including Euclidean distance, Great circle distance, Minkowski distance [38], space–time distance [12], network distance [47] and flow distance [48] can be calculated to fit various GW models. Moreover, a kernel function (e.g. Gaussian, Exponential, Box-car, Bi-square and Tricube) with a specific bandwidth (fixed or adaptive, optimized or predetermined) provide a flexible weighting scheme for various application scenarios.
(3) GW models: A number of GW modules have been incorporated into the current release of ***GWmodelS***, including GW descriptive statistics [19], GW principal components

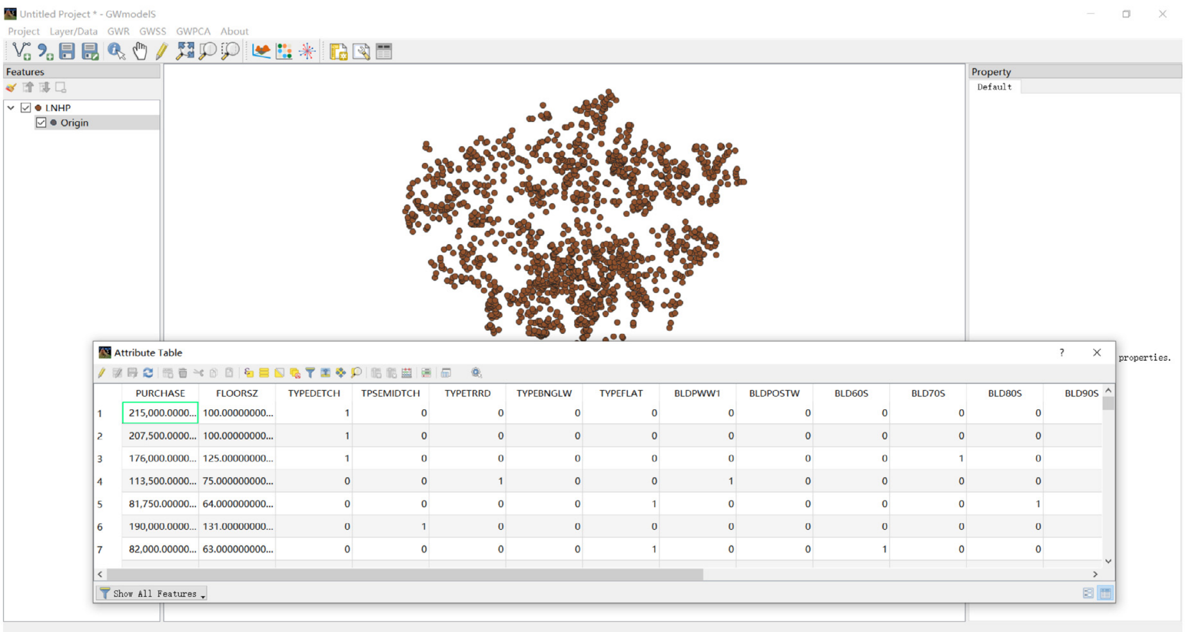

Fig. 3. Load the LNHP data into ***GWmodelS***.

analysis [21], standard GW regression [6] and a number of extensions applicable to large-scale data set [49], multi-scale GW regression [9,50,51], geographically and temporally weighted regression [12], robust GW regression [13], locally compensated GW regression, generalized GW regression [8] and GW interaction model for analysing origin–destination flows [48]. Moreover, GW discriminant analysis [23] and GW artificial intelligence techniques [25] are also tentatively coded and to be incorporated in the near future.

(4) Computational options: In practice, stand-alone computing is fine for most of the GW models, particular when the sample size is small (e.g. less than 3000) [41]. With concerning large scale data, high-performance options, including multi-core parallel and CUDA techniques are also seamlessly integrated in this software.

(5) Mapping and visualization: Apart from conventional tools, like thematic map and statistical charts, some specific visualization tools are also incorporated, like multivariate glyph plot [36] and circle view for visualizing model specification [47]. To facilitate the multivariate mapping, we developed a batch mapping tool, i.e. setting all the cartographical parameters as one-off and producing all the maps accordingly.

3. An illustrative example of standard GW regression

With the ***GWmodelS*** installed, an example data set could be found in the installation directory "sample_data", i.e. a house price data set for London, UK (LNHP) (see details in [47]). Load it into ***GWmodelS*** by clicking "Layer/Data->ESRI Shapefile", the data is ready to be examined, as shown in Fig. 3. Using QGIS libraries, navigation, editor and attribute viewer tools are also available to preliminarily explore the data, which is always necessary for further modelling with it.

In this section, we take the standard GW regression as an illustrative example, more sophisticated examples can be found in the attached video demo. By clicking the menu bar "GWR->Basic GWR" or the button in the toolbar, we can see the form to configure parameters for calibrating a basic GW regression model,

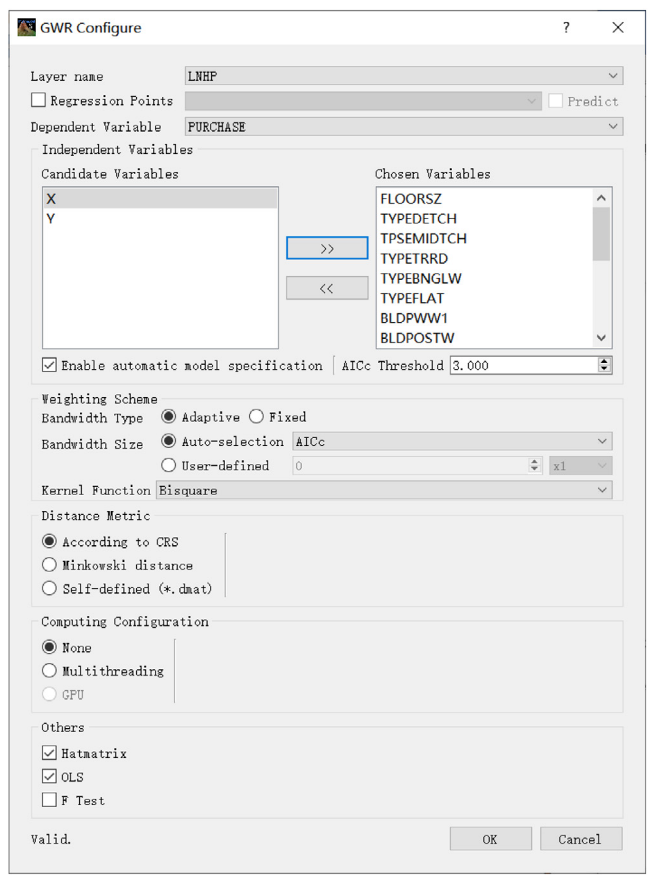

Fig. 4. Parameter configuration forms for calibrating a standard GW regression model.

as shown in Fig. 4. All the parameters could be set up in the following steps:

(1) Select the data layer from a combobox, i.e. LNHP; tick the checkbox "Regression Points" if you have a different layer as regression points, but keep unchecked in this example.

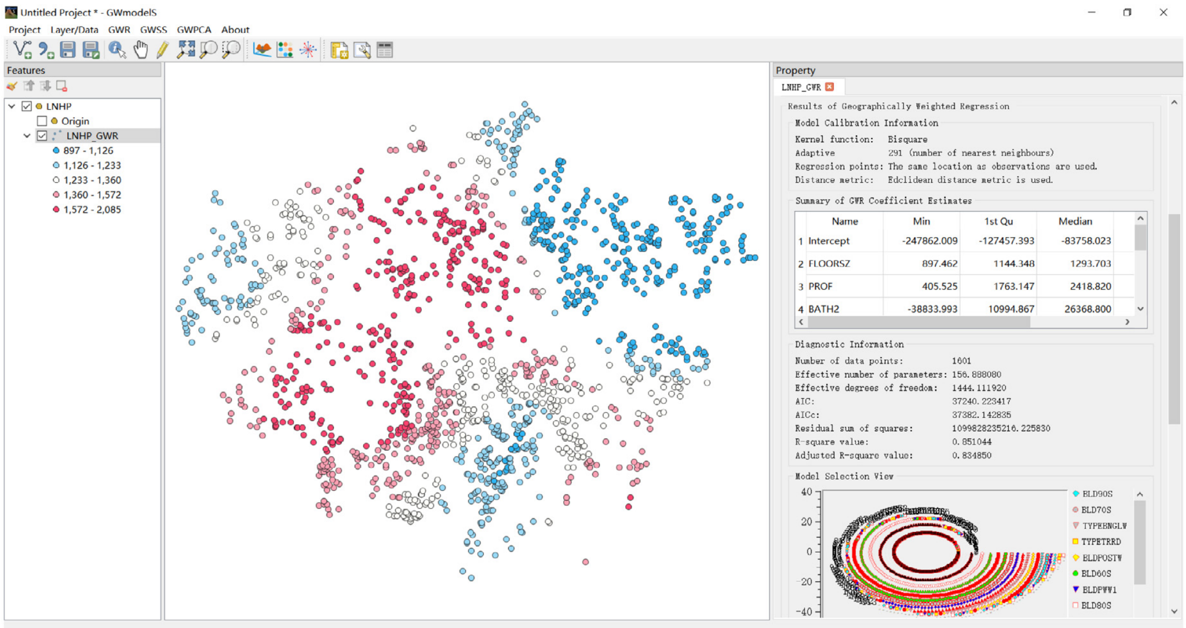

Fig. 5. Results returned from a standard GW regression model.

(2) Specify dependent variable and independent variables from the combobox and list box; if tick the checkbox "Enable automatic model specification", the independent variables will be automatically selected via a stepwise procedure (see details in [47]), and it is checked in this example.
(3) Define the weighting scheme by ticking the radio buttons, i.e. fixed or adaptive bandwidth, user-defined or optimized via the cross validation (CV) approach or corrected Akaike Information Criterion (AICc) [37], and finally choosing the kernel function from the combobox; the Bi-square kernel function with an optimized adaptive bandwidth is adopted in this example.
(4) Tick the radio button to calculate distance metric, where "According to CRS" means great circle and Euclidean distances will be calculated when the coordinate reference system (CRS) is geographic and projected, respectively; Minkowski distance or an individual distance matrix is also allowable by ticking the rest radio buttons. The first choice is taken by default in this example.
(5) Configure the computing option by ticking the radio buttons "None" or "Multithreading" to define whether parallel computing is adopted or not; Note the stand-alone mode is used in this example, as the sample size of LNHP is 1601.
(6) Extra options are available in the "Others" panel, where diagnostic information, results of ordinary least squares regression and significance tests of spatial heterogeneity [52] will be returned if the checkboxes "Hatmatrix", "OLS" and "F tests" are ticked, respectively; default choices are kept in this example.
(7) Clicking on the "OK" button, the results of the GW regression model will be returned as a new layer (including location-wise coefficient estimates) and summary information in the property panel once the computation is completed successfully, as shown in Fig. 5.

4. Impact

As a concomitant software with the R package **GWmodel**, ***GWmodelS*** will provide quite different experiences in applying GW models in practice, as,

(1) Friendly interface and fundamental GIS tools integrated allow users to easily run all the GW models, even though their programming skills might be zero.
(2) A wide range of geographically weighted models have been incorporated, and this number is still increasing quickly as more and more GW models and extensions are developed.
(3) High performance techniques, including Open-MP and CUDA are seamlessly integrated for all the GW models, which are important for modelling with large scale data.
(4) Visualization tools including batch mapping, multivariate glyph plot and circle view are specifically incorporated for mapping spatially varying results from GW models.

All in all, ***GWmodelS*** present great advantages in GUI, operability, computational efficiency and accessibility. This is how we expect ***GWmodelS*** to assist much more users from a wide range of domains in applying GW models in their quantitative studies.

5. Conclusions

GW models have become an important branch of local techniques to quantitatively explore spatial heterogeneities and non-stationarities in data relationships. The early versions of **GWmodel** have acquired more than 400 citations (according to records in google scholar) as a preferred GW toolkit in a wide range of studies, e.g. meteorology [53,54], urban studies [55,56], and environmental science [57–61]. To promote better and broader usages of GW models, we developed this user-friendly, high-performance integrated and bigeminal software. The release of ***GWmodelS*** is a good start, and follow-up updates and tutorials have been under long-term planning.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

we would like to thank and pay tribute to Dr. Martin Charlton for his extraordinary contributions in developing the earliest GWR software and R package GWmodel. This study is jointly funded by National Key Research and Development Program of China (No. 2021YFB3900904), the National Natural Science Foundation of China (42071368, 42001115) and Zhizhuo Research Fund on Spatial-Temporal Artificial Intelligence (Grant No. ZZJJ202203).

References

[1] Goodchild MF. The validity and usefulness of laws in geographic information science and geography. Ann Am Assoc Geogr 2004;94(2):300–3.
[2] Tobler WR. A computer movie simulating urban growth in the detroit region. Econ Geogr 1970;46(2):234–40.
[3] Fotheringham AS, Charlton ME, Brunsdon C. Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. Environ Plan A 1998;30(11):1905–27.
[4] Fotheringham AS, Brunsdon C, Charlton M. Geographically weighted regression: the analysis of spatially varying relationships. Chichester: Wiley; 2002.
[5] Brunsdon C, Fotheringham S, Charlton M. Geographically weighted regression-modelling spatial non-stationarity. J R Statist Soc Ser D (Statist) 1998;47(3):431–43.
[6] Brunsdon C, Fotheringham AS, Charlton ME. Geographically weighted regression: A method for exploring spatial nonstationarity. Geogr Anal 1996;28(4):281–98.
[7] Wang N, Mei CL, Yan XD. Local linear estimation of spatially varying coefficient models: an improvement on the geographically weighted regression technique. Environ Plan A 2008;40(4):986–1005.
[8] Nakaya T, et al. Geographically weighted Poisson regression for disease association mapping. Stat Med 2005;24(17):2695–717.
[9] Lu B, et al. Geographically weighted regression with parameter-specific distance metrics. Int J Geogr Inf Sci 2017;31(5):982–98.
[10] Liu Y, et al. Geographically weighted temporally correlated logistic regression model. Sci Rep 2018;8(1):1417.
[11] Leong Y-Y, Yue JC. A modification to geographically weighted regression. Int J Health Geogr 2017;16(1):11.
[12] Huang B, Wu B, Barry M. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. Int J Geogr Inf Sci 2010;24(3):383–401.
[13] Harris P, Fotheringham AS, Juggins S. Robust geographically weighted regression: A technique for quantifying spatial relationships between freshwater acidification critical loads and catchment attributes. Ann Am Assoc Geogr 2010;100(2):286–306.
[14] Fotheringham AS, Yang W, Kang W. Multiscale geographically weighted regression (MGWR). Ann Am Assoc Geogr 2017;107(6):1247–65.
[15] Dong G, Nakaya T, Brunsdon C. Geographically weighted regression models for ordinal categorical response variables: An application to geo-referenced life satisfaction data. Comput Environ Urban Syst 2018;70:35–42.
[16] Comber A, Harris P. Geographically weighted elastic net logistic regression. J Geogr Syst 2018;20(4):317–41.
[17] Cysneiros FJA, Paula GA, Galea M. Heteroscedastic symmetrical linear models. Statist Probab Lett 2007;77(11):1084–90.
[18] Geniaux G, Martinetti D. A new method for dealing simultaneously with spatial autocorrelation and spatial heterogeneity in regression models. Reg Sci Urban Econ 2018;72:74–85.
[19] Brunsdon C, Fotheringham AS, Charlton M. Geographically weighted summary statistics – a framework for localised exploratory data analysis. Comput Environ Urban Syst 2002;26(6):501–24.
[20] Harris P, Charlton M, Fotheringham A. Moving window kriging with geographically weighted variograms. Stoch Environ Res Risk Assess 2010;24(8):1193–209.
[21] Harris P, Brunsdon C, Charlton M. Geographically weighted principal components analysis. Int J Geogr Inf Sci 2011;25(10):1717–36.
[22] Foley P, Demšar U. Using geovisual analytics to compare the performance of geographically weighted discriminant analysis versus its global counterpart, linear discriminant analysis. Int J Geogr Inf Sci 2012;27(4):633–61.
[23] Brunsdon C, Fotheringham S, Charlton M. Geographically weighted discriminant analysis. Geogr Anal 2007;39(4):376–96.
[24] Dykes J, Brunsdon C. Geographically weighted visualisation: Interactive graphics for scale-varying exploratory analysis. IEEE Trans Vis Comput Graph 2007;13:1161–8.
[25] Hagenauer J, Helbich M. A geographically weighted artificial neural network. Int J Geogr Inf Sci 2021;1–21.
[26] Du Z, et al. Geographically neural network weighted regression for the accurate estimation of spatial non-stationarity. Int J Geogr Inf Sci 2020;34(7):1353–77.
[27] R Core Team. Memory limits in R. In: Seminar for statistics 2020. 2020, [cited 2020 May 21st]; Available from: https://stat.ethz.ch/R-manual/R-devel/library/base/html/Memory-limits.html.
[28] Bivand R, Yu D. Package ‘spgwr’. CRAN; 2020.
[29] Yoneoka D, Saito E, Nakaoka S. New algorithm for constructing area-based index with geographical heterogeneities and variable selection: An application to gastric cancer screening. Sci Rep 2016;6(1):26582.
[30] Murakami D. Spmoran: An r package for Moran’s eigenvector-based spatial regression analysis. 2017, arXiv:1703.04467v3.
[31] Araujo YA, Cysneiros FJA, Cysneiros AHMA. Gwer: Geographically weighted elliptical regression. 2020.
[32] Kalogirou S. Lctools: Local correlation, spatial inequalities. In: Geographically weighted regression and other tools. CRAN; 2020.
[33] Kalogirou S. Spatial analysis: methodology and applications with R. Greek, Athens: Hellenic Academic Libraries Link; 2015.
[34] Wheeler D. gwrr: Fits geographically weighted regression models with diagnostic tools. 2013, R package version 0.2-1.
[35] Lee D. CARBayes : An r package for spatial areal unit modelling with conditional autoregressive priors. 2015.
[36] Lu B, et al. The GWmodel R package: further topics for exploring spatial heterogeneity using geographically weighted models. Geo-Spat Inf Sci 2014;17(2):85–101.
[37] Gollini I, et al. GWmodel: an R package for exploring spatial heterogeneity using geographically weighted models. J Stat Softw 2015;63(17):1–50.
[38] Lu B, et al. The Minkowski approach for choosing the distance metric in geographically weighted regression. Int J Geogr Inf Sci 2016;30(2):351–68.
[39] Comber A, et al. A route map for successful applications of geographically weighted regression. Geogr Anal 2022.
[40] Maestre JDLH, Gómez MJF, Mendez S. GeoWeightedmodel: user-friendly interface for geographically-weighted models. CRAN; 2022.
[41] Lu B, et al. High-performance solutions of geographically weighted regression in R. Geo-Spat Inf Sci 2022;1–14, Online.
[42] Wickham H. Ggplot2: Elegant graphics for data analysis. New York: Springer-Verlag; 2016.
[43] ESRI. ArcGIS spatial analyst: Advanced GIS spatial analysis using raster and vector data. In: Esri white paper. New York: Esri; 2001.
[44] GDAL/OGR contributors. GDAL/OGR geospatial data abstraction software library. Open Source Geospatial Foundation; 2022.
[45] Sanderson C, Curtin R. Armadillo: a template-based C++ library for linear algebra. J Open Source Softw 2016;1:1–26.
[46] QGIS.org QGIS 322. Geographic information system developers manual. QGIS Association; 2022.
[47] Lu B, et al. Geographically weighted regression with a non-Euclidean distance metric: a case study using hedonic house price data. Int J Geogr Inf Sci 2014;28(4):660–81.
[48] Kordi M, Fotheringham AS. Spatially weighted interaction models (SWIM). Ann Am Assoc Geogr 2016;106(5):990–1012.
[49] Murakami D, et al. Scalable GWR: A linear-time algorithm for large-scale geographically weighted regression with polynomial kernels. Ann Am Assoc Geogr 2020;1–22.
[50] Lu B, et al. Improvements to the calibration of a geographically weighted regression with parameter-specific distance metrics and bandwidths. Comput Environ Urban Syst 2018;71:41–57.
[51] Lu B, et al. A response to ‘A comment on geographically weighted regression with parameter-specific distance metrics’. Int J Geogr Inf Sci 2019;33(7):1300–12.
[52] Leung Y, Mei CL, Zhang WX. Statistical tests for spatial nonstationarity based on the geographically weighted regression model. Environ Plan A 2000;32(1):9–32.
[53] Bai H, et al. Spatiotemporal change in livestock population and its correlation with meteorological disasters during 2000–2020 across inner mongolia. ISPRS Int J Geo-Inf 2022;11. http://dx.doi.org/10.3390/ijgi11100520.

[54] Nawrotzki RJ, Bakhtsiyarava M. International climate migration: Evidence for the climate inhibitor mechanism and the agricultural pathway. Popul Space Place 2017;23(4):e2033.
[55] Tu W, et al. Portraying the spatial dynamics of urban vibrancy using multisource urban big data. Comput Environ Urban Syst 2020;80:101428.
[56] Li Q, et al. The spatial varying relationship between services of the train network and residential property values in Melbourne, Australia. Urban Stud 2020;0042098019896977.
[57] Samec P, et al. Spatial lag effect of aridity and nitrogen deposition on Scots pine (Pinus sylvestris L.) damage. Environ Pollut 2020;265:114352.
[58] Li C, Managi S. Global malaria infection risk from climate change. Environ Res 2022;214:114028.
[59] Bi J, et al. Incorporating low-cost sensor measurements into high-resolution PM2.5 modeling at a large spatial scale. Environ Sci Technol 2020;54(4):2152–62.
[60] Meng X, Pi H, Gao X, He P, Lei J. A high-accuracy vegetation restoration potential mapping model integrating similar habitat and machine learning. Land Degrad. Dev. 2022;1–17. http://dx.doi.org/10.1002/ldr.4527.
[61] Dong G, et al. Developing a locally adaptive spatial multilevel logistic model to analyze ecological effects on health using individual census records. Ann Am Assoc Geogr 2020;110(739–757).