Development and application of a "spray-can" tool for fuzzy geographical analysis.

J. J. Huck[1], J. D. Whyatt[2], S. Yielding[2], H. Stanford[2], P. Coulton[3]

[1] School of Computing and Communications, Lancaster University, Lancaster, LA1 4WA
j.huck2@lancaster.ac.uk

[2] Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ
[3] Imagination Lancaster, LICA, Lancaster University, Lancaster, LA1 4YW

KEYWORDS: PPGIS, sentiment analysis, fuzzy geography, spray-can.

1. Introduction

Most information used in policymaking, whether with regard to crime, land-use planning, environmental health or other phenomena, contains a spatial component (Sieber, 2006, Carver et al., 2001). As such, most planning decisions would benefit from the collection of spatial information from stakeholders and the public, in order gauge levels of satisfaction or identify causes for concern, possibly allowing plans to be adjusted to minimise objection. Such a system, when based upon a GIS, may be referred to as PPGIS (Public Participation GIS) (Sieber 2006).

Brabham (2009) describes the web as the ideal medium for facilitating creative participation due to its speed, reach, asynchrony, anonymity, interactivity and multi-media nature. However, he also highlights the limitations of the web, chiefly the so-called 'digital divide', referring to inequitable levels of access to computing equipment and the internet. With the development of 'web 2.0', the now seemingly ubiquitous nature of web mapping services such as Google Maps (Google, 2012), and the relative ease with which such facilities may be customised and added into users' own websites, these 'non-traditional' GIS systems appear to provide the ideal platform for a PPGIS.

The use of a 'spray-can' interface for a PPGIS is not a novel one, having previously been utilised by Waters and Evans (2003) to collect information from the public regarding a 'fuzzy geography'. Evans and Waters (2008) define fuzzy locations as those that are difficult to delineate due to one of a number of characteristics. These characteristics include: 'Indifference', whereby the boundary of the location is of little or no concern (e.g. 'town centre'); 'continuousness', whereby boundaries are difficult to define (e.g. a mountain); 'multivariate classification', whereby a number of variables are binned together for descriptive convenience (e.g. 'the rough area'); and 'averaging', whereby a discrete boundary is an average of time or scale-variable boundaries (e.g. the coast).

2. Background to Study

The purpose of this paper is to build upon the work of Waters and Evans (2003), and create a spray-can mapping tool for the collection of fuzzy data. To date, two projects have been completed using such a tool, firstly with Yielding (2010), and secondly with Stanford (2011). Both of these projects were based upon a custom Google Maps (2012) spray-can interface, and users were asked to 'spray' onto the map any areas that they thought were suitable or unsuitable for a wind farm development.

The advantage of the tool, beyond that gained from basing it upon the familiar Google Maps (2012) interface, is that each of the 'blobs' of paint created by the user is stored in the database as a discrete point. Each point has associated attribute information including latitude, longitude, map scale, a timestamp, demographic information relating to the user, additional attributes required for subsequent analysis (in this case 'suitable' or 'unsuitable' locations for a wind farm) and free-text comments provided by the user. This allows the easy sub-setting of data using selected attributes or free-text comments, facilitating much more sophisticated analysis than the image-based outputs of Waters and Evans (2003), hence permitting a greater depth of understanding into the thoughts and opinions of those taking part in a survey.

A number of challenges, however, arise from the use of a 'slippy map' interface, which has become the standard platform for web mapping in recent years. The main issue relates to the ability of the user to view the map and generate data at a multitude of different scales, which can be beneficial some, where the concept of scale is understood, but has the potential to be confusing for others. This is because the density of a spray pattern will change with scale and may lead to the misrepresentation of intended patterns. It is for this reason that the scale at which each point is created is captured as an attribute of that point, and this information may be used in subsequent analysis if required.

3. Methodology

This study utilises subsets of the data collected by Yielding (2010) and Stanford (2011) in order to demonstrate some of the analytical techniques that may be applied to data collected using this spray-can application. These examples illustrate the advantages of storing fuzzy data as individual points with multiple attributes, rather than as surfaces or images.

The tool operates in a Google Maps (2012) interface using the Google Maps API (2012), with the spray-can tool added onto it in JavaScript, which handles the drawing of the 'paint' onto the map, and the capture of the geographical location and timestamp. As the user sprays data onto the map, the JavaScript transfers it asynchronously to a PHP script that loads it into a MySQL relational database, where it is linked to geo-demographic information that has been collected about the user, and comments that are submitted by the user to supplement their spray. Any subset of these data can be retrieved as CSV files using SQL queries, and then imported into a GIS for analysis.

4. Results

4.1. Quantitative Assessment

The first set of analyses relate to the patterns and distribution of the spray itself. The simplest method to illustrate such patterns is to create density surfaces from the point-based data. Such a surface would normally be created in order to illustrate areas that demonstrate a concentration of a particular attribute, which in the case of this study relates to areas that were deemed either 'suitable' or 'unsuitable' for the development of a wind farm. The process of generating these surfaces is illustrated in Figure 1, using the data collected by Yielding (2010).

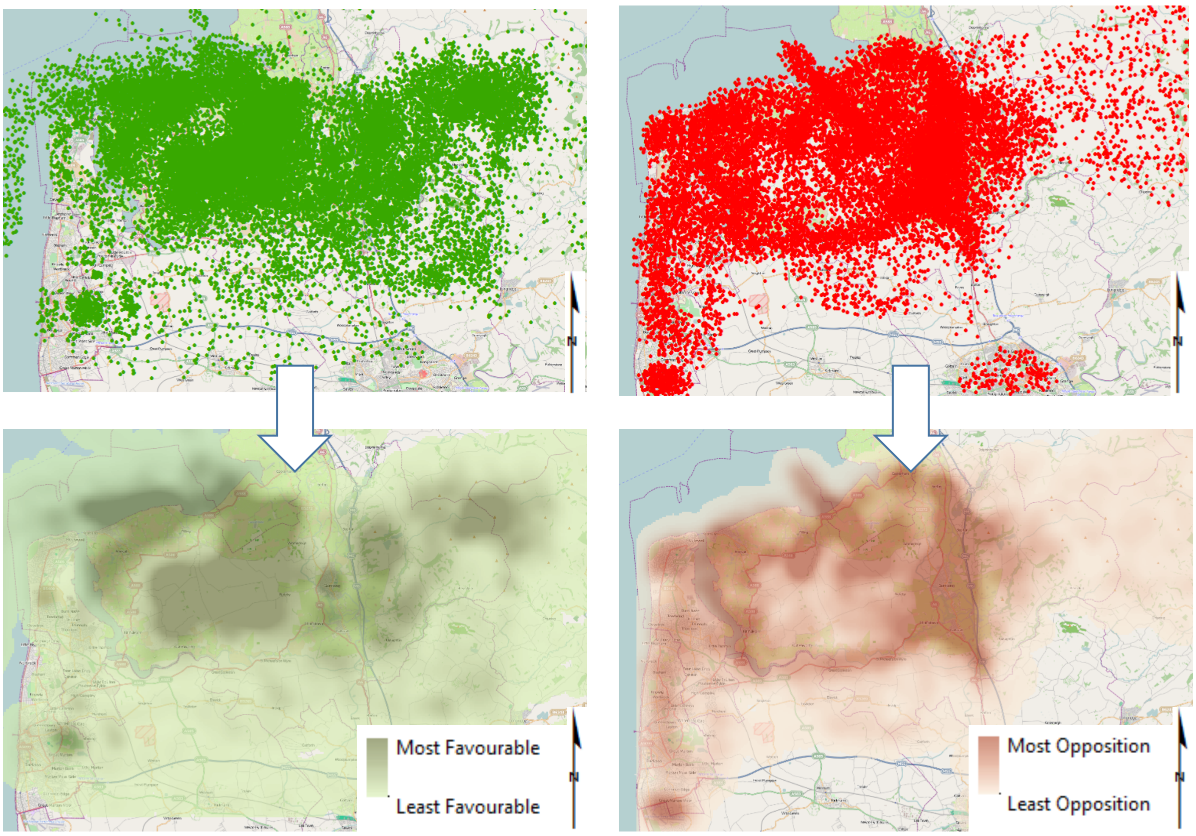

Figure 1: Density surfaces indicating favourable (left) and less favourable (right) locations for wind turbines, taken from Yielding (2010). Base map data taken from OpenStreetMap (2012).

In order to gain more information than merely where was felt to be 'suitable' or 'unsuitable' by an aggregate of survey participants, the two density surfaces can be compared in order to provide an overview of which areas are deemed 'suitable' and 'unsuitable' overall, accounting for each opposing view. In order to do this, it is first necessary to normalise the values within the two datasets, so that they both represent values of 0-1 on an equal scale of spray intensity. Overall suitability may then be calculated using Equation 1. An example of such data is illustrated in Figure 2.

$$suitability = positive\ density - negative\ density \tag{1}$$

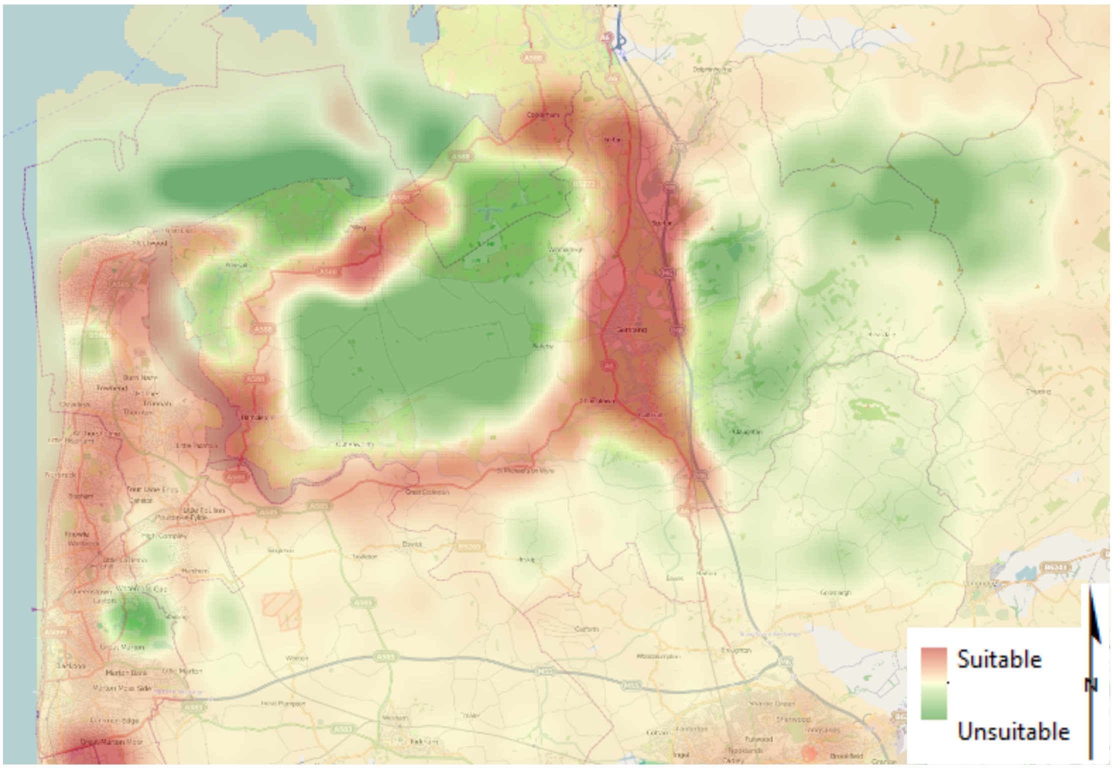

Figure 2: Map of areas of overall positive and negative spray (suitable and unsuitable for wind farm development), based upon data from Yielding (2010). *There is a notable concentration of negative spray around urban areas, roads and the coast, whereas positive concentrations occur in open countryside.* Base map data taken from OpenStreetMap (2012).

In order to add additional confidence to analysis, the 'validity' of areas defined as particularly suitable or unsuitable can be tested with cluster analysis such as the Getis-Ord local statistic (Getis and Ord, 1992), which will statistically identify hotspots of "suitable" and "unsuitable" data. Attention may then be focussed upon areas identified as being "suitable" or "unsuitable" to a given level of significance, thus lending more confidence to the analysis.

Due to the emotive nature of topics such as wind farm development, it could be considered more useful to identify not only areas that are considered positive or negative overall, but those which caused the most conflict of opinion between participants. This is important as an area of high suitability, but also high public opposition, may be less desirable for development than a less suitable area, but with less public opposition in a situation where a prospective developer is looking to avoid objections to a planning application, or minimise the perceived 'impact' of the development on the public. Conflict level can be calculated using Equation 2 with the re-scaled density surfaces. An example of such a surface is given in Figure 3.

$$conflict = positive\ density \times negative\ density \tag{2}$$

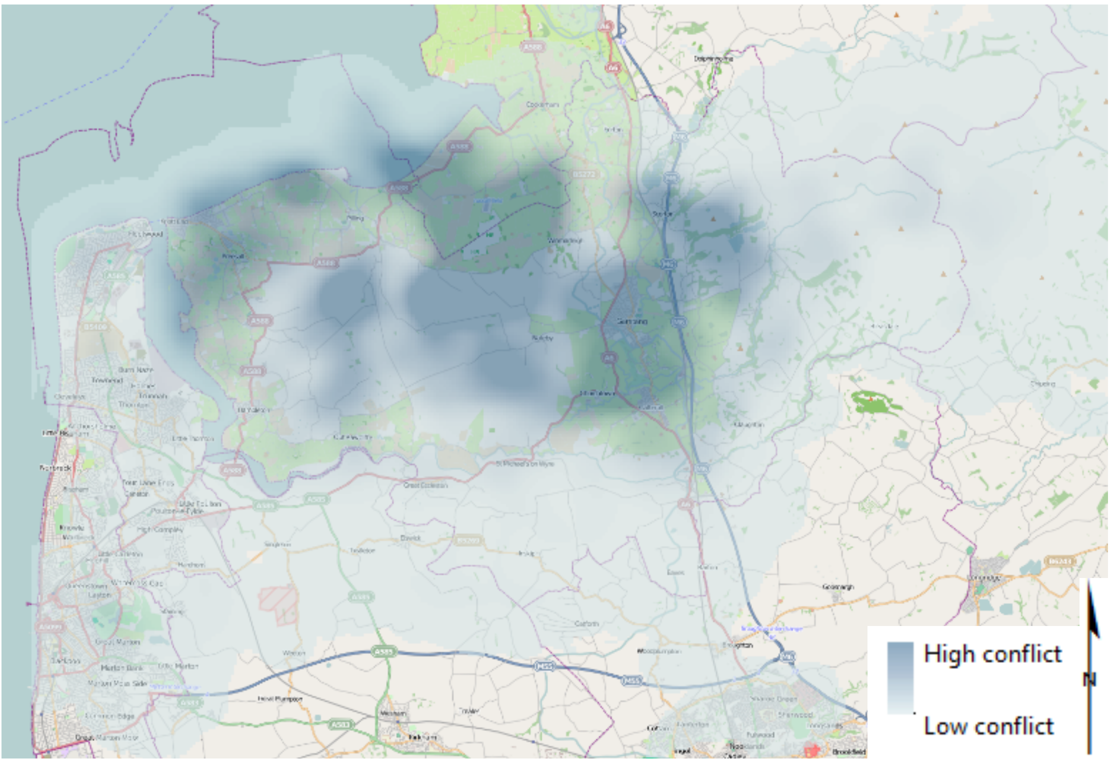

Figure 3: Map of areas of conflict of opinion regarding suitability for wind farm development, based upon data from Yielding (2010). *Conflict of opinion is concentrated around urban areas and their surroundings, possibly due to the familiarity of users with these areas.* Base map data taken from OpenStreetMap (2012).

Comments made by users who sprayed in areas of high or low conflict can be analysed in order to assess the causes of their views. The following example, taken from the data collected by Stanford (2011), demonstrates attitudes towards urban areas, and raises the following comments, both of which aim to minimise the visual impact of wind turbines, but suggesting entirely opposite approaches:

"I would put turbines largely in the hills because they are out of the way and don't cause much of a visual disturbance."

"I put the wind turbines… in cities and urban areas, as this means that they won't cause as much visual pollution there, particularly in cities as there are already industrial sites and tall buildings."

4.2. Qualitative Assessment

In order to gain further insight into public opinion, it is possible to augment findings from the patterns in the spray with analysis of the explanatory comments created by each participant in a survey. One method by which this can be achieved is to consolidate all of the comments, and perform a corpus-based noun extraction. The nouns can then be lemmatised (reduced contextually to the root word), grouped, and counted in order to establish the places that were raised most frequently by the survey participants. Noun identification, lemmatisation and database interaction were performed with Python scripts using the NLTK (Natural Language Tool Kit) (NLTK 2012) library to facilitate natural language processing tasks.

Extracted nouns can then be linked back to the spray locations associated with the comments that included them. If an average is taken of all of the spray locations related to comments including a given noun, then analysis can take place as to whether spray patterns are representative of the comments that they associated with their spray. This can provide more insight both into the quality of information given by participants, and into the effectiveness of the tool in allowing people to express their feelings onto the map.

For example, are the spray patterns of those participants who referred to the motorway or coast on their comments actually focussed around those areas? An example of this is given in Figure 4, where a location for the "*M6*" motorway has been plotted on the map based upon the spray patterns of people who referred to it in their comments. Examples of such comments are also given in Figure 4.

Figure 4: Map demonstrating the extracted focus of spray from users who discussed "M6", based upon data from Stanford (2011). *Examples of such comments are given in the speech bubbles with green bubbles representing positive comments, and red representing negative.* Base map data taken from OpenStreetMap (2012).

When based upon presence of the noun in the text alone, this technique yields variable results, with some points located very successfully (as in Figure 4), and others less so. A more robust model requires the introduction of 'sentiment analysis', whereby a naïve Bayesian classifier is trained to identify positive and negative spatial statements, and can then be used to compare nouns from positive statements with positive ('suitable') spray patterns, and nouns from negative statements with negative ('unsuitable') spray patterns. The ability to compare positive and negative spray patterns to positive and negative comments respectively will permit locations to be defined more precisely (by discounting the influence of un-related spray), and therefore provide a greater level of understanding regarding the effectiveness and usability of the tool.

5. Conclusions

This paper demonstrates the value of collecting fuzzy data as a set of discrete points with associated attributes as opposed to a surface or an image, and illustrates some of the many forms of spatial analysis that may be utilised because of the flexibility that this approach introduces. This diverse selection of analytical possibilities is the advantage of this tool as a PPGIS, affording decision makers the ability to gain a greater understanding of the opinions and feelings of stakeholders and the public, and as such make better-informed policy or development decisions.

This paper also highlights the remaining work to be done on this project; the development and training of a naïve Bayesian classifier capable of identifying and classifying positive and negative spatial statements such as "*wind farms should all be offshore*" or "*it would not be good to have wind turbines near roads*". When this work is complete, it will be possible to extract fuzzy features from the text and apply them back to the map with a much greater degree of confidence, thus permitting the achievement of an even greater level of understanding of the data collected.

Other on-going work relates to the development of the 'Map-Me' system, a website permitting users to create an online survey using a customisable spray-can interface, providing them with data that may be analysed in the above manner.

6. References

Brabham, D, C. (2009) Crowdsourcing the public participation process for planning projects. *Planning Theory.* **8.** *242-262.*

Carver, S., Evans, A., Kingston, R., and Turton, I. (2001) Public participation GIS and cyberdemocracy: evaluating on-line spatial decision support systems. *Environment and Planning B: Planning and Design.* **28**. *907-921.*

Evans, A. J. and Waters, T. (2007) Mapping Vernacular Geography: Web-based GIS tools for capturing "fuzzy" or "vague" entities. *International Journal of Technology, Policy and Management.* **7** (2), *1468-4322.*

Google (2012) Google Maps. *WWW Document.* http://maps.google.com.

Google Developers (2012) Google Maps API. *WWW Document.* https://developers.google.com/maps.

Map-Me (2012) Map-Me. *WWW Document.* http://map-me.org.

Natural Language Toolkit (2012) Natural Language Toolkit – NLTK 2.0. *WWW Document.* http://nltk.org.

OpenStreetMap (2012) OpenStreetMap. WWW Document. http://www.openstreetmap.org/.

Sieber, R. (2006) Public Participation Geographic Information Systems: A Literature Review and Framework. *Annals of the Association of American Geographers.* **96** (3) *491-507.*

Stanford (2011) Gathering public attitudes to wind farm development using the Google Maps 'Spraycan' application. *Unpublished MSc Thesis, Lancaster University.*

Waters, T. and Evans, A. J. (2003) Tools for web-based GIS mapping of a "fuzzy" vernacular geography, *in Proceedings of the 7th International Conference on GeoComputation*. Available at http://geocomputation.org/2003

Yielding, S. (2010) Application of 'neogeography' and GIS to the wind industry. *Unpublished MSc Thesis, Lancaster University.*

7. Biography

Jonny Huck is a 3rd year part-time PhD student researching geospatial computer science jointly within the School of Computing and Communications, and the Lancaster Environment Centre at Lancaster University. His interests include neo-geography, fuzzy geography and spatio-temporal visualisation.

Duncan Whyatt is a Senior Lecturer in GIS within the Lancaster Environment Centre, Lancaster University. His research interests span social and environmental applications of GIS, with specialisms in air pollution.

Simon Yielding studied MSc Environmental Informatics at the Lancaster Environment Centre, Lancaster University, co-supervised by Duncan Whyatt and Jonny Huck. His dissertation represented the first application of this spray-can tool.

Harriett Stanford studied MSc Energy and the Environment at the Lancaster Environment Centre, Lancaster University, co-supervised by Duncan Whyatt and Jonny Huck. Her dissertation represented the first qualitative analysis of comments collected using the spray-can tool.

Paul Coulton is a Senior Lecturer in Design within Lancaster Universities open and exploratory design-led research lab Imagination Lancaster and his research interests are primarily around experience design, interaction design, and design fictions. His research often encompasses an 'in the wild' evaluation methodology utilising 'app stores' and social networks as an experimental platforms.