

Spatial analysis of cultural ecosystem services using data from social media: A guide to model selection for research and practice

Andrew M. Neill[‡], Cathal O'Donoghue[§], Jane C. Stout[‡]

[‡] Trinity College Dublin, Dublin, Ireland

[§] University of Galway, Galway, Ireland

Corresponding author: Andrew M. Neill (anneill@tcd.ie)

Academic editor: Joachim Maes

Abstract

Experiences gained through in person (in-situ) interactions with ecosystems provide cultural ecosystem services. These services are difficult to assess because they are non-material, vary spatially and have strong perceptual characteristics. Data obtained from social media can provide spatially-explicit information regarding some in-situ cultural ecosystem services by serving as a proxy for visitation. These data can identify environmental characteristics (natural, human and built capital) correlated with visitation and, therefore, the types of places used for in-situ environmental interactions. A range of spatial models can be applied in this way that vary in complexity and can provide information for ecosystem service assessments. We deployed four models (global regression, local regression, maximum entropy and the InVEST recreation model) to the same case-study area, County Galway, Ireland, to compare spatial models. A total of 6,752 photo-user-days (PUD) (a visitation metric) were obtained from Flickr. Data describing natural, human and built capital were collected from national databases. Results showed a blend of capital types correlated with PUD suggesting that local context, including biophysical traits and accessibility, are relevant for in-situ cultural ecosystem service flows. Average trends included distance to the coast and elevation as negatively correlated with PUD, while the presence of major roads and recreational sites, population density and habitat diversity were positively correlated. Evidence of local relationships, especially town distance, were detected using geographic weighted regression. Predicted hotspots for visitation included urban areas in the east of the region and rural, coastal areas with major roads in the west. We conclude by presenting a guide for researchers and practitioners developing cultural ecosystem service spatial models using data from social media that considers data coverage, landscape heterogeneity, computational resources, statistical expertise and environmental context.

Keywords

cultural ecosystem services, visitation, social media, spatial modelling, geographic weighted regression, maximum entropy, InVEST, Ireland, ecosystem service assessment

Introduction

Cultural ecosystem services are defined as “the non-material outputs of ecosystems that affect physical and mental states of people”, some of which require physical (in-situ) interactions between people and ecosystems (CICES v.5.1, Haines-Young and Potschin (2018)). These outputs include benefits, such as improved physical and mental health, opportunities for recreation and social interaction, connections to socio-cultural heritage, spiritual enrichment and biodiversity appreciation (Scholte et al. 2015, Haines-Young and Potschin 2018). The flow of cultural ecosystem services at a given place is the result of the underpinning stock, condition and configuration of natural capital (natural assets including geology, hydrology, soil, air and biodiversity), human capital (knowledge, skills and social networks within a population) and built capital (human-made infrastructure and assets, such as roads and buildings) (Chan et al. 2012a, Fish et al. 2016, Costanza et al. 2017, Díaz et al. 2018, Langemeyer et al. 2018). The values associated with these benefits encompass instrumental, relational, intrinsic, economic and community-based values, that contribute to peoples’ health, happiness and well-being (Chan et al. 2016). Assessments of cultural ecosystem services are required to incorporate these benefits and values within environmental decision-making and secure their long-term provision (Daily et al. 2009, Costanza et al. 2017, Dasgupta 2021).

Assessing cultural ecosystem services is challenging for several reasons, one of which is that they vary spatially. As the mosaic of capital stocks varies across the landscape, so too does the basket of services those ecosystems provide to people (Carpenter et al. 2009, Andrew et al. 2015, Costanza et al. 2017). This is complicated further because different individuals receive different cultural ecosystem services flows, based on their own values and preferences (Chan et al. 2012b, Díaz et al. 2018, Chan and Satterfield 2020). This variation across spatial scales and between people means that cultural ecosystem services have strong perceptual characteristics and have been described as simultaneously “everywhere and nowhere” (Chan et al. 2012a, Chan et al. 2016). Investigating where people choose to visit across a landscape can lend insight into places that facilitate in-situ cultural ecosystem service flows and, therefore, provide benefits to people. This is relevant for decision-making and environmental management because, without spatially-explicit assessment, these services are vulnerable to being excluded from consideration (Daily et al. 2009, Andrew et al. 2015, Plieninger et al. 2015).

Social media platforms contain information related to in-situ environmental experiences through uploaded content and associated metadata, such as GPS coordinates, descriptive text, titles and date of content creation (Oteros-Rozas et al. 2018, Zhang et al. 2020

). Accessing these data represents a passive form of data collection at a scale that is rarely possible with alternative methods (interviews, visitor surveys) and has proven especially useful in otherwise data-scarce or inaccessible regions (Wood et al. 2013, Ghermandi and Sinclair 2019, Wood et al. 2020, Zhang et al. 2020). The popularity of social media platforms, the potential for large volumes of data collection and lower resource costs have made social media studies on the topic of people-environment interactions increasingly common (Zhang et al. 2020).

GPS-tagged content uploaded to social media is an emergent source of spatial data used as a proxy for visitation occurrence and intensity because it records a "digital footprint" of places where people have visited (Wood et al. 2013, Levin et al. 2015, Tenkanen et al. 2017, Mancini et al. 2018, Zhang et al. 2020). Researchers and practitioners can then apply spatial statistics to create models of visitation as a proxy for some form of in-situ cultural ecosystem service flow. This is an expanding field of research as cultural services are amongst the services most commonly studied (Czúcz et al. 2018) and in-situ cultural ecosystem service studies, recreation in particular, are a leading application of data from social media (Cheng et al. 2019). While not every study uses the same cultural ecosystem service framework to define their research question, the applications of GPS-tagged content to explore people-environment interactions are numerous. Examples include mapping visitor behaviour within national parks (Levin et al. 2015, Tenkanen et al. 2017, Sinclair et al. 2020) and at tourism hotspots (Fisher et al. 2019, Kim et al. 2019, Pickering et al. 2020); evaluating aesthetic preferences using photographic content and location (Figueroa-Alfaro and Tang 2017, Yoshimura and Hiura 2017); evaluating the success of restoration projects (Kaiser et al. 2021); and identifying cultural ecosystem service flow hotspots across landscapes (Richards and Friess 2015, Oteros-Rozas et al. 2018, Arslan and Örüçü 2021).

Modelling the relationships between visitor occurrence and environmental characteristics (both biophysical and human and social attributes) is a common analysis applied to spatial data from social media (Zhang et al. 2020). Examples of spatial model structures include the calculation of regional-level average relationships, local-level spatially varying relationships and predictive models of visitation suitability and occurrence. Expertise in the most up-to-date spatial statistics and modelling approaches is a potential barrier to creating such assessments given the growing volume and availability of "big data" from social media databases and machine-learning and remote sensing applications (Richards and Friess 2015, Pettorelli et al. 2017). The need for such spatially-explicit assessments of cultural ecosystem services is growing as momentum behind natural capital accounting and ecosystem service assessment continues to build (Hein et al. 2020, United Nations 2021).

We selected four different modelling approaches to spatial data from social media, based on their use within literature: 1) global regression, 2) local regression, 3) maximum entropy (MaxEnt) and 4) InVEST recreation model. Throughout this paper, we use the term global with reference to the study area in its entirety (not in the planetary sense) and corresponding models that summarise one average relationship (Tenerelli et al. 2016). While presented as four parallel workflows for clarity and comprehension, they are really a

nested set of regression analyses that share the same foundational underpinnings, but vary in their depth, complexity and method of computation. The assessment is focused on the spatial modelling of biotic, in-situ, cultural ecosystem services using visitation as a measure of potential people-ecosystem interactions (label 3.1 using the classification scheme from CICES v.5.1, (Haines-Young and Potschin 2018)). Further disaggregation within this category was not attempted given the strong perceptual nature of these services and the passive nature of data collection that does not include users' perceived benefits or values.

Regression models applied at the global scale of a study area, such as generalised linear models (GLMs), summarise the average relationship between variables of interest and social media metrics. Examples include travel-cost estimations for tourism (Sinclair et al. 2018) and analysis of USA national park visitation (Sessions et al. 2016). Local regression uses geographic weighted regression (GWR) that allows relationships to vary over space rather than calculating one single average relationship for the entire study area (Fotheringham 2020). This method has been used to consider varying visitor preferences across Europe (Tenerelli et al. 2016) and tourism patterns in South-East Asia (Kim et al. 2019). Maximum entropy (MaxEnt) modelling uses machine-learning and presence records to predict areas of high suitability for a phenomenon of interest (Phillips 2017). MaxEnt has been used to predict potential cultural ecosystem service hotspots in several case-studies, for example, in Japan (Yoshimura and Hiura 2017), Portugal (Clemente et al. 2019) and Turkey (Arslan and Örüçü 2021). Finally, the InVEST recreation model from the Natural Capital Project provides a self-contained tool to model recreation and tourism services using ordinary least squares (OLS) regression (Sharp et al. 2018) and has been used in a number of studies, such as a restoration project in China (Zhao et al. 2021) and spatial planning in Chile (Outeiro et al. 2015).

Few published studies consider more than one of the modelling approaches outlined above (exceptions include Byczek et al. (2018) who used the InVEST model to triangulate their custom model and Tenerelli et al. (2016) who first discounted a non-spatial, global GLM in favour of a GWR). To our knowledge, no study has compared the application of different social media-derived spatial models for the same case-study area. Similarly, no such model of in-situ cultural ecosystem services, based on data from social media, has been used in Ireland. We addressed this dual-knowledge gap by deploying four modelling approaches using data collected from social media in a previously untested case study, County Galway, Ireland. The research questions were as follows:

1. What environmental characteristics are correlated with social media-derived visitation as a proxy for in-situ cultural ecosystem service flows in Galway, Ireland?
2. What are the main differences in output and useability between spatial models and how can this provide information for future cultural ecosystem service assessment?

Model selection for ecosystem service assessment should be co-informed by data availability, data-processing expertise, research questions and spatial extent of the area of interest (Pettorelli et al. 2017, Meraj et al. 2021). We aim to provide a guide for future

cultural ecosystem assessment that takes into account these dimensions and that is relevant for researchers and practitioners using spatial models coupled with social media-derived data.

Methods

Study area

County Galway, Ireland, was selected as the study area because of its heterogeneous landscape, socio-cultural heritage and high visitor numbers. In 2018, visitor numbers were estimated at 1 million domestic and 1.8 million international, contributing a total of €800 million in revenue (Cunningham et al. 2015, Galway County Council 2021). Located on the west coast of Ireland (53°19' N, -9°00' W) (Fig. 1), the county covers 6,150 km² with a population of 175,000 in 2011 (CSO 2011a). The scope of this study focused on spatial trends over a continuous, semi-natural landscape. Therefore, the area of interest was restricted to exclude Galway City as an urban hotspot, Lough Corrib as an inaccessible expanse of freshwater and islands. The remaining area (5,850 km²) contains a diverse landscape of natural features including mountainous areas, grasslands, wetlands, forested areas and coastline. The region also contains areas of biological interest such as Connemara National Park, 19 Special Protection Areas (SPAs) and 77 Special Areas of Conservation (SACs) (NPWS 2022).

Social media data collection

Data were collected from the Flickr social media platform using the statistical programming language R v.4.1.2 (R Core Team 2021), the Flickr API and the R package photosearcher v.1.0 (Fox et al. 2020). Flickr was selected because of its sizeable userbase estimated to have uploaded 5.67 billion photos between 2004 and 2016 (Ding and Fan 2019), previous work that suggested the platform hosts a more diverse userbase compared with other platforms (Oteros-Rozas et al. 2018) and its use in similar studies on the topic (Wood et al. 2013, Tenerelli et al. 2016, Wood et al. 2020). Flickr has the benefit of complementary data access policies that permit the collection and use of data for academic research purposes (Fox et al. 2020). Photosearcher calls on the Flickr API to retrieve data based on user-provided parameters. All searches were conducted in February 2022 and parameterised to retrieve photo records with GPS coordinates taken between 1 January 2010 and 31 December 2019. This time range was selected to ensure a sufficiently large dataset for modelling purposes, to reflect the widespread use and accuracy of GPS devices and to exclude the disruption of travel restrictions imposed due to COVID-19 legislation in 2020. Two datasets were collected: a validation dataset and a modelling dataset. Firstly, popular locations, based on official national tourism statistics with recorded visitor numbers from Fáilte Ireland (2019a), were used to validate the relationship between social media records and visitation. Validation sites were selected, based on two criteria: 1) sites with visitor number estimates by the national tourism body for at least 4 years between 2010

and 2019 (Fáilte Ireland 2019a) and 2) solely indoor sites (e.g. concert venues, indoor museums) identified using authors' expert knowledge were removed as they were suspected to provide limited potential for in-situ ecosystem service supply. This process produced 38 sites (Suppl. material 1). The second dataset was collected to model visitation across the area of interest. A vector file defining Galway was used to retrieve photo records and the output was cleaned to contain only photo ID, user ID, date taken and GPS coordinates.

Photo-user-days calculation

The photo-user-days (PUD) metric developed by Wood et al. (2013) has been used as an indicator for visitation, based on geotagged social media data in a number of studies (Sonter et al. 2016, Lee et al. 2019, Wood et al. 2020). The PUD metric is defined as the number of users who upload at least one photograph in a day, at a given area or location and is designed to prevent the inflation of photo-counts based on extremely active users. For the validation dataset, PUD values were calculated at the site level to match official visitor statistics and provide an appropriate comparison to validate PUD as a proxy for visitor numbers (Fáilte Ireland 2019a). The relationship between visitor numbers and PUD counts was checked using Pearson's correlation statistic and the suitability of this test checked using the Shapiro-Wilk test for normality (Crawley 2015). In the modelling dataset, PUD values were calculated per 2 km grid square as users may choose to visit multiple places in a single day and it was desirable to capture these multiple visits. GPS points were assigned a 200 m buffer, based on a conservative estimate of technological accuracy and previous work that found photos uploaded in Western Europe had a mean inaccuracy of 100 m (Zielstra and Hochmair 2013). GPS points were overlaid with a 2 km² grid and assigned a grid ID based on spatial overlap. A variable representing the combination of user ID, grid ID and date was constructed. The final PUD dataset was created by randomly slicing this variable to give one data point per unique user ID, grid ID and date combination. This sampling technique prevents the inflation of data based on individuals contributing many photos from a single visit to one grid cell in one day, but allows users to contribute to multiple cells per day.

Environmental variable selection and data sources

Environmental variables were selected, based on natural, human and built capital attributes identified as factors contributing to cultural service flows (particularly recreation) in the UN System of Environmental-Economic Accounting framework (United Nations 2021) and previous studies using geotagged social media records (Tenerelli et al. 2016, Byczek et al. 2018, Tieskens et al. 2018, Chang and Olafsson 2022). Natural capital attributes included biophysical variables describing ecosystems at a given location linked to potential ecosystem service supply, built capital attributes included infrastructure associated with the accessibility and attraction of a given place and human capital was included using population density as a proxy for service demand (Table 1). While this does not capture the complexity of all factors linked to cultural service flows, this schema was designed to cover the variety of capital stocks identified in literature, given the available data for the

study area. Spatial data related to these variables were collected from existing databases and clipped to the area of interest. The raw spatial data were saved and a second set of maps were created by calculating an indicator value per 2 km grid cell using zonal statistics tools in ArcMap v.10.7.

Statistical modelling

Four modelling exercises were applied in this study: 1) global regression (using both presence and count data), 2) local GWR (presence data), 3) MaxENT and 4) InVEST (Table 2).

Global regression

A logistic GLM (model a) was computed in R using a binary response variable describing the presence of PUD records per 2 km² grid cell according to the formula:

PUD presence ~ Environmental predictors, family = binomial (link = logit) (a)

The grid size was selected to ensure a sufficient proportion of cells contained presence records and to capture the variation of environmental attributes. Model optimisation was conducted using stepwise model selection to minimise the Akaike Information Criterion (AIC). This is a standard model optimiser that balances performance and complexity to identify the most parsimonious model (Crawley 2015). Multicollinearity was checked using variance inflation factors (VIF). Dispersion, outliers and the assumption of homoskedasticity were checked using the dHARMA package (Hartig and Lohse 2018). The receiver operator curve (ROC) was used to assess model fit to the dataset. The area under the curve statistic (AUC) quantifies this trait with values of 0.5 describing a model that performs as well as a random model and a value of 1 describing a model that perfectly fits the data (Swets 1988, Yoshimura and Hiura 2017).

The output of this global model summarised average trends for the entire region. We hypothesised that these relationships may vary across the landscape due to local socio-environmental context and accessibility. Model performance across spatial scales can be assessed by testing for spatial autocorrelation of the residuals (Fotheringham 2020, Comber et al. 2022). We checked for evidence of spatial autocorrelation using Moran's I correlograms generated in the pgirmess package (Giraudoux et al. 2022). This test assesses if model performance (magnitude of residuals) is randomly dispersed across the area of interest (Comber et al. 2022). Evidence of spatial autocorrelation was detected and, therefore, the development of models that account for this spatial heterogeneity was recommended (Suppl. material 4). This was done in two ways: 1) a spatially autocorrelated mixed-model (SAM) that models global relationships, but incorporates a spatial random effect and 2) a GWR that permits locally varying relationships (see next section). The SAM model (model b) consisted of two components: environmental predictor fixed effects and a spatial random effect, based on latitude and longitude, computed using the spaMM package (Rousset and Ferdy 2014), with the following formula.

PUD presence ~ Environmental predictors + Matern(1|Lat + Long), family = binomial (link = logit) **(b)**

This procedure is a recommended approach for modelling spatial data in ecology where spatial autocorrelation is detected (Comber et al. 2022). Model checks were computed as outlined in the previous section.

The PUD count per 2 km² grid cell was used as the response variable in a third model. Count data are typically modelled using a poisson GLM (O'Hara and Kotze 2010). Upon inspection, the global poisson model was not appropriate due to overdispersion, homoskedasticity and spatial autocorrelation of residuals. Instead, a SAM (model c) was deployed in the same procedure as above, according to the formula:

PUD Count ~ Environmental predictors + Matern(1|Lat + Long), family = poisson (link = log) **(c)**

Local regression - Geographic weighted regression

A logistic GWR model was conducted to test for spatially varying relationships using the presence of PUD. GWR computes repeated regression analyses across the landscape and applies a distance-based weighting function so that data points closer to the regression point are weighted more compared to distant data points (see Fotheringham et al. (2003) for an in-depth methodological text). The maximum distance around each regression point to include a data point is referred to as the bandwidth and the shape of the weighting function is referred to as the kernel (Fotheringham et al. 2003). The model output supplies a regression coefficient and t-value for each predictor variable at each regression point. Changes in sign, magnitude and significance for any relationships between environmental attributes and PUD occurrence can then be mapped and inspected. The GWR analysis was computed in MGWR v.4.3 software using the fixed kernel setting and optimised using the CV method (Oshan et al. 2019). Local-level independence was checked using local VIF values. As with the previous models, Moran's I correlogram and the ROC were plotted and AUC value calculated. Coefficient surfaces for each environmental predictor were mapped using the tmap package (Tennekes 2018).

A GWR model using the PUD count variable was trialled, but ultimately deemed inappropriate. As a global poisson GLM was not supported due to overdispersion and heteroskedasticity concerns, it was surmised that a GWR using a poisson distribution was not recommended as this simply runs a series of repeated poisson models with the same variables, just different weighting schemes. More sophisticated model structures to address this (e.g. negative binomial, quasipoisson) are not currently available in combination with GWR techniques and lack consensus in the academic community and so were not pursued further by us. As more sophisticated model structures become available in the future, these could be considered.

MaxEnt model

MaxEnt was the third model type tested. This model predicts areas of high probability for visitation, based on characteristics associated with sampled PUD occurrence. MaxEnt differs from the regression models outlined above because it adopts a “presence-only” approach. Sampling social media records cannot prove the absence of visitation and, therefore, a value of 0 PUD does not reflect a true and tested absence of visitation (Phillips 2017). Instead, Maxent uses high-resolution environmental data to compare areas of observed presence records to a set of background pixels using machine-learning software to model the probability of PUD occurrence. All environmental predictors were converted into 100 m raster format using spatial statistics tools in ArcMap v.10.7 because MaxEnt only accepts data of identical resolution and extent. The objective of this exercise was to predict the suitability of visitation and so the PUD dataset was randomly partitioned into 75% training data and 25% test data. We ran the model 100 times using a bootstrap procedure to produce an average map of suitability and jackknife analysis was used to compare model performance (Phillips 2017). Jackknife analysis runs two models for each environmental predictor: the first includes only that predictor in isolation and the second includes all variables, except that predictor. Differences in model performance compared to the maximal model can identify predictors of greater contribution and predictive influence.

InVEST model

The final model considered was the InVEST Visitation, Recreation and Tourism v.3.10 ecosystem service model (Sharp et al. 2018). The InVEST model uses an archive of Flickr data from 2005-2017 to calculate annual PUD values and the user supplies spatial environmental predictor data. The model performs spatial analysis to create indicator values “in-house”, based on a limited number of pre-set options. It can also compute an OLS regression using those generated indicators, user specified cell size and retrieved PUD archives. We ran the InVEST model using the OLS regression option at a 2 km² grid size from 2010-2017 and supplied the shapefiles of environmental attributes (Suppl. material 2).

Results

Validation of PUD and visitation

All 38 validation sites returned PUD counts that were plotted against official visitor numbers reported by Fáilte Ireland (Fig. 2) (Fáilte Ireland 2019a). The Pearson correlation coefficient was 0.7 indicating a positive correlation and a significant relationship was observed using OLS regression ($p < 0.001$ at the 0.05 level, $R^2 = 0.3$). This result supports the use of social media-derived PUD as a visitation proxy in Ireland.

Social media data and PUD calculation

A total of 25,170 geotagged photographic records were retrieved (Suppl. material 3) and converted into 6,762 PUDs (Fig. 3a-b). The number of PUD per grid cell ranged from 0 to 413 with a mean of 5.65 and standard deviation of 20.8 (Fig. 3c). Spatial analysis using the Getis-Ord G_i^* tool in ArcMap v.10.7 showed significant clustering of PUD, and therefore, a non-random distribution (Fig. 3d). Hot spots were found in western, coastal areas and the area south-east of Galway City. Cold spots of low PUD count were concentrated in the east of the region.

Global regression

Logistic regression - PUD presence

The non-spatial, logistic GLM model of PUD presence contained 11 environmental variables (Table 3a). Land-cover type variables produced collinearity concerns ($VIF > 5$), while protected status, town distance and Shannon's diversity of land cover were not significant. These variables were not included in the final model. VIF values for the 11 remaining predictors were < 5 and, therefore, satisfied the assumption of independence (Guisan and Zimmermann 2000). Distance to the coast and elevation were found to significantly decrease the likelihood of PUD occurrence, while the remaining variables were found to increase the likelihood of PUD occurrence. The AUC value was 0.839 indicating a moderate fit to the dataset (Fig. 4). The Moran's I correlogram of the model residuals showed statistically significant values at distance classes < 25 km, supporting evidence of spatial autocorrelation (Suppl. material 4).

A logistic SAM was constructed given the spatial autocorrelation detected in the logistic GLM (Table 3b). This model identified the same variables as significantly correlated with PUD presence, although coefficients and standard errors varied marginally. The logistic SAM displayed a lower AIC value compared to the non-spatial global model (1518 compared to 1629) and so, this was identified as the more parsimonious model for PUD occurrence at the global scale. Evidence of spatial autocorrelation was not detected (Suppl. material 4).

Poisson regression - PUD count

Visitor intensity was modelled using a SAM of PUD counts (Table 3c). Results showed that all coefficients retained the same direction (sign) as the presence model, although population was not significant at the 95% level. This suggests similar characteristics are associated with places most likely for Flickr contributors to visit and places with the highest visitation rates, with the exception that places with high population density are more likely to have a PUD > 0 , but not necessarily high total PUD counts. Model performance was inspected by plotting observed PUD counts and model predicted PUD counts (Fig. 5). The results showed a significant positive correlation (slope = 0.84, $p < 0.001$) and a coefficient

of determination (R^2) value of 0.5. The PUD variable is skewed with many zero values and fewer high values. This is reflected in the model fit with greater statistical noise at lower PUD values. Evidence of spatial autocorrelation was not detected (Suppl. material 4).

Geographic weighted logistic regression - PUD presence

The global logistic model displayed evidence of spatial autocorrelation that may mask local relationships. GWR was used to investigate this and the resulting coefficients were mapped to show variation in magnitude, direction and significance (Fig. 6). Distance to the coast was found to produce collinearity problems and dropped from this model. Monte Carlo simulation produced a significant result ($p = 0$) for all predictors, further supporting the presence of spatially-varying relationships. Several variables showed changes in significance levels across the area of interest, most notably town distance that showed a negative coefficient in eastern areas and a positive coefficient in the west. In other words, areas close to towns in the east of the region are more likely to have a PUD record, but remote areas far from towns are more likely to record a PUD in the west. The presence of recreational sites was found to be significant in three hotspot areas across the landscape. The ROC for the GWR local model was plotted with an AUC of 0.909 (Fig. 4). Compared to the non-spatial global logistic model, the GWR local model showed an improved model fit as indicated by its AUC value (0.909 compared to 0.839) and AIC value (1469 compared to 1629).

MaxEnt model

The MaxEnt model used 100 m rasters of environmental predictors and PUD occurrence coordinates to predict areas of visitation suitability (Fig. 7). The mean AUC value was 0.8 indicating an improved discrimination compared to a random prediction model. The results show hot spots of high suitability for PUD visits in clusters in the east of the county, along road networks and coastal areas and surrounding Galway City. Dark blue areas indicate low likelihood of suitability for PUD occurrence and span the predominantly agricultural areas in the east and wetlands with poor connectivity and low population. Variable response curves are included in Suppl. material 5.

MaxEnt's jackknife analysis, based on average AUC values, is shown in Fig. 8. Distance to the coast, elevation, presence of major roads and presence of recreational sites contributed the most information to the model in isolation (shown in dark blue bars). Similarly, the model performance showed the greatest decrease with the removal of elevation and major roads (shown by the light blue bars) suggesting that these variables contain information that is not captured by other variables. The variables with the lowest contribution to the model include geological heritage sites, river length and slope that suggests they have limited predictive power in isolation.

InVEST recreation model

The output of the InVEST recreation model reports an OLS regression of log transformed annual PUD values (retrieved from an archived dataset) on environmental predictors (Table 4). Direct statistical comparison to other models is not appropriate given differing input data; however, we provide some observations regarding outputs and computation. The InVEST model uses log annual PUD as the response variable that is characterised by low values and produces correspondingly small absolute coefficient values. This model output included land-cover variables manually dropped from previous models due to collinearity concerns. The InVEST tool does not compute or report collinearity check. Other differences compared to previous analyses include river length and geological heritage being insignificant. All land-cover variables are reported with a negative sign suggesting that any cell dominated by one land-cover type has fewer PUD counts. This is consistent with a positive and significant coefficient for habitat diversity.

Discussion

This study investigated potential in-situ cultural ecosystem service flows across a previously untested context using data from social media as a visitation metric and a spectrum of spatial models. The following discussion is split into three major themes: (1) PUD as a proxy for potential in-situ cultural ecosystem service flows in County Galway, (2) spatial model selection and use and (3) general comments about the use of social media data.

Cultural ecosystem service assessment using PUD in County Galway

This is the first study to use a social media-derived PUD indicator in the Irish context and the positive correlation between PUD and official visitor counts is consistent with other validation studies (Sonter et al. 2016, Fisher et al. 2019, Kim et al. 2019). Application of the PUD metric successfully revealed global trends, local relationships and predictive suitability maps. These results demonstrate the applicability of social media-derived analysis for providing spatially explicit information for ecosystem service assessments. This information is useful for environmental management by providing a mechanism for decision-makers to account for these services and corresponding benefits in policy and planning. For example, this information could contribute to designing future tourism plans and local development strategies given the projected increase in visitor numbers to the area and understanding how and where those people interact with nature (Galway County Council 2021). In the long-term, spatially explicit assessment of the contributions ecosystems make to peoples' lives can support the preservation of these contributions for future generations (Díaz et al. 2018, Dasgupta 2021).

A core set of environmental attributes representing a blend of natural, human and built capital were identified as correlated with PUD across all models: coastal distance,

presence of major roads, population density, habitat diversity, elevation and presence of recreational sites. The finding that a blend of capital stocks was correlated with PUD mirrors results in other contexts (Tenerelli et al. 2016, Byczek et al. 2018, Tieskens et al. 2018), and demonstrates that natural and biophysical characteristics, socio-cultural context and accessibility are all implicated in the potential flow of cultural ecosystem services (Byczek et al. 2018). This is aligned with the natural capital approach that considers the underpinning stocks that give rise to ecosystem services in terms of their unique combination, configuration and condition at a given place (Chan et al. 2012a, Jones et al. 2016, Costanza et al. 2017, Mace 2019). Another shared result between all models used was that protected status was not correlated with the PUD indicator, unlike in other studies, for example, USA (Figueroa-Alfaro and Tang 2017) and Japan (Yoshimura and Hiura 2017), which may suggest that protected areas serve a different role in people-environment relationships compared to other contexts. The majority of protected areas in Co. Galway fall under SAC and SPA designations that are targeted at nature conservation under EU-wide nature directives (NPWS 2022). This may be a reason why the protected status does not appear to be correlated with in-person visitation in Co. Galway, as the primary purpose for their designation lies in biodiversity conservation rather than providing a societal utility. It may be that characteristics that allow for sites of high biodiversity value prevent high visitor numbers, for example, inaccessibility and lack of man-made infrastructure or sites unsuitable for such development.

The global, non-spatial logistic model was found to display spatial autocorrelation (violating the assumption of independence) and a higher AIC value compared to both spatial model alternatives (SAM and GWR). Therefore, models that account for the spatial nature of geo-tagged social media data should be used in such cases. Spatially varying local relationships were found using GWR. This is consistent with studies that found evidence of local relationships when modelling cultural ecosystem service flows (Tenerelli et al. 2016, Schirpke et al. 2018, Kim et al. 2019). In the most extreme case of town distance, the relationship was reversed across the study area with remote areas correlated with visitation in the west and places close to towns correlated with visitation in the east. By definition, this phenomenon was obscured by models that produce one single relationship for the entire region: in both the global logistic regression and the SAM town, distance was not significant, while the InVEST model suggested a positive relationship overall. Previous European studies identified natural areas close to urban sites as potential hotspots for providing cultural ecosystem services (Ridding et al. 2018, Long et al. 2020) and, in Galway, this appears to be the case for some areas, but not everywhere. The rugged, mountainous and wetland landscapes in western areas may be perceived as more attractive to visit because of their remoteness and attract visitors away from urban areas, whereas the predominantly agricultural landscapes surrounding towns in the east may not. In less pronounced examples of local-level relationships, some variables were found to be significant, but only in limited areas rather than across the entire region, for example, habitat diversity, recreational sites and geological heritage. The identification of local relationships using GWR shows that caution should be applied when downscaling global average trends (non-spatial GLM, SAM, InVEST) to local areas (Fotheringham 2020, Comber et al. 2022).

The MaxEnt model was the only method used for prediction due to its presence-only approach (as opposed to testing for correlations in GWR and SAM models). Jackknife analysis showed that elevation, coastal distance, recreational sites and town distance were the variables of greatest predictive influence in the model. Other variables were found to have limited contributions to model prediction, such as river length, water cover and geological diversity, despite showing significant correlations in regression analysis. This result can support the prioritisation of data collection when designing management interventions as some variables appear to be more informative. Results in this case-study suggest rural areas, close to the coast, of moderate elevation and with a major road should be prioritised for targeted management interventions. These areas have the potential to experience high visitor volumes through in-situ cultural ecosystem service supply and the associated anthropogenic disturbance could contribute to negative ecological consequences and compromise long-term service flows.

The InVEST model presented some differences compared to the other models, such as the inclusion of land-cover variables and different significance levels for water cover and geological diversity. This is not unexpected given that InVEST is premised on a different response variable dataset (archived Flickr database), but it does provide a comparator to triangulate with other methods. Overall, the variables of greatest significance in user-led regression techniques (coast distance, elevation, recreational sites, major roads) were also identified as significant using InVEST with less intensive processing of data required. Stepwise model optimisation and diagnostics are not provided by the default InVEST tool and so any changes must be led by the user manually inspecting model outputs, making desired changes and re-running the model in its entirety, which can be time-consuming. These characteristics may be limitations to the InVEST model depending on the research context and similar remarks have been stated in literature (Byczek et al. 2018).

This is the first social media-based spatial modelling study in the Irish context and so comparison is limited. Previous research used a stated-preference methodology to elicit aesthetic preferences of rural landscapes in Ireland, based on a nationally-representative survey of 430 individuals (O'Donoghue et al. 2020). The results showed some overlap between highly valued aesthetic characteristics and characteristics correlated with the social media-derived PUD variable, for example, freshwater (lakes, ponds, rivers), marine areas and beaches, heritage monuments and geological features (mountains and cliffs). On the other hand, built attributes (roads, fences, buildings) were assigned a low value in the stated preference study, but urban areas and roads were found to be correlated with the PUD indicator. These differences may be due to the different phenomena investigated (visitation versus aesthetics), the different cohorts sampled, differences between revealed behaviour and stated preferences and the reality that not all areas in the landscape may be equally accessible and, therefore, visitor occurrence may not reflect a true "choice" amongst all possible options. This emphasises the value of assembling a diverse suite of tools to investigate cultural ecosystem services and the contributions they make to people's lives (Zhang et al. 2020).

Model selection and useability

The results demonstrate the applicability of social media to developing spatial models of visitation. The differences in model outputs identified have the potential to create differing interpretations when a single approach is deployed in isolation. Therefore, we find merit in investigating multiple modelling tools—even if, ultimately, one is favoured for final reporting. While this study presented four workflows for clarity and comprehension, there is flexibility to customise these approaches to user needs beyond what is presented here. Some observations regarding data availability, data processing, expertise, research question and spatial extent required are presented below to provide information for future cultural ecosystem service assessment using these methodologies.

Area of interest and scale

Data from social media are widely applicable, but their geographic coverage is unequal. The InVEST model is not suited for use in data-scarce regions as it recommends at least 50% of cells contain PUD records (Sharp et al. 2018). Data availability also has a role when designing global and local regression analysis. In areas of limited data coverage, the resolution of analysis (cell size) may be constrained to ensure sufficient coverage and variation in the response variable. Both global regression and GWR can adopt model families to account for non-normal response variables (count data, presence data); however, GWR lacks consensus regarding more complex model types and optimisation that limits its application, for example, the negative binomial GLM family, is currently unsupported and multiscale (varying bandwidth) GWR cannot be combined with GLM model families (Oshan et al. 2019). MaxEnt's presence-only approach lends itself most easily to data-scarce regions as it was intended for use on species distribution data that are often characterised by a low number of observations.

Model resolution should also consider the size, the heterogeneity of environmental attributes and the suspected behaviour of the sampled population of the area of interest. For large areas with suspected variation in local socio-environmental context, the presence of local relationships may be hypothesised from study initiation (Tenerelli et al. 2016), whereas, for more constrained areas with a shared socio-environmental context, for example, visitation within a self-contained national park (Tenkanen et al. 2017, Sinclair et al. 2020), this may not be the case. The case-study of County Galway used a 2 km² resolution to create the PUD indicator and calculate environmental indicators to account for both data coverage and local knowledge that people engage in several visits to distinct locations within single daytrips. Spatial autocorrelation of model residuals should be inspected regardless of scale. In this case, Moran's I statistic was used to assess model performance by testing model residuals to see if they behave in a clustered, dispersed or random pattern. If spatial autocorrelation is detected, alternative models that account for this should be considered, such as a mixed-model using spatial random effects (SAM) or a GWR analysis (Comber et al. 2022). This was the case for the non-spatial, global, logistic GLM reported in this study that was ultimately discounted due to evidence of spatial

autocorrelation. The selection between the logistic SAM and GWR models can then be determined, based on the research question, context and statistical support for spatially varying relationships. Without these checks, there is a danger of landing on a model that performs differently in different places, but is poorly representative everywhere (Fotheringham 2020).

Variable selection and indicator calculation

The available environmental data also determine grid size and mapping outputs for all four model workflows discussed. Variable selection should be grounded in the hypothesis for the phenomenon of interest and data may be gathered from available regional or national data sources, on-the-ground sampling or remote sensing. Where appropriate data are available, indicator calculation is limited only by the users' expertise with data processing with a range of basic spatial statistics tools available in most GIS software, for example, proximity, presence, count, density, mean, min, max values. These indicators should be designed so that they vary across the landscape in a meaningful way, based on the cell resolution selected and this necessitates an interplay between indicator design and model resolution. For example, the use of "river presence" was not appropriate for modelling Co. Galway because the landscape contains many river features and, at the 2 km² scale, almost all cells contained an identical value that was not meaningful. Instead, river length was selected as the indicator. When using GWR, this should also hold true at the bandwidth scale to ensure sufficient variation around each regression point, in addition to local multicollinearity checks (Fotheringham et al. 2003). The use of GWR may also preclude the use of some proximity-based indicators that vary monotonically across the landscape and introduce collinearity problems, for example, coastal distance in the case of Co. Galway (Comber et al. 2022).

The balance between resolution and indicator selection was also apparent when creating raster files for use in MaxEnt. Input data to MaxENT must be of identical extent and resolution in raster format that matches the desired output resolution. This required adjusting indicator file formats (indicator values themselves were not recalculated in this process). A second adjustment was made to change some indicators that were informative at the 2 km² resolution, but less informative at the 100 m resolution. For example, at the larger cell size, binary presence indicators (presence of geological heritage, presence of recreational sites) captured the effect of a feature throughout the surrounding area, whereas, when using smaller cell sizes, this landscape effect was lost. Instead, a proximity or density-based indicator may be more appropriate at fine resolution where the impact of a feature is hypothesised to extend beyond the cell size. Careful consideration is required to understand these relicts of model specification and may require a back-and-forth approach to identify the most appropriate indicators for a given resolution. These considerations are significantly constrained when using the InVEST model which only contains seven pre-set options to the user for calculating environmental indicators (two raster, five vector data types) that stymies customisation (Byczek et al. 2018).

Computational and resource costs and savings

Beyond statistical details, there are practical considerations that may determine the analysis of data from social media in ecosystem service assessments, such as the availability of computational resources, time and expertise. The InVEST recreation model is designed to be accessible to any user who may be unfamiliar with advanced coding, statistics or GIS by providing a self-contained, comprehensive interface and interpretable outputs (Sharp et al. 2018). Running the model requires an internet connection, but is otherwise computationally light. It also benefits from available online tutorials and documentation provided by the Natural Capital Project to guide users (Sharp et al. 2018). These advantages must be balanced against some drawbacks, such as the use of an archived dataset of only Flickr-based records that may become outdated (2005-2017 currently) and limited user-customisation of model structure and variable calculation.

The MaxEnt model is supported through a self-contained programme (and compatible R package) with introductory online resources available (Phillips 2017). The user must supply a suitable set of raster datafiles of identical extent and resolution that requires some experience working with spatial datasets to prepare. The model then computes a form of logistic regression using machine-learning to optimise model gain. Computing this potentially complex model structure from scratch would require advanced knowledge unavailable to many potential users and, therefore, the MaxEnt tool makes otherwise inaccessible statistical analyses possible. The default output includes a heatmap of predicted suitability that is intuitive; however, advanced options (jackknife analysis, scenario comparison, bootstrap replications) require a greater depth of interpretation and expertise. The model may also take a significant length of time to compute when including replication. The MaxEnt model was often deployed amongst studies that clearly defined cultural ecosystem services within the core research question(s) (examples included Richards and Friess (2015), Yoshimura and Hiura (2017), Clemente et al. (2019), Long et al. (2020), Arslan and Örüçü (2021)), possibly due to the model's origin in biodiversity mapping and shared ecological disciplinary overlap.

The use of user-defined global regression techniques, for example, GLMs and SAMs, is the most customisable approach detailed in this study. The user should follow the usual statistical checks (outliers, dispersion, heteroskedasticity), as well as spatial autocorrelation checks. The user must prepare a data table containing the response variable and environmental indicator for each cell in the landscape, which often requires data wrangling, cleaning and manipulation using spatial statistics. Highly-advanced model structures may be theoretically possible, but inaccessible due to the expertise required to define, run and interpret them. In some cases, the desired model may not yet be computationally possible. For example, multiscale GWR is currently limited to gaussian distributions, while GWR using poisson and binomial model families can only use fixed bandwidths. GWR can be computed in a number of ways including a number of R packages, Python and the stand-alone MGWR tool, although their respective optimising algorithms vary, leading to different outputs. GWR model runs can be computationally

intensive and take long periods of time depending on the size of the dataset, optimisation criteria and use of Monte Carlo simulation (Oshan et al. 2019, Comber et al. 2022).

These considerations (summarised in Table 5) have implications for the time, resources and expertise required to conduct analysis using geo-tagged social media data. In some cases, more than one model is required necessitating a to-and-fro procedure. This level of consideration is required from project inception to create the most informative model for a given phenomenon of interest. The alternative is the creation of knowledge premised on the shaky foundation of what is most familiar and accessible, rather than what is most appropriate.

Social media data use and limitations

The focus of this study was the application of social media-derived data in different spatial models, not social media-derived data itself. The caveats of social media as a data source have been well-described in literature, but its use continues to increase, especially in studies related to cultural ecosystem services (Cheng et al. 2019, Ghermandi and Sinclair 2019, Zhang et al. 2020). We make some general comments here to contextualise the use of social media data and, despite these caveats, the results presented remain relevant as the first such application in Ireland and they provide transferable insight regarding spatial model selection for other contexts.

Potential inaccuracy in the geo-tagged coordinates was accounted for by using a 200 m buffer when calculating the PUD variable, given the suspected technological accuracy in northern Europe (Zielstra and Hochmair 2013, Tieskens et al. 2018). MaxENT requires precise GPS points and InVEST assumes GPS-tags are accurate. If there are suspected errors in GPS accuracy, a user can choose to manually validate GPS tags by inspecting photographic content (although this is costly for large samples) (Walden-Schreiner et al. 2018) or select a resolution coarse enough to render the suspected error marginal.

Social media user-groups are a self-selecting, unrepresentative sample of the general population and so data collected carries with it the bias of the userbase (Ghermandi and Sinclair 2019, Sinclair et al. 2020). Typically, this bias skews towards wealthier individuals, younger people and contains a gender-bias depending on the platform (Zhang et al. 2020), although some studies suggest Flickr is more diverse than other leading platforms (Fox et al. 2020). We acknowledge that the results of this study represent a self-selecting group of Flickr contributors and do not intend to generalise the overall population interacting with the landscape of County Galway. While skewed towards a specific userbase, the results of this pilot study provide a first-look into social media data applications in Ireland and future work can complement these findings by deploying a range of social media platforms and other data sources (visitor statistics, choice-based experiments, participatory data collection) (Wood et al. 2020).

The social media content used in this study was not screened or filtered beyond the removal of areas outside the area of interest. The volume of data collected (25,000 data points contributed by 1,866 users) and the widespread distribution of those points across

the landscape suggest that the sampled data contain a diversity of information and were successfully applied as per other similar studies exploring visitation (Sonter et al. 2016, Sinclair et al. 2020). We emphasise that this study considered in-situ cultural ecosystem services broadly without disaggregating specific services or assuming user intentions and, therefore, analysis is based on the location of the photographer, not the content captured in the photograph itself (which would require inspection) (Tenerelli et al. 2016, Langemeyer et al. 2018). There is also an assumption that the physical presence of an individual in an ecosystem represents a potential cultural ecosystem service flow. Investigators may choose to filter content to varying degrees depending on their available resources, data needs and research questions (e.g. based on sentiment, content, contributor or machine-learning) (Langemeyer et al. 2018, Oteros-Rozas et al. 2018, Lee et al. 2019, Fox et al. 2021). Even with these efforts, it is not possible to know with certainty how the intent and values behind the location content was created without the input of the individual contributor. We also do not suggest that the findings of this study reveal true “preferences” as this implies a degree of choice amongst all options that was not tested in this study and there is an inherent bias towards more “picturesque” places as an artefact of using a photo-sharing platform (Clemente et al. 2019). All social media-based studies must grapple with these problems and, despite these caveats, the popularity and breadth of social media applications continues to grow (Ghermandi and Sinclair 2019, Zhang et al. 2020).

Conclusions

Spatial models applied to data from social media revealed a blend of environmental characteristics related to visitation and potential in-situ cultural ecosystem service flows across County Galway, Ireland. These characteristics included coastal distance, elevation, major roads, recreational sites, urban distance and habitat diversity. Famously, all models are imperfect; but by discussing the workflow for each approach, we have articulated where and why different models may be useful. We hope that this exercise, zooming in on the application of spatial models using social media data to investigate cultural ecosystem services, can serve as a useful signpost for researchers and practitioners involved in ecosystem service assessments and natural capital accounting. Model selection considerations in such exercises should capture the context of the area of interest, computational demands, data availability and structure and research scope. Furthermore, for transparency and clarity, we encourage all researchers and practitioners to explain and justify their choice of model and variables in detail. The results presented are especially pertinent given the growing volume of available data from social media and the need for spatially-explicit models for natural capital accounting and ecosystem service assessments.

Acknowledgements

Thanks to L. Coleman for GIS data preparation advice and discussion and C. Farrell for cultural ecosystem service discussions prior to project commencement.

Funding program

All of the authors were funded by Science Foundation Ireland (SFI) under Grant Number 16/RC/3889 for BiOrbic Bioeconomy, SFI Research Centre, which is co-funded under the European Regional Development Fund and by BiOrbic industry partners.

Hosting institution

Trinity College Dublin, the University of Dublin

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Andrew M, Wulder M, Nelson T, Coops N (2015) Spatial data, analysis approaches, and information needs for spatial ecosystem service assessments: a review. *GIScience & Remote Sensing* 52 (3): 344-373. <https://doi.org/10.1080/15481603.2015.1033809>
- Arslan ES, Örüçü Ö (2021) MaxEnt modelling of the potential distribution areas of cultural ecosystem services using social media data and GIS. *Environment, Development and Sustainability* 23 (2): 2655-2667. <https://doi.org/10.1007/s10668-020-00692-3>
- Byczek C, Longaretti P, Renaud J, Lavorel S (2018) Benefits of crowd-sourced GPS information for modelling the recreation ecosystem service. *PloS One* 13 (10). <https://doi.org/10.1371/journal.pone.0202645>
- Carpenter S, Mooney H, Agard J, Capistrano D, DeFries R, Díaz S, Dietz T, Duraiappah A, Oteng-Yeboah A, Pereira HM (2009) Science for managing ecosystem services: Beyond the Millennium Ecosystem Assessment. *Proceedings of the National Academy of Sciences* 106 (5): 1305-1312. <https://doi.org/10.1073/pnas.0808772106>
- Chang P, Olafsson AS (2022) The scale effects of landscape variables on landscape experiences: a multi-scale spatial analysis of social media data in an urban nature park context. *Landscape Ecology* 37: 1271-1291. <https://doi.org/10.1007/s10980-022-01402-2>
- Chan KA, Satterfield T, Goldstein J (2012a) Rethinking ecosystem services to better address and navigate cultural values. *Ecological Economics* 74: 8-18. <https://doi.org/10.1016/j.ecolecon.2011.11.011>
- Chan KA, Guerry A, Balvanera P, Klain S, Satterfield T, Basurto X, Bostrom A, Chuenpagdee R, Gould R, Halpern B, Hannahs N, Levine J, Norton B, Ruckelshaus M, Russell R, Tam J, Woodside U (2012b) Where are cultural and social in ecosystem services? A framework for constructive engagement. *BioScience* 62 (8): 744-756. <https://doi.org/10.1525/bio.2012.62.8.7>

- Chan KA, Balvanera P, Benessaiah K, Chapman M, Díaz S, Gómez-Baggethun E, Gould R, Hannahs N, Jax K, Klain S (2016) Why protect nature? Rethinking values and the environment. *Proceedings of the National Academy of Sciences* 113 (6): 1462-1465. <https://doi.org/10.1073/pnas.1525002113>
- Chan KA, Satterfield T (2020) The maturation of ecosystem services: Social and policy research expands, but whither biophysically informed valuation? *People and Nature* 2 : 1021-1060.. <https://doi.org/10.1002/pan3.10137>
- Cheng X, Van Damme S, Li L, Uyttenhove P (2019) Evaluation of cultural ecosystem services: A review of methods. *Ecosystem Services* 37 <https://doi.org/10.1016/j.ecoser.2019.100925>
- Clemente P, Calvache M, Antunes P, Santos R, Cerdeira JO, Martins MJ (2019) Combining social media photographs and species distribution models to map cultural ecosystem services: The case of a Natural Park in Portugal. *Ecological Indicators* 96: 59-68. <https://doi.org/10.1016/j.ecolind.2018.08.043>
- Comber A, Brunsdon C, Charlton M, Dong G, Harris R, Lu B, Lü Y, Murakami D, Nakaya T, Wang Y, Harris P (2022) A Route Map for Successful Applications of Geographically Weighted Regression. *Geographical Analysis* <https://doi.org/10.1111/gean.12316>
- Costanza R, de Groot R, Braat L, Kubiszewski I, Fioramonti L, Sutton P, Farber S, Grasso M (2017) Twenty years of ecosystem services: How far have we come and how far do we still need to go? *Ecosystem services* 28: 1-16. <https://doi.org/10.1016/j.ecoser.2017.09.008>
- Crawley M (2015) *Statistics: An introduction using R*. 2. Wiley [ISBN 978-1-118-94109-6]
- CSO (2011a) *Census 2011 - Population Classified by Area*. Central Statistics Office, Ireland. URL: <https://www.cso.ie/en/census/census2011reports/census2011populationclassifiedbyareaformerlyvolumeone/>.
- CSO (2011b) *Census 2011 Grid data set*. Central Statistics Office. URL: <https://www.cso.ie/en/census/census2011griddataset/>
- CSO (2011c) *Census 2011 Boundary Files*. Central Statistics Office. URL: <https://www.cso.ie/en/census/census2011boundaryfiles/>
- Cunningham J, Dolan D, Kelly D, Young C (2015) *Tourism Sectoral Overview: Galway City and County Economic and Industrial Baseline Study*. URL: <https://silo.tips/download/tourism-sectoral-overview-galway-city-and-county-economic-and-industrial-baselin>
- Czúcz B, Arany I, Potschin-Young M, Bereczki K, Kertész M, Kiss M, Aszalós R, Haines-Young R (2018) Where concepts meet the real world: A systematic review of ecosystem service indicators and their classification using CICES. *Ecosystem Services* 29: 145-157. <https://doi.org/10.1016/j.ecoser.2017.11.018>
- Daily G, Polasky S, Goldstein J, Kareiva P, Mooney H, Pejchar L, Ricketts T, Salzman J, Shallenberger R (2009) Ecosystem services in decision making: Time to deliver. *Frontiers in Ecology and the Environment* 7 (1): 21-28. <https://doi.org/10.1890/080025>
- Dasgupta P (2021) *The Economics of Biodiversity: The Dasgupta review*. URL: <https://www.gov.uk/government/publications/final-report-the-economics-of-biodiversity-the-dasgupta-review>
- Díaz S, Pascual U, Stenseke M, Martín-López B, Watson R, Molnár Z, Hill R, Chan KA, Baste I, Brauman K (2018) Assessing nature's contributions to people. *Science* 359 (6373): 270-272. <https://doi.org/10.1126/science.aap8826>

- Ding, Fan (2019) Exploring the Distribution Patterns of Flickr Photos. ISPRS International Journal of Geo-Information 8 (9). <https://doi.org/10.3390/ijgi8090418>
- EEA (2017) Copernicus Land Monitoring Service - EU-DEM. European Environment Agency. URL: www.eea.europa.eu/data-and-maps/data/copernicus-land-monitoring-service-eu-dem
- EEA (2019) Corine Land Cover 2012. European Environment Agency. Release date: 2019-6-14. URL: <https://land.copernicus.eu/pan-european/corine-land-cover/clc-2012?tab=metadata>
- EPA (2012) Water Flow Network. Environmental Protection Agency. Release date: 2012-1-01. URL: <https://gis.epa.ie/GetData/Download>
- Fáilte Ireland (2019a) Research Insights - Current Tourism performance. <https://www.failteireland.ie/Research-Insights/Current-Tourism-Performance.aspx>. Accessed on: 2022-1-21.
- Fáilte Ireland (2019b) Activities and attractions data set. Fáilte Ireland. URL: <https://www.failteireland.ie/Research-Insights/Open-data.aspx>
- Figueroa-Alfaro RW, Tang Z (2017) Evaluating the aesthetic value of cultural ecosystem services by mapping geo-tagged photographs from social media data on Panoramio and Flickr. Journal of Environmental Planning and Management 60 (2): 266-281. <https://doi.org/10.1080/09640568.2016.1151772>
- Fisher D, Wood S, Roh Y, Kim C (2019) The geographic spread and preferences of tourists revealed by user-generated information on Jeju Island, South Korea. Land 8 (5): 7. <https://doi.org/10.3390/land8050073>
- Fish R, Church A, Winter M (2016) Conceptualising cultural ecosystem services: A novel framework for research and critical engagement. Ecosystem Services 21: 208-217. <https://doi.org/10.1016/j.ecoser.2016.09.002>
- Fotheringham AS, Brunson C, Charlton M (2003) Geographically weighted regression: the analysis of spatially varying relationships. Wiley [ISBN 0470855258]
- Fotheringham AS (2020) Local modelling: one size does not fit all. Journal of Spatial Information Science 2020 (21): 83-87. <https://doi.org/10.5311/JOSIS.2020.21.726>
- Fox N, August T, Mancini F, Parks K, Eigenbrod F, Bullock J, Sutter L, Graham L (2020) "photosearcher" package in R: An accessible and reproducible method for harvesting large datasets from Flickr. SoftwareX 12 <https://doi.org/10.1016/j.softx.2020.100624>
- Fox N, Graham L, Eigenbrod F, Bullock J, Parks K (2021) Enriching social media data allows a more robust representation of cultural ecosystem services. Ecosystem Services 50 <https://doi.org/10.1016/j.ecoser.2021.101328>
- Galway County Council (2021) Draft Galway County Development Plan 2022-2028: Chapter 8 Tourism and Landscape. <https://consult.galway.ie/en/consultation/draft-galway-county-development-plan-2022-2028/chapter/chapter-8-tourism-and-landscape>. Accessed on: 2022-11-28.
- Ghermandi A, Sinclair M (2019) Passive crowdsourcing of social media in environmental research: A systematic map. Global Environmental Change 55: 36-47. <https://doi.org/10.1016/j.gloenvcha.2019.02.003>
- Giraudoux P, Antonietti J, Beale C, Lancelot R, Pleydell D, Triglia M (2022) pgirmess: Spatial analysis and data mining for field ecologists. URL: <https://cran.r-project.org/web/packages/pgirmess/index.html>
- GSI (2020) Geological Survey Ireland: The Geological Heritage of County Galway. URL: <https://www.gsi.ie/en-ie/publications/Pages/The-Geological-Heritage-of-Galway.aspx>

- Guisan A, Zimmermann N (2000) Predictive habitat distribution models in ecology. *Ecological Modelling* 135 (2-3): 147-186. [https://doi.org/10.1016/S0304-3800\(00\)00354-9](https://doi.org/10.1016/S0304-3800(00)00354-9)
- Haines-Young R, Potschin M (2018) Common international classification of ecosystem services (CICES) V5.1 and guidance on the application of the revised structure. <https://doi.org/10.3897/oneeco.3.e27108>
- Hartig F, Lohse L (2018) DHARMA: Residual diagnostics for hierarchical (multi-Level / mixed) regression models. <https://cran.r-project.org/web/packages/DHARMA/index.html>
- Hein L, Bagstad K, Obst C, Edens B, Schenau S, Castillo G, Soulard F, Brown C, Driver A, Bordt M, Steurer A, Harris R, Caparrós A (2020) Progress in natural capital accounting for ecosystems. *Science* 367 (6477): 514-515. <https://doi.org/10.1126/science.aaz8901>
- Jones L, Norton L, Austin Z, Browne AL, Donovan D, Emmett BA, Grabowski ZJ, Howard DC, Jones JPG, Kenter JO (2016) Stocks and flows of natural and human-derived capital in ecosystem services. *Land Use Policy* 52: 151-162. <https://doi.org/10.1016/j.landusepol.2015.12.014>
- Kaiser N, Ghermandi A, Feld C, Hershkovitz Y, Palt M, Stoll S (2021) Societal benefits of river restoration—Implications from social media analysis. *Ecosystem Services* 50 <https://doi.org/10.1016/j.ecoser.2021.101317>
- Kim Y, Kim C, Lee DK, Lee H, Andrada RIT (2019) Quantifying nature-based tourism in protected areas in developing countries by using social big data. *Tourism Management* 72: 249-256. <https://doi.org/10.1016/j.tourman.2018.12.005>
- Langemeyer J, Calcagni F, Baro F (2018) Mapping the intangible: Using geolocated social media data to examine landscape aesthetics. *Land Use Policy* 77: 542-552. <https://doi.org/10.1016/j.landusepol.2018.05.049>
- Lee H, Seo B, Koellner T, Lautenbach S (2019) Mapping cultural ecosystem services 2.0—Potential and shortcomings from unlabeled crowd sourced images. *Ecological Indicators* 96: 505-515. <https://doi.org/10.1016/j.ecolind.2018.08.035>
- Levin N, Kark S, Crandall D (2015) Where have all the people gone? Enhancing global conservation using night lights and social media. *Ecological Applications* 25 (8): 2153-2167. <https://doi.org/10.1890/15-0113.1>
- Long P, Nogué S, Benz D, Willis K (2020) Devising a method to remotely model and map the distribution of natural landscapes in Europe with the greatest recreational amenity value (cultural services). *Frontiers of Biogeography* 13.1 (e47737). <https://doi.org/10.21425/F5FBG47737>
- Mace G (2019) The ecology of natural capital accounting. *Oxford Review of Economic Policy* 35 (1): 54-67. <https://doi.org/10.1093/oxrep/gry023>
- Mancini F, Coghill G, Lusseau D (2018) Using social media to quantify spatial and temporal dynamics of nature-based recreational activities. *PloS One* 13 (7). <https://doi.org/10.1371/journal.pone.0200565>
- Meraj G, Singh SK, Kanga S, Islam MN (2021) Modeling on comparison of ecosystem services concepts, tools, methods and their ecological-economic implications: a review. *Modeling Earth Systems and Environment* 8 (1): 15-34. <https://doi.org/10.1007/s40808-021-01131-6>
- NPWS (2022) Protected sites in Ireland | National Parks & Wildlife Service. URL: <https://www.npws.ie/protected-sites>

- O'Donoghue C, Hynes S, Kilgarriff P, Ryan M, Tsakiridis A (2020) Assessing preferences for rural landscapes: An attribute based choice modelling approach. *Bio-based and Applied Economics* 9 (2): 171-200. <https://doi.org/10.13128/bae-8287>
- O'Hara R, Kotze J (2010) Do not log-transform count data. *Methods in Ecology and Evolution* 1: 118-122. <https://doi.org/10.1111/j.2041-210X.2010.00021.x>
- Oshan TM, Li Z, Kang W, Wolf LJ, Fotheringham AS (2019) mgwr: A Python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. *ISPRS International Journal of Geo-Information* 8 (6): 269. <https://doi.org/10.3390/ijgi8060269>
- OSi (2020) OSi open data portal. Ordnance Survey Ireland. URL: <https://data-osi.opendata.arcgis.com/>
- OSM (2021) Open street map roads ireland and northern ireland. Open Street Map. URL: <http://download.geofabrik.de/europe/>
- Oteros-Rozas E, Martín-López B, Fagerholm N, Bieling C, Plieninger T (2018) Using social media photos to explore the relation between cultural ecosystem services and landscape features across five European sites. *Ecological Indicators* 94: 74-86. <https://doi.org/10.1016/j.ecolind.2017.02.009>
- Outeiro L, Häussermann V, Viddi F, Hucke-Gaete R, Försterra G, Oyarzo H, Kosiel K, Villasante S (2015) Using ecosystem services mapping for marine spatial planning in southern Chile under scenario assessment. *Ecosystem Services* 16: 341-353. <https://doi.org/10.1016/j.ecoser.2015.03.004>
- Pettorelli N, Schulte to Bühne H, Tulloch A, Dubois G, Macinnis-Ng C, Queirós A, Keith D, Wegmann M, Schrodtt F, Stellmes M, Sonnenschein R, Geller G, Roy S, Somers B, Murray N, Bland L, Geijzendorffer I, Kerr J, Broszeit S, Leitão P, Duncan C, El Serafy G, He K, Blanchard J, Lucas R, Mairota P, Webb T, Nicholson E (2017) Satellite remote sensing of ecosystem functions: opportunities, challenges and way forward. *Remote Sensing in Ecology and Conservation* 4 (2): 71-93. <https://doi.org/10.1002/rse2.59>
- Phillips SJ (2017) A brief tutorial on Maxent. http://biodiversityinformatics.amnh.org/open_source/maxent/. Accessed on: 2022-3-20.
- Pickering C, Walden-Schreiner C, Barros A, Rossi SD (2020) Using social media images and text to examine how tourists view and value the highest mountain in Australia. *Journal of Outdoor Recreation and Tourism* 29: 100252. <https://doi.org/10.1016/j.jort.2019.100252>
- Plieninger T, Bieling C, Fagerholm N, Byg A, Hartel T, Hurley P, López-Santiago C, Nagabhatla N, Oteros-Rozas E, Raymond C, van der Horst D, Huntsinger L (2015) The role of cultural ecosystem services in landscape management and planning. *Current Opinion in Environmental Sustainability* 14: 28-33. <https://doi.org/10.1016/j.cosust.2015.02.006>
- R Core Team (2021) R: A language and environment for statistical computing. 4.1.2. R Foundation for Statistical Computing. Release date: 2021-11-01. URL: <https://www.R-project.org/>
- Richards D, Friess D (2015) A rapid indicator of cultural ecosystem service usage at a fine spatial scale: Content analysis of social media photographs. *Ecological Indicators* 53: 187-195. <https://doi.org/10.1016/j.ecolind.2015.01.034>
- Ridding L, Redhead J, Oliver T, Schmucki R, McGinlay J, Graves A, Morris J, Bradbury R, King H, Bullock J (2018) The importance of landscape characteristics for the delivery

- of cultural ecosystem services. *Journal of Environmental Management* 206: 1145-1154. <https://doi.org/10.1016/j.jenvman.2017.11.066>
- Rousset F, Ferdy J (2014) Testing environmental and genetic effects in the presence of spatial autocorrelation. *Ecography* 37 (8): 781-790. <https://doi.org/10.1111/ecog.00566>
 - Schirpke U, Meisch C, Marsoner T, Tappeiner U (2018) Revealing spatial and temporal patterns of outdoor recreation in the European Alps and their surroundings. *Ecosystem Services* 31: 336-350. <https://doi.org/10.1016/j.ecoser.2017.11.017>
 - Scholte SK, van Teeffelen AA, Verburg P (2015) Integrating socio-cultural perspectives into ecosystem service valuation: A review of concepts and methods. *Ecological Economics* 114: 67-78. <https://doi.org/10.1016/j.ecolecon.2015.03.007>
 - Sessions C, Wood S, Rabotiyagov S, Fisher D (2016) Measuring recreational visitation at U.S. National Parks with crowd-sourced photographs. *Journal of Environmental Management* 183: 703-711. <https://doi.org/10.1016/j.jenvman.2016.09.018>
 - Sharp R, Tallis HT, Ricketts T, Guerry AD, Wood SA, Chaplin-Kramer R, Nelson E, Ennaanay D, Wolny S, Olwero N (2018) InVEST User Guide—InVEST 3. 7. 0 documentation. InVEST User Guide URL: <https://invest-userguide.readthedocs.io/ / downloads/en/3.5.0/pdf/>
 - Sinclair M, Ghermandi A, Sheela A (2018) A crowdsourced valuation of recreational ecosystem services using social media data: An application to a tropical wetland in India. *Science of the Total Environment* 642: 356-365. <https://doi.org/10.1016/j.scitotenv.2018.06.056>
 - Sinclair M, Mayer M, Woltering M, Ghermandi A (2020) Using social media to estimate visitor provenance and patterns of recreation in Germany's national parks. *Journal of Environmental Management* 263 <https://doi.org/10.1016/j.jenvman.2020.110418>
 - Sonter L, Watson K, Wood S, Ricketts T (2016) Spatial and temporal dynamics and value of nature-based recreation, estimated via social media. *PloS One* 11 (9). <https://doi.org/10.1371/journal.pone.0162372>
 - Sport Ireland (2021) Sport Ireland Outdoors National Trails Register. URL: https://data.gov.ie/dataset/sport-ireland-outdoors-national-trails-register?package_type=dataset
 - Swets J (1988) Measuring the accuracy of diagnostic systems. *Science* 240 (4857): 1285-1293. <https://doi.org/10.1126/science.3287615>
 - Tenerelli P, Demšar U, Luque S (2016) Crowdsourcing indicators for cultural ecosystem services: A geographically weighted approach for mountain landscapes. *Ecological Indicators* 64: 237-248. <https://doi.org/10.1016/j.ecolind.2015.12.042>
 - Tenkanen H, Di Minin E, Heikinheimo V, Hausmann A, Herbst M, Kajala L, Toivonen T (2017) Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. *Scientific Reports* 7 (1). <https://doi.org/10.1038/s41598-017-18007-4>
 - Tennekes M (2018) tmap: Thematic Maps in R. *Journal of Statistical Software* 84: 1-39. <https://doi.org/10.18637/jss.v084.i06>
 - Tieskens K, Van Zanten B, Schulp CE, Verburg P (2018) Aesthetic appreciation of the cultural landscape through social media: An analysis of revealed preference in the Dutch river landscape. *Landscape and Urban Planning* 177: 128-137. <https://doi.org/10.1016/j.landurbplan.2018.05.002>

- United Nations, et al. (2021) System of Environmental-Economic Accounting—Ecosystem Accounting (SEEA EA). White cover publication, pre-edited text subject to official editing. . URL: <https://seea.un.org/ecosystem-accounting>
- Walden-Schreiner C, Rossi SD, Barros A, Pickering C, Leung Y (2018) Using crowd-sourced photos to assess seasonal patterns of visitor use in mountain-protected areas. *Ambio* 47 (7): 781-793. <https://doi.org/10.1007/s13280-018-1020-4>
- Wood S, Guerry A, Silver J, Lacayo M (2013) Using social media to quantify nature-based tourism and recreation. *Scientific Reports* 3 (1): 1-7. <https://doi.org/10.1038/srep02976>
- Wood S, Winder S, Lia E, White E, Crowley CL, Milnor A (2020) Next-generation visitation models using social media to estimate recreation on public lands. *Scientific Reports* 10 (1): 1-12. <https://doi.org/10.1038/s41598-020-70829-x>
- Yoshimura N, Hiura T (2017) Demand and supply of cultural ecosystem services: Use of geotagged photos to map the aesthetic value of landscapes in Hokkaido. *Ecosystem Services* 24: 68-78. <https://doi.org/10.1016/j.ecoser.2017.02.009>
- Zhang H, Huang R, Zhang Y, Buhalis D (2020) Cultural ecosystem services evaluation using geolocated social media data: a review. *Tourism Geographies* 1-23. <https://doi.org/10.1080/14616688.2020.1801828>
- Zhao D, Xiao M, Huang C, Liang Y, An Z (2021) Landscape dynamics improved recreation service of the Three Gorges Reservoir Area, China. *International Journal of Environmental Research and Public Health* 18 (16). <https://doi.org/10.3390/ijerph18168356>
- Zielstra D, Hochmair H (2013) Positional accuracy analysis of Flickr and Panoramio images for selected world regions. *Journal of Spatial Science* 58 (2): 251-273. <https://doi.org/10.1080/14498596.2013.801331>

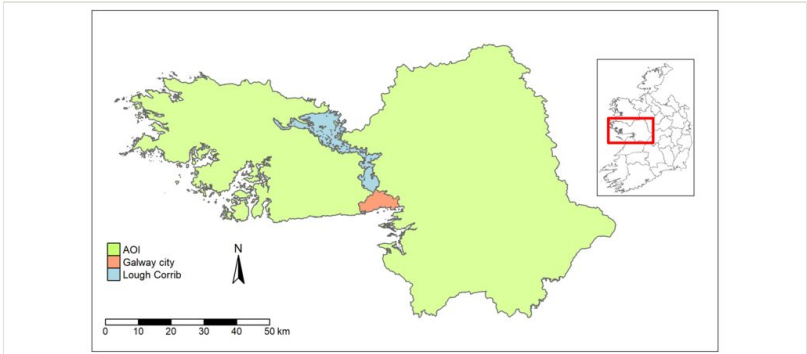


Figure 1.

Case-study area of interest, County Galway and its location on the west coast of Ireland. Areas excluded from analysis (Galway City and Lough Corrib) are highlighted.

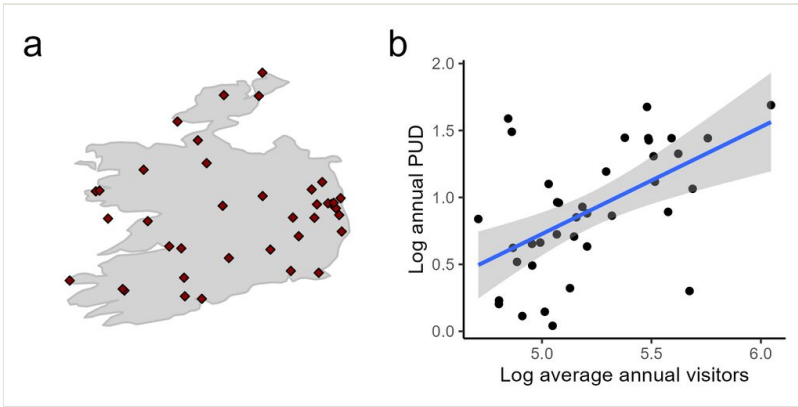


Figure 2.

a) Location of 38 tourist sites, b) scatterplot showing significant positive correlation between official visitor counts and photo-user-days visitation metric (log scale).

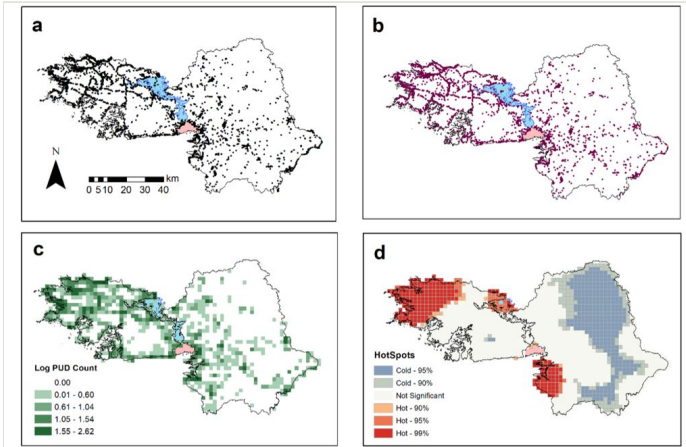


Figure 3.

a) GPS points of 25,170 photographs retrieved, b) GPS points of 6,762 photo-user-days (PUD), c) aggregation of PUD points into 2 km grid resolution (log scale), d) Getis-Ord G_i^* statistic hot and cold spots.

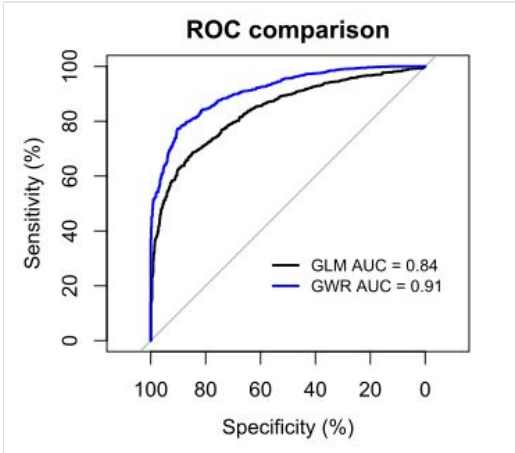


Figure 4. ROC plots for global logistic regression (black) and local GWR logistic regression (blue), based on presence of PUD at 2 km² pixel size.

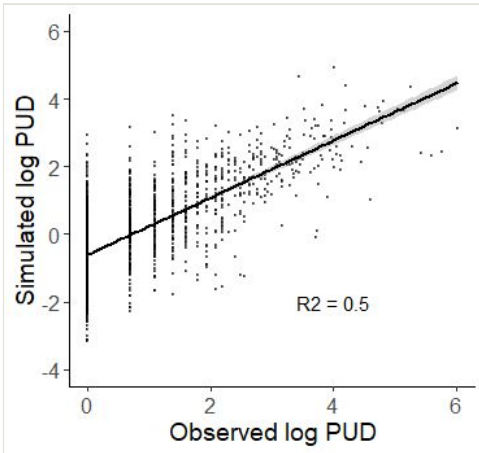


Figure 5.

Observed photo-user-day (PUD) counts plotted against modelled PUD, based on SAM output (log scale), with an R^2 value of 0.5.

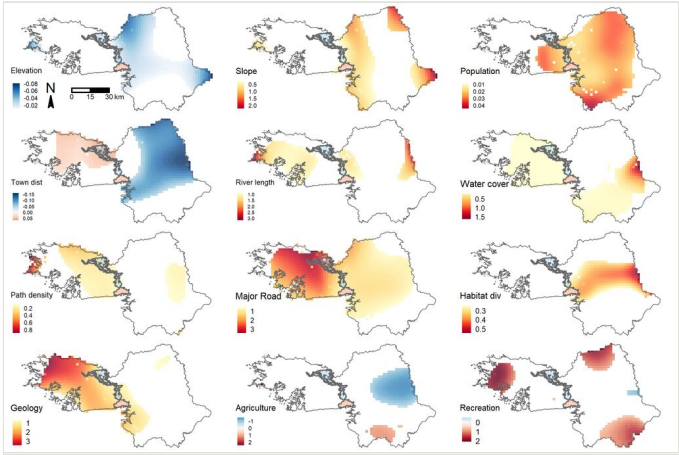


Figure 6.

GWR coefficient surfaces for 12 environmental predictor variables used to model photo-user-day occurrence. Only significant coefficients at the 95% level are shaded. Warm colours indicate a positive coefficient, while cold colours indicate a negative coefficient.

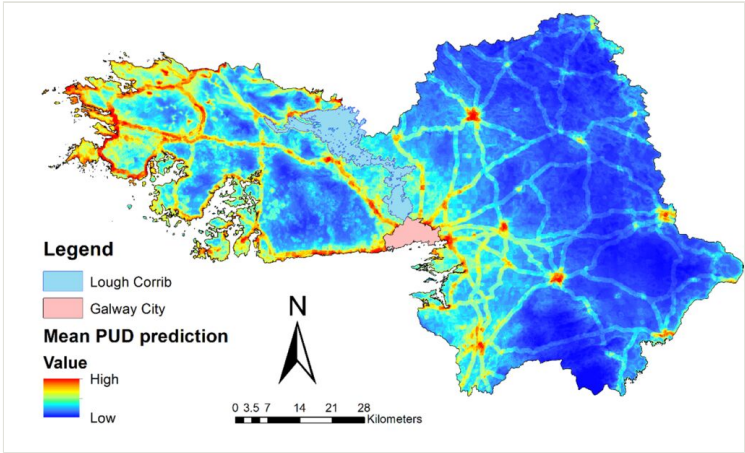


Figure 7.
Average suitability for Flickr-derived photo-user-day occurrence (100 replicates) at 100 m resolution generated by the maximum entropy model.

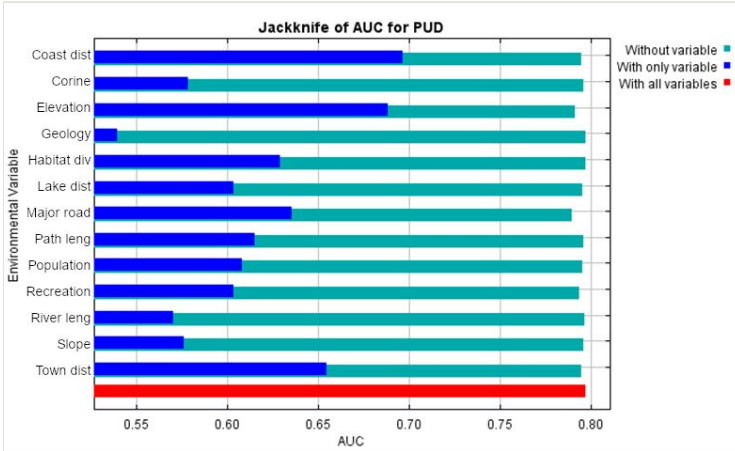


Figure 8.

Jackknife tests for environmental predictor contributions to MaxEnt prediction models. Dark blue shows models with the variable in isolation, light blue shows the full model minus the variable and red shows maximal model for comparison.

Table 1.

List of variables and indicators capturing human, natural and built capital for each 2 km² grid cell and their sources.

Predictor	Capital	Indicator	Data source and format
Elevation	Natural	Average elevation	Copernicus remote sensing DEM. 25 m raster data (EEA 2017).
Slope	Natural	Average slope	Copernicus remote sensing DEM. 25 m raster data (EEA 2017).
Rivers	Natural	Length of river	Environmental Protection Agency. Vector data of river bodies (EPA 2012).
Freshwater cover	Natural	Area of freshwater (lakes, ponds, rivers)	Water framework directive. Vector data of lake segments (EPA 2012).
Coastline	Natural	Distance to coastline	Ordnance Survey Ireland. Land mask of Ireland 250 k vector file (OSi 2020).
Habitat diversity	Natural	Number of CORINE classification types	CORINE land-cover map 2012. Raster of land-cover types at 100 m (EEA 2019).
Land cover	Natural	Area of four land-cover types (agriculture, wetlands, urban and forestry and natural areas).	CORINE land-cover map 2012. Raster of land-cover types at 100 m (EEA 2019).
Land-cover diversity	Natural	Shannon's diversity index	CORINE land-cover map 2012. Raster of land-cover types at 100 m (EEA 2019).
Geological heritage	Natural	Presence of designated geological heritage	Geological Survey Ireland. Vector data of recommended heritage sites (GSI 2020).
Protected status	Natural	Area covered under protected status	National Parks and Wildlife Service. Vector of protected areas. (NPWS 2022).
Town distance	Built	Distance to nearest town	CSO Census 2011. Boundaries of designated towns and cities (vector) (CSO 2011c).
Population	Human	Population density	CSO Census 2011. Grid of population density at 1 km ² (CSO 2011b).
Major Roads	Built	Presence of a major road	Ordnance Survey Ireland. National road network vector file. (OSi 2020).
Path density	Built	Density of path length	Open street map roads database (vector) (OSM 2021).
Amenity and recreation sites	Built	Presence of a recreational site (examples include bike rental, sports trails, boating, angling, golf courses)	Faillte Ireland activity database (point data) (Fáilte Ireland 2019b), Sport Ireland trail map (vector) (Sport Ireland 2021).

Table 2.

Outline of four models and description of their basic structure, data requirements and output. PUD refers to photo-user-days metric.

Model family	Model structure	Description	Response (GPS-tagged content)	Predictors (natural, human and built capital)	Output
Global regression (1)	Non-spatial global logistic regression (a)	Generalised linear model (binomial family and logit link).	Binary. PUD presence	Environmental indicators at 2 km ² resolution	Average relationships between environmental attributes and likelihood of PUD occurrence.
	Global logistic spatially autocorrelated mixed-model (SAM) (b)	Generalised linear mixed-model (binomial family and logit link). Environmental predictor fixed effects and spatial random effect (long + lat).	Binary. PUD presence	Environmental indicators at 2 km ² resolution	Average relationships between environmental attributes and likelihood of PUD occurrence accounting for spatial heterogeneity.
	Global poisson spatially autocorrelated mixed-model (SAM) (c)	Generalised linear mixed-model (poisson family and log link). Environmental predictor fixed effects and spatial random effect (long + lat).	Count. PUD total	Environmental indicators at 2 km ² resolution	Average relationship between environmental attributes and PUD counts accounting for spatial heterogeneity.
Local regression (2)	Logistic geographic weighted regression (GWR)	A series of logistic regressions (binomial family) computed across the landscape weighting data, based on proximity to regression point.	Binary. PUD presence	Environmental indicators at 2 km ² resolution	Local relationship between environmental attributes and PUD occurrence that vary spatially in magnitude, direction and significance.

Maximum Entropy (3)	Maximum Entropy (MaxEnt)	Predictive model that uses machine-learning and a presence-only approach, based on observed occurrence to predict areas of high suitability.	Point coordinates. PUD occurrence	Environmental variables as 100 m rasters	100 m resolution map of modelled suitability of visitation occurrence and variable contributions to model performance.
InVEST (4)	InVEST recreation model	Self-contained model that queries an archived data set and calculates spatial statistics on user-supplied spatial data to compute an OLS regression.	Log Count. Annual PUD	Raw spatial data (vector or raster) analysed using pre-set options	Average relationship between environmental attributes and log annual PUD.

Table 3.

Results of global regression models. PUD refers to the photo-user-days metric. SAM denotes a spatially-autocorrected mixed-model using a random spatial effect. Significance levels (Sig) are denoted as *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Predictor	Global logistic non-spatial (a) (PUD Presence)			Global logistic SAM (b) (PUD Presence)			Global poisson SAM (c) (PUD Count)		
	Coefficient	Standard error	Sig	Coefficient	Standard error	Sig	Coefficient	Standard error	Sig
(Intercept)	-1.303	-0.291	***	-1.767	0.582	***	-1.109	0.243	***
Elevation	-0.016	-0.002	***	-0.018	0.003	***	-0.009	0.002	***
Slope	0.278	-0.042	***	0.309	0.072	***	0.129	0.025	***
Coast distance	-0.029	0.005	***	-0.047	0.011	***	-0.037	0.006	***
River length	0.785	-0.18	***	0.943	0.353	***	0.373	0.159	***
Water cover	0.022	-0.004	***	0.031	0.007	***	0.012	0.002	***
Path length	0.082	-0.025	***	0.127	0.034	***	0.103	0.011	***
Major road	0.674	-0.138	***	1.131	0.189	***	0.767	0.08	***
Population	0.019	0.004	***	0.022	0.001	***	0	0	
Recreation site	0.608	-0.152	***	0.666	0.229	***	0.507	0.086	***
Geological heritage	0.555	-0.161	***	0.742	0.245	***	0.29	0.089	***
Habitat diversity	0.118	-0.053	**	0.190	0.079	**	0.152	0.031	***
N	1642			1642			1642		
Log likelihood	-803			-748			-3076		
AIC	1629			1518			6178		
Spatial autocorrelation of residuals detected:	Yes			No			No		

Table 4.

InVEST regression model output (ordinary least squares regression and log annual PUD response). Significance levels (Sig) are denoted as *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

	Coefficient	Standard error	t value	Sig
Intercept	0.112	0.108	1.04	
Elevation	-0.00216	0.000340	-6.35	***
Slope	0.0450	0.00599	7.51	***
Population	0.00115	0.000260	4.41	***
Habitat diversity	0.141	0.0337	4.17	***
Agriculture	-0.00462	0.000958	-4.83	***
Forest and Natural Area	-0.00538	0.00109	-4.92	***
Wetlands	-0.00331	0.000932	-3.55	***
Coast distance	-0.0000356	0.000001	-3.21	**
Town distance	0.0000871	0.000001	8.72	***
Path Length	0.0000516	0.000004	13.17	***
Recreation distance	-0.0000158	0.000004	-4.09	***
Degrees of freedom	1586			
Adjusted R2	0.487			

Table 5.

General findings from deploying four spatial model types to the same area of interest for providing information for model selection.

	Global regression (1)	Local regression (2)	Maximum entropy (3)	InVEST recreation (4)
Are of interest and resolution	<ul style="list-style-type: none"> · Determined by landscape heterogeneity and phenomenon of interest · Data coverage determines resolution 	<ul style="list-style-type: none"> · Variation at local scale required · Deployed where spatially varying relationships hypothesised · Data coverage determines resolution 	<ul style="list-style-type: none"> · Useful in data-scarce regions with low observations · Permits fine resolution 	<ul style="list-style-type: none"> · Resolution should ensure 50% cells contain > 0 annual PUD · Limited use in data-scarce regions
Social media indicator	<ul style="list-style-type: none"> · User-determined (automated or manual collection) 	<ul style="list-style-type: none"> · User-determined (automated or manual collection) 	<ul style="list-style-type: none"> · User-determined (automated or manual collection) 	<ul style="list-style-type: none"> · Flickr only 2005-2017 · Stable database
Response variable	<ul style="list-style-type: none"> · Pre-processing and filtering possible · GPS inaccuracy can be buffered · Response variable user defined such as occurrence, rate or count variables 	<ul style="list-style-type: none"> · Pre-processing and filtering possible · GPS inaccuracy can be buffered · Response variable user defined, such as occurrence, rate or count variables 	<ul style="list-style-type: none"> · Pre-processing and filtering possible · Point data required (GPS coordinates of occurrence) 	<ul style="list-style-type: none"> · PUD variable fixed and calculated automatically · GPS assumed accurate · Pre-processing and filtering not possible
Environmental predictor variables	<ul style="list-style-type: none"> · User-driven indicator calculation · May require spatial statistics and GIS · Standard procedure for variable inspection (outliers, collinearity, skewness) 	<ul style="list-style-type: none"> · User-driven indicator calculation · May require spatial statistics and GIS · Variable inspection at global and local scale (outliers, collinearity, skewness) · Proximity variables may be unsuitable 	<ul style="list-style-type: none"> · User-driven indicator calculation as raster files of identical extent · May require GIS to prepare · Limited variable inspection and diagnostics 	<ul style="list-style-type: none"> · User supplied spatial data files · Indicators calculated by pre-set spatial statistics within model run · Output returns variables calculated · Edits require re-running entire model
Computation	<ul style="list-style-type: none"> · Standard regression tools in any statistical software, for example, R · Model assumptions require manual checks (dispersion, outliers, normality, spatial autocorrelation) · If spatial autocorrelation detected, consider SAM or GWR models 	<ul style="list-style-type: none"> · Specialised tools available, for example, MGWR software · Emerging packages (R and Python) · Bandwidth and kernel set by user · Prior to run, standard checks required (dispersion, outliers, normality) · Local model checks also required, for example, local VIF, Cook's distance 	<ul style="list-style-type: none"> · Standalone model software open- access and freely available · Compatible R packages also available · Default output includes model gain, AUC and variable response curves · Optional jackknife analysis included · Optional replication and data partition 	<ul style="list-style-type: none"> · Standalone model software, open-access and freely available · Some additional tools in Python API · Automatically runs OLS regression · No available option for model inspection beyond default regression table output

<p>Accessibility and transferability</p>	<ul style="list-style-type: none"> · Basic knowledge of statistics and software of choice, for example, R · Flexible and customisable model structure that can be altered by the user with relative ease along a spectrum of complexity · Available to any user familiar with basic regression techniques 	<ul style="list-style-type: none"> · Advanced spatial modelling approach with some customisation (bandwidth, optimiser criteria, GLM family) · Requires some specialist knowledge for preparation and interpretation · Advanced model runs may take time, especially Monte Carlo simulation · Evolving field of research with some limitations, for example, multiscale bandwidth models 	<ul style="list-style-type: none"> · Available online supports · Software and machine learning permits complex modelling that otherwise may be unavailable · Advanced model runs may take time especially when replication is used · Interpretation requires expertise with statistics · Optional scenario modelling included 	<ul style="list-style-type: none"> · No coding knowledge required · User-friendly, stand-alone interface and outputs · Dedicated online support tools and tutorials available · Optional scenario modelling · Uncertain future as archive becomes outdated
<p>Summary</p>	<p>Widely applicable and customisable, but requires consideration when selecting resolution, scale, indicators and response variable calculation. May be limited depending by data availability and presence of local relationships</p>	<p>Useful for investigating spatially varying relationships at the landscape scale, but should only be used when justified. Requires knowledge of spatial statistics and mapping tools. Some limitations due to evolving research field and development of new statistical approaches</p>	<p>Applicable at high resolution and in data-scarce regions to predict areas of high suitability, but requires exact GPS coordinates and some degree of advanced statistical interpretation, computational resources and preparation of raster files</p>	<p>Useful for users to detect and visualise general trends with limited resources and expertise in regions that are sufficiently data-rich, where the use of 2005-2017 Flickr data is deemed sufficient</p>

Supplementary materials

Suppl. material 1: Sites used for validation

Authors: Andrew Neill

Data type: Table

Brief description: List of 38 sites across Ireland with official visitor statistics used to validate the PUD social media indicator.

[Download file](#) (12.18 kb)

Suppl. material 2: InVEST model configuration

Authors: Andrew Neill

Data type: Table

Brief description: Table showing InVEST parameters used for environmental indicator calculation within the model run. For full details of InVEST model configuration, see Sharp et al. (2018).

[Download file](#) (507.00 bytes)

Suppl. material 3: Sampled PUD occurrence

Authors: Andrew Neill

Data type: Occurrences

Brief description: GPS location of 6,672 PUD occurrences at the 2 km grid cell size, based on data retrieved from Flickr API query.

[Download file](#) (339.20 kb)

Suppl. material 4: Moran's I Correlograms

Authors: Andrew Neill

Data type: Model diagnostics

Brief description: Moran's I correlograms used to check for spatial autocorrelation of residuals of three models used: non-spatial logistic GLM, SAMs and GWR.

[Download file](#) (1.02 MB)

Suppl. material 5: MaxEnt Supplementary Info

Authors: Andrew Neill

Data type: Variable response curves

Brief description: MaxEnt output files showing variable response and ROC plot, based on 100 bootstrap replicates using a 75:25 training:test partition.

[Download file](#) (900.54 kb)