Geographically Weighted Structural Equation Models: spatial variation in the drivers of environmental restoration effectiveness

Alexis Comber
School of Geography, University of Leeds, Leeds, LS2 9JT, UK
a.comber@leeds.ac.uk

Ting Li
Chinese Academy of Sciences, Beijing, 100049, China
tli_st@rcees.ac.cn

Yihe Lü
Chinese Academy of Sciences, Beijing, 100085, China
lyh@rcees.ac.cn

Bojie Fu
Chinese Academy of Sciences, Beijing, 100049, China
bfu@rcees.ac.cn

Paul Harris
Rothamsted Research, EX20 2SB, UK
paul.harris@rothamsted.ac.uk

Abstract

This paper describes a methodological extension to Geographically Weighted (GW) models. It develops and applies a GW structural equation model (SEM) to understand the observed and latent drivers associated with effective landscape restoration in Northern China. The paper reviews recent landscape restoration activities in China, and describes the environmental context of these: soil loss, erosion and land degradation. Restoration effectiveness was described by changes in net primary production and fractional vegetation cover, as recorded in MODIS data for the period 2000-2012. County level census data provided information on hypothesised latent variables of *population pressure*, *off-farm economy* and *rural economy*. The GW SEM analysis allows the spatial variation in the contributions made by different socio-economic factors to restoration effectiveness to be evaluated. Although developed with ropey data – the County level population totals were perhaps unrealistically interpolated over a 2km grid and the MODIS data were aggregated over the same – the GW SEM allows the detail of the how what and where to be identified thus supporting local policy and planning. A number of future developments in refining this method are outlined.
Keywords: GWR; soil erosion; Loess Plateau; Critical Zone

1 Introduction

This paper presents a novel application of the Geographically Weighted (GW) paradigm (Brunsdon et al., 1996; Gollini et al., 2015). GW approaches seek to explore spatial heterogeneity in processes and relationships by developing a series of local analysis rather than taking a global or 'whole map' approach. Here, GW approaches are applied to Structural Equation Models (SEMs) as described in Ullman and Bentler (2003) and Kline (2015). SEMs are a series of linked regressions are typically used to identify explanatory but hidden (or latent) variables and relationships. In this paper they are used to explore the relationships associated with the effectiveness of vegetation restoration (Grace 2006) for a case study in China related to socioeconomic changes. There are many applications of SEMs in the literature and Fan et al (2016) provide a review their use in ecological applications. The rationale for developing Geographically Weighted Structural Equation Models (GW-SEM) models derives from the problem in identifying the drivers of ecological restoration effectiveness and, critically, how these vary spatially.

Ecological restoration has been found to enhance biodiversity and landscape functionality (Clewell and Aronson, 2013) in support of sustainability. Much work has sought evaluate the effectiveness of different ecological restoration programmes at different scales (eg Felton *et al.* 2010; Calmon *et al.* 2011; Meli *et al.* 2014; Li et al., submitted). One of the critical issues in evaluation is that large scale assessments of restoration effectiveness is difficult (Martin *et al.* 2014). This is due to a lack of knowledge of how local socio-economic factors interact with ecological activities and a lack of robust and agreed metrics for quantifying ecological spatio-temporal changes (Li et al., submitted). In part this is because not all restoration activities have the same temporal dimension, for example forest ecosystems recover in ~50 years whereas grassland ecosystems recover in ~25 years (Jones and Schmitz, 2009), but it is also because little research has established suitable methods for characterising restoration activities over time (Berkowitz 2013), despite a consensus of the importance of considering restoration temporal scales (Sanderson *et al.* 2008; Jones & Schmitz 2009; McAlpine *et al.* 2016). As yet little work has sought to understand and quantify the effects of socio-economic drivers (Timilsina *et al.* 2014) on restoration initiatives despite their impact on ecological processes (Petursdottir *et al.* 2013; Zhang *et al.* 2013; Lü *et al.* 2015) and ecological change (Zhang *et al.* 2013; Lü *et al.* 2015).

This paper employs a GW-SEM approach to quantify the spatial variations in ecological restoration effectiveness arising from socio-economic factors. It builds directly on the work of Li et al (submitted) by extending SEMs spatially.

2 Background: ecological restoration in China

China has undertaken a number of large-scale ecological restoration and conservation programmes in order to address the ecological and socio-economic sustainability (Lü *et al.* 2012; Sun *et al.* 2015). These include the 'Three Norths Shelter Forest System Project' (since 1978), the 'Natural Forest Conservation Program' (since 2000) and the 'Grain to Green Program' (GTGP, since 2000). Some are highly proactive in the manner in which they promote ecological restoration: the GTGP promotes the conversion of steep cultivated land to forest and grassland with a subsidy after 8 years of conversion to natural forests, 5 years to commercial forests and 2 years to grasslands (Liu *et al.* 2008; Yin & Yin 2010). But, despite successes with significant increases in vegetation cover (Wang *et al.* 2010), regional and local socio-economic drivers (population migration and industrial changes) can have a negative impact on restoration effectiveness. This is particularly evident in the Loess Plateau in central China which has seen many positive results from the GTGP (Fan *et al.* 2015; Zhai *et al.* 2015) but also wide scale socio-economic changes. Methodologically, measures of re-vegetation have been found to provide effective measures of restoration project success (Lü *et al.* 2015). These capture changes in biodiversity, carbon sequestration and soil quality (Fu *et al.* 2000; Jin *et al.* 2014) and can be derived from temporal and high spatial resolution remote sensing data such as MODIS, visa measures such as the fractional vegetation cover (*FVC*) (Wu *et al.* 2014) and net primary production (*NPP*) (Donmez, Berberoglu & Curran 2011).

3 Methods

3.1 3.1 Data and Study Area

The study area for this analysis were 8 counties in Northern Shaanxi province situated in the middle of Loess Plateau in China (see Figure 1). This region is dominated by a semi-arid and continental climate and has been extensively studied in the context of soil loss and erosion. The study area has been part of a pilot and demonstration region for the GTGP since 1999.

Figure 1. The study area in red, with a Stamen Toner / OSM backdrop

Data describing FVC and NPP were calculated from 250m MODIS imagery from 2000 to 2012, which has a 16-day return time. Mean annual FVC and NPP were calculated using the Normalized Difference Vegetation Index (Carlson & Ripley 1997) and the CASA (Carnegie-Ames-Stanford) ecosystem model, respectively. Socio-economic data for each county for 2000 to 2012 were extracted from the Shaanxi Province Statistical Yearbooks and county annual socioeconomic statistical bulletins. All the data were interpolated over a 2km grid using an implementation in R of Tobler's pycnophylactic interpolation (Tobler, 1979; Brunsdon, 2015), and changes between 2000 and 2012 was calculated for all variables for each 2km pixel. The change surfaces are shown in Figure 2.

3.2 SEM Analysis

SEMs are a complex form of nested regressions. They are used in to identify and latent model variables. A good introduction can be found in Bollen (2005) and a more detailed review in Bollen and Long (1993). The basic assumption of SEMs is that explanatory models may include hidden or *latent* variables. To handle this a series of latent equations are used to generate parameters that are passed to regression operations and residual correlation evaluations. In this case a global SEM was computed to model three latent variables: population pressure, off-farm economy and rural economy. These were conceptualised as being driven by the following socioeconomic factors: Population pressure by Total population, Rural employment; Off-farm economy by Secondary industry, Tertiary industry, Primary industry; and Rural economy by Rural per capita net income and Grain yield. Restoration effectiveness was conceptualised as being driven by the 3 latent variables and described by changes in annual Net primary production and Fractional vegetation cover from the MODIS data.

3.3 GW-SEM Analysis

GW-SEMs were applied to the interpolated socioeconomic data, the NPP data and the FVC data at each location on the 2km grid. GW models are well established, including that of GW regression, GW discriminant analysis and GW PCA (Gollini et al 2015). In brief they use a moving window or kernel to calculate a series of *local* models. At each location, data falling under the kernel are weighted by their distance to the kernel centre and then passed to whatever analysis is being undertaken: regression, PCA, etc. Their relevance to spatial problems is that they explicitly reflect Tobler's 1st law of Geography rather than generate a 'whole map' statistic, they test for the presence of local, non-stationary relationships and because they are directly concerned with modelling spatial non-stationarity they will often indirectly account for any spatial autocorrelation (e.g. Harris et al. 2010). The critical issue in GW models is the bandwidth specification (kernel size), which may be fixed as a single distance or a variable distance to include a fixed proportion of data points. Methods exist to optimise bandwidth for many GW models (see Gollini et al 2015) but have not as yet been developed for SEMs. In this case a bandwidth of 100 data points under a bi-square kernel was used to specify the GW-SEM. For each data point (P_j) under the kernel, a weight $w_{i,j}$ is calculated based on its distance to the centre of the kernel as follows:

$$w_{i,j} = 1 - ((d_{i,j})^2 / b^2) \qquad (1)$$

where $d_{i,j}$ is the Euclidean distance from the centre of the kernel to the data point Pj and b is the bandwidth. This weights data points near to the kernel centre more highly than those towards the edge.

Thus, instead of a single SEM being constructed for the entire study area, a geographically weighted one was constructed from the nearest 100 data points at each location on a 2km grid.

This allows parameters and relationships of interest to be explored by examining the spatial variation in the path coefficient estimates from the GW-SEM, and thereby how the impact of different factors on the effectiveness of restoration activities varies spatially.

4 Results and brief discussion

A SEM was used to impute the relationships between the latent variables, restoration effectiveness, FVC and NPP. The schema and results for the global (whole map) SEM coefficient estimates are shown in Figure 3. This suggests that, globally, the positive impacts on restoration effectiveness (EF) were from *off-farm economy* (OE) and *rural economy* (path coefficients of 2.81 and 0.94), whereas *population pressure* (PP) had a significant negative effect (0.78).

Figure 3. The coefficient estimates from the global structural equation model, with data inputs in indicated by squares and latent variables by circles.

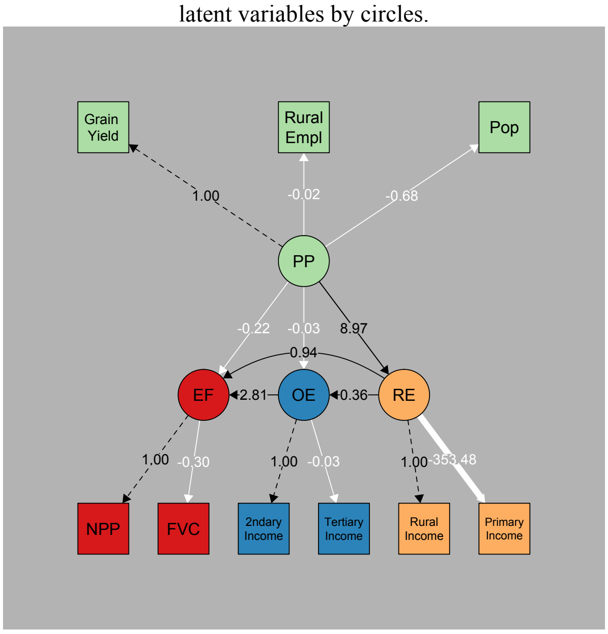

Figure 4 shows the spatial distribution of the coefficient estimates arising from the GW-SEM. These show considerable spatial variation and some degree of clustering suggesting that the drivers of restoration effectiveness vary locally, as a result of differences in the relationships between the observed variables and the degree to which they contribute to the latent variables.

The spatial distribution of the coefficients in Figure 4 suggests that the bandwidth used for the GW-SEM is less than optimal: there is more 'speckle' than one would typically see, but it is important to note that this research is a method extension still requiring work on optimal bandwidth selection. Furthermore, the socio-economic data were interpolated over large areas without consideration of the underlying land use or potential dasymetric inputs. For example, grain yield might be better interpolated across agricultural areas, population across urban areas, etc. Notwithstanding these methodological observations, the GW-SEM outputs show significant and interesting spatial patterns.

Future work will extend this analysis in the following ways: it will develop methods for optimally selecting bandwidth for GW-SEMs; socioeconomic inputs will be more robustly interpolated; and, the approach will be used to develop a spatially distributed method for modelling restoration effectiveness in the entire Shaanxi Province, a Critical Zone for sediment loss and soil erosion.

However, this research shows that GW-SEM analyses support a deeper understanding of the interaction of environmental and socioeconomic factors in relation to efforts to re-green this area of China with global implications.

5 Acknowledgements

This research was supported the China-UK bilateral collaborative research on critical zone science (the Natural Environment Research Council Newton Fund NE/N007433/1, and the National Natural Science Foundation of China NO. 41571130083) and by the National Key Research and Development Program of China (No. 2016YFC0501601). All of the data preparation, analyses and mappings were undertaken in R 3.3, the open source software. The code and data will be provided to interested researchers on request.

6 References

Berkowitz, J.F. (2013) Development of restoration trajectory metrics in reforested bottomland hardwood forests applying a rapid assessment approach. *Ecological Indicators*, **34**, 600-606.

Bollen, Kenneth A., and J. Scott Long. *Testing structural equation models*. Vol. 154. Sage, 1993.

Bollen, K. A. (2005). Structural Equation Models. *Encyclopedia of Biostatistics*. 7.

Brunsdon C, Fotheringham AS, Charlton M (1996). Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis*, 28, 281–289.

Brunsdon, C. (2015). *pycno: Pycnophylactic interpolation. R package version 1.2*. https://cran.r-project.org/web/packages/pycno/pycno.pdf

Calmon, M., Brancalion, P.H.S., Paese, A., Aronson, J., Castro, P., da Silva, S.C. & Rodrigues, R.R. (2011) Emerging Threats and Opportunities for Large-Scale Ecological Restoration in the Atlantic Forest of Brazil. *Restoration Ecology*, **19**, 154-158.

Clewell, A.F. & Aronson, J. (2013) *Ecological Restoration: Principles, Values, and Structure of an Emerging Profession.* Island Press, Washington, DC

Donmez, C., Berberoglu, S. & Curran, P.J. (2011) Modelling the current and future spatial distribution of NPP in a Mediterranean watershed. *International Journal of Applied Earth Observation and Geoinformation*, **13**, 336-345.

Carlson, T.N. & Ripley, D.A. (1997) On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sensing of Environment*, **62**, 241-252.

Fan, Y., Chen, J., Shirkey, G., John, R., Wu, S. R., Park, H., & Shao, C. (2016). Applications of structural equation modeling (SEM) in ecological studies: an updated review. *Ecological Processes*, *5*(1), 19.

Felton, A., Knight, E., Wood, J., Zammit, C. & Lindenmayer, D. (2010) A meta-analysis of fauna and flora species richness and abundance in plantations and pasture lands. *Biological Conservation*, **143**, 545-554.

Fu, B.J., Chen, L.D., Ma, K.M., Zhou, H.F. & Wang, J. (2000) The relationships between land use and soil conditions in the hilly area of the loess plateau in northern Shaanxi, China. *Catena*, **39**, 69-78.

Gollini, I., Lu, B., Charlton, M., Brunsdon, C., & Harris, P. (2015). GWmodel: an R package for exploring spatial heterogeneity using geographically weighted models. *Journal of Statistical Software*, 63 (17), 1–50.

Grace, J.B. (2006) Structural equation modeling and natural systems. Cambridge University Press.

Harris, P., A.S. Fortheringham, R. Crespo, & M. Charlton. 2010. The use of geographically weighted regression for spatial prediction: an evaluation of models using simulated data sets. Mathematical Geosciences 42:657-680.

Jin, Z., Dong, Y., Wang, Y., Wei, X., Wang, Y., Cui, B. & Zhou, W. (2014) Natural vegetation restoration is more beneficial to soil surface organic and inorganic carbon sequestration than tree plantation on the Loess Plateau of China. *Science of the Total Environment*, **485**, 615-623.

Jones, H.P. & Schmitz, O.J. (2009) Rapid Recovery of Damaged Ecosystems. *Plos One*, **4**.

Kline, RB (2015). *Principles and practice of structural equation modeling*. Guilford publications.

Li T, Lü Y, Bojie F, Comber AJ, Harris P and Wu L (submitted). Gauging policy-driven large-scale vegetation restoration programmes under a changing environment: their effectiveness and socioeconomic relationships. Paper submitted to the *Journal of Applied Ecology* (January 2017)

Liu, J.G., Li, S.X., Ouyang, Z.Y., Tam, C. & Chen, X.D. (2008) Ecological and socioeconomic effects of China's policies for ecosystem services. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 9477-9482.

Lü, Y.H., Fu, B.J., Feng, X.M., Zeng, Y., Liu, Y., Chang, R.Y., Sun, G. & Wu, B.F. (2012) A Policy-Driven Large Scale Ecological Restoration: Quantifying Ecosystem Services Changes in the Loess Plateau of China. *Plos One*, **7**.

Lü, Y.H., Zhang, L.W., Feng, X.M., Zeng, Y., Fu, B.J., Yao, X.L., Li, J.R. & Wu, B.F. (2015) Recent ecological transitions in China: greening, browning, and influential factors. *Scientific reports*, **5**.

McAlpine, C., Catterall, C.P., Mac Nally, R., Lindenmayer, D., Reid, J.L., Holl, K.D., Bennett, A.F., Runting, R.K., Wilson, K., Hobbs, R.J., Seabrook, L., Cunningham, S., Moilanen, A., Maron, M., Shoo, L., Lunt, I., Vesk, P., Rumpff, L., Martin, T.G., Thomson, J. & Possingham, H. (2016) Integrating plant- and animal-based perspectives for more effective restoration of biodiversity. *Frontiers in Ecology and the Environment*, **14**, 37-45.

Meli, P., Benayas, J.M.R., Balvanera, P. & Ramos, M.M. (2014) Restoration Enhances Wetland Biodiversity and Ecosystem Service Supply, but Results Are Context-Dependent: A Meta-Analysis. *Plos One*, **9**.

Petursdottir, T., Arnalds, O., Baker, S., Montanarella, L. & Aradottir, A.L. (2013) A Social-Ecological System Approach to Analyze Stakeholders' Interactions within a Large-Scale Rangeland Restoration Program. *Ecology and Society*, **18**.

Sanderson, E.W., Redforda, K.H., Weber, B., Aune, K., Baldes, D., Berger, J., Carter, D., Curtin, C., Derr, J., Dobrott, S., Fearn, E., Fleener, C., Forrest, S., Gerlach, C., Gates, C., Gross, J.E., Gogan, P., Grassel, S., Hilty, J.A., Jensen, M., Kunkel, K., Lammers, D., List, R., Minkowski, K., Olson, T., Pague, C., Robertson, P.B. & Stephensom, B. (2008) The ecological future of the north American Bison: Conceiving long-term, large-scale conservation of wildlife. *Conservation Biology*, **22**, 252-266.

Sun, W.Y., Song, X.Y., Mu, X.M., Gao, P., Wang, F. & Zhao, G.J. (2015) Spatiotemporal vegetation cover variations associated with climate change and ecological restoration in the Loess Plateau. *Agricultural and Forest Meteorology*, **209**, 87-99.

Timilsina, N., Escobedo, F.J., Staudhammer, C.L. & Brandeis, T. (2014) Analyzing the causal factors of carbon stores in a subtropical urban forest. *Ecological Complexity*, **20**, 23-32.

Tobler, W.R. (1979) *Smooth Pycnophylactic Interpolation for Geographical Regions*. Journal of the American Statistical Association, v74(367) pp. 519-530.

Ullman, J. B., & Bentler, P. M. (2003). *Structural equation modeling*. John Wiley & Sons, Inc.

Wang, L., Fu, B.J., Lü, Y.H. & Yuan, Z. (2010) Spatio-temporal Variation of Vegetation Cover in Northern Shaanxi Province Under the Background of Ecological Restoration *chinese Jouanl of Applied Ecology*, **21**, 2109-2116.(In Chinese)

Wu, D., Wu, H., Zhao, X., Zhou, T., Tang, B., Zhao, W. & Jia, K. (2014) Evaluation of Spatiotemporal Variations of Global Fractional Vegetation Cover Based on GIMMS NDVI Data from 1982 to 2011. *Remote Sensing*, **6**, 4217-4239.

Yin, R.S. & Yin, G.P. (2010) China's Primary Programs of Terrestrial Ecosystem Restoration: Initiation, Implementation, and Challenges. *Environmental Management*, **45**, 429-441

Zhai, J., Liu, R.G., Liu, J.Y., Huang, L. & Qin, Y.W. (2015) Human-Induced Landcover Changes Drive a Diminution of Land Surface Albedo in the Loess Plateau (China). *Remote Sensing*, **7**, 2926-2941.

Zhang, B.Q., Wu, P.T., Zhao, X.N., Wang, Y.B. & Gao, X.D. (2013) Changes in vegetation condition in areas with different gradients (1980-2010) on the Loess Plateau, China. *Environmental Earth Sciences*, **68**, 2427-2438.

Figure 2. The changes in the interpolated socioeconomic variables per 2km^2 grid cell: tonnes of Grain yield (GY), Rural per capita net income in Yuan (RI), Tertiary industry (10^4 Yuan) (TI), Secondary industry (10^4 Yuan) (SI), Primary industry (10^4 Yuan) (PI), Rural employment (REm), Population (10^4 people) (Pop), annual Net primary production (in g C/m^2/year) (Npp) and mean percentage Fractional vegetation cover (Fvc).

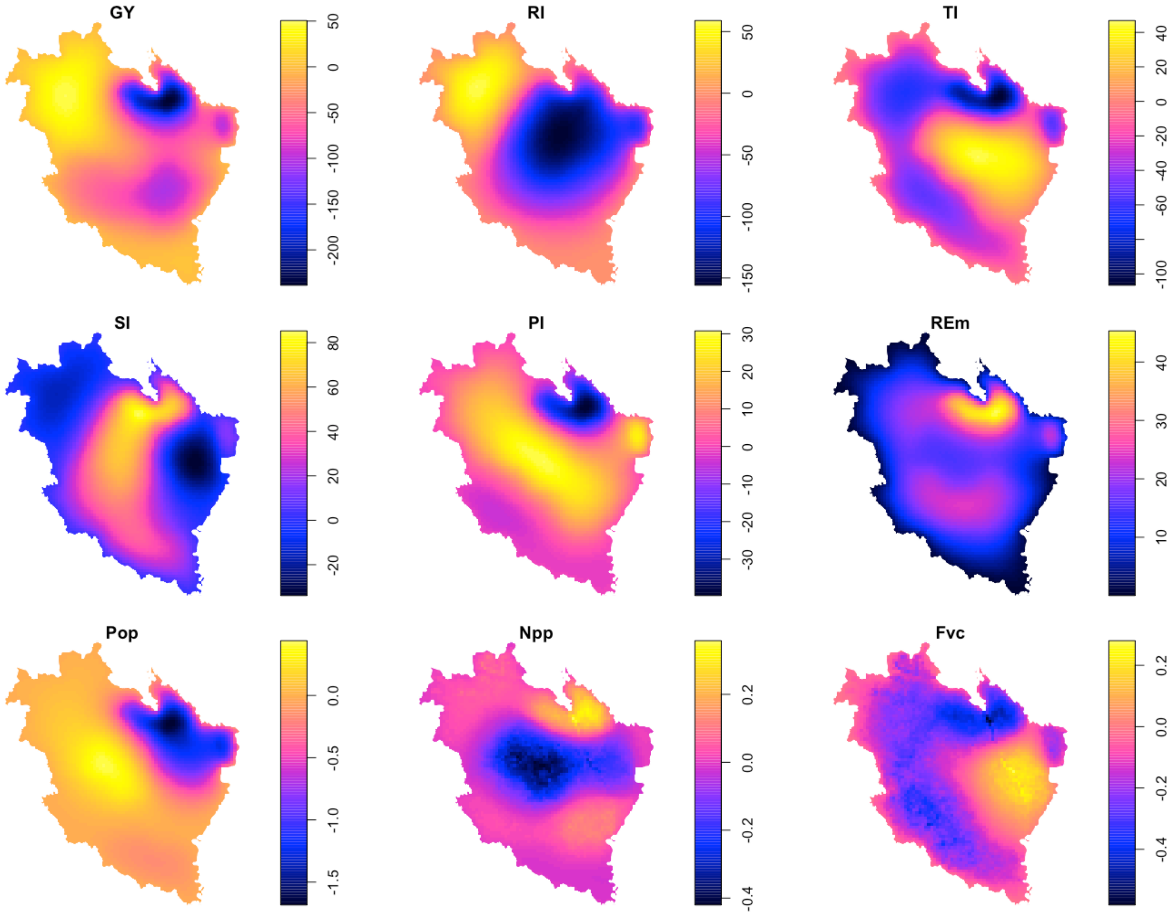

Figure 4. The spatial distribution of the coefficient estimates arising from a GW-SEM and the degree to which they predict different latent variables.