Where (we think) the wild things are: comparing citizen generated and formal measures of wilderness

Alexis Comber\textsuperscript{*}, Steve Carver\textsuperscript{1}

\textsuperscript{1}School of Geography, University of Leeds, UK

*Corresponding author: a.comber@leeds.ac.uk

Abstract
This paper compares mapped data describing the degree of wilderness form a formal data analysis – the Wilderness Quality Index (WQI) – against crowdsourced estimates of wilderness collected by the Geo-Wiki initiative. The analysis examines the quality of the crowdsourced wilderness reporting, and how that quality varies spatially, using a geographically weighted model. The results indicate globally a positive relationship between the 2 datasets, but with wilderness being under-estimated by the crowdsourced data. However, when local patterns are examined it is evident that in more wild areas (the Alps, eastern Norway, etc) the crowdsourced data have a stronger relationship with the formal data. This suggests that, while citizens may be good at identifying locales that at the extremes of a continuum of wilderness (\textit{wild} and \textit{not-wild}), they may not be so reliable in describing the wilderness in areas with a wilderness character between these extremes.

Keywords
Geographically weighted model; citizen science; wildness; Geo-Wiki

I BACKGROUND
This paper compares citizen measures of human impact from Geo-wiki with formal wilderness data to determine the degree to which citizens perceive wilderness. Geo-Wiki contributors record the degree of human impact [0-100\%] and the land cover they perceive to be present. Different approaches have been used to map wilderness and activities at continental / global scales use a continuum / environmental modification spectrum (Kuiters et al. (2011). Other approaches such as the 2012 Google data that applied a 10km buffer to its road dataset under the assumption that roadless areas represent natural ecosystems that are undisturbed. The aim of this work was to compare two different datasets describing wilderness and wild land, one collected under the Geo-Wiki initiative (Perger et al., 2012; Fritz et al., 2012) and the other as part of a pan-European wilderness project described in Kuiters et al. (2011).

Geo-Wiki uses Google Earth imagery and asks contributors to label features at randomly selected points. There have been a number of campaigns and interface developments. While much work has examined the utility and quality of the land cover information that can be derived from Geo-Wiki land cover data (See et al 2103; Foody et al 2014, Comber et al 2013), where users had to allocate scenes in Google Earth imagery to one of a predefined number of classes, as yet little work has considered the uncertainties, accuracies and errors associated with user contributed measures within a continuum, such as wildness. Specifically this work seeks to determine whether the effort involved in developing traditional maps of wilderness (as a suitability layer) results in significantly higher quality data products. Or whether, more
informal approaches, such as are facilitated by crowdsourced data, are able to support the development of data of quality to be scientifically useful.

II DATA & ANALYSIS

Figure 1 shows ~13,000 Geo-Wiki data points and formal data the wilderness quality index (WQI, Kuiter et al. (2011). It shows that, broadly, both datasets are describing similar distributions of wildness. Iceland, the Alps and North western Scotland has all have high wildness values in these areas.

The WQI data reports wildness measure in the range [0, 1] at 1km scale. The Geo-Wiki data has a number of attributes including land cover class (from a pre-defined set of 10 classes), a Human Impact score which was treated as the inverse of a wildness measure and subtracted from 1, and a self-reported measure of Geo-Wiki volunteer confidence for each point that was labelled.

The analysis used a standard OLS regression to model the relationship between Geo-Wiki volunteer measures of Human Impact, under the assumption that this describes the inverse of wildness, with the formal measure of wilderness quality index. For each Geo-Wiki data point, the coincident wilderness measure was extracted from the WQI data. The model describing the degree to which Geo-Wiki Human Impact predicts WQI data was constructed using the following:

\[ y = \beta + \beta_1 x_1 \]  
(Eqn 1)

where \( y \) is the WQI at each data point and \( x_1 \) is the Geo-Wiki wildness value derived from the Human Impact value, and \( \beta_1 \) is the coefficient estimate for Geo-Wiki wildness. This identified constructed a global model of the relationship between these 2 variables.

Next a Geographically Weighted model was constructed to examine whether and how the coefficient estimates vary. The idea here is that global statistical models implicitly make an
assumption that the relationship between the dependent and independent variables is constant over the study area. In reality this is frequently not the case. In such situations, where spatial non-stationarity is thought to occur, local models can provide a more detailed and informative analysis. The Geographically Weighted (GW) framework (Brunsdon et al., 1996) supports a suite of local models in order to account for local variation in statistical relationships between variables (Gollini et al, 2015). A GW regression (GWR) was used to examine how the coefficient estimates of the degree to which Geo-Wiki measure predict WQI measures from of a standard OLS regression vary spatially. A GWR version of the OLS (Equation 1) is defined as follows

\[ y(u_i, v_i) = \beta_0(u_i, v_i) + \beta_1 x_1(u_i, v_i) \]  

(Eqn 2)

where \((u_i, v_i)\) is a vector of two dimensional co-ordinates describing the location of \(i\) over which the coefficient estimates are assumed to vary. The outputs provide spatially varying estimates of the coefficients an the aim of generating these was to better understand how informal, citizen generated measures of landscape wildness relate to formal measures, how these vary spatially and how measures of wildness interact with the self-reported confidence value attached by volunteers to their labelling.

### Results

Figure 2 shows the Geo-Wiki wildness scores (inverted from the Human Impact value) against the WQI data. The OLS model suggested a coefficient estimate \(\beta_1\) of 0.356 for the Geo-Wiki wildness value, indicating that globally, each increase of 0.1 in the WQI data is associated with an increase of 0.036 in the Geo-Wiki data. The results of the all the regressions are summarised in Table 1. The distribution of the GWR coefficient estimates indicates that there is considerable local variation from the global model.

|              | OLS Estimate | Std. Error | t value | Pr(>|t|) | Sig |
|--------------|--------------|------------|---------|----------|-----|
| (Intercept)  | 0.3400       | 0.0031     | 108.500 | <2e-16   | *** |
| Geo-Wiki wildness | 0.3563       | 0.0051    | 69.9700 | <2e-16   | *** |

Figure 2: WQI against rescaled Geo-wiki wildness scores, with the slope of regression coefficient.
Next the coefficient estimates were mapped to show how the coefficient estimates linking wildness score from Geo-Wiki volunteers relate to formal WQI wildness index scores. This is shown in Figure 3 and indicates distinct areas and gradients where volunteers found it difficult to accurately assign wilderness values. Clearly there is high agreement in the Alps and in eastern Norway but large differences between formal measures and volunteered ones in for example eastern France. Under the assumption that the formal WQI data are correct, this suggests that there are locales where people’s perceptions of wilderness differ hugely from the way that it is formally recorded.

This paper presents the first analysis of cognitive aspects of how citizens perceive wilderness. Future work will seek to unpick the origins of these differences in relation to landscape context (for example the land cover and land use in these areas), user self-assigned confidence in their scores and their interaction. It will also consider the use of Geo-Wiki data to validate other datasets as has been done for many global land cover initiatives, and how to objectively quantify the accuracy of either dataset.

![Figure 3](image-url)

Figure 3: The spatial variation of degree to which informal Geo-Wiki measures of wildness predict formal measures from the WQI dataset.

### References


