

# Geographically Weighted Partial Correlation for Spatial Analysis

GI\_Forum 2017, Issue 1

Page: 36 - 43

Short Paper

Corresponding Author:

percival.joseph.82m@st.kyoto-u.ac.jp

DOI: 10.1553/giscience2017\_01\_s36

Joseph Percival and Narumasa Tsutsumida

Graduate School of Global Environmental Studies, Kyoto University, Japan

## Abstract

Spatial correlation between variables may exist if the observed data exhibits spatial variation in a manner that is described by Tobler's first law of geography. Partial correlation is useful when considering multivariate data as it can highlight the effects of certain control variables on the correlation between any two other variables. Techniques for estimating spatial correlation have been developed based on a geographically weighted scheme. However, a partial correlation technique for spatial data has not yet been considered. Hence, we describe a technique for obtaining geographically weighted partial correlation coefficients between three variables. This approach is then applied, as an example, to global climate data in order to explore the relationship between terrestrial vegetation (by NDVI proxy), land surface temperature, and precipitation in the year 2014. Spatial variations of those variables are observed and the geographically weighted correlation and partial correlation coefficients (along with associated levels of statistical significance) are compared.

## Keywords:

partial correlation, spatial statistics, geographically weighted approach

## 1 Introduction

Correlation between variables is often observed when exploring spatial data and it is important to carefully inspect such characteristics before analyses. Furthermore, correlation itself is sometimes found to be a valuable descriptive statistic. For example, it may be useful in describing relationships between climate and vegetation growth, temperature and greenhouse gases, or among various socio-economic data. Quantifying these relationships using Pearson's or Spearman's coefficients is well known and well practised when observing two variables. However, calculations for partial correlation are necessary to find the degree of association between two variables when three or more are involved. Partial correlation measures the correlation between two variables whilst removing the effects of any number of control variables. This has been used to describe meaningful relationships in several disciplines, ranging from the statistical dependencies between regions in the brain (Marrelec et al., 2006), to indicators of economic growth across countries (Barro, 1991). Such

correlative or partially correlative relationships are generally summarized as static global measures and therefore fail to incorporate any underlying spatial structure in the data. Various geographically weighted (GW) approaches have been developed to deal with such underlying spatial structure, many of which are summarized in the R `GWmodel` package (Gollini et al., 2015; Lu et al., 2014). In this paper, we describe a novel technique for obtaining localized geographically weighted partial correlation coefficients among three variables. Spatial relations between temperature, precipitation and terrestrial vegetation activity as measured by a Normalized Difference Vegetation Index proxy (NDVI) are then demonstrated as a case study.

## 2 Background

Geographically weighted analyses are conducted using a moving window weighting technique that operates under the assumption of Tobler's first law of geography. Namely, that 'everything is related to everything else, but near things are more related than distant things' (Tobler, 1970). This approach is useful when describing statistical features and relationships that are inadequately described by global models, and instead are more accurately described by local ones. If the local model outputs vary across space, it can be taken that spatial non-stationarity exists in the data. These techniques differ from non-weighted moving window techniques where localized analyses are done independent of 'nearness' at each localized scale (see Shuttleworth & Lloyd (2005), for example). To date, several GW techniques have been developed and are applicable to, among many other things, statistical measures of mean, standard deviation and correlation coefficients, statistical models of regression, and principle component analysis (Gollini et al., 2015; Tsutsumida & Comber, 2016). Our technique is a small addition to this growing list of GW statistical methods.

## 3 Data

This study uses temperature and precipitation observation data from the Climate Research Unit Time Series (CRU TS, version 3.23) data set of the University of East Anglia (Harris et al., 2014) and NDVI data from the GIMMS NDVI3g.v1 data set (Pinzon & Tucker, 2014) to demonstrate local relations. The CRU TS observation data is considered one of the best-organized and most complete data sets of temperature and precipitation, covering all available stations across the world. For a complete description of the data set, see Harris et al. (2014). The GIMMS NDVI3g.v1 is the latest release in the NDVI3g time series data set. This data set is well-organized and contains bimonthly NDVI values obtained from a series of Advanced Very High Resolution Radiometer (AVHRR) instruments at a scale of 8km, extending from 1981 to 2015 (Pinzon & Tucker, 2014). Despite its coarse resolution, this data is frequently used for macro-scale global environmental research and is often analysed together with the CRU TS gridded data (see for instance Gonsamo et al. (2016)). For this study, the average annual temperature, precipitation, and NDVI values for the year 2014 are

calculated and used as inputs. After omitting locations with missing data, 737 sample data are created.

## 4 Methodology

A GW technique operates by creating a spatial weights matrix  $W(\mathbf{u}, \mathbf{v})$  at coordinates  $(\mathbf{u}, \mathbf{v})$  for all observations in each window.  $W(\mathbf{u}, \mathbf{v})$  is an  $n \times n$  diagonal matrix of spatial weights ( $\omega$ ) of  $n$  sample size which are calculated by some defined kernel. Any kernel function may be used. However, typically a distance-decay kernel function is employed. More complete descriptions of administering kernel functions for GW models may be found in Gollini et al. (2015). For simplicity, in this paper we apply a bi-square kernel function  $\omega_{\mathbf{k}}$  for calculating the weight at any point  $\mathbf{k}$  from an observation point situated at  $(\mathbf{u}, \mathbf{v})$ :

$$\omega_{\mathbf{k}}(\mathbf{u}, \mathbf{v}) = \begin{cases} \left(1 - \left(\frac{d_{\mathbf{k}}}{b}\right)^2\right)^2 & \text{if } |d_{\mathbf{k}}| < b, \\ 0 & \text{otherwise} \end{cases}$$

where  $d_{\mathbf{k}}$  is the distance between the observation at the point centred at  $(\mathbf{u}, \mathbf{v})$  and an observation  $\mathbf{k}$ , and  $b$  is the 'bandwidth'. The bandwidth determines the size of the moving window. Smaller bandwidths contribute to inspections of local variations in space, while larger bandwidths approach the results of the global model.

Here, we develop a technique in the R programming language to calculate localized partial correlation coefficients between two variables and a set of control variables in a way that is consistent with the GW model approach using either Pearson's or Spearman's method. This process operates by first calculating a spatially weighted co-variance matrix at each given data point under the moving kernel defined above. With this weighting scheme, the GW co-variance matrix at each observation point may be found by the following calculation:

$$\Sigma(\mathbf{u}, \mathbf{v}) = \mathbf{P}^T \mathbf{W}(\mathbf{u}, \mathbf{v}) \mathbf{P}$$

where  $\Sigma(\mathbf{u}, \mathbf{v})$  is the local co-variance matrix at coordinates  $(\mathbf{u}, \mathbf{v})$ ,  $\mathbf{P}$  is the data matrix of  $n$  rows by  $m$  variables, and  $W(\mathbf{u}, \mathbf{v})$  is the spatial weights matrix (see Gollini et al. (2015)). It follows, then, that the geographically weighted partial correlation between two variables,  $m_i$  and  $m_j$  of the set  $M$  with cardinality  $m$ , given all others of the set, at a location  $(\mathbf{u}, \mathbf{v})$ , may be given by:

$$\rho_{m_i m_j \cdot M \setminus \{m_i, m_j\}}(\mathbf{u}, \mathbf{v}) = \frac{-c_{ij}}{\sqrt{c_{ii} c_{jj}}}$$

where  $\mathbf{c}$  is the inverted GW co-variance matrix,  $\mathbf{c} = (\Sigma(\mathbf{u}, \mathbf{v}))^{-1}$ . In the case of three variables ( $m = 3$ ),  $\Sigma(\mathbf{u}, \mathbf{v}) = \mathbf{P}^T \mathbf{W}(\mathbf{u}, \mathbf{v}) \mathbf{P}$ , where  $\mathbf{P}$  is now the data matrix made up of three variables,  $x, y, z$ . Therefore, the geographically weighted partial correlation coefficient at coordinates  $(\mathbf{u}, \mathbf{v})$  between  $x$  and  $y$ , while controlling for  $z$ , can be written as  $\rho_{xy \cdot z}(\mathbf{u}, \mathbf{v}) = \frac{-c_{xy}}{\sqrt{c_{xx} c_{yy}}}$ . The outputs of this function may then be mapped for further exploratory analysis.

The geographically weighted partial correlation function (called `gwpcor`) was developed in R by extending the geographically weighted summary statistics (`gwss`) function of the `GWmodel` package (Gollini et al., 2015; Lu et al., 2014) to incorporate operations of the `corpcor` package (Schäfer et al., 2015). A simple comparison is also made between GW correlation and GW partial correlation using `gwpcor` in order to demonstrate `gwpcor`'s utility in being able to easily calculate both coefficients. In its current form, `gwpcor` can also calculate and provide levels of statistical significance for each local coefficient.

For simplicity, in this paper we used a fixed bandwidth of 3,000km. This was chosen after observing the results of several fixed bandwidths. Although a number of approaches have been developed for automated optimal bandwidth selection, these require that an objective function exists and use cross-validation (CV) techniques (Gollini et al., 2015). However, this cross-validation technique has been shown to return the saturation point (corresponding to values near the maximum bandwidth) when the input variables are correlated, thus resulting in the failure to identify local variation (Tsutsumida et al., in press). Thus, there are currently no such optimized bandwidth selection tools for `gwpcor`. Automated optimal bandwidth selection for `gwpcor` (and other `gwss` functions) is an important research area that needs addressing.

## 5 Results and Discussion

Before proceeding with the results of the GW functions, it is important to observe the global measures of correlation and partial correlation coefficients among the three variables. These are calculated without regard to spatial location and structure and are displayed in Tables 1 and 2 below.

**Table 2:** Pearson's global correlation matrix

	Temperature	Precipitation	NDVI
Temperature	1.00	0.36 ***	0.26 ***
Precipitation	0.36 ***	1.00	0.62 ***
NDVI	0.26 ***	0.62 ***	1.00

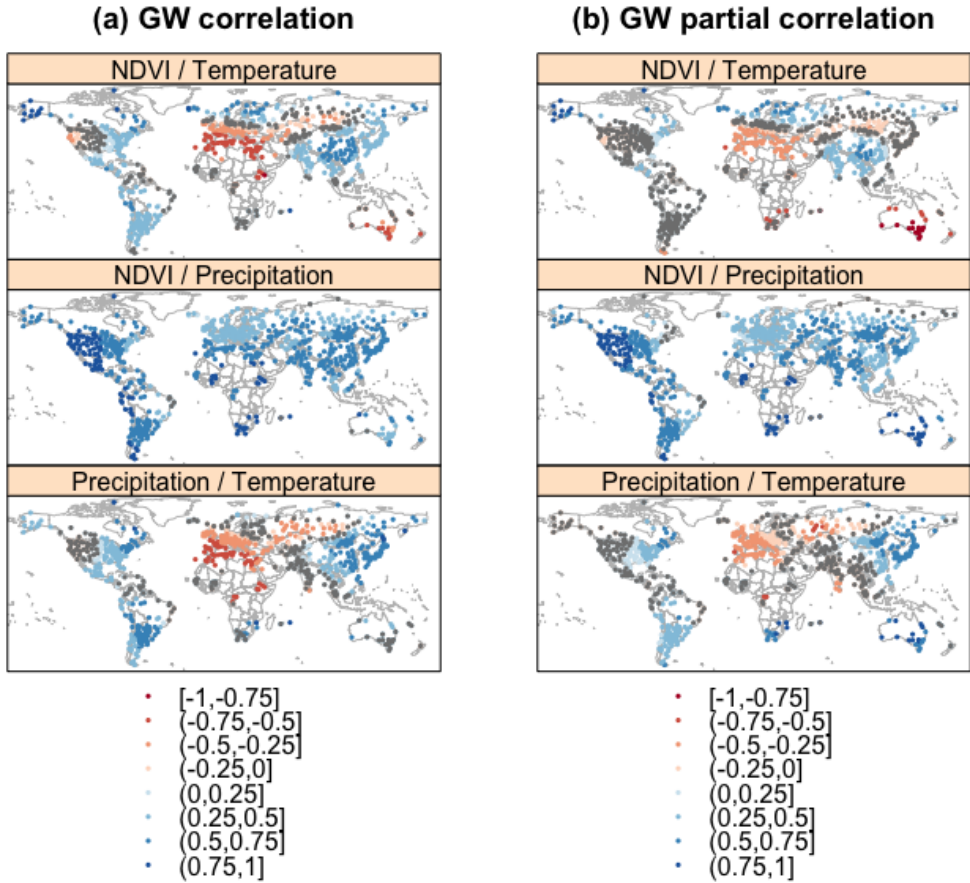
**Table 1:** Pearson's global partial correlation matrix

	Temperature	Precipitation	NDVI
Temperature	1.00	0.26 ***	0.05
Precipitation	0.26 ***	1.00	0.58 ***
NDVI	0.05	0.58 ***	1.00

P-values:  $0 < *** < 0.01 \leq ** < 0.05 \leq * < 0.10$

This global model suggests a weak positive correlation (0.26) between yearly averaged global NDVI and temperature for the year 2014 (Table 1). Controlling for precipitation effects results in an even weaker and less significant positive correlation (0.05) between the two (Table 2). By holding this very simplistic view, it can be said that at the global scale there is almost no correlation between NDVI and temperature. Equally weak correlations and partial correlations are observed between temperature and precipitation (holding NDVI constant), and a moderate correlation is observed between precipitation and NDVI (with little difference when controlling for temperature).

However, as Clinton et al. (2014) observe, terrestrial vegetation can have either strong positive or strong negative temporal correlations with temperature and precipitation, depending on the geographic location. They review studies that show that plant growth may have a strong positive correlation with precipitation in both arid and temperate regions, and experience a negative one in some tropical regions (Clinton et al., 2014). In the same manner, the correlation between temperature and plant growth varies across space and is often both strong and positive in mid to high latitudes (Clinton et al., 2014). Thus it is clear how a GW approach may be beneficial in this study. Figure 1 demonstrates these relationships for the year 2014.



**Figure 1:** Maps of spatial relations among average NDVI, Temperature and Precipitation in 2014 calculated by (a) geographically weighted Pearson's correlation and (b) geographically weighted Pearson's partial correlation. Coefficients with p-values greater than 0.05 are coloured in grey.

As expected, there is high variation in the correlation and partial correlation coefficient values in both maps when compared to the global results from Tables 1 and 2. The calculation of GW partial correlation coefficients demonstrates how the relationships between temperature, precipitation and NDVI change when controlling for effects from one of the three variables. This is particularly useful in order to check what factors might be affecting the GW correlation between any of the two variables. For instance, both maps suggest interesting local variations in the correlation between NDVI and temperature. The two tend to be negatively correlated in dry regions and positively correlated in other regions. However, the magnitude of the negative correlations found in North Africa using GW correlation may in fact be overestimated when compared with the results of *gwpcor*, which controls for the effect of precipitation. Furthermore, several areas with weak positive coefficients between temperature and NDVI are found to be statistically insignificant when controlling precipitation. Focusing now on the relation between NDVI and precipitation, we find that the GW correlation does not seem to change very much, even after controlling for

the effect of temperature, and most values retain a high level of statistical significance. This suggests that the correlation between NDVI and precipitation might be largely consistent, regardless of temperature values.

## 6 Conclusion

The estimation of local GW partial correlation coefficients is valuable when data exhibits spatial non-stationarity, and such spatial information is important across many disciplines related to GIS. Partial correlation coefficients provide a useful descriptive statistic, and the extension of this to include geographic weights is helpful if spatial relations in data might be described by Tobler's law. More could be done to improve the functionality of *gwpcor*, including adding methods for Kendall's correlation coefficients and optimizing bandwidths. Interesting future research could also be done to extend this and other GW summary statistic techniques into the time dimension, in order to allow for more comprehensive spatio-temporal analyses.

## Acknowledgements

The authors would like to thank anonymous referees for their insightful comments on this article. This work was supported by the JSPS KAKENHI under Grant number 15K21086; and the KU SPIRITS project.

## References

- Barro, R. J. (1991). Economic growth in a cross section of countries. *Quarterly Journal of Economics*, 106(2), 407. <https://doi.org/10.2307/2937943>
- Clinton, N., Yu, L., Fu, H. H., He, C. H., & Gong, P. (2014). Global-Scale Associations of Vegetation Phenology with Rainfall and Temperature at a High Spatio-Temporal Resolution. *Remote Sensing*, 6(8), 7320–7338. Retrieved from <https://doi.org/10.3390/rs6087320>
- Gollini, I., Lu, B., Charlton, M., Brunson, C., & Harris, P. (2015). *GWmodel* : an R package for exploring spatial heterogeneity. *Journal of Statistical Software*, 63(17), 1–50. Retrieved from <https://doi.org/10.1080/10095020.2014.917453>
- Gonsamo, A., Chen, J. M., & Lombardozzi, D. (2016). Global vegetation productivity response to climatic oscillations during the satellite era. *Global Change Biology*, 1–13. Retrieved from <https://doi.org/10.1111/gcb.13258>
- Harris, I., Jones, P., Osborn, T., & Lister, D. (2014). Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. *International Journal of Climatology*, 34(3), 623–642. Retrieved from <https://doi.org/10.1002/joc.3711>
- Lu, B., Harris, P., Charlton, M., & Brunson, C. (2014). The *GWmodel* R package: further topics for exploring spatial heterogeneity using geographically weighted models. *Geo-Spatial Information Science*, 17(2), 85–101. Retrieved from <https://doi.org/10.1080/10095020.2014.917453>
- Marrelec, G., Krainik, A., Duffau, H., Péligrini-Issac, M., Lehericy, S., Doyon, J., & Benali, H. (2006). Partial correlation for functional brain interactivity investigation in functional MRI. *NeuroImage*, 32(1), 228–237. Retrieved from <https://doi.org/10.1016/j.neuroimage.2005.12.057>

- Pinzon, J. E., & Tucker, C. J. (2014). A Non-Stationary 1981–2012 AVHRR NDVI3g Time Series. *Remote Sensing*, 6, 6929–6960. Retrieved from <https://doi.org/10.3390/rs6086929>
- Schäfer, J., Opgen-Rhein, R., Zuber, V., Ahdesmäki, M., Silva, A. P. D., & Strimmer, K. (2015). Corpcor: Efficient estimation of covariance and (partial) correlation. Retrieved from <https://CRAN.R-project.org/package=corpcor>
- Shuttleworth, I., & Lloyd, C. (2005). Analysing average travel-to-work distances in Northern Ireland using the 1991 census of population: The effects of locality, social composition, and religion. *Regional Studies*, 39, 909–921. Retrieved from <https://doi.org/10.1080/00343400500289895>
- Tobler, W. R. (1970). A Computer Movie Simulation Urban Growth in the Detroit Region. *Economic Geography*, 46, 234–240. Retrieved from <https://doi.org/10.2307/143141>
- Tsutsumida, N., & Comber, A. (2016). Spatial accuracy measures of soft classification in land cover. In J. Miller, D. O’Sullivan, & N. Wiegand (Eds.), *Ninth international conference on GIScience short paper proceedings* (pp. 340–343). Montreal.
- Tsutsumida, N., Harris, P., & Comber, A. (In Press). The application of a geographically weighted principal components analysis for exploring 23 years of goat population change across Mongolia. *Annals of the American Association of Geographers*.