The Journal of Open Source Software

XLUR: A land use regression wizard for ArcGIS Pro

Anna Molter[1]

1 Department of Geography, School of Environment, Education and Development, The University of Manchester

DOI: 10.21105/joss.02177

Software

- Review
- Repository
- Archive

Editor: Hugo Ledoux

Reviewers:

- @Athelena
- @fortinma

Submitted: 19 September 2019
Published: 11 June 2020

Summary

Introduction

XLUR is a Python toolbox for ArcGIS Pro (v2.2.4 or higher, Environmental Systems Research Institute (ESRI), Redlands, CA) that enables the development and application of land use regression models via a wizard style interface. Land use regression (LUR) is a commonly used technique in environmental sciences to analyse factors influencing pollutant levels and to predict pollutant levels at unmeasured locations. LUR is extensively used in studies on air pollution exposure (Molter et al., 2010a, 2010b), but it is widely applicable and has been used in fields ranging from water pollution (Kelsey, Porter, Scott, Neet, & White, 2004) to urban climatology (Heusinkveld, Steeneveld, Hove, Jacobs, & Holtslag, 2014).

Purpose of the software

The traditional approach to the development of a LUR model requires several steps: 1. The creation of a point dataset of known observation data for a variable of interest (such as nitrogen dioxide concentrations or air temperatures), which will be used as outcome variables. 2. Carrying out various spatial analyses of additional geospatial data with these point locations using geographic information systems (GIS) to extract potential predictor variables. 3. Data wrangling of extracted data from steps 1 and 2 into a format that can be readily used by statistical software packages. 4. Carrying out multiple regression analysis to obtain a best fit parsimonious model.

If done manually, steps 2 to 4 are repetitive and time consuming, making them inefficient and prone to error. XLUR provides a wizard style interface that guides a user through the development of a LUR model without the need to access and run multiple tools and additional software packages. XLUR largely automates steps 2 to 4, which significantly speeds up the model development process and reduces user error. Depending on the type and amount of data to be processed, and the available hardware, models can be developed in less than one hour. Furthermore, the output files produced by XLUR ensure that the model development is well documented and reproducible. Lastly, XLUR also makes the method available for a wider range of users.

In addition to developing LUR models, XLUR can also apply a previously developed model to a new set of locations within the same study area. New locations can be defined by the user or can be based on a dataset containing regularly located points or randomly located points. Again, XLUR largely automates this process with minimal effort from the user.

XLUR is aimed at GIS specialists. It uses the ArcGIS Pro software, which is the most widely used commercial GIS software worldwide. XLUR is based on the LUR methodology used in the European Study of Cohorts for Air Pollution Effects (ESCAPE) (Beelen et al., 2013;

Eeftens et al., 2012) as set out in the ESCAPE Exposure assessment manual (2010). Within air pollution research the ESCAPE methodology is used as the standard for developing LUR models. XLUR also allows hybrid LUR models to be developed, based on an extension of the ESCAPE methodology that included the addition of satellite derived data and data from chemical transport models (Hoogh et al., 2016).

Current application

XLUR has been developed through the NERC Newton-DIPI funded Urban hybriD models for AiR pollution exposure Assessment (UDARA) study, which is a collaboration between the University of Manchester and Institut Teknologi Bandung. This study aims to develop air pollution prediction models for Indonesian urban areas and to analyse the effects of air pollution on health indicators provided by the Indonesian Family Life Survey.

State of the field

Currently, only a small number of LUR software packages have been developed. RLUR (Morley & Gulliver, 2018) has been developed in the R programming language and is aimed at users with a background in statistical analyses, such as exposure scientists or epidemiologists. PyLUR (Ma, Longley, Salmond, & Gao, 2020) is written in Python, but not implemented within the ArcGIS software. Instead, it uses GDAL/OGR libraries for spatial analysis. The authors report that PyLUR currently does not have a user-friendly graphic user interface (GUI) and at the time of writing it is not available in an open source repository. OpenLUR (Lautenschlager et al., 2020) is designed to develop LUR models exclusively based on OpenStreetMap data. Unlike XLUR, RLUR and PyLUR it does not use the ESCAPE methodology, but an unsupervised machine learning process featuring automated hyper-parameter tuning.

One major difference between these LUR software packages and XLUR is that they are designed specifically for air pollution models, i.e. their spatial analyses only extract potential predictors relevant for air pollution modelling. In contrast, XLUR is more widely applicable and its wizard style interface can be used to extract potential predictor variables for a range of environmental phenomena. Another major difference is that XLUR can develop classic LUR models and hybrid LUR models, that add a measure of global variability to the measures of local variability modelled in LUR. Furthermore, XLUR is the only software that is implemented within ArcGIS Pro.

Availability and implementation

XLUR is a Python toolbox for use within ArcGIS Pro (2020). An ArcGIS Pro Project file (XLUR.aprx) containing the XLUR toolbox is available on the GitHub repository (https://github.com/anmolter/XLUR). The repository also provides the source code of the tools, a user manual, an example dataset for the tutorial in the user manual, and example outputs. The user manual contains instructions for installing additional Python packages (wxpython, statsmodels, seaborn, patsy) required by XLUR via ArcGIS Pro's Python Package Manager.

Acknowledgments

XLUR was reviewed and tested by Prof S Lindley, Department of Geography, School of Environment, Education and Development, The University of Manchester. This work is supported via the NERC Newton-DIPI Urban hybriD models for AiR pollution exposure Assessment

(UDARA) project, PIs: Prof G McFiggans, Faculty of Science and Engineering, The University of Manchester, UK, and Dr D Driejana, Faculty of Civil and Environmental Engineering, Institut Teknologi Bandung, Indonesia, NE/P014631/1. It builds on work carried out in the European Union's Seventh Framework Programme Theme ENV.2007.1.2.2.2. European cohort on air pollution.

References

Beelen, R., Hoek, G., Vienneau, D., Eeftens, M., Dimakopoulou, K., Pedeli, X., Tsai, M. Y., et al. (2013). Development of no2 and nox land use regression models for estimating air pollution exposure in 36 study areas in europe - the escape project. *Atmospheric Environment*, *72*, 10–23. Journal Article. doi:10.1016/j.atmosenv.2013.02.037

Eeftens, M., Beelen, R., Hoogh, K. de, Bellander, T., Cesaroni, G., Cirach, M., Declercq, C., et al. (2012). Development of land use regression models for pm(2.5), pm(2.5) absorbance, pm(10) and pm(coarse) in 20 european study areas; results of the escape project. *Environ Sci Technol*, *46*(20), 11195–205. Journal Article. doi:10.1021/es301948k

Heusinkveld, B. G., Steeneveld, G. J., Hove, L. W. A. van, Jacobs, C. M. J., & Holtslag, A. A. M. (2014). Spatial variability of the rotterdam urban heat island as influenced by urban land use. *Journal of Geophysical Research-Atmospheres*, *119*(2), 677–692. Journal Article. doi:10.1002/2012jd019399

Hoogh, K. de, Gulliver, J., Donkelaar, A. van, Martin, R. V., Marshall, J. D., Bechle, M. J., Cesaroni, G., et al. (2016). Development of west-european pm2.5 and no2 land use regression models incorporating satellite-derived and chemical transport modelling data. *Environmental Research*, *151*, 1–10. Journal Article. doi:10.1016/j.envres.2016.07.005

Kelsey, H., Porter, D. E., Scott, G., Neet, M., & White, D. (2004). Using geographic information systems and regression analysis to evaluate relationships between land use and fecal coliform bacterial pollution. *Journal of Experimental Marine Biology and Ecology*, *298*(2), 197–209. Journal Article. doi:10.1016/S0022-0981(03)00359-9

Lautenschlager, F., Becker, M., Kobs, K., Steininger, M., Davidson, P., Krause, A., & Hotho, A. (2020). OpenLUR: Off-the-shelf air pollution modeling with open features and machine learning. *Atmospheric Environment*, *233*, 117535. Journal Article. doi:10.1016/j.atmosenv.2020.117535

Ma, X., Longley, I., Salmond, J., & Gao, J. (2020). PyLUR: Efficient software for land use regression modeling the spatial distribution of air pollutants using gdal/ogr library in python. *Frontiers of Environmental Science & Engineering*, *14*(3), 44. Journal Article. doi:10.1007/s11783-020-1221-5

Molter, A., Lindley, S., Vocht, F. de, Simpson, A., & Agius, R. (2010a). Modelling air pollution for epidemiologic research–part i: A novel approach combining land use regression and air dispersion. *Sci Total Environ*, *408*(23), 5862–9. Journal Article. doi:10.1016/j.scitotenv.2010.08.027

Molter, A., Lindley, S., Vocht, F. de, Simpson, A., & Agius, R. (2010b). Modelling air pollution for epidemiologic research–part ii: Predicting temporal variation through land use regression. *Sci Total Environ*, *409*(1), 211–7. Journal Article. doi:10.1016/j.scitotenv.2010.10.005

Morley, D. W., & Gulliver, J. (2018). A land use regression variable generation, modelling and prediction tool for air pollution exposure assessment. *Environmental Modelling & Software*, *105*, 17–23. Journal Article. doi:10.1016/j.envsoft.2018.03.030

(2010). Web Page. Retrieved from http://www.escapeproject.eu/manuals/ESCAPE_Exposure-manualv9.pdf

(2020). Web Page. Retrieved from https://pro.arcgis.com/en/pro-app/get-started/get-started.htm