

# Geographically weighted correspondence matrices for change analyses and error reporting

*Alexis Comber<sup>1</sup>, Chris Brunsdon<sup>2</sup>, Martin Charlton<sup>2</sup>, Paul Harris<sup>3</sup>*

<sup>1</sup> School of Geography, University of Leeds, Leeds, LS2 9JT, UK

Email: a.comber@leeds.ac.uk

<sup>2</sup> NUI Maynooth, Maynooth, Co Kildare, Ireland

Email: {christopher.brunsdon; martin.charlton} @nuim.ie

<sup>3</sup> Rothamsted Research, North Wyke, EX20 2SB, UK

Email: paul.harris@rothamsted.ac.uk

## Abstract

This letter describes and applies generic approaches for generating local measures from the correspondence table. These were developed by integrating the functionality of 2 R packages: `gwxtab` and `diffeR` in order to demonstrate the development of spatially explicit accuracy measures comparing comparison and reference data (predicted and observed) and change measures if the data represent classes at time  $t$  and  $t + 1$ . The approach builds on on earlier work that considered the measures derived from the correspondence matrix in the context of generalized linear models and probability, and instead computes local geographically weighted correspondence matrices, from which local statistic can be calculated. In this case a selection of the overall and categorical difference measures proposed by Pontius and Milones (2011) and Pontius and Santacruz (2014) were examined spatially.

## Key Words:

geographically weighted; accuracy and error; correspondence, validation, error matrix

# Introduction

One of the under-explored areas in research of remote sensing land cover and land use is the investigation of local statistical models. Most remote sensing methods (classification, validation, change detection, etc) apply global approaches under the assumption that models and relationships between the variables or data under consideration remain constant over geographical space. Any spatial ‘variation’ in the results is driven by variation in data or variable values. Global models describe processes or patterns which are assumed to be location independent of stationary. However, this assumption of process spatial stationarity and invariance is contradicted by the many observations of spatial auto-correlation in landscape processes, and in coincident remotely sensed data, especially classification error (from Campbell 1981 to Comber in press 1) and more widely in geographic analyses under Tobler’s 1st law of geography (Tobler, 1970). Spatial auto-correlation occurs when changes in properties of nearby features in geographic space are found to be correlated, contradicting the underlying assumption of independence in statistical analysis and inference. The result is process spatial non-stationarity when the statistical pattern or relationship observed in one region differs from that in another. An example of this *global* rather than *local* philosophy in remote sensing is the persistence of the use of the correspondence matrix. In error reporting this summarises the spatial coincidence of a classified dataset with a reference dataset, sometimes referred to as *predicted* and *observed* data, and in change analyses it summarises the class to class transitions between data collected at different times.

Recent research has started to address this fundamental statistical blind spot in remote sensing, and spatially sensitive approaches have been proposed for remote sensing classification (Comber et al, 2016 in press 2), data fusion (Lisev et al., 2016) and in applications where remote sensing data provides one of the input variables, for example mapping above ground biomass (Propastin, 2012), population segregation (Yu and Wu, 2004), net primary production (Wang et al., 2005) and in epidemiology (Khormi and Kumar, 2011).

In remote sensing error analysis Foody (2005) calculated geographically distributed correspondence matrices and interpolated between them to generate surfaces of error. Extensions by Comber et al. (2012) and Comber (2013) developed geographically weighted (GW) measures accuracy from a GW logistic regression. which were further extended to examine the spatio-temporal characteristics of classification accuracy by Tsutsumida and Comber (2015). All of these approaches provide spatially distributed measures of error that can be easily quantified using a simple logistic regression of the data in the correspondence matrix as described in Comber (2013). However, many other measures may be calculated from the correspondence matrix including Kappa estimates (Congalton, 1991) and the quantity and allocation disagreements suggested by Pontius and Millones (2011). Indeed, within statistics there is a long literature describing statistical measures that can be derived from the family of contingency tables (e.g. Hartigan and Kleiner, 1981; Friendly, 1994), of which the correspondence matrix is member. These are not straightforward to describe or formalise in a logistic regression, geographically weighted or not.

This letter describes a generic method for calculating spatially distributed correspondence matrices that support a much wider set of geographically weighted analyses of error, accuracy and correspondence. It uses the `gwxtab` R package (Brunsdon et al, 2016) as a framework for calculating local correspondence matrices. Then it uses these to calculate *local* difference metrics as described in Pontius and Milones (2011) and Pontius and Santacruz (2014) and implemented in the `diffeR` R package (Pontius and Santacruz, 2015). For good measure we calculate a local Kappa estimate as well. This letter highlights how thinking locally rather than globally can result in more spatially nuanced reportings of accuracy and other comparative measures such as change. It illustrates how generic tools such as `gwxtab` can be used to calculate local versions of any correspondence table derived metric, which can in turn be mapped, to generate novel, spatially distributed measures of accuracy.

# Data and Methods

## Data

The data used in this analysis was developed IIASA (Austria) and is included in the `gwxtab` package. It describes land cover at 2,439 locations in the British Isles (minus the islands!) from 4 sources collapsed to 10 classes as described in Comber et al (2013): volunteered land cover data collected by the Geo-Wiki initiative (Perger et al, 2012), the GLC-2000 database (Fritz et al., 2003), the MODIS land cover product (Loveland et al., 2000) and GlobCover (Bicheron et al., 2008). The analyses in this letter compare the Geo-Wiki and Modis data. Table 1 shows the correspondence matrix of the Modis (rows) against the Geo-Wiki data (columns).

Table 1. The Geo-Wiki (columns) and Modis data (rows) correspondence matrix

	1	2	3	4	5	6	7	8	9	10
1. Forest	31	3	37	33	10	1	55	0	1	1
2. Shrub	0	0	0	1	0	0	31	0	0	0
3. Grass	11	4	26	62	15	1	19	0	5	1
4. Crop	13	1	117	355	10	1	78	0	2	3
5. Mosaic	53	1	293	135	1	0	89	0	2	1
6. Wetland	0	0	0	0	0	0	0	0	0	0
7. Urban	21	1	104	49	0	0	749	0	5	4
8. Snow	0	0	0	0	0	0	0	0	0	0
9. Barren	0	0	0	0	0	0	0	0	0	0
10. Water	0	0	1	1	0	0	1	0	0	0

## Geographically weighted correspondence matrices

A correspondence matrix summarises the spatial intersection of 2 datasets and in a remote sensing error analysis, it compares the classified data with higher quality reference data at sample locations. In full a correspondence analysis, for example examining change over time, it summarises the spatial intersection of all data points or pixels. The problem with the correspondence matrix and the measures that can be derived from it from a remote sensing perspective are that it provides no information about the spatial distribution of intersecting values and the global measures may mask local variations (McGwire and Fisher, 2001).

The basic idea of geographically weighted approach is that local correspondence matrices are computed from subsets of the full datasets at predefined locations, in order to quantify the spatial variation in the relationship between the 2 datasets, for example to generate spatially distributed measures of error. At each location, a subset of the data falling under a kernel are weighted by their distance to that location and then used to construct the correspondence matrix. This local correspondence matrix can be used to calculate the statistic of interest.

The kernel size or bandwidth, can be fixed (e.g. 20km) or it can be adaptive to subset the nearest  $n$  data points (e.g. 15%) and different kernel functions (shapes) can be used for the distance weighting. Generally larger bandwidths result in a greater degree of spatial smoothing. Gollini et al (2013) describe some of these and methods for determining bandwidth optimally. In this case, an adaptive bandwidth of 15% was and a bisquare kernel were applied. For a given bandwidth  $h$ , this is defined by:

$$f(d) = \begin{cases} \left(1 - \left(\frac{d}{h}\right)^2\right)^2 & \text{if } d < h; \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

where  $d$  is the distance of the data point to the kernel centre.

Local, geographically weighted correspondence matrices were constructed to compare volunteered land cover data collected by the Geo-Wiki initiative (Perger et al, 2012) with the 3 global land cover products at each location on a XYZ grid covering the study. The study area, grid and data points are shown in Figure 1.

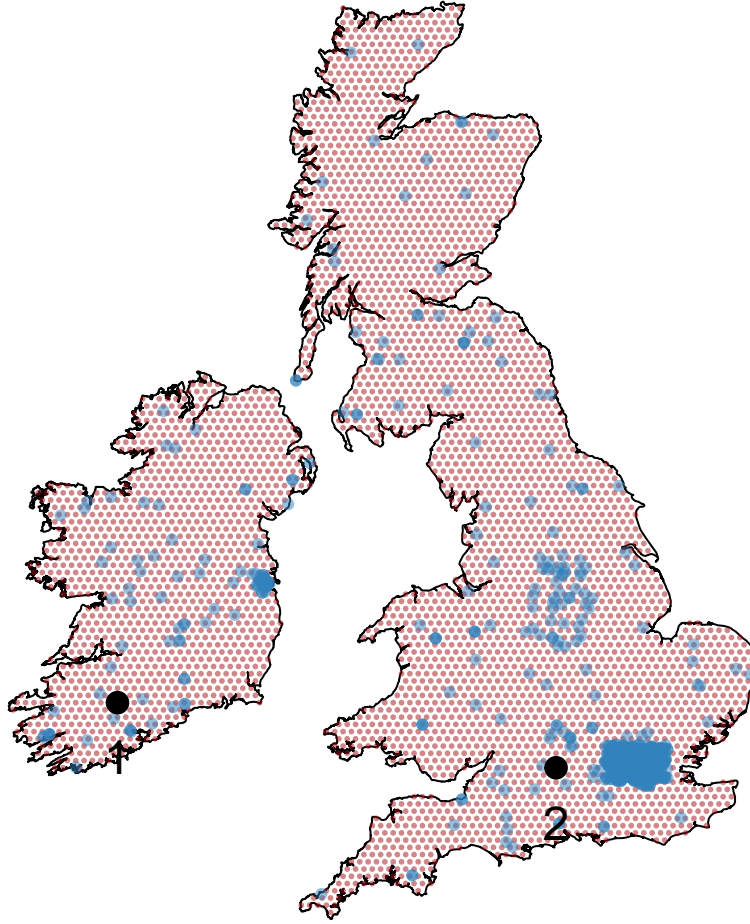


Figure 1: The study area, analysis grid in red and the data points, in blue with a transparency term, and 2 example locations.

Table 2. The local Geographically Weighted correspondence matrix at Location 1

	1	2	3	4	5	6	7	8	9	10
1. Forest	1.58	0	3.30	0.00	0.01	0	0.00	0	0	0
2. Shrub	0.05	0	2.81	0.74	0.00	0	0.76	0	0	0

	1	2	3	4	5	6	7	8	9	10
3. Grass	0.60	0	5.01	0.16	0.01	0	0.00	0	0	0
4. Crop	3.94	0	32.24	7.44	0.58	0	0.23	0	0	0
5. Mosaic	2.25	0	5.51	1.55	0.03	0	0.00	0	0	0
6. Wetland	0.00	0	0.41	0.24	0.00	0	0.00	0	0	0
7. Urban	0.91	0	4.46	0.79	0.09	0	18.66	0	0	0
8. Snow	0.00	0	0.00	0.00	0.00	0	0.00	0	0	0
9. Barren	0.94	0	2.95	0.00	0.00	0	0.00	0	0	0
10. Water	0.00	0	0.00	0.24	0.00	0	0.00	0	0	0

Table 3. The local Geographically Weighted correspondence matrix at Location 2

	1	2	3	4	5	6	7	8	9	10
1. Forest	1.12	0	0.00	0.11	1.12	0	0.52	0	0	0
2. Shrub	0.18	0	0.00	0.00	0.07	0	0.00	0	0	0
3. Grass	0.83	0	0.11	2.08	2.10	0	0.62	0	0	0
4. Crop	1.15	0	0.11	11.27	3.73	0	0.94	0	0	0
5. Mosaic	0.88	0	0.04	1.57	0.24	0	0.00	0	0	0
6. Wetland	0.00	0	0.00	0.00	0.00	0	0.00	0	0	0
7. Urban	0.95	0	0.28	0.62	1.71	0	3.90	0	0	0
8. Snow	0.00	0	0.00	0.00	0.00	0	0.00	0	0	0
9. Barren	0.00	0	0.00	0.47	0.31	0	0.00	0	0	0
10. Water	0.00	0	0.00	0.00	0.15	0	0.00	0	0	0

## Accuracy measures: Quantity and Allocation Disagreements

Pontius and Millones (2011) and Pontius and Santacruz (2014) describe a number of methods for calculating difference metrics and composite measures of accuracy from the correspondence matrix. These are based around map-to-map cross-tabulations or correspondence matrices and can be used to compare mapped land cover data for error or change. For example, *quantity disagreement* is defined as the amount of difference between the *Observed* reference data and the *Predicted* classified data relative to the proportions of the classes in the *Observed* and *Predicted* data. It is computed from the sum of the row totals (the *Predicted* data) minus the sum of the column totals (the *Observed* data) divided by 2. Similarly, the *allocation disagreement* is defined as the amount of difference between the *Observed* data and the *Predicted* data that are due mis-matches in the spatial allocation of classes, relative to the class proportions. It is computed from the total number of pixels minus the diagonal agreement, minus the *quantity disagreement*. In all cases the measures can be computed from correspondence tables of counts of coincident pixels or proportions. Some of the measures from the `diffeR` package that were applied in this analysis are summarised in Table 2.

Table 4: A summary of the difference and disagreement measures in the **diffeR** package

Measure	Function	Return value	Descriptions
Overall allocation difference	<b>overallAllocD</b>	Single value	The amount of difference between <i>observed</i> and <i>predicted</i> data due to the *less than maximum match* in the spatial allocation of the categories, given the proportions of the categories in both datasets.
Overall difference	<b>overallDiff</b>	Single value	The overall difference between tabulated <i>observed</i> and <i>predicted</i> data calculated from the sum of the quantity and allocation components of difference.
Overall exchange difference	<b>overallExchanged</b>	Single value	Exchange is the transition from category <i>i</i> to category <i>j</i> in some observations and a transition from category <i>j</i> to category <i>i</i> in an identical number of other observations.
Overall quantity difference	<b>overallQtyD</b>	Single value	This is the amount of difference between the <i>observed</i> variable and a <i>predicted</i> variable that is due to the less than maximum match in the proportions of the categories.
Overall shift difference	<b>overallShiftD</b>	Single value	Shift describes to the difference remaining after subtracting Quantity difference and Exchange from the Overall difference.
Category exchange difference	<b>exchangeDj</b>	Value for each class	The exchange value as above at the class level.
Category overall difference	<b>overallDiffCatj</b>	Value for each class	The overall difference as above at the class level.
Category quantity difference	<b>quantityDj</b>	Value for each class	The quantity difference as above at the class level.
Category shift difference	<b>shiftDj</b>	Value for each class	The shift difference as above at the class level.

## Addendum: Kappa Estimates

Although now widely discredited, many remote sensing analyses still use the Kappa estimate,  $\hat{\kappa}$ . Its derivation is below:

$$\hat{\kappa} = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} * x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} * x_{+i})} \quad (2)$$

where  $N$  is the total number of observations in the matrix,  $r$  is the number of rows,  $x_{ii}$  is observations in row  $i$  and column  $i$ ,  $x_{i+}$  and  $x_{+i}$  are the marginal totals of row  $i$  and column  $i$ , respectively. Essentially what this does is:

1. Multiply the sum of the diagonals by the table sum.
2. Then subtract from this the sum of the row and column marginal totals products.

3. Next, divide this by the sum of the table squared - minus the sum of the row and column marginal totals products as above The top part of the equation gives a measure of chance agreement and the bottom part a measure of the expected disagreements.

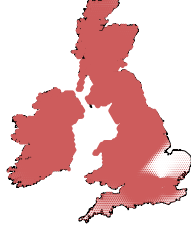
## Code

The `R`markdown script used to produce this manuscript, including all the code used in the analysis and to produce the mapped figures, can be found at <https://github.com/lexcomber/RSLcode>

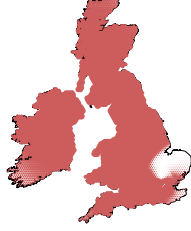
## Results

A fixed bandwidth of 15% of the data points under a bisquare kernel was specified, and for each grid location the Kappa was estimated. A series of geographically weighted analyses were undertaken to compare the Modis and Geo-Wiki data using the metrics described in Table 1. Figure 2 shows maps of the spatial distribution some of the overall comparative measures and Figure 3 shows the spatial distribution of some class specific ones. Note that measures of agreement are *not* related to the density of the data points, rather they reflect local measures calculated from local correspondence matrices at the locations mapped in Figure 1.

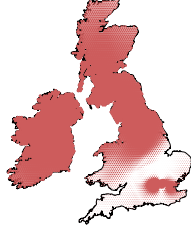
**a) Overall allocation difference**



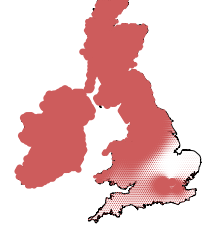
**c) Overall exchange difference**



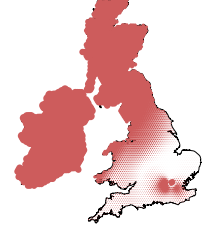
**e) Overall quantity difference**



**b) Overall difference**



**d) Overall quantity difference**



**c) Kappa Estimate**

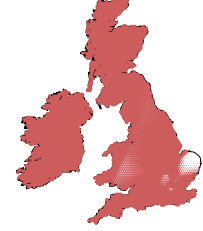


Figure 2: The spatial distribution of the difference measures, scaled to  $[0, 1]$ , comparing the Geo-Wiki crowdsourced data with MODIS modis data.



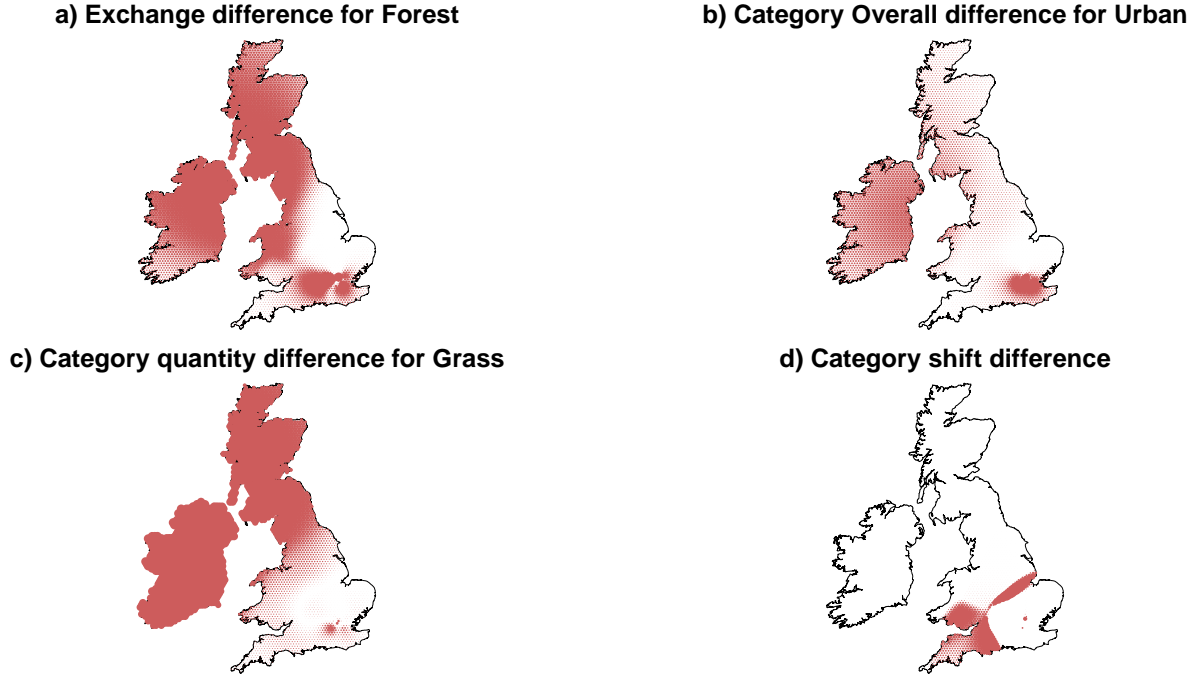


Figure 3: The spatial distribution of the class level, or categorical difference measures, scaled to  $[0, 1]$ , comparing the Geo-Wiki crowdsourced data with MODIS modis data.

## Discussion

The correspondence matrix is the *de facto* method in remote sensing for reporting comparisons between classified data, habitats, land cover and land use, for accuracy and error reporting and for summarise the results of change detection analyses. Local, spatially distributed, geographically weighted cross-tabulations matrices provide a framework for examining *how* and *where* measures derived from the correspondence matrix vary. They are an advance on the logistical regression methods suggested in Comber et al (2012) and Comber (2013) because they support the generation of any local measure that the user wishes to specify. In this case the analyses integrated functions from 2 R packages: the `gwxtab` package to create local geographically weighted correspondence matrices and the overall and categorical difference measures in the `diffeR` package. The availability of open, free and transparent code provides a dynamic and rich research environment within which method extensions can be developed.

There are some important considerations related to the application of GW models, including bandwidth specification and kernel shape as described in Gollini et al (2013). Exploration using a range of bandwidths is valid and recommended. Here a bandwidth of 15% of the data points was selected but bandwidth optimisation tools exist for standard GW approaches are being developed for cross-tabulation statistics.

The GW framework for correspondence tables presented here supports greater understanding of the spatial process and statistical relationships under investigation. This is important as the number and diversity of remote sensing derived products and applications increases and reflects the original aims of geographically weighted regression (Brunsdon et al., 1996). For example, it could be used to dynamically visualize accuracy, to characterise error, to provide local distributions of Chi-squared statistics, to explore the implications of different to locally focus additional ground-truth sampling, to assess the value of the imagery itself locally (e.g. LANDSAT versus MODIS), to explore the utility of

different class definitions (e.g. Cropland vs. Managed grassland, which can be particularly difficult to classify) and to identify locales with missing and misaligned data.

Finally, spatially explicit approaches such as the GW correspondence matrices allow some of the dominant assumptions of spatial non-stationarity of processes within remote sensing methods to be examined and tested. They accommodate the spatial auto-correlation found remote sensing data and analyses of many landscape processes.

## References

- Bicheron, P., Defourny, P., Brockmann, C., Schouten, L., Vancutsem, C., Huc, M., Bontemps, S., Leroy, M., Achard, F., Herold, M., Ranera, F., Arino, O., 2008. Glob- Cover: Products Description and Validation Report, 18, Toulouse, France. URL: [http://due.esrin.esa.int/files/GLOBCOVER\\_Products\\_Description\\_Validation\\_Report\\_I2.1.pdf](http://due.esrin.esa.int/files/GLOBCOVER_Products_Description_Validation_Report_I2.1.pdf)
- Brunsdon CF, Charlton M and Harris P (2016). Geographically Weighted Cross-Tabulation, <https://github.com/chrisbrunsdon/gwxtab>, available 2 July 2016.
- Campbell J, (1981). Spatial correlation effects upon accuracy of supervised classification of land cover. *Photogrammetric Engineering of Remote Sensing*, 47: 355–364.
- Comber A, Mooney P, Purves RS, Rocchini D and Walz A (in press 1). It matters who the crowd are: mapping semantic variations crowd sourced land cover. Paper accepted for publication by PlosONE, (June 2016) DOI: 10.1371/journal.pone.0158329
- Comber,A., See, L., Fritz, S., Van der Velde, M., Perger, C., Foody,G.M. (2013). Using control data to determine the reliability of volunteered geographic information about land cover. *International Journal of Applied Earth Observation and Geoinformation*, 23: 37–48.
- Comber A, Harris P and Tsutsumida N (in press 2). Improving land cover classification using input variables derived from a geographically weighted principal components analysis. Paper accepted for publication in *ISPRS Journal of Photogrammetry and Remote Sensing* (June, 2016).
- Comber A.J., (2013). Geographically weighted methods for estimating local surfaces of overall, user and producer accuracies. *Remote Sensing Letters*, 4(4): 373-380.
- Comber, A., Fisher, P.F., Brunsdon, C. and Khmag, A. (2012). Spatial analysis of remote sensing image classification accuracy. *Remote Sensing of Environment*, 127: 237–246.
- Foody, G.M., 2005. Local characterization of thematic classification accuracy through spatially constrained confusion matrices. *International Journal of Remote Sensing* 26, 1217–1228.
- Friendly, M. (1994). Mosaic displays for multi-way contingency tables. *Journal of the American Statistical Association*, 89(425), 190-200.
- Fritz, S., Bartholomé, E., Belward, A., Hartley, A., Stibig, H. J., Eva, H., ... & Roy, P. S. (2003). Harmonisation, mosaicing and production of the Global Land Cover 2000 database (Beta Version). Publication of the European Commission EUR, 20849.
- Gollini, I., Lu, B., Charlton, M., Brunsdon, C., & Harris, P. (2013). GWmodel: an R package for exploring spatial heterogeneity using geographically weighted models. arXiv preprint arXiv:1306.0413.
- Hartigan, J. A., & Kleiner, B. (1981). Mosaics for contingency tables. In *Computer science and statistics: Proceedings of the 13th symposium on the interface* (pp. 268-273). Springer US.
- Khormi, H. M., & Kumar, L. (2011). Modeling dengue fever risk based on socioeconomic parameters, nationality and age groups: GIS and remote sensing based case study. *Science of the Total Environment*, 409(22), 4713-4719.

- Lesiv M, Moltchanova E, Schepaschenko D, See L, Shvidenko A, Comber A and Fritz S (2016). Comparison of data fusion methods using crowdsourced data in creating a hybrid forest cover map. *Remote Sensing*, 8, 261; doi:10.3390/rs8030261
- Loveland, T.R., Reed, B.C., Brown, J.F., Ohlen, D.O., Zhu, Z., Yang, L., Merchant, J.W., 2000. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *International Journal of Remote Sensing* 21 (6–7), 1303–1330.
- McGwire, K. C., & Fisher, P. (2001). Spatially variable thematic accuracy: Beyond the confusion matrix. In C. T. Hunsaker, M. F. Goodchild, M. A. Friedl, & T. J. Case (Eds.), *Spatial uncertainty in ecology: Implications for remote sensing and GIS applications* (pp. 308–329). New York: Springer–Verlag.
- Perger C, Fritz S, See L, Schill C, Van der Velde M, McCallum I et al. A campaign to collect volunteered geographic Information on land cover and human impact. In: Jekel T, Car A, Strobl J and Griesebner G, editors. *GI\_Forum 2012: Geovizualisation, Society and Learning*. Berlin: Verlag; 2012. pp 83-91.
- Pontius Jr, R.G. and Millones, M., 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, 32(15), pp.4407-4429.
- Pontius Jr, R.G., Santacruz, A. 2014. Quantity, exchange and shift components of difference in a square contingency table. *International Journal of Remote Sensing* 35 (21), 7543-7554.
- Pontius Jr, R.G., Santacruz, A. 2015. Package ‘diffeR’: Metrics of Difference for Comparing Pairs of Maps. <https://cran.r-project.org/web/packages/diffeR/diffeR.pdf> [available 18 July 2016]
- Propastin, P. (2012). Modifying geographically weighted regression for estimating aboveground biomass in tropical rainforests by multispectral remote sensing data. *International Journal of Applied Earth Observation and Geoinformation*, 18, 82-90.
- Tobler, W.R., 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46(sup1), pp.234-240.
- Tsutsumida N. and Comber A.J. (2015). Measures of spatio-temporal accuracy for time series land cover data. *International Journal of Applied Earth Observation and Geoinformation*, 41: 46-55
- Wang, Q., Ni, J., & Tenhunen, J. (2005). Application of a geographically weighted regression analysis to estimate net primary production of Chinese forest ecosystems. *Global Ecology and Biogeography*, 14(4), 379-393.
- Yu, D., & Wu, C. (2004). Understanding population segregation from Landsat ETM+ imagery: a geographically weighted regression approach. *GIScience & Remote Sensing*, 41(3), 187-206.