Article:

Comber, A, Fonte, C, Foody, G et al. (4 more authors) (2016) Geographically weighted evidence combination approaches for combining discordant and inconsistent volunteered geographical information. GeoInformatica, 20 (3). pp. 503-527. ISSN 1384-6175

https://doi.org/10.1007/s10707-016-0248-z

Geographically weighted evidence combination approaches for combining discordant and inconsistent volunteered geographical information

Alexis Comber[1], Cidália Fonte[2], Giles Foody[3], Steffen Fritz[4], Paul Harris[5], Ana-Maria Olteanu-Raimond[6] and Linda See[4]

[1] School of Geography, University of Leeds, LS2 9JT, UK
[2] Department of Mathematics, University of Coimbra, Coimbra, Portugal / Institute for Systems and Computers Engineering at Coimbra, Coimbra, Portugal
[3] School of Geography, University of Nottingham, Nottingham, UK
[4] Ecosystems Services and Management Program, International Institute for Applied Systems Analysis, Laxenburg, Austria
[5] Sustainable Soil and Grassland Systems, Rothamsted Research, North Wyke, Okehampton, EX20 2SB, UK
[6] COGIT Laboratory, French Mapping Agency, 73 Avenue de Paris, 94160 Saint-Mandé, France

Abstract:
There is much interest in being able to combine crowdsourced data. One of the critical issues in information sciences is how to combine data or information that are discordant or inconsistent in some way. Many previous approaches have taken a majority rules approach under the assumption that most people are correct most of the time. This paper analyses crowdsourced land cover data generated by the Geo-Wiki initiative in order to infer the land cover present at locations on a 50km grid. It compares four evidence combination approaches (Dempster-Shafer, Bayes, Fuzzy Sets and Possibility) applied under a geographically weighted kernel with the geographically weighted average approach applied in many current Geo-Wiki analyses. A geographically weighted approach uses a moving kernel under which local analyses are undertaken. The contribution (or salience) of each data point to the analysis is weighted by its distance to the kernel centre, reflecting Tobler's 1st law of geography. A series of analyses were undertaken using different kernel sizes (or bandwidths). Each of the geographically weighted evidence combination methods generated spatially distributed measures of belief in hypotheses associated with the presence of individual land cover classes at each location on the grid. These were compared with GlobCover, a global land cover product. The results from the geographically weighted average approach in general had higher correspondence with the reference data and this increased with bandwidth. However, for some classes other evidence combination approaches had higher correspondences possibly because of greater ambiguity over class conceptualisations and / or lower densities of crowdsourced data. The outputs also allowed the beliefs in each class to be mapped. The differences in the soft and the crisp maps are clearly associated with the logics of each evidence combination approach and of course the different questions that they ask of the data. The results show that discordant data can be combined (rather than being removed from analysis) and that data integrated in this way can be parameterised by different measures of belief uncertainty. The discussion highlights a number of critical areas for future research.

Key words: Crowdsourcing, land cover, data quality, VGI, data mining

1. Introduction

One of the critical areas of research in information sciences is how to combine data or information that are discordant or inconsistent in some way. Discord may arise from different spatial frameworks, different measurement devices, different classifications or measurement units. However there is much scientific interest in being able to combine data from different sources to enhance information value and utility, to measure change if the data are temporal, to compare different treatments and so on. This research develops, applies and compares a number of geographically weighted evidence combination approaches to a crowdsourced land cover problem. The aim was to use the sometimes conflicting crowdsourced data to describe or infer the land cover present at specific locations and to generate some measure of confidence or belief in the inference.

There has been much recent interest in the potential opportunities for extracting information contained in large datasets arising through big data initiatives and the many crowdsourcing activities. Much of this interest relates to the high data volumes and low cost of crowdsourced data and to the opportunities for extracting useful and novel knowledge by integrating information from the many data silos under big data. A number of publications have described the potential opportunities arising from analysis of crowdsourced data [1-4] but one of the critical and as yet unaddressed issues is how to deal with conflicting information. Most of the current solutions to this problem adopt a Linus law or majority rules approach, where the majority view of contributors is deemed to be correct. In many situations this approach is intuitive and logical. However, it implicitly assumes that every contribution (or datum) has equal salience, that the majority of observations are correct and pragmatically treats the observers and their observations as independent. In reality some observations and some observers may be more reliable than others [4, 5], which may matter in some circumstances but not in others [5, 6] and observations may be exhibit spatial non-stationarity and strongly influence by local processes reflecting Tobler's First law of geography[1] [7].

---

1 "Everything is related to everything else, but near things are more related to each other"

Critically Tobler's First law infers that local information and context should be considered when combining spatial data and Geographically Weighted approaches have been proposed to do this [8]. In this paper geographically weighted evidence combination approaches based on Bayesian Probability, Dempster-Shafer, Fuzzy Sets and Possibility theories were developed and applied. Geographically weighted approaches use a moving kernel to weight data inputs by their distance to the location being considered [8]. The geographically weighted evidence combination approaches were applied to crowdsourced data describing land cover to infer the presence of land cover at each location in a study area. Other research has shown how different evidence combination methods partition uncertain evidence in different ways. This is due to their different underlying assumptions and logics and consequently they ask different questions of the evidence they combine [9 -11]. The paper develops and applies these methods for handling uncertain evidence under a geographically weighted framework for the first time. In so doing, it explores their utility for generating useable spatial information from crowdsourced data, parameterised by uncertainty, and considers their performance under different sized kernels overall and for specific land cover classes.

2. Background

There has been much interest in crowdsourced data with applications ranging from astronomy to zoology [12]. Data are varied and include all kinds of digital information from microblogs, tagged photographs to web based interfaces. These data are easily shared with others via dedicated servers to which information is uploaded or through informal social networks. The result is a very dynamic data environment and there is huge scientific interest in opportunities afforded by the high data volumes at low cost. Of note is that data contributed by citizens are increasingly spatially referenced due to the increasing number of portable GPS- and web-enabled digital devices (e.g. smartphones, tablets, etc.).

The terms ‘crowdsourced’, ‘citizen science’ and ‘volunteered geographic information’ have their own nuanced meanings and recent work has developed a typology of crowdsourced data to try to capture some of this [13]. The term *‘crowdsourcing’* originally referred to the ability of citizens to validate and correct the errors that an individual might make and to potentially arrive at some truth [14]. A recent example is the Geo-Wiki project. This web-based interface to Google Earth was initially used to validate a global biofuels availability dataset [15, 16] and has subsequently been used for a number of other campaigns. Typically Geo-Wiki campaigns captures data describing the land cover class at a series of locations. In some campaigns additional information is captured for example on the amount of human disturbance in the scene, the user confidence in their class allocation, etc, and, depending on the campaign, the points may be selected at random or they may be repeatedly sampled. Currently many Geo-Wiki analyses apply a geographically weighted averaging approach as described in Comber et al [4] to combine crowdsourced land cover data.

In contrast to much of the initial research using crowdsourced data to validate other data products, recent work has focussed on methods for assessing the quality of the crowdsourced data itself. Studies of the data generated by the Geo-Wiki campaigns have compared them with external data [4], with control data [5] and have applied internally based latency measures [3]. Overall quality, in terms of the correctness of land cover class allocation, has been found to vary only marginally between experts and non-experts [5] but significant differences have been found for specific land cover types in specific areas [6] and in relation to the 'experiential distance' of the observer to the phenomena being considered [17].

One of the critical issues in citizen science and crowdsourcing is how to deal with conflicting information, opinions or versions. Platforms and activities such as Wikipedia and OpenStreetMap benefit from crowdsourcing in its original sense: that of different citizens arriving at a collective view of what the truth is (with apologies to Pickles [18]). This consensual 'Linus's Law' approach generally produces acceptable outcomes, although with some known socio-scientific problems, for example on occasion the versions of the 'super user' or views of the user with the greatest edit persistence will win out despite conflicting or contrary edits in so called 'tag wars' [19].

In many cases, the capture of crowdsourced data is much less nuanced and detailed than Wikipedia or OpenStreetMap. Consider the Geo-wiki initiative. Anyone can login and evaluate land cover at discrete locations. This may result in people with different backgrounds and experiences identifying different land cover types as being present at the same location [20]. As a result one of the critical issues is how to manage conflicting user evaluations of the land cover class. In many citizen science research conflicts over what different observers consider to be there are managed by taking the majority view, many eyes or Linus's Law approach. Haklay et al [2] used this approach to determine feature locations and found that locational accuracy increased as more people contributed to the solution. However, this was not to classify or identify the feature itself and in such contexts other authors have noted that Linus's Law may not be as effective for establishing geographic facts [14, 21]. Goodchild and Li [14] comment that Linus Law approaches may be most suited to what they refer to as 'prominent' geographic facts – that is those that are not obscured by being in a sparsely populated or under-explored location, that persist over time and interest many people. The obvious gap in this list are geographic features that are inherently subjective such as land cover classification where few natural kinds exist [22].

Other research from the information sciences has noted some of the problems associated with multiple crowdsourced opinions. As Zook et al [23, pp 27-28] note "duplication is not necessarily a bad thing as it can provide multiple avenues to access information. It can, however, make interpretation of a situation more complicated as multiple sources can provide conflicting versions of the built and natural environments". Welinder and Perona [24] identified the challenges related to conflicting crowdsourced labels for images, for example when different users assert A and other users not-A. Some research has suggested that the crowd itself is used to deal with conflicts, by supplying rules to resolve disagreements and to merge conflicting inputs and to generate automatic solutions by weighting user scores [25, 26]. Other approaches propose that users contribute to an overall knowledge base in order to generate weighting probabilities [27]. Thus there is a belief that data from multiple sources and data from multiple but imperfect sources are desirable, because they offer the opportunity of insight and knowledge. However many of the solutions that have been proposed thus far are aspatial – they do not consider geographic context – and are not generic – they rely on specific, local conflict resolution strategies to determine whether one assertion should override another [28].

Evidence combination is a strong and long standing area of research within the information sciences. Many methods are available to combine conflicting or uncertain information – Bayesian Probability, Dempster-Shafer, Fuzzy Sets, Possibility theory, etc. – each with different underpinning logics. A number of reviews exists and the interested reader is directed to [29-33]. As yet no research has:

1) considered how approaches that explicitly facilitate reasoning under uncertainty may be used to integrate conflicting crowdsourced data, or
2) applied these methods under explicitly geographical frameworks such as geographically weighted kernels [8].

This paper develops a number of evidence combination approaches that are typically used to handle information uncertainty [11] under a geographically weighted framework in order to combine potentially conflicting crowdsourced data on land cover. The aim was to combine this data and then to describe or infer the land cover present at discrete locations with some indication of the degree of belief in that inference. The evidence combination methods operate in different ways, answering different questions in relation to the data, and the use of geographic kernel explicitly addresses Tobler's first law of geography by considering geographic context. In this way the methods explicitly address the 'geographic approach' to quality assurance of data created by citizens suggested by Goodchild and Li [14].

3. Methods

In overview, crowdsourced land cover data from the Geo-Wiki initiative [15] were analysed using geographically weighted evidence combination approaches to infer the actual land cover class present at each location on a 50km grid. These were applied under a series of different kernel sizes. At each location on the grid, data falling under the kernel were weighted by their distance to the kernel centre and this evidence was then combined using different methods.

3.1 Data and study area

This analysis used Geo-Wiki land cover data. The Geo-Wiki initiative [15] collects volunteered data on land cover in order to support a number of activities that range from land cover data validation to land cover data creation [16]. Geo-Wiki has web and smartphone app interfaces, is open to anyone and volunteers can contribute to different campaigns. In these they allocate what they observe from Google Earth imagery at a series of randomly selected locations, to one of a predefined set of 10 land cover classes. This legend was chosen to be consistent with the generalized land cover classes proposed by Herold et al. [34], which allows for comparison of different land cover products. More details of these competitions can be found in See et al. [35]. In order to provide some form of validation, the inferred land cover at each location on the 50km sampling grid was compared with GlobCover 2009[2], reclassified into the 10 classes and resampled to 50km. The reclassification was the same as that reported in [4]. It was devised by the Geo-Wiki team led by Steffen Fritz and the class to class relations were agreed by consensus by 3 experts using images and discussing them together. The GlobCover thematic aggregations are shown in Table 1. In this research, data from two Geo-Wiki campaigns were combined from 2011 and 2012 for a South American case study containing some 13,738 data out of a global datasets of 100,808 points. The study area, spatial distribution of Geo-Wiki data and the reclassified, spatially aggregated GlobCover data are shown in Figure 1. Note that geographically weighted approaches, as described below, develop local analyses of data that fall under a moving window or kernel. Thus the 50km grid provide a series of locations at which the local analyses take place, irrespective of the number of data points within each grid cell.

Geo-Wiki class	**GlobCover class**
(1) Tree cover	40, 50, 60, 70, 90, 100, 110, 160, 170
(2) Shrub cover	130
(3) Herbaceous / Grassland	120, 140
(4) Cultivated / Managed	11, 14
(5) Mosaic of cultivated & natural	20, 30
(6) Flooded / wetland	180
(7) Urban	190
(8) Snow and ice	220
(9) Barren	150, 200
(10) Open Water	210

Table 1. The Geo-Wiki land cover classes and the GlobCover aggregations

[2] http://due.esrin.esa.int/page_globcover.php

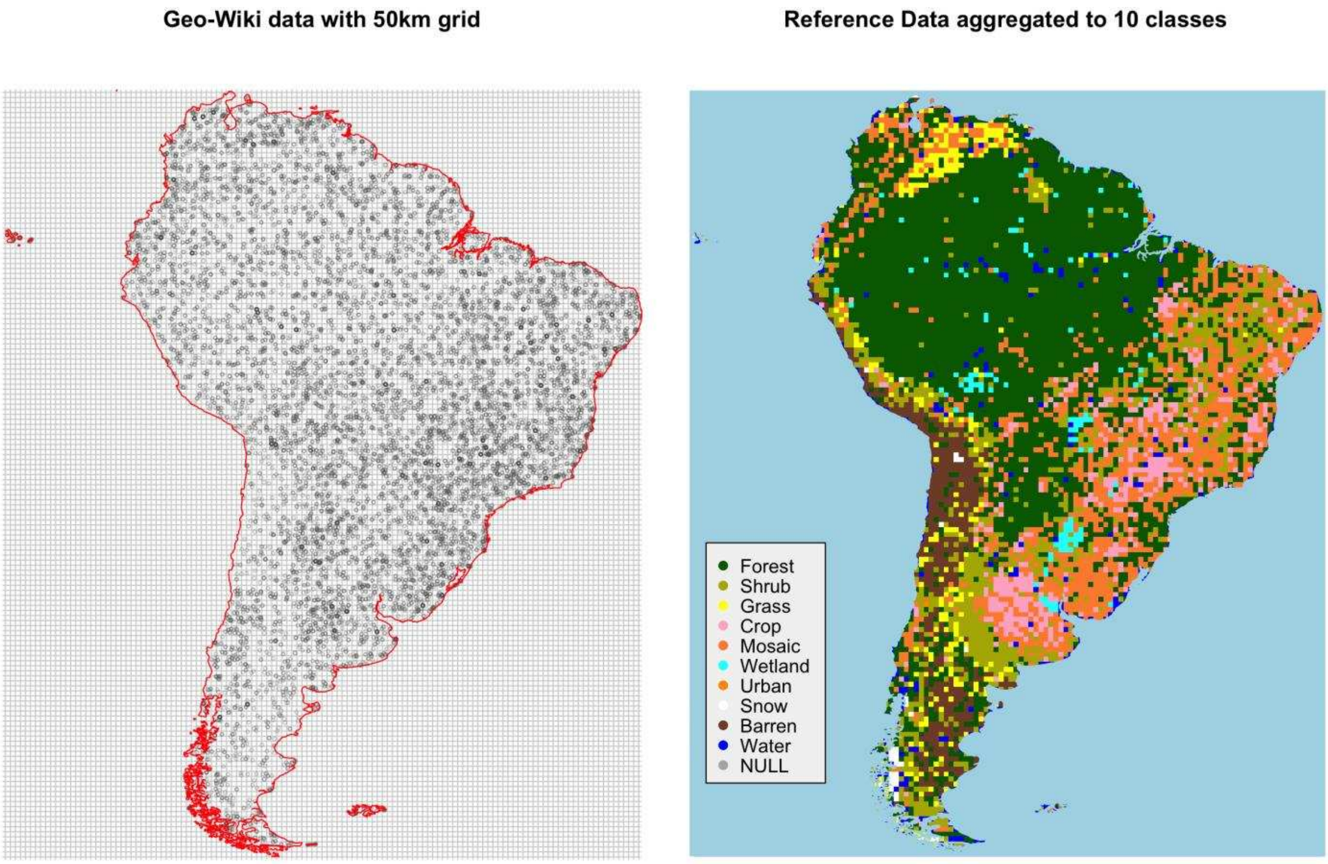


Figure 1. The study area, showing the crowdsourced data locations and the sampling grid (left) and the reclassified GlobCover data.

3.2 Geographically weighted crowdsourced data

A discontinuous geographic kernel was used to select crowdsourced data at each location on a 50km grid as shown in Figure 1. The idea was to use the land cover data points falling under the kernel to infer the land cover present at the centre of the kernel (i.e. a grid point), with data points further away from the kernel centre contributing less to the overall. The geographical extent of the kernel is determined by its bandwidth. In this analysis bandwidths from 5km to 150km at intervals of 5km were examined to explore the interactions of the evidence combination approaches with different scales of aggregation. Each crowdsourced data point records the contributor's opinion of the land cover at that location. It provides evidence in support of a hypothesis of the presence of that land cover class at the centre of the kernel. Each single piece of evidence was weighted according by its distance to the location under consideration (centre of the kernel) to produce a distinct geographically weighted crowdsourced data subset at every grid point. For the smallest bandwidths, these localised data sets will be at their most local but relaying the least information. For the largest bandwidths these localised data sets will be at their least local but relaying the most information. This is the common bias-variance trade-off encountered in any geographically weighted approach [8].

A number of discontinuous kernel functions can be specified as discussed in Gollini et al [36]. In this study a tri-cube function was applied, rather than weights derived under Gaussian or linear functions, as this generates a greater plateau of higher weights near to the kernel centre with a sharp drop off at approximately half the bandwidth. For each crowdsourced data point (P_j) under the kernel (with a given bandwidth), a weight $w_{i,j}$ was calculated based on its distance to the centre of the kernel (K_i) as follows:

$$w_{i,j} = 1 - ((d_{i,j})^3 / b^3) \qquad \text{(Eqn 1)}$$

where $d_{i,j}$ is the distance in metres from the centre of the kernel Ki to the crowdsourced data point Pj and b is the bandwidth at that location. (Note that the tri-cube kernel function specified here is not identical to that in Gollini et al [36] as they incorrectly specified their function in that publication - Paul Harris, pers com). The way that distances to the kernel centre are rescaled by Equation 1 to create distance based weightings is shown in Figure 2 for a bandwidth of 1000.

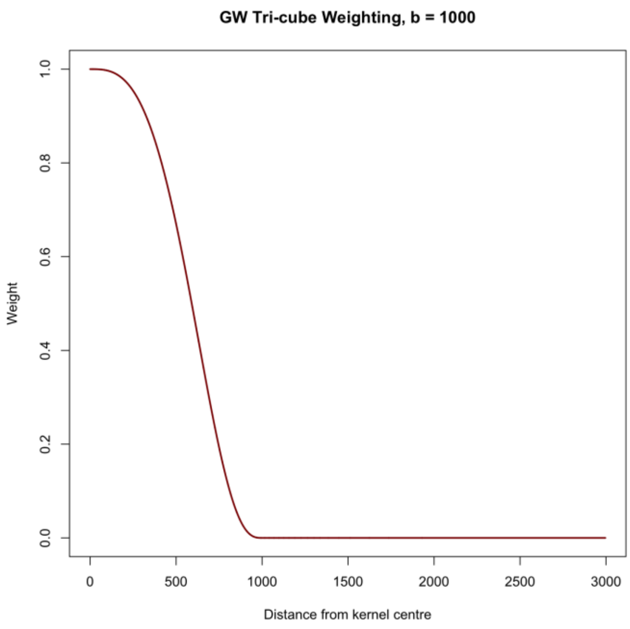


Figure 2. A plot of the weights arising from the Tri-cube kernel function, with the bandwidth b = 1000.

3.3 Evidence combination methods

Four classic approaches for combining uncertain evidence were applied to the weighted data under the kernel: Bayesian Probability, a modified Dempster-Shafer, Fuzzy Sets and Possibility theory. These were compared with a geographically weighted average that simply calculated the proportions of all weights associated with each land cover class. The choice of evidence combination approaches was driven by their ability to handle / partition evidence uncertainty, and by the different ways that they do that, as well as the inability of classic statistical inferential approaches to reason under uncertainty. Some of these have been explored in the context of generating maps from Geo-Wiki data [37] including nearest neighbour, Naive Bayes, logistic regression, classification and regression trees.

In each analysis, the set of classes allocated to the crowdsourced data points under the kernel represent a set of hypotheses to be evaluated. The evidence combination approaches were applied and the degree to which each hypothesis of was supported at the location k_i under consideration was assessed.

A Bayesian approach provides a quantitative estimate of how much belief in a particular proposition or hypothesis h_c corresponding to class c increases (or decreases) when a new piece of evidence, e, corresponding to the data provided by the crowdsourced data points, becomes available. Mathematically this is described for class c and point K_i by:

$$p(h_c|e)(i) = \frac{p(e|h_c)p(h_c)}{\sum_{k=1,m} p(e|h_k)p(h_k)} \qquad \text{(Eqn 2)}$$

where, $p(h_c)$ is the prior probability of hypothesis h_c and $p(h_c|e)$ is the posterior probability of hypothesis h_c, the evidence is e, and $p(e|h_c)$ is the probability of observing this evidence given that h_c is true and m is the number of classes. In this case, $p(e|h_c)$ was given by the weighted mean of the normalised distances $w_{i,j}$, between the location under consideration and the data points under the kernel that where assigned to class c. In this way Bayesian Probability computes a degree of belief in an uncertain hypothesis given the numerical evidence for itself and competing hypotheses. In this instance, the competing hypotheses relate to other land cover classes and posterior probabilities are computed for each land cover for location.

Dempster-Shafer is an extension of Bayes that allows for the situation where weak support for a proposition does not have to imply strong support for its negation. It assesses the belief that a proposition is provable given the evidence with some modifications (see below). Mathematically, for 2 pieces of evidence A and B, such as the weights assigned to data points of the same class, this is expressed as, the mass assignment, $m''(C)$ as follows

$$m''(C) = \sum_{A_i \cap B_j = C} m(A_i) \times m'(B_j) \, \frac{|C|}{|A_i||B_j|} \qquad \text{(Eqn 3)}$$

where $m''(C)$, is equal to the sum of the product $m(A_i)$ and $m'(B_j)$ for all i and j such that that the intersection of sets A_i and B_j equals C. The Fixsen and Mahler modification [38], the |C| etc in Equation 3, are the prior probabilities of the respective evidence sets. Dempster-Shafer does not consider the evidence hypothesis by hypothesis as Bayes does, rather the evidence is considered in light of all of the hypotheses. It generates two measures: Belief – the extent to which the evidence supports the hypothesis in this case a particular land cover class – and Plausibility – the extent to which the evidence does not refute the hypothesis, i.e. Belief with Uncertainty. In this analysis, individual pieces of evidence were created from each of the crowdsourced data points under the kernel. The weight, w generated by Equation 1 was assigned to the class recorded in the individual crowdsourced data point and (1 – w) was allocated to the set of all possible hypotheses for which there is evidence, the frame of discernment. The evidence was then combined using Equation 3.

Fuzzy Set theory develops models of uncertainty based on the degree to which the combined evidence indicates membership to the set under consideration (e.g. the membership to a land cover class). The support for different hypotheses can be evaluated using a suite of methods in fuzzy theory from (simple) weighted linear or convex combination of evidence to (more complex) ordered weighted averaging. Fisher [39] noted that the minimum interval is the standard approach for combining information in fuzzy sets but is counter-intuitive when it is used to compare different land cover classes – it only makes sense in the context of fuzzy land cover when comparing fuzzy sets of the same. For these reasons a number of alternative operators have been suggested. In this case the fuzzy memberships were defined using the weights, $w_{i,j}$, defined in Equation 1 which were transformed into fuzzy memberships, $F_c(i)$ for each land cover class c at the centre i of each kernel, in the following way:

$$F_c(i) = \sum_{\substack{k=j \\ Pj\ classified\ as\ c}} w_{i,k}/n_c \qquad \text{(Eqn 4)}$$

where $w_{i,j}$ are the weights derived from the Gaussian transformed distances described in Equation 1 considering the crowdsourced points P_j classified as class c by the crowd and n_c is the number of crowdsourced data points of class c.

Possibility Theory examines the maximum amount of support for a hypothesis (e.g. the membership to a particular land cover class) using a supremum or Possibility function and an associated uncertainty measure given by a Necessity function [40]. Possibility Theory uses a supremum function (or least upper bound) that relates to the maximum support for any given hypothesis, x. The possibility function, Poss(X) is the supremum of Poss({x}), where x are the set of elements of X and X is the set of all hypotheses.

$$Poss(\boldsymbol{h}) = 1 - Max(w_{i,j}) \qquad \text{(Eqn 5)}$$

The uncertainty associated with X is given by the corresponding necessity function (Nec). The relationship between Necessity and Possibility, in relationship to an hypothesis, **h**, is defined as:

$$Nec(\boldsymbol{h}) = 1 - Max(Poss(\neg\boldsymbol{h})) \qquad \text{(Eqn 6)}$$

where (¬**h**) describes 'not *h*'. In this way the Necessity function (Nec) gives a simple measure of the certainty of the Possibility measure relative to competing hypotheses. In this case the Possibility of the location under the kernel being any given class (i.e the hypothesis, **h**) was determined from the maximum weight value for that class and the Necessity was calculated as above.

Finally a simple geographically weighted average (GW Average) was applied which summed the weights for each class and divided these by the sum of all weights under the kernel.

In summary, at each location in the sample grid each of the geographically weighted evidence approach approaches generated a belief in a hypothesis of the presence of each land cover class at that location. At each location the land cover with the greatest belief was identified for each method and was compared with the GlobCover 2009 land cover class. A correspondence matrix was created and the overall correspondence for that method was calculated from the diagonals. Any locations where no land cover class was allocated (ie they were classified as NULL) were omitted for the correspondence analyses. This was done for each of the kernel bandwidths.

All of the analyses were undertaken in R 3.2.1, the open source software (http://cran.r-project.org) with extensive spatial analysis and mapping functionality. The code and data used in this analysis will be provided on request.

3.4 Worked Example

Consider the following single location on the 50km sample grid described above. There are 20 data points under a 50km kernel at that location as in Figure 3. Each of these has been labelled with a land cover class and contributes evidence in support of an inference about the land cover at that location. The evidence for each point is weighted by its distance to the kernel centre, and the weights are combined using the formalisms described above. Table 1 shows the support or belief associated with the candidate hypotheses.

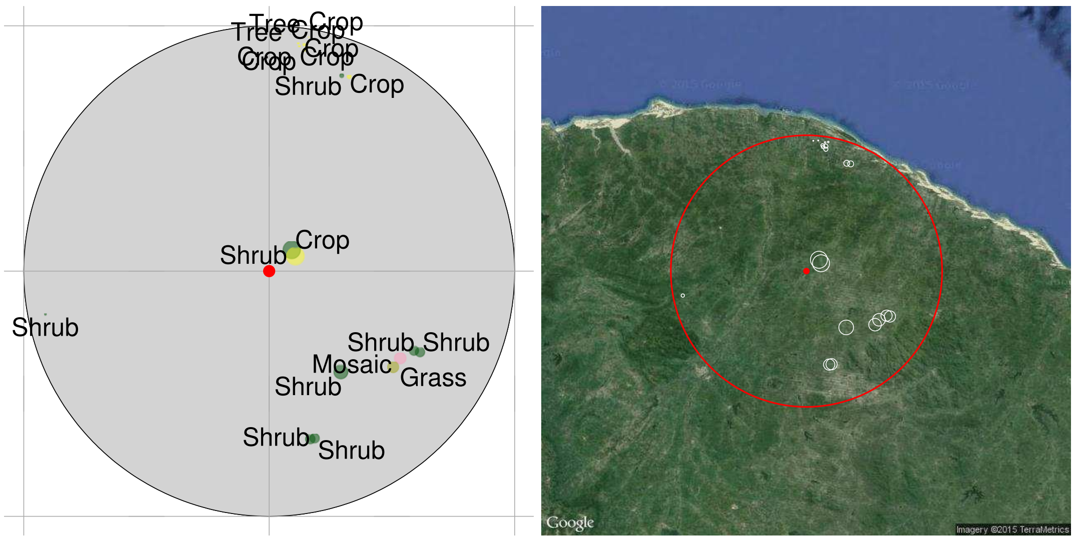


Figure 3. The location being considered in red, the crowdsourced land cover data points with class labels and a small random term added to their location (left figure) and with a Google image as context (right figure). In both figures the size of the data points are related to their distance to the kernel centre and therefore their weight.

Class	Class frequency	Dempster Shafer	Bayesian Probability	Fuzzy Sets	Possibility Theory	GW Average
Tree	2	0.000	0.000	0.000	0.001	0.000
Shrub	8	1.000	1.000	0.273	0.996	0.599
Grass	1	0.000	0.000	0.311	0.402	0.085
Crop	8	0.000	0.000	0.105	0.996	0.230
Flood	1	0.000	0.000	0.311	0.402	0.085
Urban	0	0.000	0.000	0.000	0.000	0.000
Snow	0	0.000	0.000	0.000	0.000	0.000

Barren	0	0.000	0.000	0.000	0.000	0.000
Water	0	0.000	0.000	0.000	0.000	0.000

Table 1. The inferences (beliefs, probabilities, memberships, possibilities) of the land cover present derived from the evidence combination approaches for the worked example.

This generates an interesting set of results: Dempster-Shafer and the Bayes fully support the hypothesis that the class is Shrub, and this is driven by the greater weights compared to the Crop class. For Fuzzy Sets approach the inference of Crop or Shrub class is weakened by the presence weaker weights associated with the data points at the edge of the kernel. Possibility has very strong belief in a hypothesis of Crop and of Shrub but also has some belief in the possibility of the class being Flood or Grass. The GW Average approach identifies Shrub as having the greatest mass of weighted evidence. Possibility Theory, Fuzzy Sets and GW Average indicate some degree of uncertainty in the inference with more than 1 hypothesis have a high degree of support in this example.

4. Results

4.1 Comparison with reference data

The analyses were run for 30 bandwidths of 5km to 150km in intervals of 5km. At each location on a 50km grid and for each evidence combination method, the class with the greatest support was identified and compared with the land cover class at the same location from the aggregated Globcover 2009 dataset. The overall correspondences for each combination method were calculated and the results are shown in Figure 4. This shows a general trend of increasing rates of correspondence with increasing kernel bandwidth for Dempster-Shafer, Bayes and GW Average. The geographically weighted average has the greatest correspondence regardless of bandwidth and Fuzzy Sets decreases with increased bandwidth (and therefore the number of data points being considered at each location) and Possibility plateaus around a kernel bandwidth of 75km.

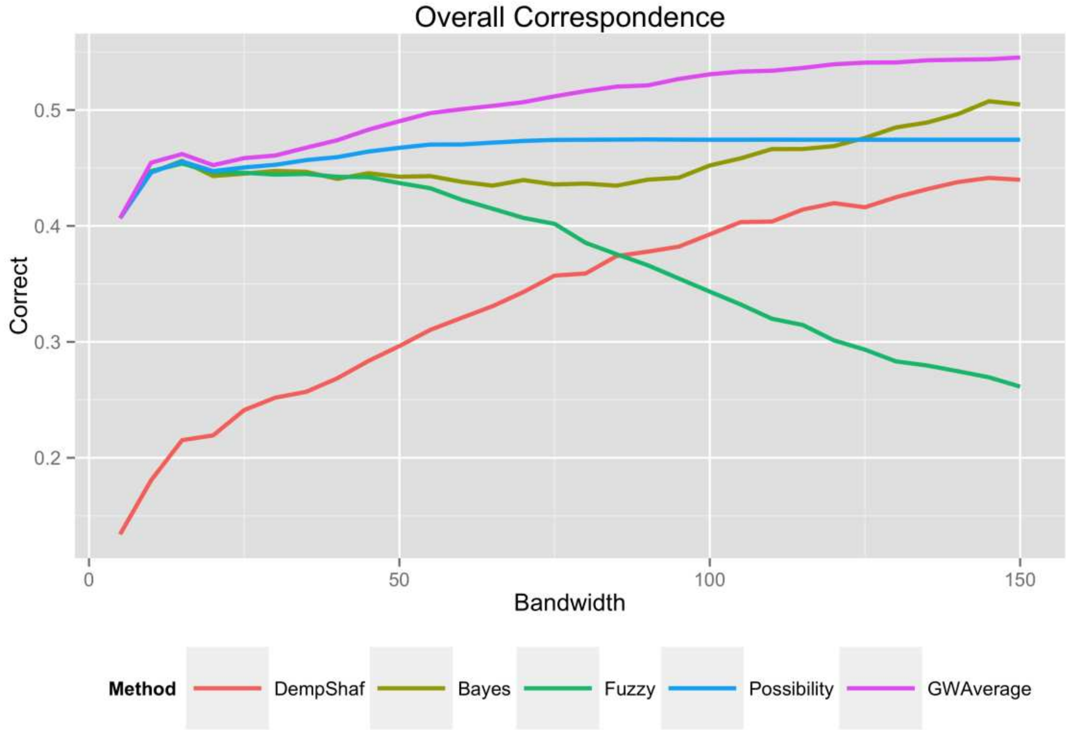


Figure 4. The trends in correspondence with GlobCover for the different evidence combination approaches as kernel bandwidth increases.

It is also instructive to consider how the correspondences vary for different land cover classes: the case study area has a particular mix of cover types and scales of ecological and anthropogenic processes. The per class correspondences were calculated from collapsed correspondence matrices (ie describing binary classes of Class and Not-Class) to determine the degree to which an inference about the presence or absence of a particular land cover class was reflected in the reference data. The results are not therefore directly comparable with the data in Figure 4. The per class correspondences are shown in Figure 5, plotted with the same Y-axis. This clearly shows that for some classes bandwidth does not matter (Urban, Flood, Snow) regardless of evidence combination method. It also shows that for some classes bandwidth is important (Mosaic, Grass) and that for some classes the importance of bandwidth depends on the evidence combination method. For example under Dempster-Shafer, Grass corresponds well at small bandwidths and Water poorly at except larger bandwidths. Table 2 summaries bandwidth and evidence combinations that generate the highest correspondences for each class.

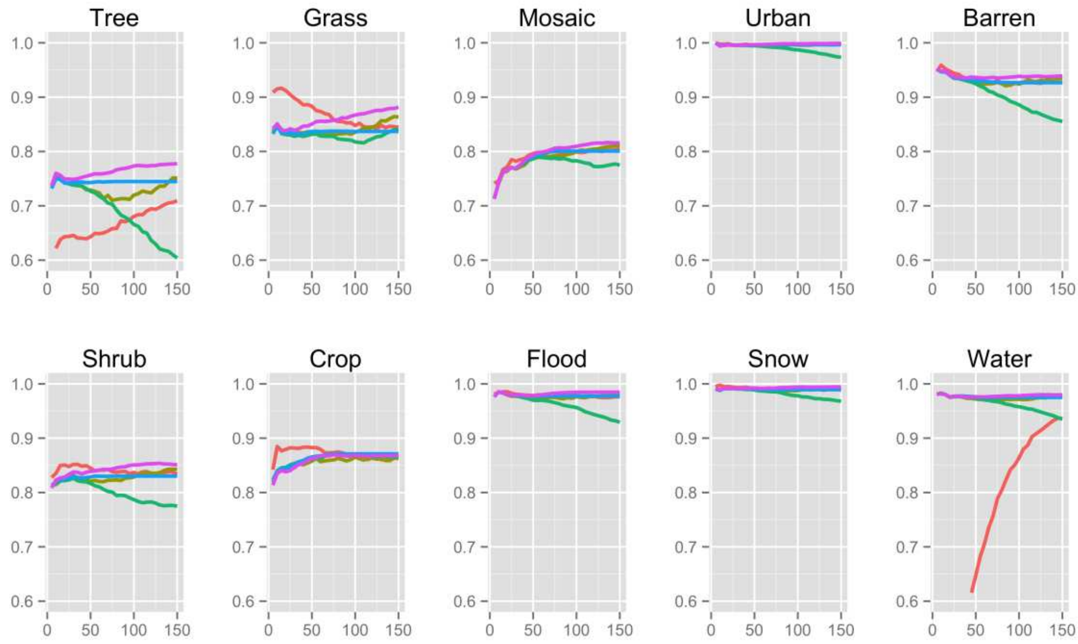


Figure 5. The per class correspondences for the different evidence combination approaches with increasing kernel bandwidth.

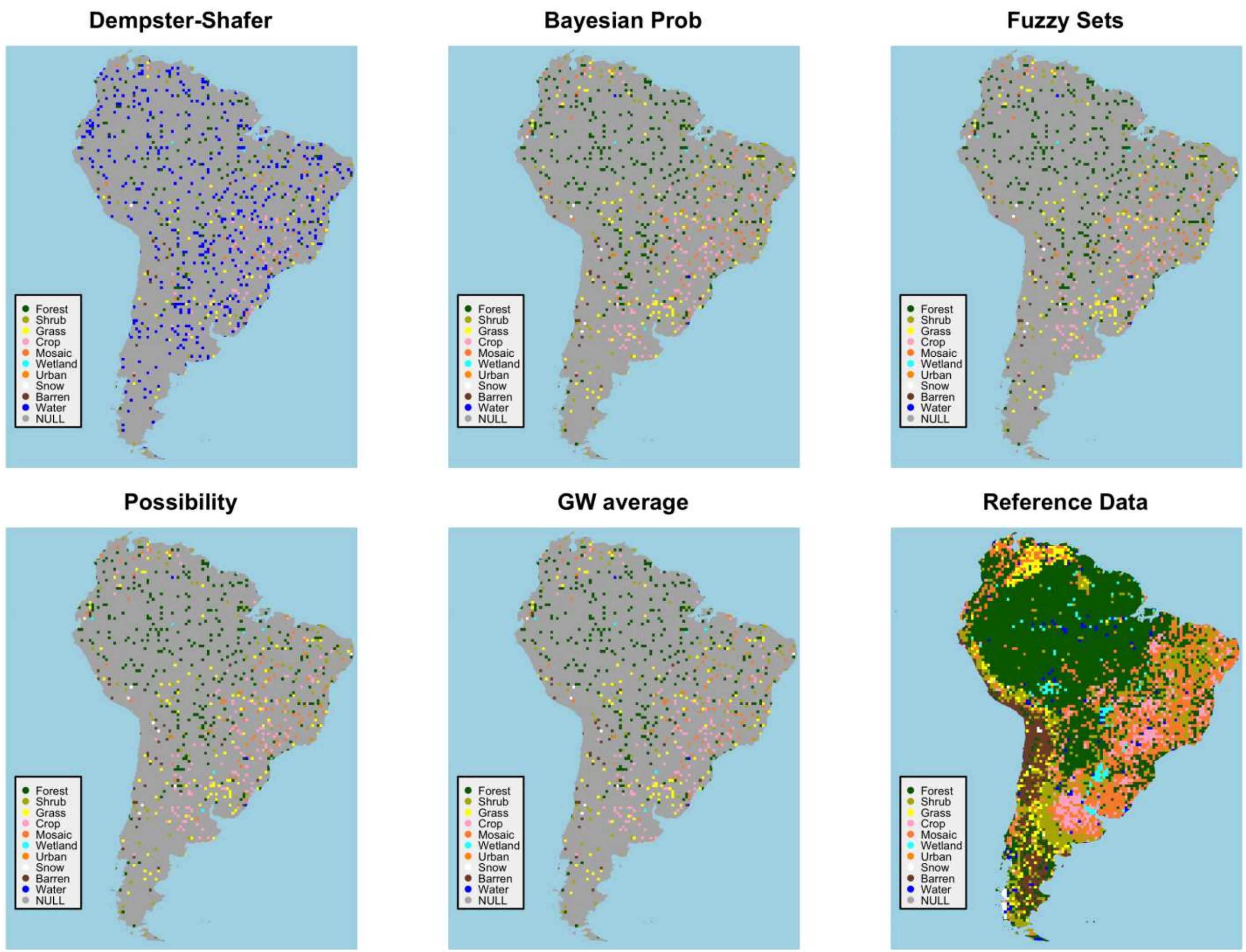
Class	Class	Bandwidth (km)	Method
Tree	1	150	GW Average
Shrub	2	130	GW Average
Grass	3	15	Dempster-Shafer
Crop	4	10	Dempster-Shafer
Mosaic	5	135	GW Average
Flood	6	20	Dempster-Shafer
Urban	7	5	Dempster-Shafer
Snow	8	10	Dempster-Shafer
Barren	9	10	Dempster-Shafer
Water	10	10	Bayesian Probability

Table 2. The bandwidth and evidence combination combinations that produce the highest correspondences with the reference data for each class

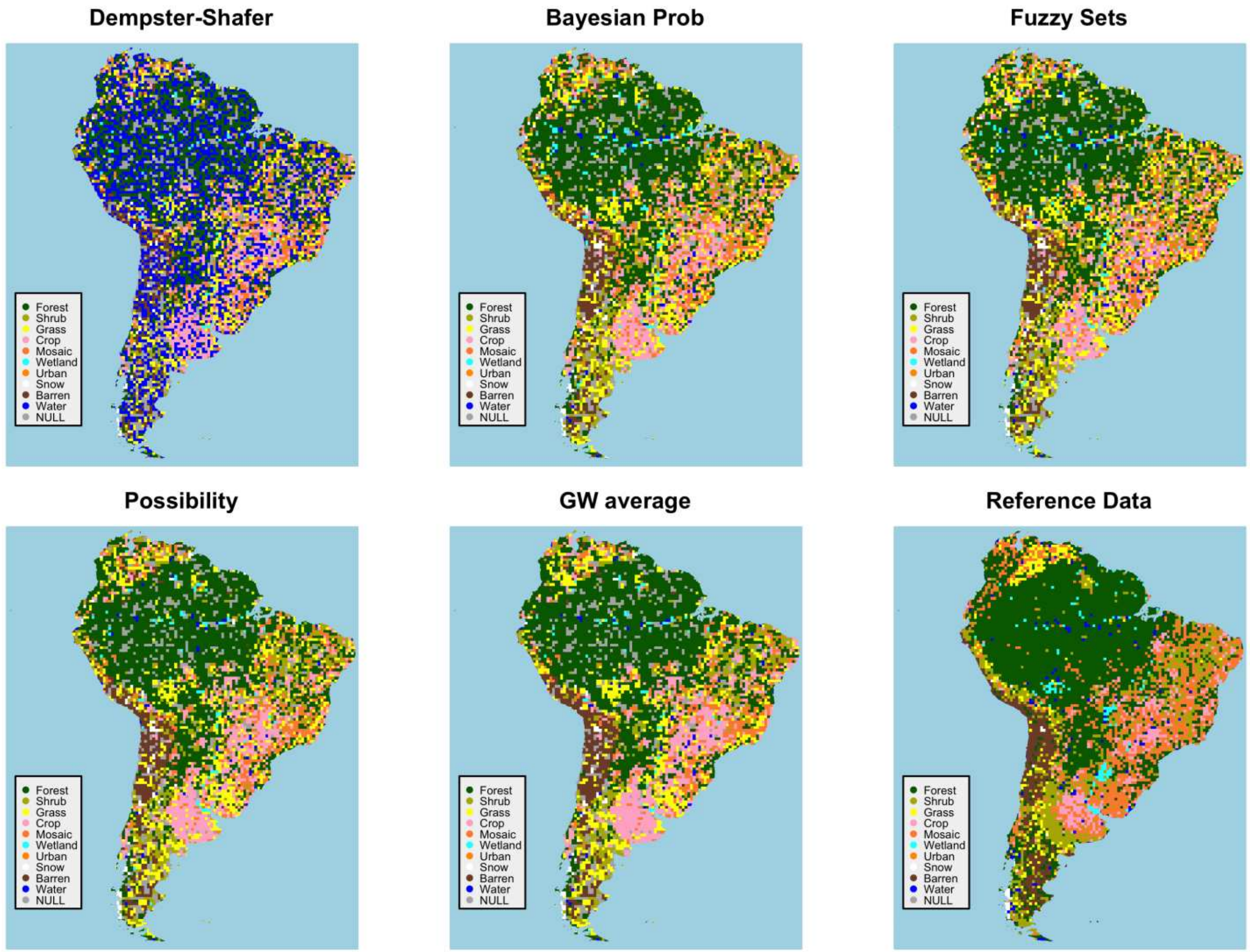
It is possible to generate maps of the land cover with the highest levels of belief generated by each approach. Figure 6 a) to d) shows the maps of these under 10km, 50km 100km and 150km kernels. These illustrate the interaction of the evidence combination approach and bandwidth (and of course the reference data). A number of trends are evident:

- 10km kernel: there is a large amount of land classified as NULL (in grey) because of an insufficient number of crowdsourced data points falling under the kernel. As the kernel size increases the NULL data decreases (note that any areas classified as NULL were omitted for them correspondence analyses above).

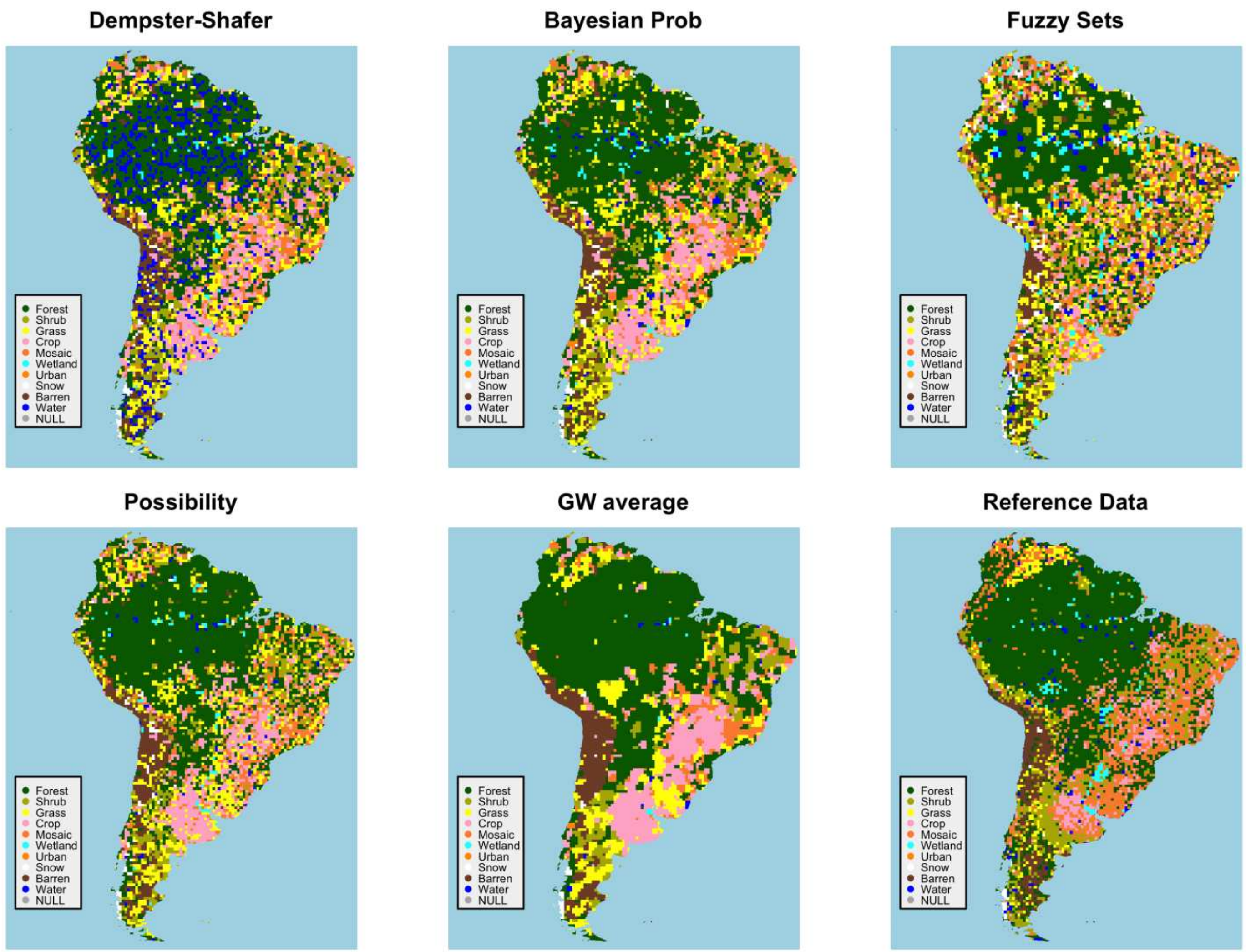
- 50km kernel: the land cover patterns (spatial distributions) start to become more similar to the reference data, with fewer NULL areas, with some spatial heterogeneity. In many areas the Dempster-Shafer analysis identified Water as the class with the greatest degree of belief (in blue). However, it is important to remember that the Dempster-Shafer results are only considering the support for the singleton class hypotheses. The maps from the other approaches start to resemble the reference data but still with some large unclassified areas.
- 100km kernel: the differences between land cover are apparent with this kernel. All of the approaches are converging on the reference data, but with Fuzzy sets showing a much greater degree of divergence than the others and GW Average a much greater degree of smoothing.
- 150km kernel. This trend continues with the largest kernel. Fuzzy generates numerous large, heterogeneous, aggregated areas and GW Average may be over-generalising (smoothing) the data. Visually the Possibility and Bayes approaches have the closest spatial similarly to the reference data.


Dempster-Shafer
Bayesian Prob
Fuzzy Sets
Possibility
GW average
Reference Data
Forest
Shrub
Grass
Crop
Mosaic
Wetland
Urban
Snow
Barren
Water
NULL

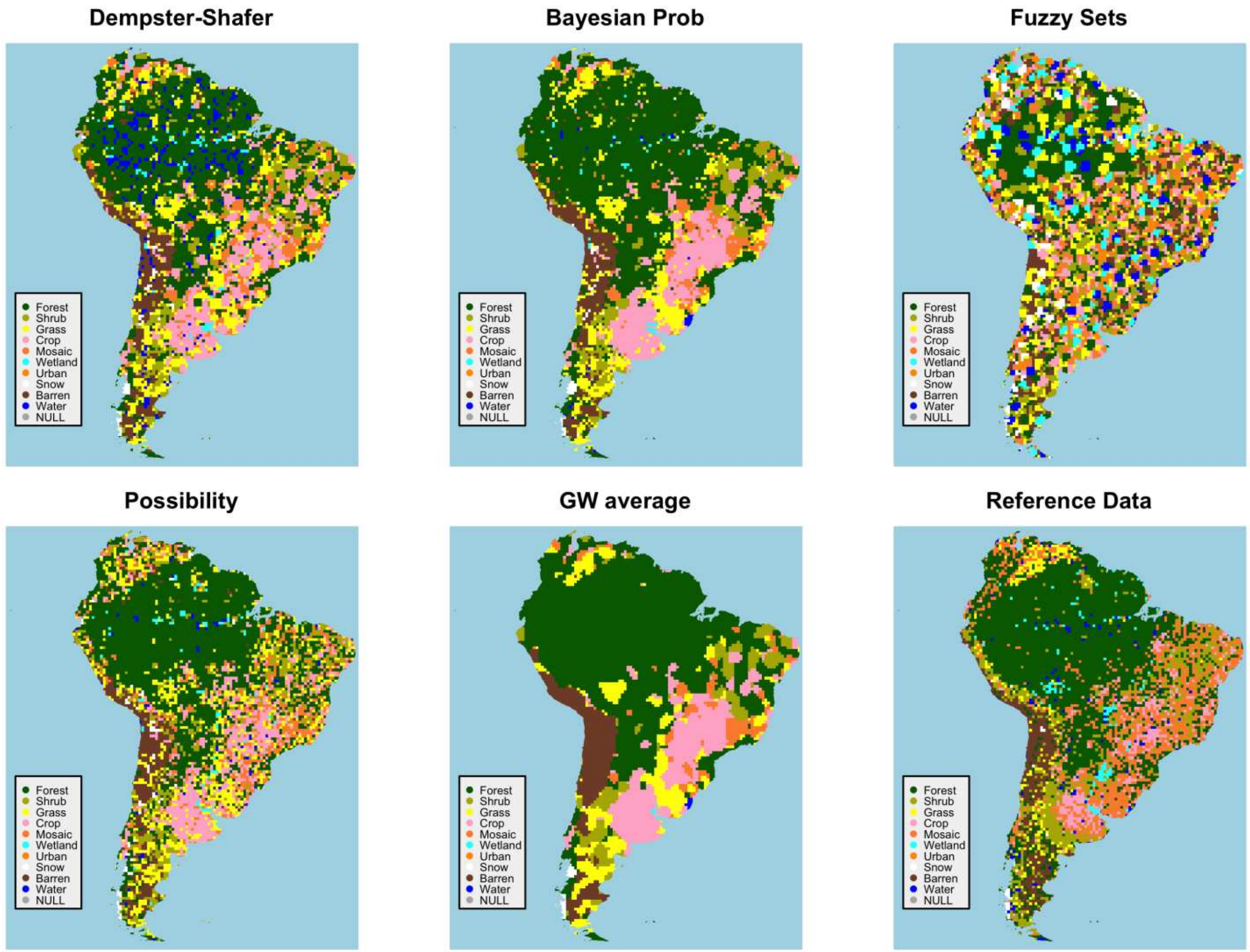
a) 10km


Dempster-Shafer
Bayesian Prob
Fuzzy Sets
Possibility
GW average
Reference Data
Forest
Shrub
Grass
Crop
Mosaic
Wetland
Urban
Snow
Barren
Water
NULL

b) 50km


Dempster-Shafer
Bayesian Prob
Fuzzy Sets
Possibility
GW average
Reference Data
Forest
Shrub
Grass
Crop
Mosaic
Wetland
Urban
Snow
Barren
Water
NULL

c) 100km

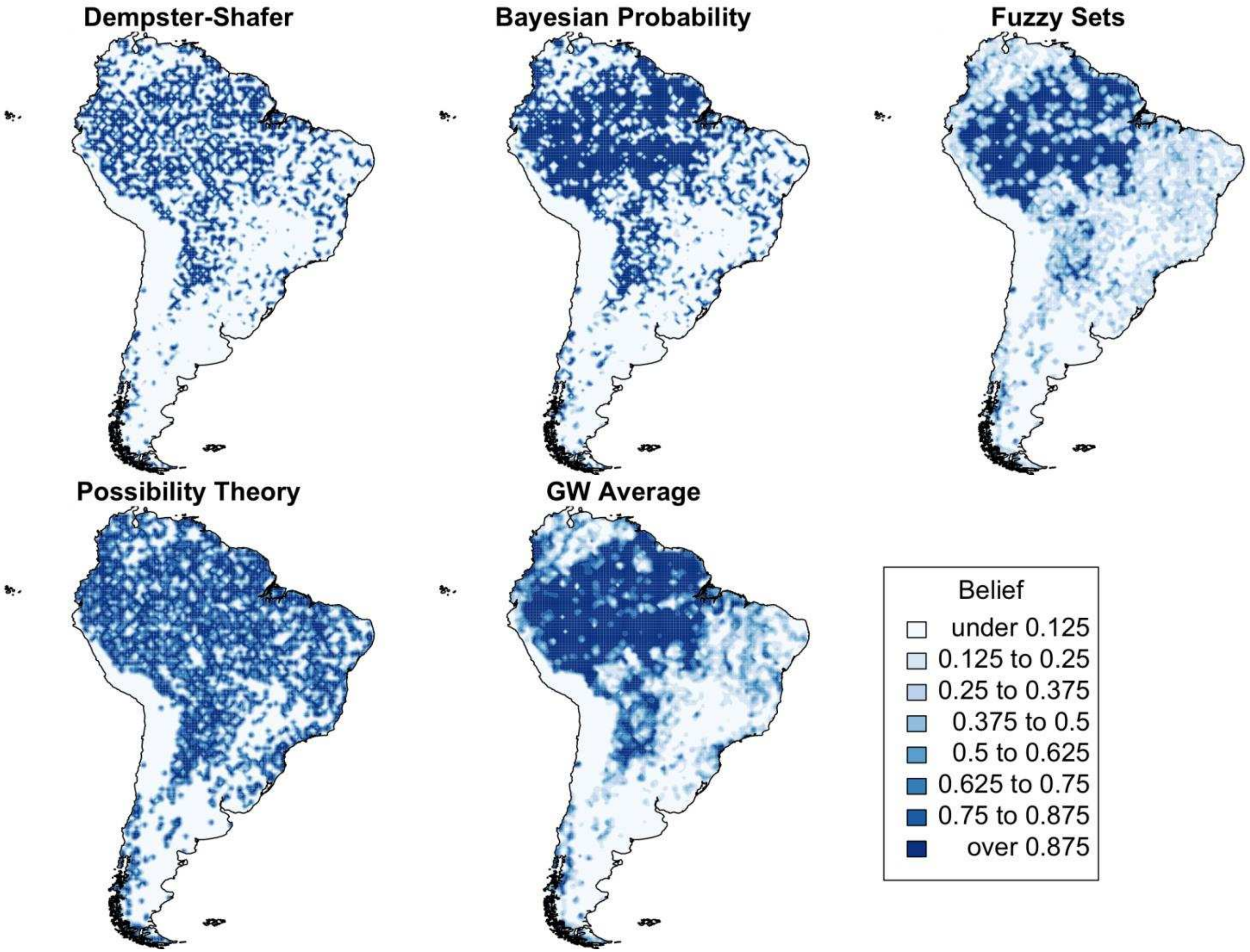


d) 150km

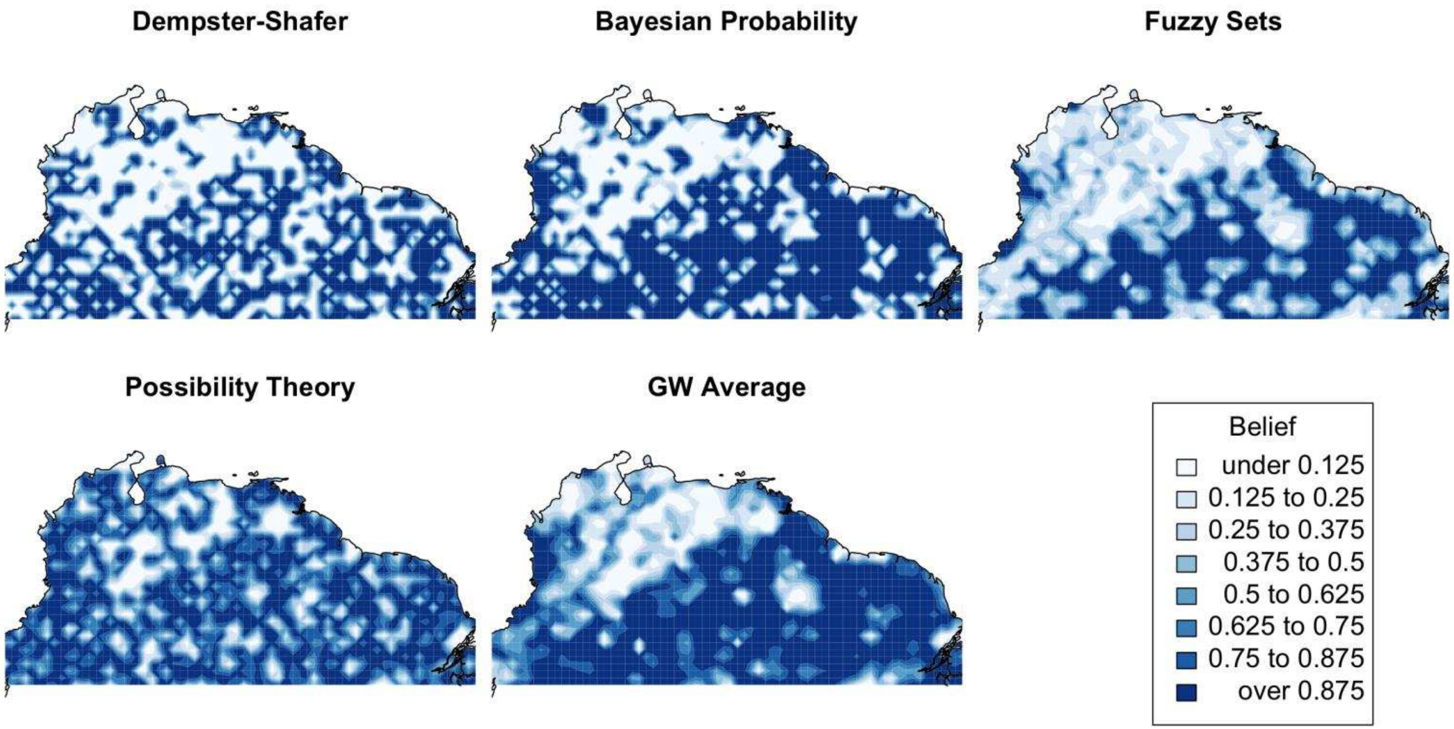
Figure 6 a to d). Maps of the land cover with the greatest belief under different evidence combination approaches and kernel sizes.

4.2 Soft classifications

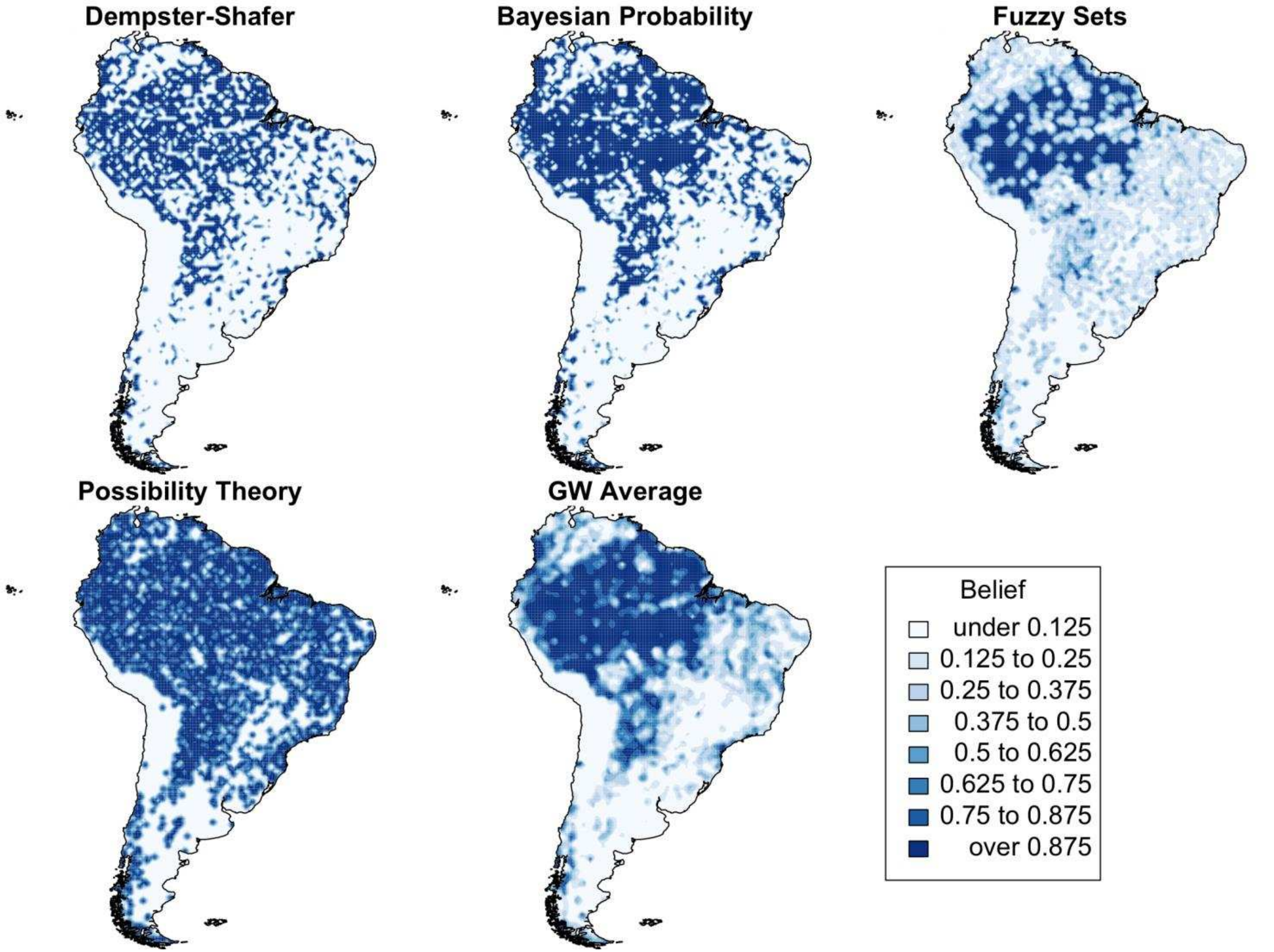
It is important to remember that all of the evidence approaches generate measures of belief in the hypotheses in the interval [0,1]. These beliefs (or memberships) in land cover can be mapped to indicate the uncertainty associated with the competing inferences about the land covers that are present. To illustrate the outputs of such classifications, the maps in Figure 7 shows the belief in the presence of the Tree land cover class arising from different approaches under kernels of 80km and 100km with local detail in the north of the study area. This class was chosen to illustrate the soft, uncertain classifications because it is the most numerous in the crowdsourced data with some 5,466 points. Other classes exhibit similar but sparser spatial structures under the different evidence combination approaches.


Dempster-Shafer
Bayesian Probability
Fuzzy Sets
Possibility Theory
GW Average
Belief
under 0.125
0.125 to 0.25
0.25 to 0.375
0.375 to 0.5
0.5 to 0.625
0.625 to 0.75
0.75 to 0.875
over 0.875

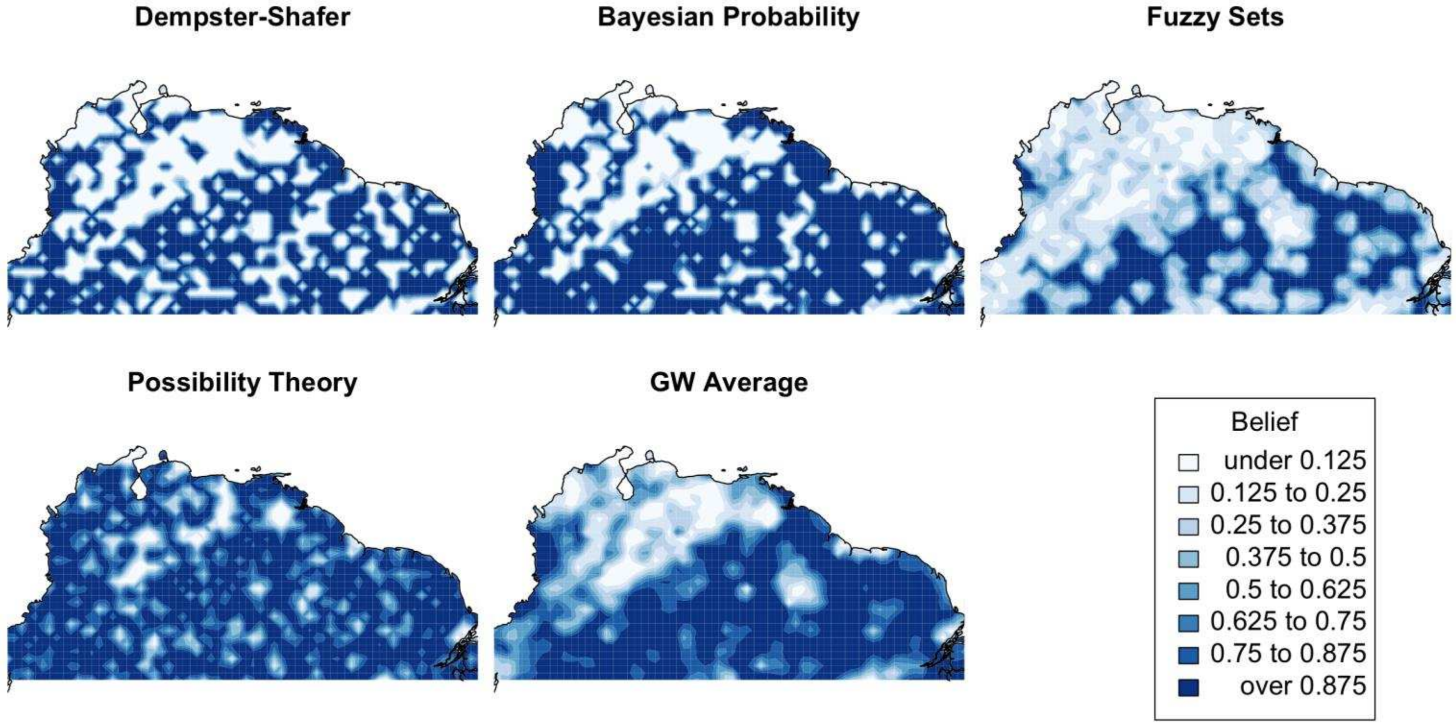
a) 80km



b) 80km detail


Dempster-Shafer
Bayesian Probability
Fuzzy Sets
Possibility Theory
GW Average
Belief
under 0.125
0.125 to 0.25
0.25 to 0.375
0.375 to 0.5
0.5 to 0.625
0.625 to 0.75
0.75 to 0.875
over 0.875

c) 100km



d) 100km detail

Figure 7. a) to d) Belief in the presence of Tree land cover inferred from the crowdsourced data under bandwidths of 80km and 100km, with local detail.

There are consistent spatial trends across the different bandwidths and the different evidence combination approaches. In each case a pattern similar to the distribution of Tree land cover in Figure 1 is evident. As bandwidth increases there are also some discernable trends in the patterns of inference associated with each of the methods. Generally, as the kernel size increases the clusters of inferred land cover increase in size – they become more clumped. This is to be expected because of the way that data under the kernel are brought together, in a way similar to a smoothing operation. Within this, the different approaches aggregate evidence in different ways which relate to their logics. The modified Dempster-Shafer generates discrete clumps of high belief, with clear areas of low belief between. Bayes shows similar patterns, with larger clusters at wider bandwidths. There are fewer extreme beliefs, in this case memberships to the set of Tree, in the Fuzzy Sets approach and therefore many fewer clusters. This is interesting: the Fuzzy approach does not cleave the evidence into belief and disbelief as does Bayes and Dempster-Shafer when there is high consensus among the data points. This is because the belief in any hypothesis of fuzzy set membership is diluted by the presence of other data points of the same class points near the edge of the kernel. The Possibility function generates beliefs that relate to the distance from the kernel centre of the nearest data point of the class being considered, and thus the class with maximum Possibility is the one that has the data point closest to the kernel centre. This is effectively “nearest point wins”. Possibility produces very large clusters of high possibility especially at the larger bandwidths as a consequence.

5. Discussion and Conclusions

The main aims of this paper were two-fold. The first was to develop, apply and evaluate novel geographically weighted evidence combination approaches for integrating conflicting information. The mapped results how these methods are able to make inferences about the presence of specific land cover types at each location and how each approaches describes the distribution of land cover, but with different spatial characteristics and interactions with the kernel bandwidth. Generally there was much greater smoothing of land cover clusters under Possibility Theory, more cleaving of the evidence into high and low belief under Dempster-Shafer and fewer clusters of extreme Fuzzy Set memberships especially at larger kernel bandwidths.

Analysing the correspondence with the reference data suggests that GW Average is the most reliable approach regardless of bandwidth, although this statement has to be qualified with 2 observations. First that we do not know how reliable the reference data were: a different global dataset may well have generated different results. In a similar vein, the approach here was to ‘crisp off’ combined belief (essentially soft classifications) into Boolean classes. This is to ignore much of the uncertainty embedded in the belief in different land cover class hypotheses. Further work is needed to compare the inferences arising from combining crowdsourced data with soft references data in the manner suggested by Fisher et al [39] and Comber et al [41] and to evaluate frames or subsets of classes such as are generated by Dempster-Shafer approaches. Second, that for some land cover classes other evidence combination approaches performed better as in Table 2. This may be because of the greater ambiguity over what, for example, the class of Barren means and uncertain conceptualisations of that class by different contributors, or as a result of different densities of crowdsourced data or because of the interaction of these factors. On-going work is exploring some of the semantic and spatial issues. However, the overall performance of the GW Average approach is an important finding as this method is used in most of the reported and current applications analysing Geo-Wiki data [eg 35, 37].

There is a paucity of approaches for dealing with conflicts in data generated by crowdsourcing and citizen science activities. Previous research has dealt with inconsistent data by applying a majority rules approach [13] with more recent work exploring, for example, latency analyses [12] outlier identification classification [42] as well as classic inferential statistics [37]. These approaches have sought to identify inconsistent data so that it can be excluded from the analysis. However, the outliers may tell us something about the uncertainty of the pure, crisp land cover in that location. Thus a second aim of this research was explore how methods that explicitly reasoned under uncertain information handled conflicting data. If crowdsourced data are to be used in scientific analyses then a full panoply of approaches for handling information uncertainty are needed in the absence of formal experimental design in data collection, training, calibration and validation. This paper was very much concerned with methods for managing conflict and not just excluding conflicting data from analyses. This is a critical issue in the context of non-expert crowdsourced citizen science since it deals with the issue of how to move beyond simple majorities if, for example (most of) the crowd are all similarly confused. The results of this work suggest that all of the approaches under an appropriate bandwidth are able to accommodate such conflicts. The maps and other figures illustrates how they treat conflicting evidence: Bayesian Probability pushes evidence into belief and disbelief indicating what is there, Dempster-Shafer pushes it into belief and plausibility indicating what is not excluded from being there, Fuzzy Sets generates a set memberships to the hypothesis under consideration and indicates what combinations of things is there and Possibility Theory indicates what could be there. The GW Average provides a distance weighted majority rules approach.

There are a couple of further issues to note. First, here a number of bandwidths were explored. Methods exist for selecting the bandwidth automatically, for example using a leave-one-out cross validation procedure which optimises the prediction probability for each individual data point when it is removed from the dataset. As a result, optimal bandwidths may therefore be class-specific which will be explored in future work. However these approaches are based on variance evaluations and may not be appropriate for this kind of analysis with this kind of data. Second, the results only indicate the belief in the hypotheses, that is the extent to which the evidence supports the different hypotheses. Whereas, any uncertainty arising from the evidence is implicitly included in the Fuzzy Sets memberships and is absent in Bayesian Probability which partitions evidence into belief and disbelief (what is there), this is not the case for Dempster-Shafer and Possibility Theory. The former generates a measure of Plausibility describing the extent to which the evidence does not refute the hypothesis – that is belief plus uncertainty. The latter generates a Necessity measure describing the certainty of the belief measure relative to competing hypotheses, with the effect that the lower the Necessity value, the more competition there is. Finally a number of scale factors interact to generate the results under each evidence combination approach. These include bandwidth, sampling grid and the granularity of the data that are collected through a Google Earth interface. Future work will address these issues and will explore the analysis of user confidences in their contributions to provide a second weighting to complement distance weights. Future work will also explore the development of a contributed R package to provide a generic geographically weighted framework to support other research activities. In this work bespoke code had to be developed but the provision of a suite of functions that returned geographically weighted data or even geographically weighted functions would be of great interest to the research community.

In conclusion, this work has suggested that the GW Average approach provides the most reliable oval approach for combining crowdsourced land cover data such as are collected by the Geo-Wiki initiative. There are some caveats to this statement relating to the need for comparisons with soft (e.g. fuzzy) reference data and the examination of the inherent conceptual and semantic ambiguity of some classes at specific grains of analysis. However, the methods and results demonstrate the opportunities for generating localised measures of belief to support assessments of crowdsourced data quality and uncertainty. The geographically weighted evidence combination methods (Dempster-Shafer, Bayesian Probability, Fuzzy Sets and Possibility Theory, GW Average) provide a suite of approaches for assessing belief and for combining conflicting information when mining large crowdsourced datasets, whether the data are contributed actively such as in Geo-Wiki or passively like much social network data. The application of a geographically weighted kernel explicitly addresses the need to consider Tobler's first law of geography when mining and combining crowdsourced data, reflecting the expectation that similar features and process will be clustered and not randomly distributed. The approach of a geographically weighted framework with evidence combination approaches allow more nuanced inferences about the quality of volunteered information to be generated than simple majorities and support the exploitation of large volumes of crowdsourced data about all kinds of phenomenon.

Acknowledgements

This work was undertaken under the EU COST TD1202 'Mapping and the citizen sensor'. The authors would like to thank the anonymous reviewers whose comments helped significantly improve this article.

References

1. Goodchild M.F. (2007). Citizens as sensors: the world of volunteered geography. Geojournal 69: 211-221.
2. Haklay, M., Basiouka, S., Antoniou, V., and Ather, A. (2010). How many volunteers does it take to map an area well? The validity of Linus' law to

volunteered geographic information. The Cartographic Journal, 47(4), 315-322.
3. Foody, G. M., See, L., Fritz, S., Van der Velde, M., Perger, C., Schill, C., and Boyd, D. S. (2013). Assessing the accuracy of volunteered geographic information arising from multiple contributors to an internet based collaborative project. Transactions in GIS, 17(6), 847-860.
4. 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.
5. See, L., Comber, A.J., Salk, C., Fritz, S., Van der Velde, M., Perger, C., Schill, C., McCallum, I., Kraxner, F. and Obersteiner M. (2013). Comparing the Quality of Crowdsourced Data Contributed by Expert and Non-Experts. PLoS ONE 8(7): e69958.
6. Comber, A., Brunsdon, C., See, L., Fritz, S. and McCallum, I. (2013). Comparing expert and non-expert conceptualisations of the land: an analysis of crowdsourced land cover data. Lecture Notes in Computer Science: Spatial Information Theory, 8116: 243-260
7. Tobler, W., (1970). A computer movie simulating urban growth in the Detroit region. Economic Geography, 46(2), 234-240.
8. Brunsdon, C.F., Fotheringham, A.S. and Charlton M. (1996). Geographically Weighted Regression - A Method for Exploring Spatial Non-Stationarity, Geographical Analysis, 28, 281-298.
9. Comber, A., Fisher, P., Wadsworth, R., (2004). Integrating land cover data with different ontologies: identifying change from inconsistency. International Journal of Geographical Information Science, 18(7): 691-708.
10. Comber, A.J., Fisher, P.F., Wadsworth, R.A., (2004). Assessment of a Semantic Statistical Approach to Detecting Land Cover Change Using Inconsistent Data Sets. Photogrammetric Engineering and Remote Sensing, 70(8): 931-938.
11. Comber, A.J., Carver, S., Fritz, S., McMorran, R., Washtell, J. and Fisher, P. (2010). Different methods, different wilds: evaluating alternative mappings of wildness using Fuzzy MCE and Dempster Shafer MCE. Computers, Environment and Urban Systems, 34: 142-152.
12. Foody, G.M., See, L., Fritz, S., Van der Velde, M., Perger, C., Schill, C., Boyd, D.S. and Comber, A., (2014). Accurate attribute mapping from volunteered geographic information: issues of volunteer quantity and quality. The Cartographic Journal doi: http://dx.doi.org/10.1179/1743277413Y.0000000070
13. Haklay, M. (2013). Citizen Science and Volunteered Geographic Information – overview and typology of participation. Pp 105-122 in Sui, D.Z., Elwood, S. and M.F. Goodchild (eds.), 2013. Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice. Berlin: Springer.
14. Goodchild, M. F., and Li, L. (2012). Assuring the quality of volunteered geographic information. Spatial Statistics, 1: 110-120.
15. Perger, C., Fritz, S., See, L., Schill, C., Van der Velde, M., McCallum, I. and Obersteiner, M. 2012. A campaign to collect volunteered geographic Information on land cover and human impact. In: Jekel, T., Car, A., Strobl,

J. and Griesebner, G. (Eds.) GI_Forum 2012: Geovizualisation, Society and Learning. Herbert Wichmann Verlag, VDE VERLAG GMBH, Berlin/Offenbach, pp.83-91.

16. Fritz, S., McCallum, I., Schill, C., Perger, C., See, L., Schepaschenko, D., van der Velde, M., Kraxner, F., and Obersteiner, M. (2012). Geo-Wiki: An online platform for improving global land cover. Environmental Modelling and Software, 31: 110-123.
17. Comber, A., See, L., and Fritz, S. (2014). The Impact of Contributor Confidence, Expertise and Distance on the Crowdsourced Land Cover Data Quality. GI_Forum 2014-Geospatial Innovation for Society, http://goo.gl/nJnzwo
18. Pickles, J. (1995). Ground truth: The social implications of geographic information systems. Guilford Press.
19. Mooney, P., 2011. The evolution and spatial volatility of VGI in OpenStreetMap. Paper presented at the Hengstberger Symposium Towards Digital Earth: 3D Spatial Data Infrastructures, Heidelberg, September 7–8.
20. Comber, A., Mooney, P., Purves, R. S., Rocchini, D., & Walz, A. (2015). Comparing national differences in what people perceive to be there: mapping variations in crowd sourced land cover. ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 1, 71-75.
21. Comber, A., Mooney, P., Purves, R., Rocchini, D. and Walz, A. (2015). Comparing national differences in what the people perceive to be there: Mapping variations in crowd sourced land cover. In Proceedings of International Symposium on Spatial Data Quality, Montpellier 29-30th September 2015.
22. Elwood, S., Goodchild, M. F., and Sui, D. (2013). Prospects for VGI research and the emerging fourth paradigm. In Crowdsourcing Geographic Knowledge (pp. 361-375). Springer Netherlands.
23. Comber, A.J., Fisher, P.F., Wadsworth, R.A., (2005). What is land cover? Environment and Planning B, 32:199-209.
24. Zook, M., Graham, M., Shelton, T., and Gorman, S. (2010). Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake. World Medical and Health Policy, 2(2), 7-33.
25. Welinder, P., and Perona, P. (2010, June). Online crowdsourcing: rating annotators and obtaining cost-effective labels. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference (pp. 25-32). IEEE.
26. McCann, R., Shen, W., and Doan, A. (2008, April). Matching schemas in online communities: A web 2.0 approach. In Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference (pp. 110-119). IEEE.
27. McCann, R., Doan, A., Varadaran, V., Kramnik, A., and Zhai, C. Building data integration systems: A mass collaboration approach. In Sixth International Workshop on Web and Databases (WebDB 2003) (pp. 25-30).
28. Richardson, M., and Domingos, P. (2003). Building large knowledge bases by mass collaboration. In Proceedings of the 2nd international conference on Knowledge capture (pp. 129-137).

29. Doan, A., Ramakrishnan, R., and Halevy, A. Y. (2011). Crowdsourcing systems on the world-wide web. Communications of the ACM, 54(4): 86-96. ACM.
30. Cohen, P. R. (1985). Heuristic reasoning about uncertainty: an artificial intelligence approach. Univ. of Massachusetts.
31. Klir G.J. and Yuan B. (1995). Fuzzy Sets and Fuzzy Logic: Theory and Applications. Prentice Hall, Englewood Cliff
32. Shafer, G., and Pearl, J. (1990). Readings in uncertain reasoning. San Mateo: Morgan Kaufmann.
33. Parsons, S., and Hunter, A. (1998). A review of uncertainty handling formalisms. In A. Hunter, and S. Parsons (Eds.), Applications of uncertainty formalisms (pp. 8–37). Berlin: Springer-Verlag.
34. Herold, M., Mayaux, P., Woodcock, C. E., Baccini, A., and Schmullius, C. (2008). Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets. Remote Sensing of Environment, 112(5), 2538-2556.
35. See, L., Fritz, S., Perger, C., Schill, C., McCallum, I., Schepaschenko, D., Duerauer, M., Sturn, T., Karner, M., Kraxner, F. and Obersteiner, M. (2015). Harnessing the power of volunteers, the Internet and Google Earth to collect and validate global spatial information using Geo-Wiki. Technological and Social Forecasting. doi:10.1016/j.techfore.2015.03.002
36. Gollini, I., Lu, B., Charlton, M., Brunsdon, C., and Harris, P. (2013). GWmodel: an R Package for Exploring Spatial Heterogeneity using Geographically Weighted Models. arXiv preprint arXiv:1306.0413.
37. Lesiv M., Moltchanova E., Schepaschenko D., See L., Shvidenko A., Fritz S. and Comber A. (in press). Comparison of data fusion methods using crowdsourced data in creating a hybrid forest cover map. Remote Sensing 7, 1-x manuscripts; doi:10.3390/rs70x000x
38. Fixsen, D. and Mahler, R. P. S. (1997), The modified Dempster-Shafer approach to classification, IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, 27, 96-104
39. Fisher, P., Arnot, C., Wadsworth, R., and Wellens, J. (2006). Detecting change in vague interpretations of landscapes. Ecological Informatics, 1: 163−178.
40. Dubois, D. and Prade, H., (2001). Possibility theory, probability theory and multiple-valued logics: a clarification. Annals of Mathematics and Artificial Intelligence, 32: 35–66.
41. 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.
42. Ali, A. L., Schmid, F., Al-Salman, R., and Kauppinen, T. (2014). Ambiguity and plausibility: managing classification quality in volunteered geographic information. In Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (pp. 143-152). ACM.