A spatial analysis of how measures of landscape wilderness correlate with crowdsourced perceptions of scenicness

Yi-Min Chang Chien
University of Leeds
School of Geography,
University of Leeds, Leeds,
LS2 9JT, UK
gyymcc@leeds.ac.uk

Alexis Comber
University of Leeds
School of Geography,
University of Leeds, Leeds,
LS2 9JT, UK
a.comber@leeds.ac.uk

Steve Carver
University of Leeds
School of Geography,
University of Leeds, Leeds,
LS2 9JT, UK
s.j.carver@ leeds.ac.uk

Abstract

The term “scenicness” refers to the perceived aesthetic and scenic beauty of observed landscapes. Crowdsourced measures of scenicness from the *Scenic-Or-Not* website provides an opportunity to explore public perceptions of landscape beauty and how they relate to other measures of landscape naturalness. This paper compares crowdsourced measures of scenicness with formal measures used to construct landscape wildness and wilderness metrics using both global regressions and local geographically weighted regressions (GWR). The global regression results indicate positive relationships between crowdsourced scenic ratings and wilderness measures. The local measures exhibit considerable spatial variation and that some of the relationships are negative in some locations and that the relationships between the landscape wilderness variables and perceptions of scenicness varied spatially in different parts of the study area (Wales). The results also suggest that features more commonly associated with the built environments could be included in scenicness assessments. A number of other areas for further work are suggested.

Keywords: Scenicness, Wilderness, *Scenic-Or-Not*, Geographically Weighted Regression.

1 Introduction

Scenicness derives from an interaction between combined features from both natural and cultural elements of the landscape and perceptual appreciation of observers. Biophysical features, such as wilderness character (Carver et al., 2013), may intuitively indicate landscape scenic quality, which can be characterised and modelled in a tangible manner via expert-based approaches. These methods quantitatively examine well-defined visual properties and biophysical features of the landscape but can, in some cases, have low local relevance (Daniel, 2001) as “beauty is in the eye of the beholder”. Difference in aesthetic preferences may be caused by varied cultural backgrounds (Zube and Pitt, 1981), age, gender, social stratum (Dramstad et al., 2006, Hunziker et al., 2008, Tveit, 2009) and perceptual analyses and surveys can lead to highly localised results with little generalisability (Beza, 2010).

Recent years have witnessed the proliferation of crowdsourcing practices including the large-scale collection of data describing human perceptions of the environment by the *Scenic-Or-Not* website (http://scenic.mysociety.org/). This is based on an online game set up to collect and share crowdsourced data across Great Britain. Participants are invited to rate random geotagged photos on an integer scale according to scenic beauty (from 1 = least beautiful to 10 = most beautiful). Each photo is located within a 1 km grid square of Great Britain and sourced from *Geograph* (http://www.geograph.org.uk/), an open online project collecting geotagged photographs for each square kilometre of Great Britain and Ireland. The original aims of collecting scenic quality was to provide to support quantitative analyses of the scenicness impacts on human wellbeing (Seresinhe et al., 2015). The scenicness dataset has been used in a number of studies, for instance, to examine the links between scenicness and land cover (Stadler et al., 2011) as well as the improved scenicness estimation and prediction combined with data from *Flickr* and *OpenStreetMap* (Seresinhe et al., 2017a).

However as with many crowdsourced data there can be issues with data-sparsity: 5% of the grid squares from *Scenicness* database are still incomplete, due to their inaccessibility which implies that true wilderness areas may be under-represented in measurements of scenicness. In wilderness research has been addressed by incorporating both ground-level and overhead imagery evaluate the scenicness of a region (Workman et al., 2017). Thus far, there is no scientific evidence on how measures of perceived scenicness relate to wilderness measures, which focus on the natural state of the environment and the lack of human artefacts (Carver et al., 2002).

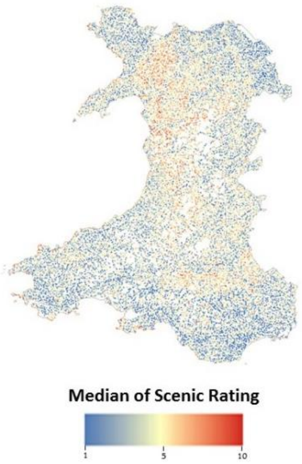
This paper evaluates the extent to which wilderness measures (i.e. remoteness from settlement, remoteness from access, apparent naturalness, and biophysical naturalness) (Carver et al., 2012) are related to scenicness assessments in order to predict the beauty of scenes for new places for which we either do not have crowd-sourced scenicness data.

2 Materials and Methods

Scenic-Or-Not data for Wales were downloaded where each point includes the scenic votes for each one square kilometre (Figure 1). Rather than using the average scenicness scores, the median of each ratings was taken as the dependent variable of the regression model to ensure that the ratings were less likely to be skewed by unusual opinions about scenic beauty.

The wilderness attributes of landscape in Wales are derived from incorporating information such as built-up areas, transport network and woodlands datasets, combined with other ancillary datasets via a weighted linear summation model. More detail can be found in (Carver and Fritz, 1995). The factor maps are shown in figure 2. Further, the values of four wilderness attributes for each image location were extracted as independent variables for the regression model.

Figure 1: The median of *Scenic-Or-Not* rating in Wales



First of all, a global regression model (i.e. Ordinary Least Squares, OLS) was applied to detect the relationships between predictor and outcome variables. The OLS model can be expressed as follows:

$$y_i = \beta_0 + \sum_{j=1}^{m} \beta_j x_{ij} + \varepsilon_i \qquad (1)$$

where for observations indexed by $i = 1, \ldots, n$, y_i is the response variable, x_{ij} is the value of the j^{th} predictor variable, m is the number of predictor variables, β_0 is the intercept term, β_j is the regression coefficient for the j^{th} predictor variable and ϵ_i is the random error term.

Next a Geographically Weighted Regression (GWR) was undertaken to explore spatial heterogeneity (Brunsdon et al., 2002). This is similar in form to linear regression, except that GWR calculates a series of local linear regressions rather than one global one. A GWR model has locations associated with the coefficient terms and can be expressed as:

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^{m} \beta_j(u_i, v_i) x_{ij} + \varepsilon_i \qquad (2)$$

where (u_i, v_i) is the spatial location of the i^{th}observation and $\beta_j(u_i, v_i)$is a realization of the continuous function $\beta_j(u, v)$ at point i. The geographical weighting results in data nearer to the kernel centre making a greater contribution to the estimation of local regression coefficients at each local regression calibration point k. For this study, the weights were generated using a *bisquare* kernel, which for the bandwidth parameter r_k is defined by:

$$\begin{aligned} w_{ik} &= \left(1 - \left(\frac{d_{ik}}{r_k}\right)^2\right)^2 \quad \text{if } d_{ik} \le r_k \\ w_{ik} &= 0 \; otherwise \end{aligned} \qquad (3)$$

Where the bandwidth can be specified as a constant distance value, or in an adaptive, varying distance way, where the number of nearest neighbours is fixed. In this case, fixed bandwidths were chosen as the data are on a regular grid.

3 Results

Initially, a global Ordinary Least Squares (OLS) linear regression was performed to check the global relations between the scenic rating and the four wilderness attributes, and then a GWR was undertaken to explore the spatial variation.

3.1 OLS Regression

The result of the coefficient estimates reveals that all the wilderness variables are positively correlated with the scenicness, also indicating both variables of biophysical and apparent naturalness are most associated with the increase of scenic ratings shown in Table 1.

The p-values indicate that the effect of all the wilderness attributes on the rating of scenicness could be declared extremely significant relationship with aesthetic preference of landscapes except the variable of remoteness from settlement ($p > 0.05$). This corresponds with the recent research findings on individual notions of scenicness including natural features as well as man-made features (Seresinhe et al., 2017b). The R-squared values indicate a loose fitting model.

Table 2 Ordinary Least Squares (OLS) results.

Parameter	Estimate	Std. Error	p-value
Intercept	2.143e+00	4.050e-02	<2e-16
Biophysical naturalness	1.425e+00	6.029e-02	<2e-16
Apparent naturalness	6.212e-01	1.686e-02	<2e-16
Remoteness from access	8.365e-02	3.837e-03	<2e-16
Remoteness from settlement	3.050e-05	1.606e-05	0.0576

3.2 GWR regression

3.2.1 Bandwidth Selection

One of the key parts of any GW analysis is to determine an optimal kernel size or bandwidth, as this controls how much data are included in each local model and the degree of smoothing in the GW model. Gollini et al (2015) provide a full discussion but in essentially the bandwidth determines the scale at which each localised model operates (Gollini et al., 2015). Smaller bandwidths result in greater local variation in the outputs and larger ones result in outputs that are increasingly closer to the global measure.

In this study, the GWR model bandwidth was calibrated by minimising the corrected Akaike Information Criterion (AICc) score for a fixed bisquare kernel. The kernel bandwidth was 13.0514 km.

3.2.2 Local coefficient estimates

The descriptive statistics for all variables derived from GWR are presented in Table 3. The measure of dispersion within interquartile range (IQR) for each coefficient estimate suggests the spatial variability of these variables. The higher IQR exhibits stronger reginal variation. The IQR for the variable of biophysical naturalness also exhibits the highest range of 1.1021, thereby indicating the most influence of localness over scenic beauty.

Table 3 Descriptive statistics for all explanatory variables.

Variable	1st Quartile	Median	3rd Quartile
Biophysical naturalness	8.119e-01	1.357e+00	1.914e+00
Apparent naturalness	2.839e-01	4.724e-01	6.345e-01
Remoteness from access	4.097e-02	8.441e-02	1.266e-01
Remoteness from settlement	-7.255e-05	8.969e-05	2.994e-04

The GWR coefficient estimates can be used to generate surfaces for each model parameter. Each surface depicts the spatial variation of the relationship with the target variable shown as Figure 2.

Table 4 presents the indicators of goodness of fit for comparing the OLS and GWR models. The value of AICc declined from 70257.045 in OLS to 69048.244 in GWR whilst the value of adjusted R^2 increased from 0.244719 in OLS to 0.2963275 in GWR, which indicates the local model resulted in improved model fit.

Table 4 Diagnostic statistics for local and global models.

	Local Model GWR	Global Model OLS
AICc	69048.244	70257.045
Adjusted R^2	0.2963275	0.244719
R^2	0.3099835	0.244876

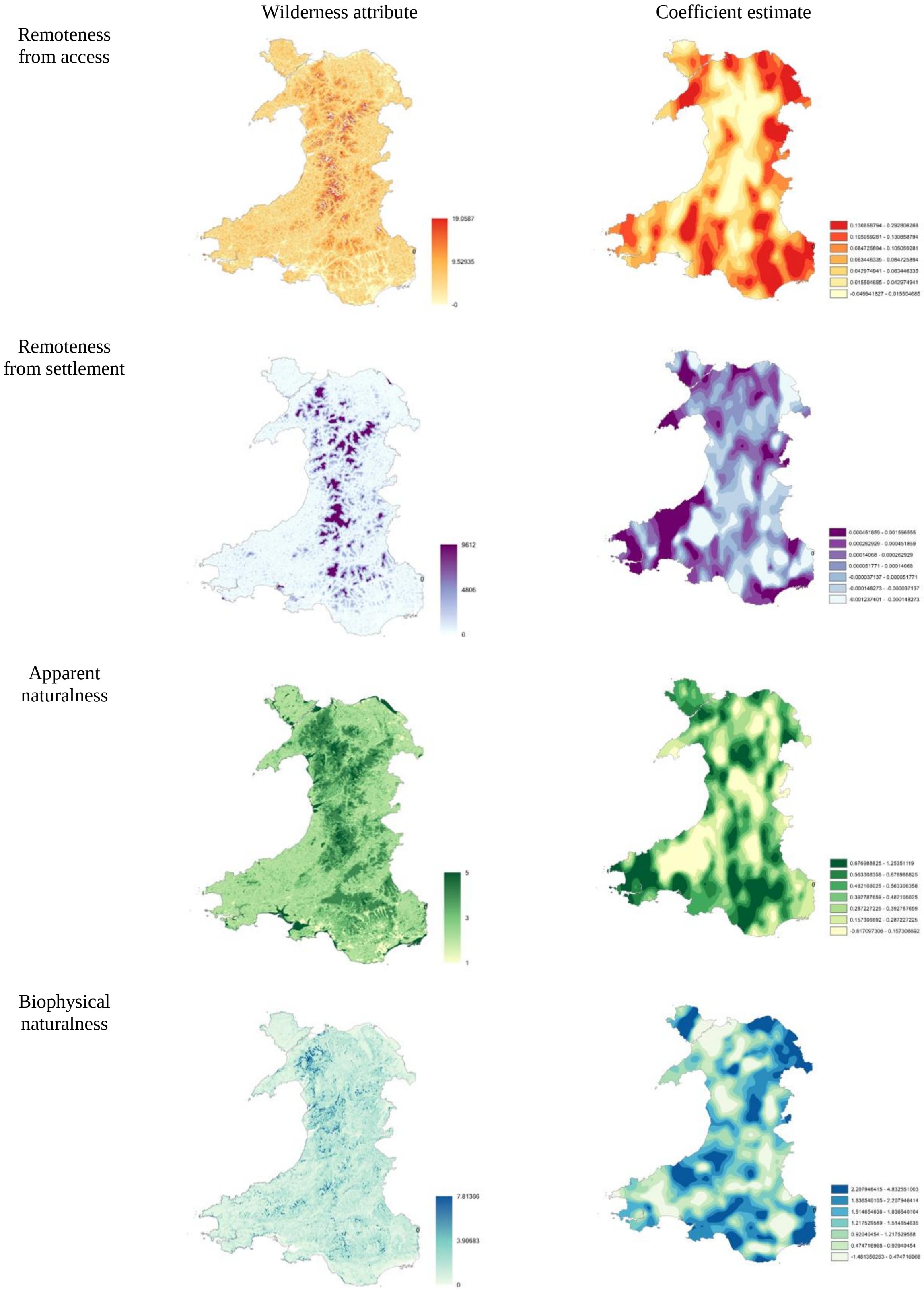
4 Discussion

Wilderness definitions focus on the natural state of the environment with no human settlement and low related impacts (Carver et al., 2002) and as yet it is unclear how wilderness measures could contribute to scenicness assessments. For these reasons this study sought to evaluate to what extent the beauty of scenes coincides with the wilderness attributes.

To be noted, the result shows that the naturalness factors of wildernesss measures contribute more than the remoteness factors to human aesthetic perception of the landscapes as demonstrated by the recent finding of some built environment features (e.g. “Viaduct” and “Aqueduct”) in high correlation with scenic preference (Seresinhe et al., 2017b). However, the more significant impact of perceived naturalness on landscape aesthetics was implied by the language model of scenicness (Chesnokova et al., 2017). Particularly proven in our result is the apparent naturalness quality is less profound effects on the scenicness than the biophysical aspect of naturalness. Further research thus is needed to find out the main reasons behind these different influences on scenic preference scores between biophysical and apparent naturalness.

A number of issues suggest areas for future research. In some squares, only one photo may be present. This may be insufficient to adequately represent a particular landscape of each square kilometre grid. To date, 5% of the grid squares from *Geograph* database are still lacking photos, probably due to their inaccessibility which implies that true wilderness areas may be under-represented in measurements of scenicness. Additionally, some images are fulfil professional photographic criteria (composition, framing, colour, etc.), which may introduce bias in the scenicness ratings. A larger number of photos may be required to insure adequate representation of each square kilometre grid. Previous research findings have suggested the need to incorporate more crowdsourced images from other photo-sharing website (e.g. *Flickr*) into estimation and prediction of scenicness (Seresinhe et al., 2017a). Second, the variation in vantage point and orientation in the square also has impact on measures of scenic beauty. The actual physical locations of scene pictures usually differ from the published location leading to different wilderness character. Thus, detailed information such as EXIF is needed to be taken into account to derive an accurate prediction model. Third, new variables of aesthetic man-made objects (e.g. viaduct and aqueduct) suggested by the recent study (Seresinhe et al., 2017b) could be included to develop a better model that accommodates the issue for better estimation and prediction of scenicness. Last, future work will extend to apply Geographically Weighted LASSO (GWL) to address the local collinearity between the wilderness attributes.

Figure 2: Four wilderness variables and their GWR coefficient estimates in Wales



Acknowledgements

This work is supported by grants from the Ministry of National Defence of Taiwan.

References

BEZA, B. B. 2010. The aesthetic value of a mountain landscape: A study of the Mt. Everest Trek. *Landscape and Urban Planning,* 97**,** 306-317.

BRUNSDON, C., FOTHERINGHAM, A. S. & CHARLTON, M. 2002. Geographically weighted summary statistics — a framework for localised exploratory data analysis. *Computers, Environment and Urban Systems,* 26**,** 501-524.

CARVER, S., COMBER, A., MCMORRAN, R. & NUTTER, S. 2012. A GIS model for mapping spatial patterns and distribution of wild land in Scotland. *Landscape and Urban Planning,* 104**,** 395-409.

CARVER, S., EVANS, A. & FRITZ, S. 2002. Wilderness attribute mapping in the United Kingdom. *International Journal of Wilderness,* 8**,** 24-29.

CARVER, S., TRICKER, J. & LANDRES, P. 2013. Keeping it wild: Mapping wilderness character in the United States. *Journal of Environmental Management,* 131**,** 239-255.

CARVER, S. J. & FRITZ, S. 1995. Mapping the wilderness continuum. *Proceedings of the GIS Research UK***,** 15.

CHESNOKOVA, O., NOWAK, M. & PURVES, R. S. A crowdsourced model of landscape preference. LIPIcs-Leibniz International Proceedings in Informatics, 2017. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

DANIEL, T. C. 2001. Whither scenic beauty? Visual landscape quality assessment in the 21st century. *Landscape and Urban Planning,* 54**,** 267-281.

DRAMSTAD, W. E., TVEIT, M. S., FJELLSTAD, W. J. & FRY, G. L. A. 2006. Relationships between visual landscape preferences and map-based indicators of landscape structure. *Landscape and Urban Planning,* 78**,** 465-474.

GOLLINI, I., LU, B., CHARLTON, M., BRUNSDON, C. & HARRIS, P. 2015. GWmodel: An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models. *Journal of Statistical Software; Vol 1, Issue 17 (2015)*.

HUNZIKER, M., FELBER, P., GEHRING, K., BUCHECKER, M., BAUER, N. & KIENAST, F. 2008. Evaluation of Landscape Change by Different Social Groups. *Mountain Research and Development,* 28**,** 140-147.

SERESINHE, C. I., MOAT, H. S. & PREIS, T. 2017a. Quantifying scenic areas using crowdsourced data. *Environment and Planning B: Urban Analytics and City Science***,** 0265813516687302.

SERESINHE, C. I., PREIS, T. & MOAT, H. S. 2015. Quantifying the Impact of Scenic Environments on Health. *Scientific Reports,* 5**,** 16899.

SERESINHE, C. I., PREIS, T. & MOAT, H. S. 2017b. Using deep learning to quantify the beauty of outdoor places. *Royal Society Open Science,* 4.

STADLER, B., PURVES, R. & TOMKO, M. Exploring the relationship between Land Cover and subjective evaluation of scenic beauty through user generated content. Proceedings of the 25th international cartographic conference, 2011.

TVEIT, M. S. 2009. Indicators of visual scale as predictors of landscape preference; a comparison between groups. *Journal of Environmental Management,* 90**,** 2882-2888.

WORKMAN, S., SOUVENIR, R. & JACOBS, N. Understanding and mapping natural beauty. IEEE International Conference on Computer Vision, 2017.

ZUBE, E. H. & PITT, D. G. 1981. Cross-cultural perceptions of scenic and heritage landscapes. *Landscape Planning,* 8**,** 69-87.