Outdoor Recreation Valuation (ORVal) Data Set Construction

Version 1.0

June 2016

Land, Environment, Economics and Policy (LEEP) Institute

College of Social Sciences and International Studies

University of Exeter

Authors:

Brett Day, brett.day@exeter.ac.uk

Greg Smith, g.smith4@exeter.ac.uk

Contents

Contents

1. Introduction

Underpinnning LEEP's Outdoor Recreation Valuation Tool (ORVal) is a detailed spatial dataset describing the location and characteristics of accessible greenspace across England. This document describes how that dataset has been compiled through the combination and manipulation of a large number of primary data sources.

The data have been organised using the open source software PostGIS, the spatial extension to the relational database PostgreSQL, and the manipulations described in this document carried out using scripts written in spatially enabled SQL and procedures and functions written in PLpgsql. The scripts are extensive, allowing for complete reconstruction of the dataset from the raw data.

2. Data Layers

2.1 Recreation Sites

The final dataset comprises four complimentary spatial data layers. The first layer defines the locations and extents of the recreation greenspaces. We describe these as recreation sites. In our dataset, sites come in three basic forms;

- **parks** which consist of areas of accessible greenspace within well-defined boundaries over which visitors usually have freedom to wander at will,
- **paths** which consist of accessible, walkable routes that pass through the landscape, often traversing a variety of different greenspaces and tending to restrict visitors to defined routes of passage.
- **beaches**.

The distinction between paths and parks is somewhat blurred. Nature reserves, for example, which have well-defined extents such that we classify them as parks often restrict visitors to a limited paths network. In contrast, large areas of England are designated as open country under the Countryside and Rights of Way Act (2000) (CRoW Act) affording visitors the right to roam across their full extent. In our dataset, it makes more sense to capture those recreational areas through documenting the extensive network of paths with which they are served.

As we explain in more detail subsequently, paths form connected networks. Some of those networks may be very extensive, for example, covering the whole of the South Downs from Eastbourne to Winchester. Rather than treating this extensive network as one recreation site, in our dataset we identify the locations at which visitors access the paths network and define those **path access points** as recreation sites. The assumption here is that the nature of the greenspace that individuals experience from traversing some particular path network will be determined by the point at which they access the network.

2.2 Land cover and use

The second spatial data layer defines the nature of the greenspace provided by a recreation site, we describe this as the site's 'land cover and use' or more succinctly **land cover**.

As detailed below, the process of describing land cover begins by defining an array of points over each recreation site positioned on the vertices of a 25 metre square grid. Each grid point, therefore represents an area of 25 square metres or 0.0625 ha. We call this the **land cover grid**.

In the case of paths, we expand each path by 25 metres to the left and right and then create a land cover grid over that buffered area. Accordingly, the land cover for a path is described by the nature of the greenspace along the 50 metre wide strip through which it passes.

Since an integral part of the recreation experience offered by a park or path may arise from the fact that they border a river, lake or sea, we extend the land cover grid to incorporate such bordering water features.

The land cover and use for each point on the land cover grid is identified through a series of overlay operations that draw on a wide variety of digital map layers identifying the locations of, amongst others, various forms of natural habitat, grass, sports pitches, allotments, agriculture, playgrounds, car parks, buildings etc.

We summarise the land cover of a park by summing the number of grid points in that park attributed to the various types of land cover and multiplying by 0.0625ha. As explained in more details subsequently, things are done slightly differently for access points on the paths network. For each of these path sites, we calculate a distance-weighted sum of the grid points where the distance weights decline from a value of one for grid points on the path adjacent to the access point to a value of zero at distances of 10 km or more. This approach to summarising the land cover of paths is an attempt to approximate the likelihood of interaction with different landscapes for visitors accessing the network from different locations.

2.3 Designations

The third spatial data layer developed for ORVal is one that identifies areas of parks with special designations, for example, areas designated as National Trail, Heritage Coast, Historic Parks or SSSI. Again we make use of the land cover grid, overlaying a large variety of digital maps describing the locations of designated areas and summing (or, in the case of paths, distance-weighted summing) the points in each site's grid with a particular designation.

Notice that while each point in a site's grid must be attributed exclusively to a particular land cover, it is quite possible that the same grid point may be attributed to multiple designations. For example, a point can be both in a National Park and be on Natura 2000 designated land. As a result, summing up areas of different land cover for a park will return that park's total area, the same will not be true when summing areas of different designations for paths.

2.4 Points of Interest

The final spatial data layer developed for ORVal is one that identifies particular natural or man-made structures in sites. That data layer is drawn from the Ordnance Survey's Points of Interest (OS-POI) database and locates four different type of feature:

- Historic structures (e.g. castles, country houses etc.)
- Archaeological sites (e.g. standing stones, burial mounds etc.)
- Scenic features (e.g. waterfalls, caves etc.)

- Viewpoints

The presence in a recreation site of each of these four different types is established by overlaying the recreation sites layer with the OS-POI data. For a park, that presence is a binary variable; the park either contains a point of interest of that description or does not. For paths, however, we again use distance weighting such that points of interest near to a path access point receive a higher score than those further away.

3. Data Sources

The development of the ORVal data layers has drawn on a large number of data sources. We describe some of the key sources in this section.

3.1 OpenStreetMap

One major data source exploited in the creation of ORVal's recreation dataset is OpenStreetMap (https://www.openstreetmap.org/). OpenStreetMap provides a searchable and editable digital map of the world that has been created through the efforts of millions of registered users adding more, and more detailed, digital descriptions of the world's geographic features.

Users, identify geographic features as points, lines or polygons (known respectively as nodes, ways and closed ways in the terminology of OpenStreetMap) and describe the nature of that feature through the use of an extensive system of tagging. So, for example, a user might outline a municipal park as a polygon and tag that polygon as landuse=park. Unfortunately, another user might have chosen the tag leisure=park, such that OpenStreetMap data, while incredibly extensive, is not entirely consistent. OpenStreetMap provides a searchable database that allows users to explore commonly applied tags (http://taginfo.openstreetmap.org.uk/) so as to quickly find combinations of tags that identify relevant features on the map.

For the purposes of this research, a complete download of map features for Great Britain was taken in early 2016 using the service provided by GeoFabrik (http://download.geofabrik.de/). The command line utility Osmosis was used to extract features from that data source that had been tagged with relevant labels and then uploaded to the PostGIS database using the command line utility osm2pgsql.We shall describe the extracts and tags used to identify them subsequently.

3.2 UK Government Open Data

A second key source of data was the UK government's open data website (https://open.gov.uk) which provides access to a large number of geographic datasets recording the locations of environmental features across England and the UK. That resource provides details of the extents of a large variety of areas under different environmental designations, including national parks, areas of outstanding natural beauty and Ramsar sites. It also provides details of the locations of numerous park features such as local and national nature reserves as well as more local parks such as those formed under the millennium and doorstep green and areas over which visitors have rights of access under the CroW Act. Finally, the government open data website affords access to digital resources detailing the locations of areas under different habitats, particularly those recorded under the extensive Priority Habitats Inventory (England).

3.3 Land Cover Map 2007 (LCM 2007)

Another key resource for establishing the land cover of recreation sites is provided by the CEH product Land Cover Map 2007 (LCM 2007) (http://www.ceh.ac.uk/services/land-cover-map-2007). Derived from satellite images and digital cartography, LCM 2007 provides land cover data for the whole of the UK. The data used in this research was the raster version of LCM 2007 that gives the most likely broad habitat in every cell of a 25 metre square grid across the UK.

3.4 Ordnance Survey

A number of important data sets were sourced from the Ordnance Survey (OS). Boundaries of built-up areas were taken from the 2011 OS census product identifying areas dominated by 'bricks and mortar' over a 50 metre square grid across the UK. In addition, information on points of interest were taken from the OS Points of Interest data set. A custom extract of that dataset was provided by Edina (http://digimap.edina.ac.uk/), providing a data layer identifying the features listed under the following codes:

3.5 Woods for People

Accessible woodland areas were identified by reference to the 2011 version of the 'Woods for People' dataset compiled by the Woodlands Trust with support from the Forestry Commission. The data does not include woodlands where the only access is on public rights of way, such that features drawn from this source fall into ORVal's parks rather than paths category.

4. Parks

The ORVal parks data layer is constructed through the progressive overlaying of digital maps that identify different forms of park-like greenspace (see section 2.1). As described subsequently, each different general form of greenspace is designated as a particular _type_ of park.

The process of overlaying digital maps works well when the new features being added to the ORVal parks data layer are spatially distinct from those features already identified. Things get more difficult when the new features adjoin or overlap existing features. In that case, decisions have to be made regarding whether adjoining or overlapping features should be merged into existing features, or identified as a new independent park.

Notice that the order in which new feature layers are added to the dataset is chosen so as to progress from more formally-defined park features to less formally-defined features. As a result, when merging new features into existing features, we generally assume that the new features extend already identified parks. We will come back to describe the logic through which new features are added to the map in more detail subsequently. First, however, we describe the different digital datasets used to build the parks data set presenting them in the order in which they are added layered onto the map.

4.1 Country Parks

Country Parks are described by Natural England as, "public green spaces often at the edge of urban areas which provide places to enjoy the outdoors and experience nature in an informal semi-rural park setting. Country Parks normally have some facilities such as a car park, toilets, perhaps a cafe or

kiosk, paths and trails, and visitor information. There is not necessarily public right of access, although most are publicly accessible; some charge entry others do not. Most are owned and managed by Local Authorities. Many Country Parks were designated in the 1970s by the then Countryside Commission, under the Countryside Act 1968. More recently Country Parks have been created under a less formal arrangement and Natural England is working with partners to encourage a renaissance and accreditation of parks which meet certain criteria."

Boundaries for the 415 Country Parks in England published as a GIS data layer by Natural England, was downloaded from the UK government's open data website (data.gov.uk) and were the first set of greenspace features to be added to the ORVal parks data layer. Those features were given the type designation 'country park'.

4.2 National Nature Reserves

The second set of features added to the parks data layer were national nature reserves. These are described by Natural England as containing "examples of some of the most important natural and semi-natural terrestrial and coastal ecosystems in Great Britain. They are managed to conserve their habitats or to provide special opportunities for scientific study of the habitats communities and species represented within them. In addition they may be managed to provide public recreation that is compatible with their natural heritage interests. National Nature Reserves are declared by the statutory country conservation agencies under the National Parks and Access to the Countryside Act 1949 and the Wildlife and Countryside Act 1981."

A digital map of boundaries for the 858 National Nature Reserves in England was downloaded from the UK government's open data website and added as the second layer of park features added to the ORVal parks data layer. Those features were given the designation 'nature'.

4.3 Local Nature Reserves

Local Nature Reserves are described by Natural England as being "a statutory designation made under Section 21 of the National Parks and Access to the Countryside Act 1949. Local Nature Reserves are for people and wildlife. They are places with wildlife or geological features that are of special interest locally. They offer peopled opportunities to study or learn about nature os simply to enjoy it."

Again, digital boundaries for the 2,379 Local Nature Reserves in England were sourced from the government's open data website. Those features were given the type designation 'nature'.

4.4 OSM Parks

A crucial source of data on the location of park-like greenspace was provided by OSM. OSM features (ways and relations) in which the keys *Landuse* or *Leisure* had been tagged as;

- 'park'
- 'recreation_ground'
- 'village_green'
- 'common'

were extracted and uploaded to the PostGIS database.

The data were reduced by removing all features for which access was private or otherwise restricted. In particular, features were only retained if the *Access* key was null or tagged as;

- 'public'
- 'yes'
- 'permissive'
- 'destination'

An issue with OSM data is that it is neither exhaustive nor tagged in an entirely consistent manner. Undoubtedly, numerous greenspaces that should be identified as parks have either not been mapped by OSM users or have not been tagged in a way that allows them to be identified as a park. Accordingly, while OSM is a very rich source of information on locations that people in England consider to be park-like greenspaces, we make no claim that it has allowed us to identify all such locations. In processing the data we made attempts to deal with some of the inconsistencies that arise due to users' choices of tagging.

Figure 1: Processing of OSM Parks Layer
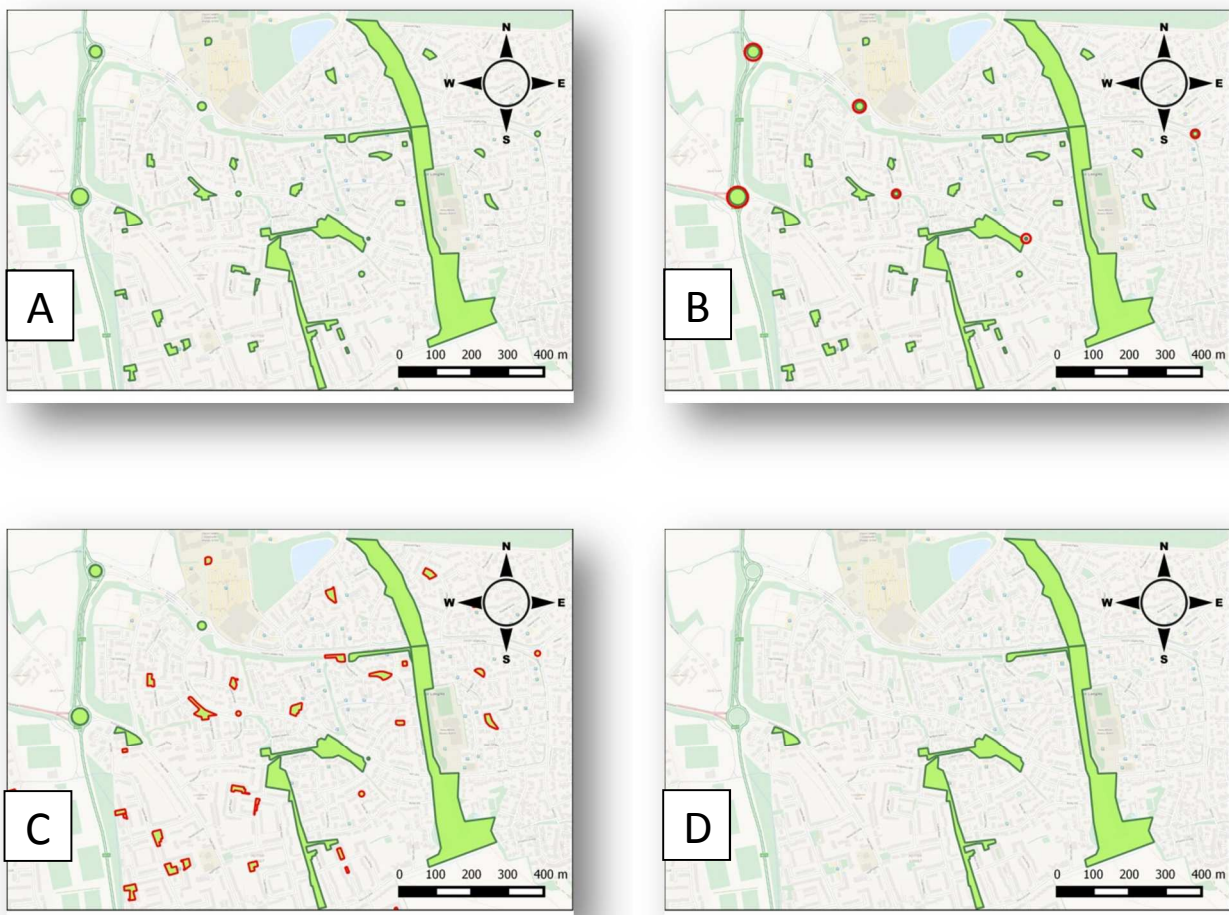
Some of those issues can be seen in Figure 1 which shows the park features extracted from OSM for an area of east Harlow in Essex. In this area an enthusiastic OSM user has geo-coded a diversity of grassed areas (panel A) including those in the middle of roundabouts (Panel B) and a variety of small grass verges and amenity spaces (Panel C) and then tagged these as parks. To address the wrong identification of these greenspaces as parks, information on the spatial extent of roundabouts was

taken from OSM and used to identify and then delete roundabouts wrongly-tagged as parks. Likewise we remove from the OSM Parks layer features that are smaller than 0.4 ha, an area about two-thirds the size of a football pitch (Panel D).

Again drawing data from OSM, features that had been labelled as parks that were found to be school grounds or in hospitals or supermarkets were also deleted from the OSM parks layer. Finally, sports facilities likely not to be open access greenspaces were identified as those in which the feature name contained any of the following;

- 'FC'
- 'sports club'
- 'sports centre'
- 'leisure centre'
- 'club'

Again these were removed before features from the OSM parks layer were added to the ORVal parks data layer each feature being given the type designation 'park'.

4.5 OSM Nature Reserves

OSM features in which the keys *Landuse* or *Leisure* had been tagged as 'nature_reserve' were extracted and uploaded to the PostGIS database. Features in that data layer for which access was private or otherwise restricted were deleted. OSM nature reserve features were added to the ORVal parks data layer with the type designation 'nature'.

4.6 OSM Public Gardens

OSM features in which the keys *Landuse*, *Leisure* or *Amenity* had been tagged as 'garden' were extracted and uploaded to the PostGIS database. Features in that data layer for which access was specifically labelled as 'public', 'yes', 'permissive' or 'destination' were assumed to be public gardens and were added to the ORVal parks data layer with the type designation 'park'.

4.7 OSM Golf Courses

OSM features in which the *Leisure* key had been tagged as 'golf_course' were extracted and uploaded to the PostGIS database. Features where the feature name contained any of the following;

- 'nets'
- 'driving range'
- 'putting'
- 'crazy'
- 'adventure'
- 'mini'

were taken to be leisure attractions that did not involve interaction with greenspace and were deleted from data layer. The remaining features were added to the ORVal parks data layer with the type designation 'golf'.

4.8 Doorstep and Millennium Greens

Doorstep Greens and Millennium Greens are community-managed spaces which have received Lottery funding to create, improve or restore areas of green space close to population centres. According to Natural England they were to be "breathing spaces - places for relaxation, play and enjoyment of nature and pleasant surroundings. They could be small or large, and in urban or rural locations". The Doorstep Green initiative ran from 2001 to 2006 and was the successor to Millennium Greens initiative. Natural England state that the "initiative is aimed at targeting communities who experience disadvantage and where regeneration of the local environment and outdoor recreation provision is sorely needed."

Digital boundaries for the 184 Doorstep and Millennium Greens in England were sourced from Natural England via the government's open data website and added to the ORVal parks data layer with the type designation 'park'.

4.9 Forestry Commission Recreation Areas

Data on recreation areas run by the Forestry Commission (FC) were constructed from the FC's digital map of 'National Forest Estate England Recreation Routes' which plots out walking and bike trails on FC land. Panel A of Figure 2 depicts the paths network that exists in Thetford Forest in East Anglia and is representative of other networks of recreation trails maintained by the FC. Observe that the trail form discrete networks that prescribe well-defined areas of woodland that can be interpreted as park-like greenspaces. To construct those parks, we first use a recursive algorithm to walk through the trails network identifying which trails intersect with which other trails and building up the set of connected trails that serve each discrete recreation area. The recreation areas themselves were constructed as the concave hull of each of those separate connected trails networks (Panel B of Figure 2).

The FC Recreation Areas were added the ORVal parks data layer with the type designation 'woods'.

Figure 2: Processing Forestry Commission Routes Data Layer
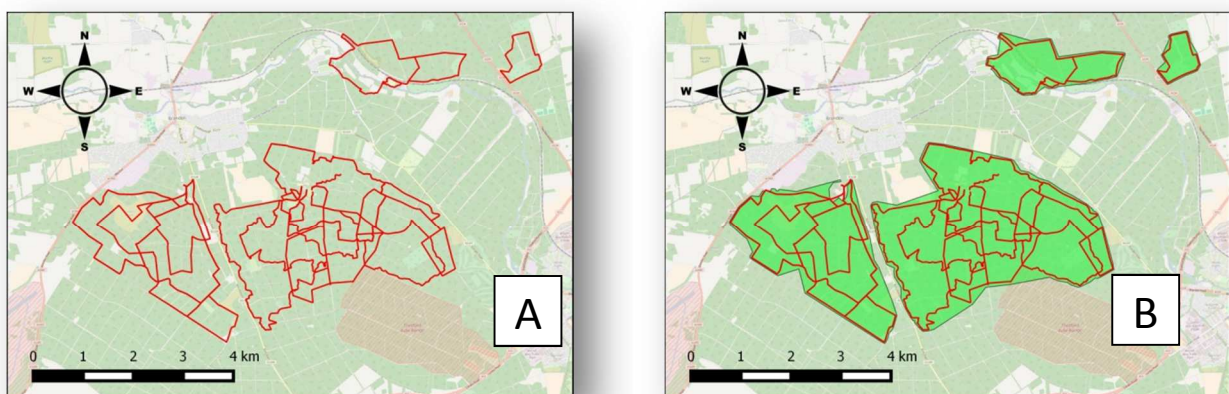
4.10   Woods for People

As described in Section 3.5 the Woodland Trust's 'Woods for People' dataset provides information on the locations of accessible woodland across England. One issue with the Woods for People dataset is that it identifies some areas of woodland that are too extensive to be considered as a single park. Panel A of Figure 3 shows one such feature; Epping Forest in Essex.

Before adding to the ORVal parks data layer, these large woodland features were subdivided using the procedure illustrated in Figure 3. First, observe Panel B in which an overlay of trunk roads has been superimposed on the woods for people data layer. We assume that these major roads form a significant obstacle to visitors, such that the large woodland is actually divided into a series of smaller woodland recreation sites whose boundaries are defined by those roads. To progress, a bounding box is drawn around each area of woodland that is divided by a major road and a series of tessellating polygons created whose boundaries are defined by the edges of that box and the network of roads (Panel C). Finally, the original woodland feature is cut using the edges of those polygons to generate the set of smaller woodland parks (Panel D).

At the other extreme, the Woods for People dataset also identifies very small patches of woodland. To mirror our treatment of small park features identified from OSM, we remove all woodlands that are smaller than 0.4ha.

The processed Woods for People features were added to the ORVal parks data layer with the type designation 'woods'.

Figure 3: Processing the Woods for People Data Layer
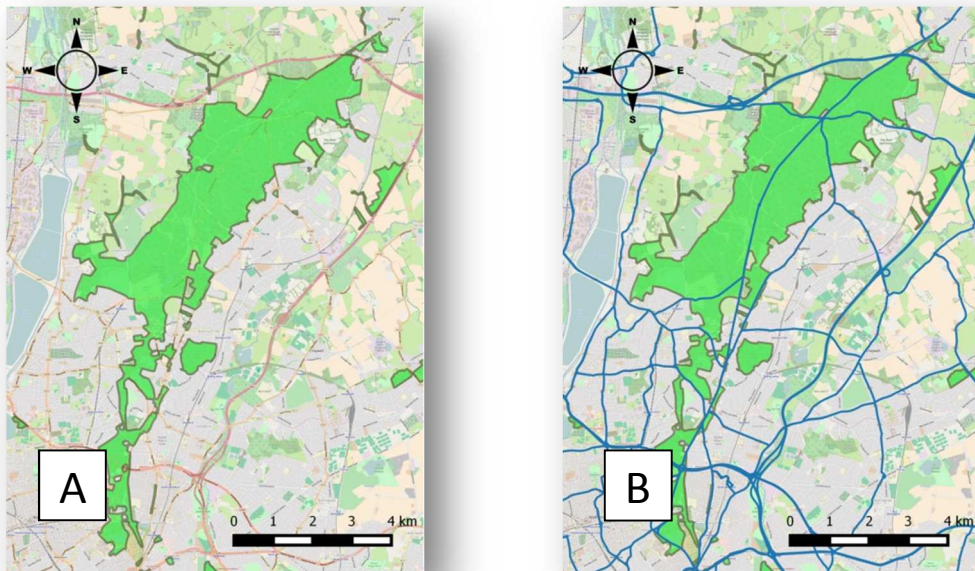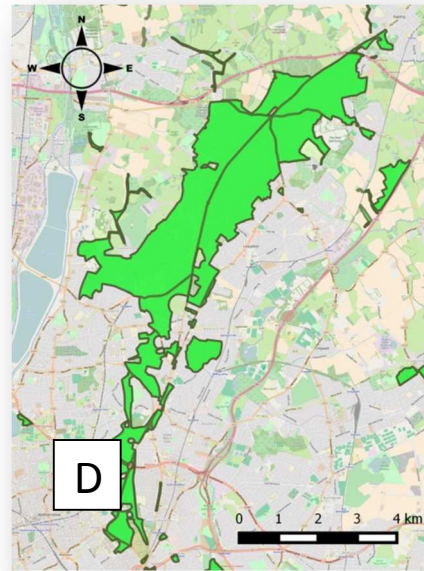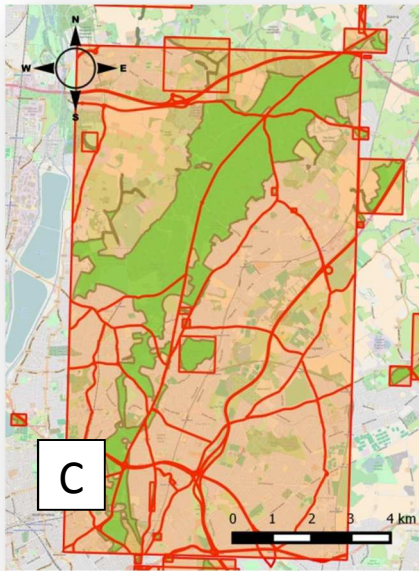
4.11   OSM Natural Areas

OSM features in which the keys *Natural* had an entry that was not 'water', 'beach' or 'sand' were extracted and uploaded to the PostGIS database. Features in that data layer for which access was specifically labelled as 'public', 'yes', or 'permissive' or had a name that included one of the following;

- recreation
- common
- park
- heath
- open access
- community
- play area
- play space

were assumed to be publicly accessible natural areas. Those accessible natural areas were added to the ORVal parks data layer giving features with an OSM *Natural* key of 'wood' or 'forest' the type designation 'wood' and the remaining natural features attributed the type 'nature'.

4.12   OSM Cemeteries

OSM features in which the keys *Landuse* or *Amenity* had been tagged as 'cemetery', 'churchyard' or 'grave_yard' were extracted and uploaded to the PostGIS database. Features in that data layer for which access was private or otherwise restricted were deleted. OSM cemetery features were added to the ORVal parks data layer with the type designation 'cemetery'.

4.13   OSM Allotments

OSM features in which the keys *Landuse* or *Leisrue* had been tagged as 'allotment' were extracted and uploaded to the PostGIS database. Features in that data layer for which access was private or

otherwise restricted were deleted. OSM allotment features were added to the ORVal parks data layer with the type designation 'allotment'.

4.14　OSM Playgrounds, Parking and Picnic Sites

The final set of features added to the ORVal parks data layer were areas tagged in OSM as being playgrounds, public parking or picnic sites. Where those features bordered a park in the ORVal parks data layer they were merged into that existing feature.

4.15　Dealing with Overlapping Features

In building the ORVal parks data layer, difficulties arise when a feature being added to the map overlaps or adjoins one or more of the park features already identified.

The simplest case occurs when the new feature is completely contained by an existing feature. In that case, the new feature can simply be ignored; the recreation area it describes has already been captured in the ORVal parks data layer.

When the new feature is only partially contained by an existing feature we proceed by first snipping off the overlapping area from the new feature and assuming that the overlapping area should be considered part of the existing park (an assumption which follows from our careful choice of the order in which features are added to the ORVal parks data layer). The remaining part of such new features do not now overlap existing features but they do share a common border. The decision as to whether they should be merged into the existing feature or identified as an independent park is determined through a calculation based on a feature's size, shape and the degree of overlap that feature originally shared with the bordering existing features. In particular, we calculate an index as follow;

$$merge\ index = area\ \times compactness \times border^{-1} \times overlap^{-1}$$

Where *compactness* measures the degree of compactness of the new feature by comparing the length of its perimeter to that of a circle (the most compact shape) with the same area as a feature;

$$compactness = \frac{perimeter\ of\ circle\ with\ same\ area\ as\ feature}{perimeter\ of\ feature},$$

border measures the proportion of the new feature perimeter that borders the existing feature;

$$border = \frac{length\ of\ perimeter\ bordering\ existing\ feature}{perimeter\ of\ feature}$$

and *overlap* measures the proportion of the area of the new feature that overlapped the existing feature;

$$overlap = \frac{area\ of\ overlap\ with\ existing\ feature}{area\ of\ feature},$$

Observe that the index takes higher values the larger the area and compactness of the new feature and lower values the greater the length of shared border and the degree of overlap. The merge index was tested against a training dataset and a threshold value of 10,000 was chosen over which a

new feature was deemed to represent an independent park. New features with indexes less than the threshold were merged into the existing feature with which they bordered.

The process of merging new features into existing features could be further complicated when a new feature bordered more than one existing feature. In that case, the new feature was split between existing features. The logic used in that process is illustrated by the example shown in Figure 4.

In this case, the existing features are two parks identified from the OSM parks layer shown as the light green areas in the centre of Panel A. The new features come from the Woods for People data set and these are shown as the dark green areas overlaying the existing features in Panel B. Notice in Panel C that the new features to the North and West do not border any existing features and are immediately accepted as new independent woodland parks (Panel C).

In contrast, two areas of woodland are observed that border and overlap with the two existing parks. The areas of overlap (hashed areas in Panel C) are snipped off the woodland features and the length of perimeter that the snipped features hold in common with the existing parks calculated (highlighted red in Panel D). The merge index is calculated for both features. Since these are relatively small, elongated woodlands that share a long common border with existing parks, it is no surprise that the merge index for both features falls below the threshold.

Things are relatively simple for the woodland to the east. Since this only borders one existing park, the woodland is immediately merged into that feature, extending its boundaries to the east (Panel E).

The procedure for splltting the remaining woodland feature between the two existing parks that it borders are illustrated in Panels F, G and H. First, the two common borders are reduced to a series of node points (Panel F). Next Voronoi cells are constructed centred on those node points and clipped to the area of the new feature (Panel G). The outcome of that procedure is to break the new woodland feature up into a series of small cells such that the area in each cell is closer to its node point than any other node point on the two borders. Finally, the cells belonging to node points along the border with each existing park are merged into that park, dividing up the woodland and extending the area of the two parks (Panel H).

4.16   Cutting out Lakes and Reservoirs

The final step in preparing the ORVal parks data layer involves cutting out areas of lakes and reservoirs from all features

Figure 4: Splitting and Merging New Features into Existing Features

16

4.17   The ORVal Parks Data Layer

The final ORVal parks data layer consists of 48,245 identified park-like recreation areas. Table 1 show the number of features identified in each of our broad type classifications.

Table 1: Feature Types in Orval Parks Data Layer

Type	Number of Features
Park	19,388
Cemetery	9,494
Woods	7,359
Allotment	6,865
Nature	2,846
Golf	1,880
Country Park	413
Total	48,245

5. Paths

The primary source of data for identifying paths is again OpenStreetMap (OSM). While there is no guarantee that data on the paths network taken from OSM is comprehensive, it is undoubtedly very extensive identifying over 1 million individual stretches of path across the UK. In addition, there is reason to believe that the paths in which we have most interest, those used for the purposes of outdoor recreation, are more likely to be recorded by OSM users. Moreover, OSM provides a tag allowing users to indicate a path's accessibility characteristics which helps in identifying stretches of path that are open for public use.

As described in this section, we build the ORVal paths data layer by first identifying publicly accessible stretches of paths from the OSM data. Each individual stretch of path is a line feature whose start and end may either connect to another path in the paths network or not. We describe ends of path stretches that do not connect as dangling ends. Dangling ends more often than not are the access points to the paths network often being the point of connection of a paths network with the roads network.

Since our interest is in paths that pass through greenspace, we restrict out attention to path stretches that do not pass through built up areas. Likewise we remove path stretches that are in parks already defined in the ORVal parks data layer.

Of course, it is not the case that each individual path stretch is a separate recreation location. Rather groups of individual stretches link-up forming connected networks of paths. We take those connected networks to be the recreation features that attract visitors. The next step in building the

ORVal paths data layer was to link connected stretches together in order to identify the separate path networks.

Since those networks can be very extensive, perhaps extending across an entire national park, the final step in constructing the ORVal paths data layer was to identify the locations at which visitors access the paths network and define those *path access points* as recreation sites. The assumption here is that the nature of the greenspace that individuals experience from traversing some particular path network will be determined by the point at which they access the network.

5.1 OSM Paths

Path stretches were extracted from OSM by selecting features in which the key *Highway* had been tagged as;

- 'track'
- 'footway'
- 'path'
- 'cycleway'
- 'byway'
- 'trail'
- 'bridleway'

The data were reduced by removing all features for which access was private or otherwise restricted. In particular, features were only retained if the *Access* key was null or tagged as;

- 'public'
- 'yes'
- 'permissive'
- 'destination'

Likewise, features were dropped if the *Foot* key was tagged as;

- 'private'
- 'official'
- 'no'
- 'customers'
- 'customer'
- 'pay'

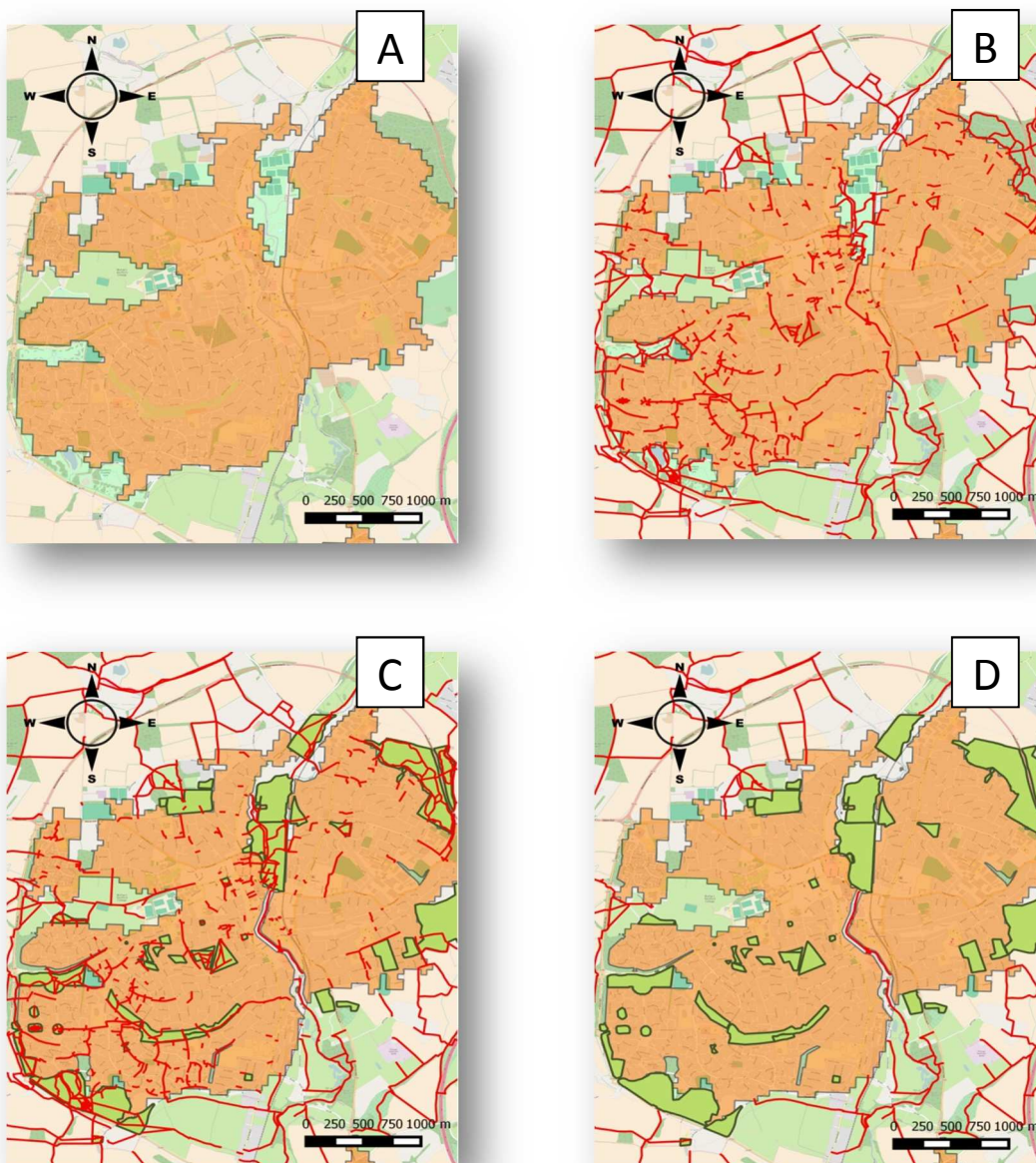## 5.2 Identifying Path Stretches in Greenspace

Reducing the OSM paths data to stretches that traverse greenspace requires deleting stretches that are part of the urban paths network. To identify such stretches we used the OS Built-Up Area 2011 dataset, illustrated, by way of example, by the built up area of Bishop Stortford in Panel A of Figure 5. Observe from Panel B that many of the path stretches identified in the OSM data fall entirely within the boundary of the built-up area. While many of these are indeed paths through built up areas, others may provide interaction with urban greenspaces. To account for that possibility, we take datasets of the coast, lakes, reservoirs, rivers, canals and natural areas (all extracted from OSM),

buffer them by 25m and then cut those out of the built-up area data layer. In Panel C of Figure 5, for example, the course of the River Stortford through Bishop Stortford has been removed from the built-up area making it clear that one of the urban paths is actually a riverside walk. Panel C also overlays the built-up area with the ORVal parks data layer making it clear that many of the paths in the urban area are actually paths through urban parks.

To identify paths through greenspace that might be used for recreation, we delete all those stretches that start and end in the urban area. We also delete all stretches that are majority within a park already identified in the ORVal parks data layer.

Panel D illustrates the path stretches that remain following these deletions.

Figure 5: Identifying Path Stretches in Greenspace

## 5.3 Combining Path Stretches into Networks

To combine path stretches into connected networks, we created a simple recursive function. The function began by selecting any stretch and identifying the other path stretches with which it

connected. Then for each connected path stretch the function recursed, identifying each other path stretch to which it connected. The recursion continued until all path stretches that were connected to the originally chosen stretch had been identified. All those stretches were merged into a single path network and the function moved on, choosing a new stretch and building the path network to which that stretch connected.

5.4 Path Access Points

To complete the ORVal paths data layer, the access points to each path network were identified by finding dangling ends of paths in a network within 50m of roads. Figure 6 continues our Bishop Stortford example (Panel A) by overlaying the OS Integrated Transport Network roads data layer (Panel B). Dangling ends in proximity to the roads network are identified in Panel C and the final set of access points shown in Panel D.

Figure 6: Identifying Path Access Points
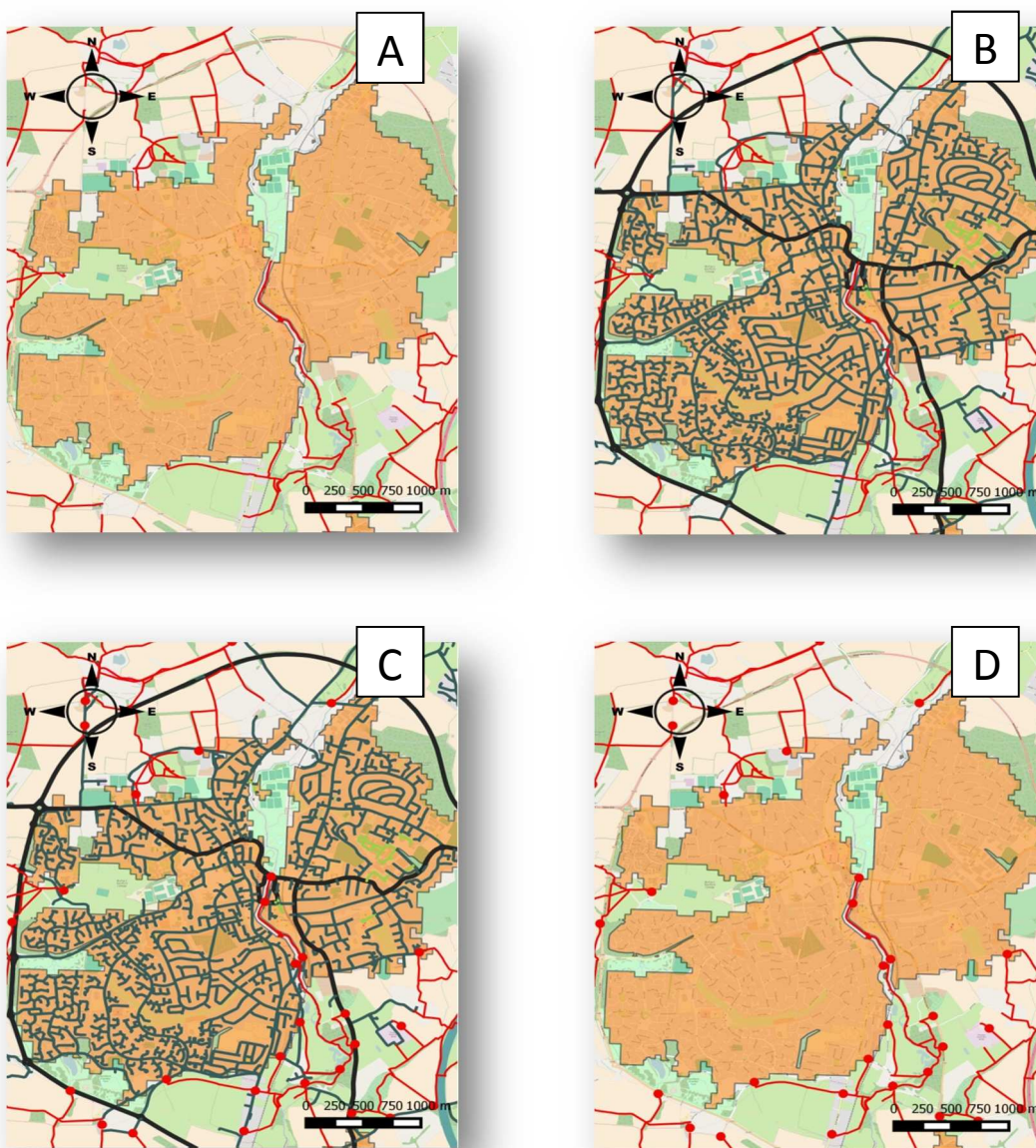
In some cases a paths network was served by a number of access points in close proximity. Since those access points present visitors with a very similar recreational opportunity we replaced groups of two or more access points that were not further than 250m from each other with a single approximating access point at the centroid of the group. For that reason, some access points in the ORVal paths data layer do not lie exactly on the intersection of a path and road.

5.5 The ORVal Paths Data Layer

The final ORVal paths data layer consisted of 82,115 access points serving 22,259 networks of paths. The number of access points serving each network ranged from a minimum of 1 up to a maximum of 246.

6. Beaches

The locations and names of beaches in England were extracted from http://britishbeaches.info and then cross-checked and supplemented by Environment Agency data downloaded from the UK governments open data website. As shown in Figure 7, the beach locations provided by these sources are point locations. They also provide information on whether the beach is predominantly sand, shingle, rock,mud or some combination of those littoral sediments.

Figure 7: Identifying Path Access Points
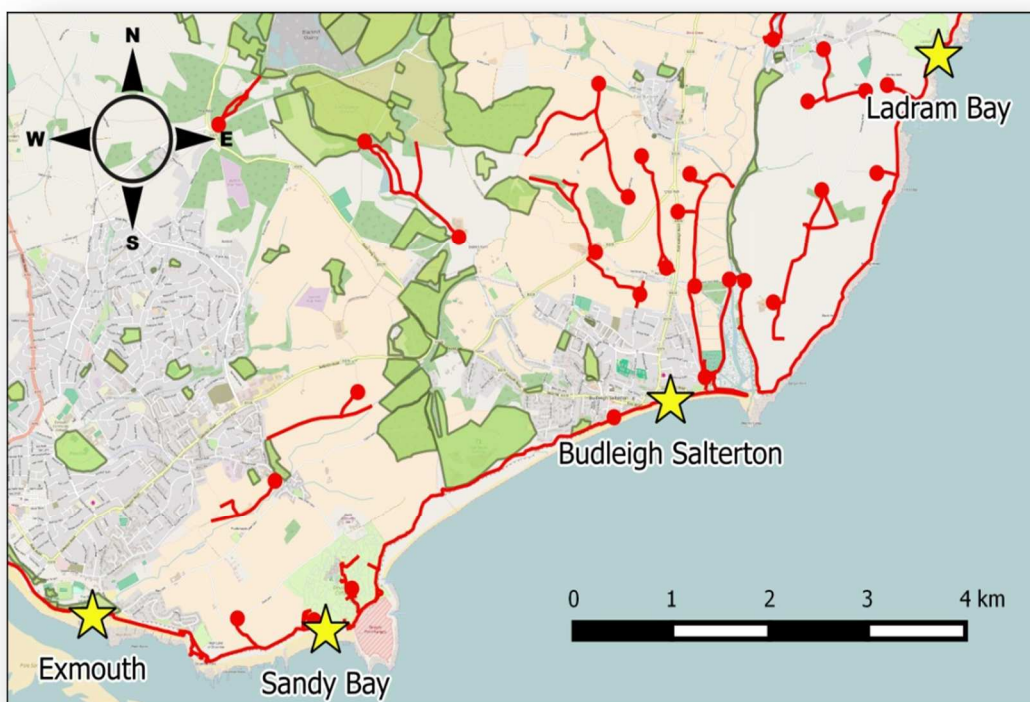
6.1 The ORVal Beaches Data Layer

The final ORVal beaches data layer consisted of 630 beach features. As listed in Table 2, sandy beaches were the most frequently observed beach type.

Table 2: Beach Types in Orval Beaches Data Layer

| Type	Number of Features
Sand	313
Sand/Shingle	110
Shingle	89
Not Known	37
Sand/Rock	34
Sand/Shingle/Rock	12
Mud-Sand/Shingle	10
Harbour	9
Shingle/Rock	8
Rock	8
Total	630

7. Land Cover: Terrestrial Features

With the park, path and beach sites defined, the next task was to understand the nature of the recreation experience provided to visitors by each location. To that end, we combined multiple datasets in order to establish each site's land cover and land use (abbreviated, henceforth, simply to land cover).

Figure 8 illustrates the procedure used to define land cover for the Country Park of Whitlingham Broad just south of Norwich (Panel A). The first step was to create a point grid across each recreation site, where each point represented the centre of the 25m square grid used by the Land Cover Map 2007 (LCM 2007) data layer. Indeed the first step in defining land covers for each recreation site was to attribute each point on a site's point grid with the land cover defined for that location in the LCM 2007 data layer. As shown in Table 3, LCM 2007 defines some 24 different land covers.

Figure 8: Defining Land Cover for a Park
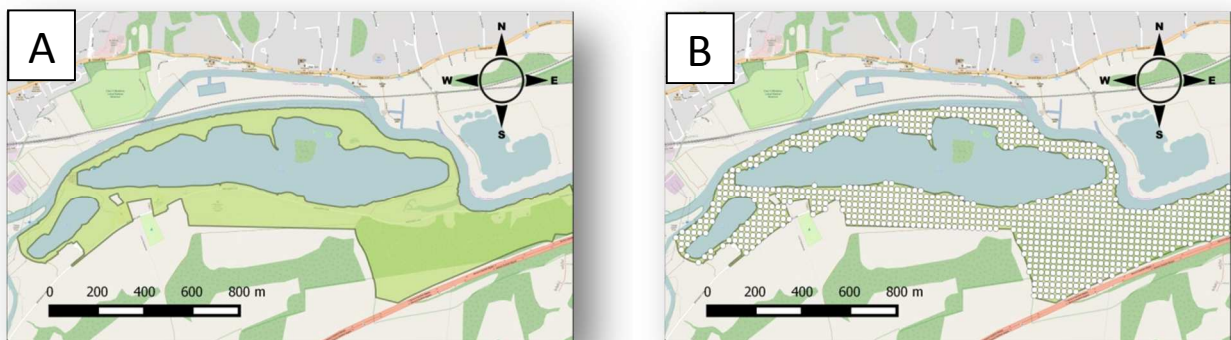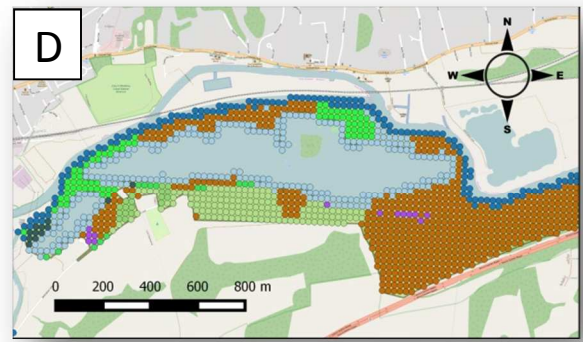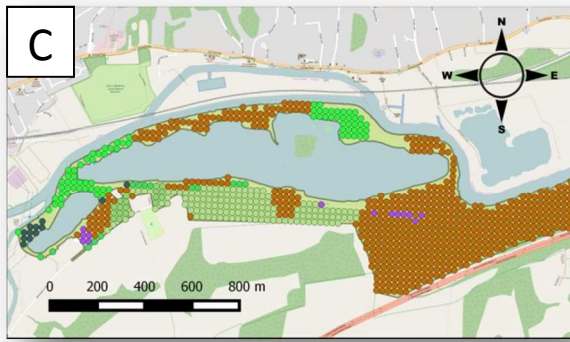
Next, the land cover defined for each grid point was updated to reflect information provided by Natural England's Priority Habitat Inventory data layers. As shown in Table 3, overlaying those data layers added another 26 land cover categories to those defined from LCM 2007.

After that, land cover information for each site was further updated to reflect information on the location of 6 different land uses extracted from OSM. As shown in Table 3, those land uses included allotments, cemeteries, golf, sports pitches, gardens, parking and playgrounds.

Finally, data provided by the OS and extracted from OSM was used to establish the locations of 7 different types of water feature within the boundaries of recreation sites. As per Table 3 those included, seawater, estuaries, lakes, reservoirs, tidal rivers, inland rivers and canals.

The procedure resulted in each grid point being attributed to one of 64 different land cover types. For the sake of simplicity, those 64 land covers were simplified to 24 broader land cover categories (see final column of Table 3).

The final pattern of land use for Whitlingham broad is depicted in Panel C of Figure 8 where the brown areas are woodland, the dull green area are wood pasture (open woodland) and the bright green are natural grass.

Since the recreation experience enjoyed at a site may be enhanced by the fact that a site borders a water feature, one final step was to buffer each park by 30m and create grid points over that buffer that corresponded to the locations of water features. In Panel D of Figure 8 that procedure results in the addition of grid points along the margins of the broad which are attributed the 'lake' land cover category. Likewise, where the River Yare runs along the northern boundary of Whitlingham Country Park, an additional line of dark blue grid points indicates the existence of a bordering river feature.

Table 3: Land Cover and Land Use Categories

Source and Type	Code	Category Description	ORVal Code	ORVal Category
	0	Sea Water	0	Sea_Water
LCM 2007	1	Broadleaved Woodland	1	Woods
	2	Coniferous woodland	1	Woods

3	Arable and Horticulture	2	Agrculture
4	Improved Grassland	3	Managed_Grass
5	Rough Grassland	4	Natural_Grass
6	Neutral Grassland	4	Natural_Grass
7	Calcareous Grassland	4	Natural_Grass
8	Acid Grassland	4	Natural_Grass
9	Fen, Marsh and Swamp	5	Fen_Marsh
10	Heather	6	Moors_Heath
11	Heather grassland	6	Moors_Heath
12	Bog	6	Moors_Heath
13	Montane Habitats	7	Mountain
14	Inland Rock	8	Rock
15	Saltwater	0	Sea_Water
16	Freshwater	24	Lakes_Reservoirs
17	Supra-littoral Rock	10	Coastal
18	Supra-littoral Sediment	10	Coastal
19	Littoral Rock	10	Coastal
20	Littoral Sediment	10	Coastal
21	Saltmarsh	11	Saltmarsh
22	Urban	12	Built_Up
23	Suburban	12	Built_Up

	100	Wood Pasture	13	Wood_Pasture
	101	Blanket bog	6	Moors_Heath
	102	Calaminarian grassland	4	Natural_Grass
	103	Coastal & floodplain grazing marsh	5	Fen_Marsh
	104	Coastal sand dunes	10	Coastal
	105	Coastal vegetated shingle	10	Coastal
	106	Deciduous woodland	1	Woods
	107	Limestone pavements	8	Rock
	108	Lowland calcareous grassland	4	Natural_Grass
	109	Lowland dry acid grassland	4	Natural_Grass
Priority	110	Lowland fens	5	Fen_Marsh
Habitats	111	Lowland heathland	6	Moors_Heath
	112	Lowland meadows	4	Natural_Grass
	113	Lowland raised bog	5	Fen_Marsh
	114	Maritime cliff & slope	8	Rock
	115	Mountain heath & willow scrub	6	Moors_Heath
	116	Mudflats	22	Estuary
	117	Purple moor grass & rush pastures	6	Moors_Heath
	118	Reedbeds	5	Fen_Marsh
	119	Saline lagoons	0	Sea_Water
	120	Coastal saltmarsh	11	Saltmarsh
	121	Traditional orchards	1	Woods

|				
	122	Upland calcareous grassland	4	Natural_Grass
	123	Upland hay meadows	4	Natural_Grass
	124	Upland heathland	6	Moors_Heath
	125	Upland flushes, fens & swamps	5	Fen_Marsh
OSM Land Uses	201	Sports Pitches	14	Sports_Pitches
	202	Public Gardens	15	Formal_Garden
	203	Parking	16	Parking
	204	Playground	17	Playground
	205	Golf	18	Golf
	206	Allotments	19	Allotments
	207	Cemeteries	20	Cemeteries
Water	301	Coast	21	Seaside
	302	Estuary	22	Estuary
	303	Tidal River	23	Rivers_Canals
	304	Inland River	23	Rivers_Canals
	305	Canal	23	Rivers_Canals
	306	Reservoirs	24	Lakes_Reservoirs
	307	Lakes	24	Lakes_Reservoirs

The procedure for defining land covers for a path feature is almost identical. As shown in Figure 9, we start with a path network (Panel A), then created a 25m buffer around that path (Panel B). Next we create a 25m point grid over that area (Panel C) and identify land covers using the same procedure as for park features (Panel D).

Since beaches are point features we simply define their land cover as being half coastal sea water and half coastal sea side.

To summarise a park's land cover, we simply sum the number of points in that park's grid that have been attributed to the different land cover categories. That quantity can then be expressed as a percentage of the entire park area by dividing by the total number of grid points in that park or as an absolute area by multiplying by 0.0625, the area of a 25m square cell.

For paths, we assume that there is a declining likelihood that a visitor will encounter some particular stretch of a paths network the farther that stretch is from the path access point. We calculate the land cover characteristics associated with accessing a paths network from some particular access point as being the weighted sum over grid points. The weights associated with each grid point decline linearly from a value of 1 for grid points at the access point to 0 at a distance of 10km. In Figure 9, therefore, the land cover associated with accessing this path network from the north-western access point would indicate larger quantities of wood (brown-coloured points) and heath (purple-coloured points) than the land cover associated with accessing that same network from the south-eastern access point which would indicate greater quantites of grassland (green-coloured points).

Figure 9: Defining Land Cover for a Path
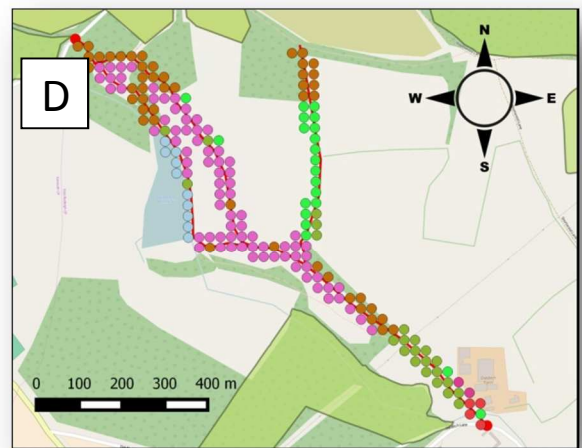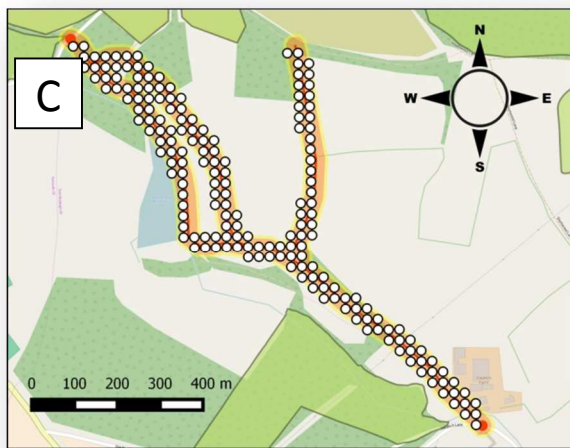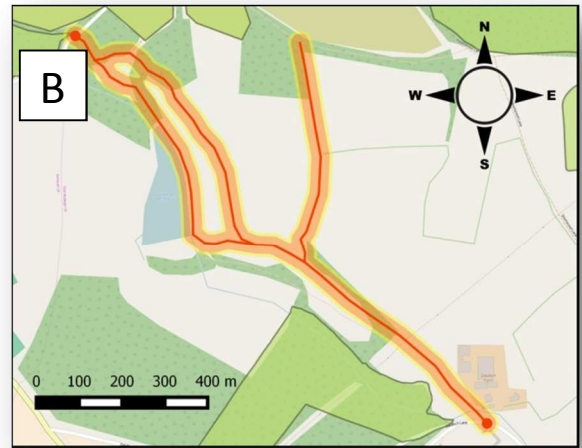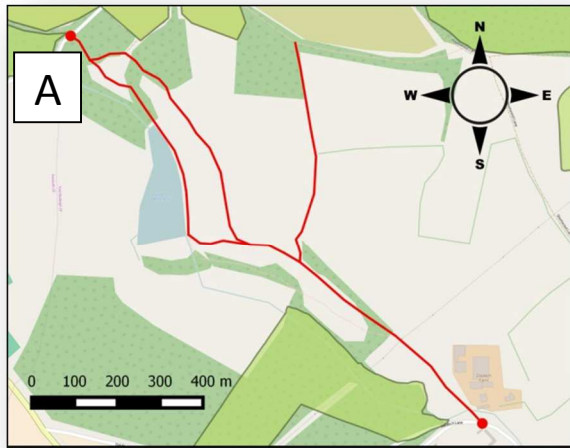
8. Designations

Identifying areas of recreation sites that are under particular designations proceeds in a similar manner to that for identifying land cover. We collated 17 different designation data layers;

- National Park
- Area of Outstanding Natural Beauty (AONB)
- Country Park
- Forestry Commission
- Historic Park
- Local Nature Reserve
- National Nature Reserve
- Natura 2000
- RAMSAR
- SSSI
- Nature Improvement Areas

- Ancient Woodland

- Doorstep and Millennium Greens

- CROW Open Access Land

- National Trail

- Heritage Coast

- National Trust

Each designation data layer in turn was overlain on the recreation site points grid and counts for parks and weighted counts for paths used to establish the area of each site under each type of designation.

9. Points of Interest

Finally, the OS Points of Interest (POI) dataset was used to identify if recreation sites afforded access to any special natural or historical features. The features selected from the OS POI dataset were as follows:

- Scenic features (POI code: 03190257)

- Historic buildings (POI code: 03170244)

- Archaeological sites (POI code: 03170240)

to which was added another OSM extract identifying the point location of viewpoints.

A park was identified as containing a feature of a certain type if that feature was found to lie within its boundaries. In contrast a path network was identified as giving access to a feature of a certain type if that feature was within 25m of the path. Again a weighting approach was used to take account of differing distances from path access points. Where a feature was within 1km of an access point it was given a score of one but that score declined linearly after that point down to a score of zero at a distance of 10km or more.