# The following Annexes were taken from Bateman et al (2014), full reference as follows:

Bateman, I., Day, B., Agarwala, M., Bacon, P., Baďura, T., Binner, A., De-Gol, A., Ditchburn, B., Dugdale, S., Emmett, B., Ferrini,S., Carlo Fezzi, C., Harwood, A., Hillier, J., Hiscock, K., Hulme, M., Jackson, B., Lovett, A., Mackie, E., Matthews, R., Sen, A., Siriwardena, G., Smith, P., Snowdon, P., Sünnenberg, G., Vetter, S., & Vinjili, S. (2014) 'UK National Ecosystem Assessment Follow-on. Work Package Report 3: Economic value of ecosystem services', UNEP-WCMC, LWEC, UK.

# 1. Annex 1: Land use, land cover and livestock data.

## 1.1. Summary


A dataset describing classes of non-overlapping land use has been generated which has utility for research at a range of spatial scales. To our knowledge, it is the most comprehensive definition of the physical stock of land types in Great Britain for the purposes of ecosystem assessment.

Inconsistent correspondence between land cover and land use datasets and concerns over their thematic, temporal and spatial accuracy called into question the fitness of individual off-the-shelf datasets. In response, several datasets were combined to generate a custom product. In brief, satellite-derived land cover data and ancillary spatial data were used to locate areas that are likely to be functional e.g. used for agricultural production or urban activities. Results from agricultural survey data were used to refine the spatial distribution of arable and grassland and subdivide categorisation where appropriate. A Geographical Information System (GIS) was used to interrogate and integrate data to a base resolution of a 2km by 2km cell (a 1km resolution dataset was also produced for use in Section 12 only). The process was undertaken for two target years.

Rather than a complete land use definition, the resultant dataset is more adequately described as a high resolution database depicting potential land cover or land use area across Great Britain. Due to uncertainties with input data, there is greater confidence in relative magnitudes of areas (i.e. shares of land types) than absolute totals. However, as the level of spatial aggregation increases, the absolute area totals become more accurate. Also, as the timeframe of study increases, to say three to five years, data become more representative of that period, rather than a single target year.

Output from the agricultural production model (Section 5) was used to predict a baseline and changes in agricultural land use. The land use definition discussed here was used a) for estimation of models for other ecosystem components and b) as a baseline for non-agricultural land use.


## 1.2. Objectives

The land use dataset was developed to serve the following roles:

- To provide a complete picture of the spatial distribution of land use
- To generate spatially consistent land use data across time (i.e. apply a reliable methodology)
- To include England, Scotland and Wales
- To be fit for purpose at multiple levels: 2km, regional, hydrometric area, national-level

- To be used in conjunction with other data to allow the derivation of trends and indicators of change
- To be consistent with the demands of an interdisciplinary project
- To be used for the spatial re-distribution of other data e.g. heads of livestock

## 1.3. Data

Data from multiple source geographies (Table 15.1) were translated into a common spatial unit which described general classes of non-overlapping land use and land cover. Two main data types were used: satellite-derived digital land cover maps and survey data on agricultural land use practices. Ancillary datasets (e.g. road networks and political boundaries) were employed to identify areas of non-agricultural land use to refine the classification. Using a GIS, data were integrated to a common spatial unit (2km × 2km cell), with this choice of resolution being a lowest common denominator given the highest detail at which agricultural land use data could be obtained. Following initial scoping of data availability and temporal resolution, this was performed for two target years: 2000 and 2010.

The physical material at the surface of the earth, land cover, can be observed through field survey or via analysis of remotely sensed imagery. The Centre for Ecology and Hydrology (CEH) has produced Land Cover Maps for the UK: e.g. *LCM2000* (Fuller et al., 2002) and *LCM2007* (Morton et al., 2011). For each Land Cover Map, imagery taken over several years is reclassified on a pixel-by-pixel basis into land cover types (remotely sensed data were acquired between November 1996 and May 2001 for LCM2000 and between September 2005 and July 2008 for LCM2007). Land use reflects the functional dimension of Earth's surface. Land use in the UK is dominated by agriculture which accounts for 18.3 million hectares or 74.8% of the total surface area Defra (2011b). The June Survey of Agricultural and Horticultural Activity is a source of high quality land use data with national coverage. The June Survey is undertaken as a full census every ten years and as a sample survey in intervening years. The June Survey is undertaken independently in England, Scotland and Wales and results are released in aggregated spatial units. These data can either be obtained in the form of a regular grid known as the '*agcensus*' (available at 2km, 5km and 10km resolutions from JAC, 2013) or for administrative boundaries such as counties and regions (see details in Table 15.1).

Due to protection against the disclosure of information on individual holdings, there are caveats associated with the use of these 'ready-made' datasets for spatially explicit research. Broadly speaking, agcensus data can be inaccurate at fine resolutions due to spatial reworking and re-distribution of holding data, and while statistics for administrative boundaries are more accurate, many data are suppressed to preserve anonymity or released at a higher level geography where the resolution is too coarse. To combat these shortfalls, both data formats were used.

## 1.4. Methodology

A GIS was used to interrogate and integrate land use and livestock data to a base resolution of a 2km by 2km cell (see Section 15.3 for a discussion of base unit). The process was undertaken for two target years and is summarised below. Further methodological detail, including a critical discussion of underlying methodological issues, can be obtained from the authors by request.

### 1.4.1. Overall land use

The stages of data integration can be summarised as:

- Stage 1: Reclassify existing Land Cover Maps and examine summary statistics;

- Stage 2: Augment reclassified Maps with other data pertaining to non-agricultural land cover and land use (e.g. urban or forestry);
- Stage 3: Test for correlation between agricultural land cover and land use data;
- Stage 4: Perform redistribution of agricultural land use using available georeferenced data and statistics.

**Table 15.1: Raw data sources and temporal data available to describe target years 2000 and 2010.**

| Land cover and land use | Data description | Data type | Extent | Data source(s) | Target year 2000 | Target year 2010 |
|---|---|---|---|---|---|---|
| General land cover | Land Cover Map | 25 m raster grid | GB | CEH | *c.*2000 | *c.*2007 |
| Coniferous or deciduous land cover | National Inventory for Woodland and Trees | GIS polygon file | GB | Forestry Commission | 2002 | 2002 |
| Urban and developed land use | Developed Land Use Areas | GIS polygon file | GB | OS Meridian | 2009 | 2009 |
| | Roads and railways | GIS polyline files | GB | OS Meridian | 2009 | 2009 |
| Agriculture | Processed June Agricultural Survey(s) | 2 km *agcensus* | GB | EDINA | 2004 | 2010 |
| | | Spreadsheet of county-level statistics | E | Defra | 2000 | 2010 |
| | | Table for agricultural region statistics | S | ERSA | 2001 | 2010 |
| | | Spreadsheet of Small Area statistics | W | National Assembly for Wales | 2003 | 2010 |
| | OS Open Data (county and region boundaries) | GIS polygon files | E & S | OS OpenData | 2011 | 2011 |
| | Small Area boundaries | GIS polygon file | W | National Assembly for Wales | 2001 | 2001 |

Abbreviations used: CEH = Centre for Ecology and Hydrology; E = England; ERSA = Economic Report on Scottish Agriculture; GB = Great Britain; OS = Ordnance Survey; S = Scotland; W = Wales

#### *1.4.1.1. Stages 1 and 2*

Stages 1 and 2 enabled the creation of LCUP1 (2000, 2007) and LCUP2 (2000, 2007) datasets. First, the 25m resolution raster products for LCM2000 (Fuller et al., 2002) and LCM2007 (Morton et al., 2011) were used as raw land cover data for target years 2000 and 2010 respectively. Ten land cover categories, broadly corresponding to LCM2000 and LCM2007 Aggregate Classes and also consistent with habitat mapping as part of the first phase of UK-NEA (2011), were created from combining subclasses of land cover (Table 15.2). Next, a simple cross-tabulation was performed to look at land cover change on a cell-by-cell basis across the two time periods. Reasonable correlation with small changes in land cover were expected, e.g. due to development and small differences in the methodology between LCM2000 and LCM2007. However, the results of the comparison did not always perform as anticipated and there was considerable movement across many classes. To combat this, reclassified Land Cover Map data (for both target years) were augmented with Forestry Commission boundaries of existing woodland FC (2002), Ordnance Survey data on Roads and Railways and Developed Land Use Areas (OS, 2013b) (Table 15.1). These updates enabled a more reliable indication of non-agricultural land use extent (e.g. LCUP2, 2000, 2007).

**Table 15.2: Classes of land cover (after Fuller et al., 2002; Morton et al., 2011)**

| Broad land cover class | LCM2000 subclass | code | LCM2007 subclass | code |
|---|---|---|---|---|
| Deciduous | Broad-leaved / mixed woodland | 1.1 | Broadleaved woodland | 1 |
| Coniferous | Coniferous woodland | 2.1 | Coniferous woodland | 2 |
| Enclosed Farmland | Arable cereals<br>Arable horticulture<br>Arable non-rotational<br>Setaside grassland | 4.1<br>4.2<br>4.3<br>5.2 | Arable and Horticultural Land | 3 |
| Improved Grassland | Improved Grassland | 5.1 | Improved Grassland | 4 |
| Semi-natural Grass | Acid grassland<br>Neutral grassland<br>Calcareous grassland<br>Fen, marsh, swamp (rush pasture) | 8.1<br>6.1<br>7.1<br>11.1 | Acid Grassland (Bracken)<br>Neutral Grassland<br>Calcareous Grassland<br>Fen / swamp<br>Rough Grassland | 8<br>6<br>7<br>9<br>5 |
| Mountains, moors and heaths | Bog (deep peat)<br>Montane habitats<br>Inland bare ground<br>Dense dwarf shrub heath<br>Open dwarf shrub heath<br>Bracken | 12.1<br>15.1<br>16.1<br>10.1<br>10.2<br>9.1 | Bog<br>Montane habitats<br>Inland rock<br>Heather<br>Heather grassland | 12<br>13<br>14<br>10<br>11 |
| Coastal Margins | Saltmarsh<br>Littoral rock<br>Littoral sediment<br>Supra-littoral rock<br>Supra-littoral sediment | 21.2<br>20.1<br>21.1<br>18.1<br>19.1 | Saltmarsh<br>Littoral rock<br>Littoral sediment<br>Supra-littoral rock<br>Supra-littoral sediment | 21<br>19<br>20<br>17<br>18 |
| Freshwater, Wetlands | Water (inland) | 13.1 | Freshwater | 16 |
| Marine | Sea / Estuary | 22.1 | Saltwater | 15 |
| Urban and developed land | Continuous urban<br>Suburban / rural developed | 17.2<br>17.1 | Urban<br>Suburban | 22<br>23 |

### *1.4.1.2. Stage 3*

In some cases land cover classes may be synonymous with land use. Often, however, variability of land use is greater than the variability of land cover because one land cover can fulfil different functions, i.e. the relationship is not one-to-one (Gong and Weber, 2009). Nevertheless, land cover data can provide a useful framework within which to map agricultural land use (e.g. Posen et al., 2011).

Initially, relevant land areas from land cover derived data were compared with national-level June Survey statistics for agriculture (SEERAD, 2001, and SGRPID, 2011). Considerable disparities in total areas were observed. For example, the total area of Temporary and Permanent grassland land use in the June Survey (SGRPID, 2011) was greater than the Improved Grassland land cover category (LCUP2, 2007); in contrast, Arable, horticulture & fallow (LCUP2, 2007, and SGRPID, 2011) was less than the Enclosed Farmland land cover (LCUP2, 2007).

A second round of correlation testing was performed to provide an indication of the strength of the relationship between land use and land cover at the 2km level (*agcensus*). The theory was that if a set of simple rules could establish the link between land cover and land use then there would be no real need to implement more sophisticated methodologies. A cell-by-cell comparison was performed for >2,000 randomly sampled cells across Great Britain. However, from this product, it is possible for observations of agricultural land to exceed the physical area of zones (see discussion in Comber et al., 2008; Posen et al., 2011). Our testing found particular problems in Scotland and Wales. For example, in 2010 (JAC, 2013) data for Scotland approximately a quarter of 2km cells are reported with an area > 400ha. We attribute this to sprawling grass and grazing land allocated to a single farm holding. Subsequent results and analyses informed the following decisions:

- The 2 km level *agcensus* data could be used to subdivide total arable land in a corresponding 2km cell into different types of crops (fine resolution data were used to maintain local cropping patterns)
- Higher level geographies (i.e. administrative-level) were needed to define the total arable land in a 2km cell and refine the distribution of types of grassland and grazing. Greater confidence was given to the administrative-level statistics as although these are aggregated for farms within an area, they are not subject to redistribution algorithms used in the production of the *agcensus*.

Further details are available from the authors on request.

#### *1.4.1.3. Stage 4*

Stage 4 enabled the creation of creation of LCUAP1 (2000, 2010) and LCUAP2 (2000, 2010) datasets. County- and Unitary Authority-level June Survey data for 2000 and 2010 were downloaded as a spreadsheet for England. Similar summaries were obtained for Welsh Agricultural Regions. Scottish regional data were obtained as PDF files from the Economic Report on Scottish Agriculture (ERSA) (ERSA, 2013). These administrative-level data were amalgamated into one dataset of 81 zones, each with six broad land use categories compatible in definition across time and for each country: Arable, horticulture & fallow; Temporary grassland; Permanent grassland; Sole-right rough grazing; Farm woodland; All other land on farm. Next, these tabulated data were joined to spatial boundary data in a GIS. At this stage, the implicit assumption was that the variables of interest (land use types) had a homogenous spatial distribution across source zones (administrative areas).

It was then necessary to redistribute the above source zone data within the locations constrained by appropriate land cover classes. In other words, the high resolution (25m × 25m grid) reclassified land cover data (used to create e.g. LCUP2) were used to restrict probable locations for agricultural land use within each administrative area. Geographic boundaries for the administrative areas were overlain on the land cover grid. Given that the area of land use in each source zone was known, we satisfied these observations by scaling the 25m resolution land cover-derived classes. Then, each broad land use type (at 25m resolution) was summed for a set of final target zones – a regular grid of 2km cells. Target zones of 1km were used for estimation of models in Section 12 only.

In the final step of processing, relevant crop types were extracted from the 2004 and 2010 *agcensus* (2km resolution) datasets. Total Arable, horticulture & fallow land in the 2km target zones were refined into different crop types using overlying *agcensus* data (by apply corresponding areal proportions). Therefore, the final dataset could be aggregated thematically or spatially to suit different research applications (e.g. LCUAP1, 2000; LCUAP2, 2000).

Definitions of the finest thematic resolution (25 classes) are provided in Table 15.3. Further methodological details are available from the authors by request.

**Table 15.3: Disaggregated land use definitions (caveats/ restrictions in parentheses)**

| Name | Description |
|---|---|
| COAST | coastal margins |
| FWATER | freshwater |
| MARINE | sea and estuary |
| URBAN | urban and other developed land |
| PERMG | permanent grassland i.e. >5 yrs |
| TEMPG | temporary grassland i.e. <5 yrs |
| RGRAZ | rough grazing |
| GRSNFRM | semi-natural grass or mountains, moors and heaths where NOT used for farming |
| FWOOD | farm woodland |
| NFWOOD | woodland NOT used for farming |
| WHEAT | wheat |
| WBARLEY | winter barley (England and Scotland only) |
| SBARLEY | spring barley (England and Scotland only) |
| OTHCER | other cereals (includes oats and other cereals for combining) |
| POTS | potatoes |
| WOSR | winter oilseed rape (where available) |
| SOSR | spring oilseed rape (where available) |
| MAIZE | maize (Scotland 2004 is within 'othcrps') |
| SBEET | sugarbeet |
| OTHCRPS | other crops and bare fallow (includes oilseed rape for Wales; includes maize for Scotland 2004) |
| HORT | total horticulture |
| TBARLEY | total barley (Wales only) |
| TOSR | total oilseed rape (where seasonal data unavailable) |
| OTHFRM | other farmland e.g. roads, buildings, yards, ponds and, where appropriate, setaside |
| OCEAN | ocean (area that is not covered by land is given 'ocean' by default) |

### 1.4.2. Distribution of livestock

The distribution of livestock was used as a proxy for the distribution of animal excreta and manures (Section 9).

Livestock were distributed over agricultural land using stocking densities at administrative-level (head counts of livestock are available from the June Survey via the sources described in the land use section and Table 15.1). Initial analysis and a review of literature (e.g. see Lyons, 2010, and Posen et al., 2011) informed the following rules:

- Cattle were distributed at administrative-level across grassland (Temporary and Permanent).
- Sheep were distributed at administrative-level across grassland (Temporary and Permanent) and Sole-right rough grazing.
- Pigs and poultry were distributed at administrative-level across intensive agriculture (Arable, horticulture & fallow; and All other land on farm).

Then, each livestock type (at 25m resolution) was summed for the set of target zones – a regular grid of 2km cells (e.g. Livestock2, 2010).

We prepared poultry datasets to aid the estimation of nutrient export coefficients (Section 9); however, the agricultural model (Section 5) did not predict poultry numbers due to lack of temporal data. Indoor or outdoor distinction of pigs and poultry is important (e.g. for water quality, see Posen et al., 2011), but this was not possible due to a lack of spatial and temporal data.

### 1.4.3. Using an agricultural model to predict land use change

The land use definition (LCUAP2) was used to provide estimation data for models described in Sections 9 to 12. The agricultural model (Section 5) used aggregated classes of *agcensus* data over a greater time period. Other models (Sections 9 to 12) aggregated classes within LCUAP2 as appropriate.

The SEER agricultural model predicts, for an amount of farmland with a set of physical and environmental characteristics, the shares of likely land use given that a farmer will try to optimise profits (Section 5). The output land use share system has six categories: cereal; oilseed rape; sugar beet and potatoes; temporary grassland; permanent grassland; and rough grazing. A seventh category 'other farmland' included horticulture, other arable crops, farm woodland and set aside. Under changing scenarios (Section 13), these shares will change.

LCUAP2 (2010) was used to define the total agricultural area for the baseline year. The agricultural model (Section 5) provided the baseline for cropping under the seven shares above. Where other models required finer thematic resolution (e.g. amount of barley within cereals), land areas under analogous categories in target year 2010 (LCUAP2, 2010) at Landscape Character Area (LCA) level were used to proportionally adjust agricultural model output. Each LCA is defined by a unique combination of physical environment and social conditions and therefore their boundaries follow natural lines in the landscape rather than administrative areas (MAGIC, 2012; CCFW, 2012; Scottish Government, 2012b). Subdivision of the seventh land use category 'other farmland' was performed on a coincident cell-by-cell basis (farm woodland, within 'other farmland', was treated as a special case as the spatial distribution was frequently heterogeneous across a LCA). Further details are provided in Section 16.

Further adjustments required for amalgamation of the modelling components can be found in Section 16.

## 1.5. Results

Final output was a set of 2km x 2km raster grids representing a percentage of total area of each land type. Maximum thematic resolution of this dataset is 25 classes covering a spectrum of land use and land cover categories (Table 15.3). This output was produced for each target year.

Due to the regular gridded nature of the dataset, each 2km grid cell can be assigned a geographic reference (e.g. British National Grid Easting and Northing for cell centroid) and exported to spreadsheet format for use outside of a GIS. Data can also be aggregated to be used at different spatial and thematic scales. In Table 15.4, land use is aggregated to eleven broad categories at a national scale.

**Table 15.4: Potential land use in the target years 2000 and 2010**

| Land use | Area (ha) 2000 | % 2000 | Area (ha) 2010 | % 2010 | %change |
|---|---|---|---|---|---|
| | | | | | |
| Crops and bare fallow (including horticulture) | 4,623,394 | 19.9 | 4,560,095 | 19.6 | **-0.3** |
| Rough grazing (sole right) | 4,211,367 | 18.1 | 3,913,729 | 16.8 | **-1.3** |
| Permanent grassland (> 5yrs) | 4,754,225 | 20.4 | 5,259,400 | 22.6 | **2.2** |
| Temporary grassland (< 5yrs) | 1,060,984 | 4.6 | 1,107,626 | 4.8 | **0.2** |
| Farm woodland | 492,743 | 2.1 | 764,063 | 3.3 | **1.2** |
| Other farmland (roads, buildings, yards etc.) | 648,298 | 2.8 | 492,424 | 2.1 | **-0.7** |
| **ESTIMATED TOTAL AGRICULTURAL AREA** | **15,791,011** | **67.9** | **16,097,337** | **69.1** | **1.2** |
| | | | | | |
| **Urban and developed land** | **2,607,465** | 11.2 | **2,747,848** | **11.8** | **0.6** |
| **Marine and coastal** | **352,306** | 1.5 | **382,222** | **1.7** | **0.2** |
| **Freshwater** | **211,833** | 0.9 | **248,539** | **1.1** | **0.2** |
| **Non-farm grass, mountains, moors and heath** | **1,709,945** | 7.3 | **1,658,405** | **7.1** | **-0.2** |
| **Non-farm wood** | **2,609,203** | 11.2 | **2,147,413** | **9.2** | **-2** |
| | | | | | |
| **TOTAL** | **23,281,763** | **100** | **23,281,763** | **100** | |

## 1.6. Discussion

Caveats associated with the land use dataset are briefly discussed.

### 1.6.1. Interpretation of the land use dataset

Disaggregation of source data into target zones potentially generates spatial distributions that are unrepresentative of real-world phenomena. This is known as the Modifiable Areal Unit Problem (MAUP) (Openshaw, 1984). Practically, the derived dataset has limitations for use at a very local scale due to the inherent uncertainties in the base data layers and the assumptions required during integration. Furthermore, assumptions have been made about the stability of land uses and land covers within the time periods for different data sources.

For these reasons, the land use definition is more adequately described as a dataset representing the potential distribution of land use and land cover for a particular timeframe. Confidence in the absolute values increases as the 2km resolution spatial data are aggregated to higher level geographies. Greatest confidence is given in the national-level summaries of broad land use categories (Table 15.4).

The definition of land use can be manipulated easily into different thematic resolutions. While not entirely consistent with international standards (e.g. System of Environmental-Economic Accounting (SEEA), see details in Gong and Weber, 2009), the classification has maximised the suitability for Great Britain land use (with a possible extension to UK extent).

### 1.6.2. Integration of different systems

Flexibility in thematic resolution has meant that different model components were able to aggregate land use categories in different ways to improve fitness for purpose on an individual basis, thereby increasing confidence in the suitability of modelled variables (i.e. each modelled system was able to include the most significant variables, or combinations, and reduce error). However, this presented a difficulty for application of land use predictions governed by the agricultural model (Section 5). Further assumptions were needed to subdivide these seven broad categories.

Farm woodland was a special case. A lack of temporal data (and modelled insignificance in farmer decisions) meant that woodland on farms was subsumed within the 'other' land on farm category in Section 5. However, as this 'other' land changes under agricultural predictions, so does the amount of farm wood and hence total trees, which are important for other systems, e.g. water modelling. While the disparity of thematic resolution used by the different systems during modelling was not restricted to farm woodland, different assumptions were required to replicate the distinctiveness of the spatial distribution of this land use (i.e. cell-by-cell adjustments).

Broad land cover classes were often incompatible with (agricultural) land use. Grass and grazing land use was particularly problematic and led to the relinquishment of a mountains, moors and heaths habitat category (as used in the UK-NEA, 2011). However, land cover was still used in the estimation of some modelled ecosystem components with the proviso that extra assumptions would be needed for prediction (e.g. Section 12).

Finally, every 2km cell was modelled as an individual farm (for further details see Section 5).

## 1.7. Summary

- Probable land use and cover has been estimated for the purpose of spatially explicit modelling of multiple ecosystem components.
- Inconsistencies between land use and land cover datasets, and issues regarding compatibility of data from different devolved administrations of England, Wales and Scotland, present problems for generating a national land use database.
- Assumptions are required to modify the spatial units.
- Adjustments were needed to agricultural land use predictions to meet differing demands of components of an interdisciplinary project.

# 2. Annex 2: Supporting data

## 2.1. Overview

An internal digital data depository was established, providing access to a suite of datasets that described the spatially and temporally explicit components of natural and human systems. Unless otherwise stated, these data were processed to a 2km base resolution. Following introduction of raw data sources, processing steps are discussed for the core datasets. A Geographical Information System (GIS) was used for spatial data handling and processing.
Where datasets have been used that were not developed exclusively for this project, references can be found in individual sections of this report.

## 2.2. Objectives

Supporting data serve a range of specific objectives and individual sections provide more detail. General objectives can be summarised:

- To provide Great Britain-wide descriptors for natural environment and socio-economic phenomena.
- To provide a common spatial unit for analysis.
- To facilitate the testing of models that seek better understanding of natural and human systems which are related to land use.

## 2.3. Data

Spatial data were gathered from multiple sources to be processed in a GIS ESRI (2013). Often these were off-the-shelf, but in many cases agencies extracted bespoke datasets to cater to the needs of this ambitious project. Full details of all the main data sources are provided.

### 2.3.1. Elevation

Elevation data were gathered from the 50m resolution Integrated Hydrological Digital Terrain Model (IHDTM) licensed from the Centre for Ecology and Hydrology (CEH) (see Morris and Flavin, 1994, and IHDTM, 2002). This dataset, with a 0.1m vertical resolution, was originally derived from Ordnance Survey 1:50,000 mapping and vector data. This dataset was selected for its high quality and anticipated hydrological consistency.

### 2.3.2. Soil and underlying geology

The Harmonised World Soil Database (HWSD) is a 30 arc-second (approximately 1km resolution) raster (regular gridded) database with over 16,000 different soil mapping units. The HWSD is a composite dataset using existing regional and national updates of soil information with the information contained within the 1:5,000,000 scale FAO-UNESCO Soil Map of the World (FAO/IFA/IIASA/ISRIC-WSI/ISSCAS/JRC, 2009) Particularly relevant for a UK-based study, the areas covered by SOTER, including Central and Eastern Europe, are considered to have the highest reliability in the (World Soil and Terrain Digital Database project, which has an intended 1: 1,000,000 scale).

In practice, the HWSD is composed of a GIS raster image file linked to an attribute database in Microsoft Access format (freely accessible, subject to acknowledgement). There are three broad categories of data: (1) general information on the soil mapping unit composition; (2) information related to phases; (3) physical and chemical characteristics of topsoil (0—30cm) and subsoil (30—100cm). Example soil attributes include, but are not limited to, organic carbon, pH, water storage capacity, soil depth and textural class.

The ability to store and transport water through underlying rocks is important for determining the quantity of water (and ergo nutrients) that enter a river via groundwater. Spatial boundary data for superficial deposits and hydrogeology were taken from 1:625,000 scale national BGS data (BGS-DiGMapGB-625, 2013).

### 2.3.3. Climate

#### 2.3.3.1. Baseline Climate Data

Climate variables were derived from 5km grid baseline data for UKCP09 held by the Met Office (UKCP09, 2009a, 2009b). Monthly data for total precipitation (mm) and mean air temperature (°C) and were acquired for 1961–1990. The datasets were provided as space-delimited text files and were available for scientific research, subject to registration.

#### 16.3.3.2. Scenario Climate Data

Details of monthly mean daily maximum temperature mean daily minimum temperature and total precipitation projections from the UKCP09 project for the 2020s, 30s, 40s and 60s were obtained from UKCP09 (2009a, 2009b). The selected projections were for the medium emissions scenario and were on a 25km grid (2028 cells) aligned at an angle to the UK National Grid. The estimates extracted were 50% ‘change only’ values from a cumulative distribution function. This meant that there was a 50% probability of the change from the 1961-90 baseline being greater than the value specified (in °C or mm).

### 16.3.4. Water

Numerical and categorical quality and hydrometric data were gathered for a target period between 2000 and present. Unless otherwise stated, these raw data represent the finest spatial, temporal and thematic resolution data available.

#### 16.3.4.2. River quality and water chemistry

General Quality Assessment (GQA) Headline Indicators of Water Courses (nutrients) were obtained for England and Wales under license from the Environment Agency (EA-AfA163, 2012; see also details in IfRR, 2012). These data are classified concentrations of nitrates ($NO^3$ mg/l) and phosphates (P mg/l) with grades from 1 (very low) to 6 (excessively high); grades thus represent ranges of concentrations, not absolute values (Table 16.1). This project used data from 2000 and 2009 (most recent).

**Table 16.1: Environment Agency grading framework for GQA Headline Indicators of Water Courses**

| | Nitrate (NO3) concentration (mg/l) | Phosphate concentration (mg/l) |
|---|---|---|
| Grade 1 | <5 | <0.02 |
| Grade 2 | >5 to 10 | >0.02 to 0.06 |
| Grade 3 | >10 to 20 | >0.06 to 0.1 |
| Grade 4 | >20 to 30 | >0.1 to 0.2 |
| Grade 5 | >30 to 40 | >0.2 to 1.0 |
| Grade 6 | >40 | >1.0 |

Absolute concentration data were extracted by the Environment Agency for all sampling sites used for regular reporting for freshwater environments across England and Wales. These bespoke datasets, monthly resolution, included a range of determinants (e.g. NH4+, oxidised N, NO3-, NO2-, suspended solids, ortho-phosphate, TP, inorganic N, TN) and incorporated a time period from 2000 to the present day (EA-AfA194, 2012). Similar water chemistry data for Scotland were extracted for this project by the Scottish Environment Protection Agency (SEPA, 2012a).

A further set of categorical data were used as descriptors of river quality. The European Water Framework Directive (WFD, 2000) was transposed into UK law in 2003. Member States must aim to reach good chemical and ecological status in inland and coastal waters by 2015. Classification status and environmental objectives, for surface water bodies across England and Wales, have been published in the River Basin Management Plans and are publicly available from the Environment Agency (EA-WFD, 2011). Each water body had a unique identifier with attributes including a georeference and classification status (High, Good, Moderate, Poor, or Bad). Additionally, this project made use of spatial data for WFD waterbody catchments (a series of non-overlapping polygon catchments). These were obtained directly from the Environment Agency (EA-WFD, 2013) and SEPA

#### 16.3.4.3. Flow data

National River Flow Archive (NRFA) hydrometric metadata and statistics are published in the UK Hydrometric Register (Marsh and Hannaford, 2008) for over 1,500 gauging stations. These data present an average of all samples held on the archive for the station over the full period of record (up to the end of 2005). This project made particular use of mean flow ($m^3/s$) and Base Flow Index (BFI). BFI is measure of the proportion of the river runoff that derives from stored sources; for example, the more permeable the rock, superficial deposits and soils in a catchment, the higher the base flow. The UK Hydrometric Register dataset was provided in spreadsheet format by CEH. Only open stations (correct as of 2005) were used.

Finer resolution, daily mean river flow data (EA-AfA186, 2012) were exported from the Environment Agency database for a range of catchments across England and Wales (196 sampling sites). Similar daily mean flow data were extracted by the Scottish Environment Protection Agency (SEPA, 2012b) for Scotland.

#### 16.3.4.4. Freshwater boundary data

Hydrometric Areas, HA, (digital spatial boundary data licensed from the Centre for Ecology and Hydrology) are either integral river catchments having one or more outlets to the sea (or tidal estuary) or groupings of such catchments which have topographical similarity (Marsh and Hannaford, 2008). For convenience and consistency, these boundaries were used to define hydrologically similar areas (total = 97 in mainland Great Britain).

CEH's 1:50,000 Watercourses were used to identify rivers, canals and surface pipes (man-made channels for transporting water e.g. aqueducts and mill leets) (CEH, 2012; Moore et al., 1994).

### 16.3.5. Land designations

Various different types of land designations (legal or less formal) were used by different modules of this research project during model development. Brief descriptions of the types of designation follow. Unless otherwise stated, digital boundary data were downloaded from: Natural England (MAGIC, 2012), Countryside Council for Wales (CCFW, 2012), SNH (2012) or (Scottish_Government, 2012). Temporally variable data were obtained where available and appropriate (i.e. new designations or changes to boundaries).

#### 16.3.5.2. Conservation and land management legislation

National Parks are protected areas of the countryside and, although the land is often privately owned and worked (e.g. for agriculture), National Parks welcome visitors. Formal designation of land into National Parks has been staggered since the first Parks in the 1950s (see further details at Natural

England, 2013a). There are currently 15 National Parks across Great Britain. English and Welsh spatial boundary data were downloaded from aforementioned sources; Scottish data were acquired from the Scottish Government (Scottish_Government, 2012).

An Area of Outstanding Natural Beauty (AONB) is an area of high scenic quality which has statutory protection in order to conserve and enhance the natural beauty of its landscape. AONBs have equivalent status to National Parks as far as conservation is concerned, but AONBs have more limited opportunities for extensive outdoor recreation. This research takes the Scottish equivalent of an AONB as the National Scenic Area (designated by Scottish Natural Heritage).

A Site of Special Scientific Interest (SSSI) is designated for its unique, varied and often threatened habitat, wildlife and/or geology. Public bodies own only about 20% of land designated as SSSIs and they are actively managed (and legally protected) to maintain their conservation interest. Many SSSIs provide opportunities for recreation, although this is not their primary purpose. Many SSSIs are also National Nature Reserves (NNRs) or Local Nature Reserves (LNRs) and these have greater recreational potential. An NNR is a site that is recognised for its wildlife and/or geology and is run by approved bodies, including Natural England, Scottish Natural Heritage, Forestry Commission, RSPB, and many Wildlife Trusts. Almost all NNRs are accessible and provide opportunities for people to experience nature. LNRs are sites for both people and wildlife and these are maintained by district and county councils. To qualify for LNR status, a site must be of importance for wildlife, geology, education or public enjoyment.

#### 16.3.5.3. Public access, parks and gardens

Under the Countryside and Rights of Way Act 2000 (CROW), the public (England and Wales) can walk freely on mapped areas of mountain, moor, heath and down, and registered common land. Two datasets were obtained from Natural England for CROW: Access Layer Data, which consists of all conclusive open country and registered common land, and the Conclusive Register of Common Land. Spatial data for CROW land in Wales was unavailable. For Scotland, the Land Reform Act gives a right of responsible access to almost all land.

Country Parks are significant areas of accessible natural greenspace and were originally established as a result of the 1968 Countryside Act (England and Wales) and in Scotland under the Countryside (Scotland) Act 1967. They are primarily intended for recreation and leisure opportunities close to population centres and do not necessarily have any nature conservation importance. They typically deliver core facilities and services e.g. toilets and daily staff presence) but this is only a requirement for Country Parks with accredited status. There are over 400 Country Parks in England, 52 in Scotland and 35 in Wales. Due to the difficulty in obtaining a spatial dataset of Welsh Country Parks, they are excluded from the analysis.

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. The Doorstep Greens initiative ran from 2001 to 2006 and was the successor to Millennium Greens. These areas were designed to be 'safe, secure and accessible to all' (see Natural England, 2013b).

The Woods for People project (led by the Woodland Trust) has created a UK-wide inventory of accessible woodland (FC, 2012). This data source provides a good representation of recreational woodland sites. Other attributes about the type of trees and amenities Ancient Woodlands are areas that have had continuous woodland cover for at least 400 years. These woodlands are typically more ecologically diverse. The Ancient Woodland Inventory was available through MAGIC (2012), Forestry Commission (for Welsh data) (FC, 2013) and SNH (2012).

Spatial boundary files for Registered Parks and Gardens (England) were available from English Heritage (EH, 2013). Areas of land maintained by the National Trust and the National Trust for Scotland were identified using a National Trust point file (downloaded as a 'points of interest' file in Keyhole Markup Language form (GPSDT, 2013)).

#### 16.3.5.4. Environmental land management and restrictions on development

Greenbelt is a policy for controlling urban growth. Spatial data for English greenbelt (c. 2011) were licensed by Defra from Ordnance Survey (OS, 2011). Presently, there is no national digital spatial boundary dataset for Scottish greenbelt. Each council was contacted for spatial information and PDF maps or ESRI shapefiles were received for all areas of Scottish greenbelt (present and historic). Additionally, there is currently one area of greenbelt in Wales; information on this was found in local development plans (i.e. Newport and Cardiff).

The Environmentally Sensitive Areas Scheme was introduced to offer incentives to encourage farmers to adopt agricultural practices which would safeguard and enhance parts of the countryside. Although the scheme is now closed, existing agreements can run until 2014. The agricultural production module of this project has used historic digital spatial data for Environmentally Sensitive Areas, these are zones in which farmers could apply for funding and do not therefore necessarily reflect agreements taken.

Nitrate pollution prevention regulations bring into force the European Commission nitrates directive (91/676/EEC). The regulations mean that all land which drains into waters polluted by nitrates is designated as a Nitrate Vulnerable Zone (NVZ) and farms within these areas must meet a set of NVZ requirements. For example, farmers must adequately store livestock manure, plan and produce a risk map for its redistribution as a fertiliser to comply with NVZ rules.

#### 16.3.5.5. Descriptors: land type, cover and use

Land cover and land use in the UK have been described in a previous section of this report (Section 16).

Landscape can also be defined based on a unique combination of physical environment and social conditions. These natural areas are taken from National Character Areas in England (159), groupings of Landscape Character Assessment study areas in Scotland (25) and landscape character areas in Wales (48) (data sources were: MAGIC, 2012, SNH, 2012, and CCW, 2012 respectively). In this report, these natural areas are collectively referred to as 'Landscape Character Areas' (LCAs). Although they are regional-scale, the groupings are defined based on natural features of the landscape, rather than political boundaries.

### 16.3.6. Beaches, coast and coastal resorts

Under EC Directive 76/160/EEC, designated beaches are monitored for compliance to bathing water quality standards. Bathing water status for all popular UK bathing places beaches is available from the European Environment Agency (EEA, 2013). Additionally, the Environment Agency monitors and maintains a record of bathing waters in England and Wales and these data were downloaded via Open Government Licence (Geostore, 2013). Scottish designated bathing waters were available from Scottish_Government (2012). The locations and names of additional beaches visited by the public were extracted from: http://britishbeaches.info/.

Registers for Blue Flag status and Seaside Awards were used as indicators of beach quality or tourist appeal. Beaches are awarded the Blue Flag based on their conformity with 32 criteria covering: environmental education and information; water quality; environmental management; and safety and services (Blue Flag, 2013). The Blue Flag Programme is an international award scheme for labelling sustainable beaches and is maintained by a non-profit NGO. Fifty-five beaches in England were awarded Blue Flag status for the 2013 season. In Wales, 39 beaches were awarded the Blue Flag for the 2012 season (latest available data). Scotland had three Blue Flag beaches in 2012 and 2013. The UK national Seaside Award recognises and rewards beaches which achieve the highest standards of beach management (Keep Britain Tidy, 2013). There were 112 Seaside Awards presented in England in 2013 and 108 in Wales. Keep Scotland Beautiful's National Beach Award recognised 59 of Scotland's well-managed beaches in 2013 (Keep Scotland Beautiful, 2013).

Definitions for coastal resorts (or seaside towns) in England were taken from official government publications (Beatty et al., 2008; 2011; Humby, 2013) and from authors of the original reports. A seaside destination is any seaside settlement to which people travel for the beach and associated activities. Three tiers of resorts were distinguished based on their size: small (population below 10,000), medium (population 15,000 to 39,999) and large (population greater to or equal to 40,000). The spatial areas for these resorts were defined using Lower Super Output Area (LSOA) boundaries. A list of LSOA was provided for the former through personal communication with Sheffield Hallam University and LSOAs were defined from supplementary data provided by Humby (2013) for the latter two categories of resort. The definition of Welsh seaside towns came from Beatty et al. (2009) and represents coastal resorts with a population of approximately 1,500 to 66,000.

Ordnance Survey Open Data (Strategi) were used to define the coastline (high water) (OS, 2013c).

### 16.3.7. Recreational areas

The OpenStreetMap (OSM) project creates and distributes free geographic data (OSM, 2013). OSM data were used to provide an initial spatial definition of parks, paths, sports pitches, playgrounds, recreational lakes and recreational rivers (different to Section (water data)). There are several reasons for choosing OSM data for this research. First, these data are highly detailed, especially surrounding urban areas, and coverage across the UK is good. Second, the OSM project is an open-source resource and as such the (spatial-literate) public can upload data representing areas known to them. As such, the final product is likely to be updated frequently and a truer reflection of what is on the ground. However, as with any publicly sourced data, quality control is more sporadic.

### 16.3.8. Socio-economic and associated data

Socio-economic information was gathered to ascertain impact (or potential impacts) on human systems and natural systems. Key datasets are as follows:

- Demographic data
  - A range of population summaries (e.g. total usual resident, adult, ethnic minorities, retired) were sourced from Census data (Casweb, 2013) or mid-year estimates (GROfS, 2013). These demographic data were taken at the intermediate geography level (i.e. aggregates of Census Output Areas, known as Lower Super Output Areas (LSOA) in England and Wales and Data Zones in Scotland). Corresponding boundary data were downloaded from UKBorders (2013).
  - LSOAs are a geographic hierarchy designed to improve the reporting of small area statistics. They were designed in 2004 from groups of 2001 Output Areas (typically 4 to 6). LSOAs have a minimum population of 1,000 (with an average of 1,500) and a minimum resident household of 400 (with an average of 630) (ONS, 2013). There have been some (minor) changes in boundaries of LSOA from 2001 to 2011.

- Higher-level geography (urban centre) population figures were taken from the ONS (2001) and GROfS (2001). Also, boundaries for DLUA were taken from the OS (2013b).
- Household-level economic data
  - Median household income was extracted from (Experian, 2008).
- Travel/ connectivity data
  - The Meridian 2 road network (OS, 2013b) and travel times from Jones et al. (2010) were used as raw data to derive travel times.

## 16.4. Methodology

Using a GIS, raw source data were translated into a common spatial unit for analysis.

### 16.4.1. Defining the extent of Great Britain

Great Britain includes all surfaces enclosed by inland borders. A definition of total area may be restricted to land only or include inland water in the littoral zone. The Extent of the Realm usually refers to the Mean Low Water Mark but in some cases boundaries extend beyond this to include offshore islands. A definition 'clipped to the coastline' (Mean High Water Mark) gives the Great Britain a more orthodox appearance; this area is over 23 million hectares (ONS, 2012). This total area can be subdivided into different geographical hierarchies, based on arbitrary zones, administrative or political areas, or based on natural land attributes.

The aforementioned spatial and spatio-temporal datasets, describing physical and social phenomena, originate from multiple source geographies. A common spatial unit was desirable for consistency across the different systems. Choice of this spatial unit (fitness for purpose) was a compromise between resolution, processing time and quality. The effects of scale and aggregation of spatial data (MAUP) have been introduced in Section 15 and further detail can be found in Openshaw (1984).

These data were integrated to a 2km grid, with this choice of resolution being a lowest common denominator given the highest detail at which agricultural land use data could be obtained. A non-overlapping continuous 2km grid across Great Britain encompasses approximately 57,000 individual cells.

When overlaying multiple spatial datasets, there will inevitably be some partitioning of grid cells. The following sections of the methodology discuss data interpolation and manipulation required to derive variables for the different models in this report. Where necessary, spatial data are re-projected to the standard OSGB 1936 British National Grid spatial reference system.

### 16.4.2. Elevation and slope

Elevation and slope variables were derived from the 50m resolution IHDTM (obtained as an ASCII raster and manipulated in a GIS). Average elevation for a 2km cell was simply the aggregate of all 1,600 elevation values in the corresponding IHDTM grid divided by the sum of cells.

Slope (degrees inclination) was calculated from the 50m IHDTM as the maximum rate of change in value from a cell to its eight neighbours. An average slope value was then taken for an entire cell.

Further to these two standard average-per-2km-cell variables, Section 5 required farmland-specific variables (here, farmland is inclusive and defined as all crops, grasses and other land on farms). Average elevation on farmland was calculated as a weighted average from a 25m resolution base definition of farmland (LCUAP2, 2010); in practice, this operation was calculated as a sum, for each 2km square cell, of the following: (elevation × (area farmland/area of land)). The approach was similar for slope. A final terrain variable was the proportion of land that is farmland and greater than six degrees inclination. All variables were calculated in a GIS and output as TERRAIN (2012). This dataset is used by models in Sections 5 and 7.

The 2km resolution average elevation data described above is further used to define a 2km resolution Digital Elevation Model (DEM). This DEM is used to calculate flow direction (path of steepest descent), flow length downstream and flow accumulation. In turn, flow accumulation defines a stream network (where number of cells draining to a cell > 25). The calculation of these variables is performed using standard hydrology tools in a GIS; however, it is also an iterative process in this case as minor modifications were made to the DEM to ensure that the river network and drainage basins showed reasonable correspondence with river and boundary data (Marsh and Hannaford, 2008, and CEH, 2012). Final variables were specifically developed for Section 10 (water quality; WATER, 2012).

### 16.4.3. Soil and hydrogeology

All soil variables were derived from HWSD (FAO/IFA/IIASA/ISRIC-WSI/ISSCAS/JRC, 2009) and pertain to the topsoil (0-30cm) unless stated otherwise. The source raster data was converted into vector format to allow the addition of an attribute table and the intersection of the 2km grid. Percentage total are in each class of interest in the 2km cell was then taken, or area-weighted averages were taken if more appropriate. Processing was carried out in a GIS and exported to MS Excel (SOIL, 2012). See variable descriptions in individual sections for further details.

Superficial deposits were reclassified into permeable (blown sand, Crag Group, glacial sand, landslip, raised marine, river terrace, other sand/ gravel) or low permeability (alluvium, Brickearth, clay-with-flints, Drift geology, lacustrine, peat, till). For hydrogeology, the Class 3 attribute was simplified (highly and moderately productive aquifers, low productivity aquifer and rocks with essentially no groundwater). Newly classified superficial deposits and hydrogeology layers were combined based on a classification scheme (Table 16.2). The 2km grid was then overlain on the reclassified surface and the minimum value in a 2 km cell was taken.

**Table 16.2: Classification scheme for hydrogeology data**

| Class | Rule |
|---|---|
| 1 | High/Moderate productive aquifer AND permeable cover |
| 2 | High/Moderate productive aquifer AND low permeable cover |
| 3 | low productivity aquifer AND permeable cover |
| 4 | low productivity aquifer AND low permeable cover |
| 5 | No groundwater |

Observations for BFI originate from Marsh and Hannaford (2008) at gauging station level. Summary statistics are then taken at the HA-level (average and minimum BFI). The 2km grid was then overlain, taking the minimum BFI from summarised HA-level data if a 2 km cell crosses the boundaries of more than one HA.

Both hydrogeology and BFI variables can be found in GROUNDWATER (2013).

### 16.4.4. Climate

#### *16.4.4.1. Baseline Climate Data*

The grids were subsequently summed to create annual and growing season (April to September) totals for each 5km grid cells and these values were subsequently bilinearly interpolated to estimate values for each of the 57,230 points on the 2km resolution mesh covering Great Britain. Initially there were 239 points in coastal locations with missing data values for the climate variables so further processing was undertaken in ArcGIS to assign each of these points with the value of the closest point with complete estimates. None of these assignments involved using data from points more than 2,850m distant. The final dataset was stored as CLIMATE (2012).

#### *16.4.4.2. Scenario Climate Data*

Individual monthly values were further summarised in an Excel workbook as follows:

- Calculate the average of the daily maximum and minimum temperature changes for each month;
- Average these monthly mean values for the six growing season months in each year;
- Average the monthly precipitation change values for the six growing season months in each year.

These growing season totals were joined onto the polygon grid of 2,028 cells. Many of these had null values (e.g. areas of sea) so a second version was extracted with the 440 cells of 'non-null' values. A central point was then generated for each cell and the coordinates re-projected to the UK National Grid. Processing was then carried out in ArcGIS to assign each of the 57,230 points in the baseline 2km climate mesh with the change values for the nearest location in the 440 point scenario data. Once this integration had been achieved it was straightforward to calculate new absolute values of average growing season temperature (°C) and total precipitation (mm) for the for future decades (PROJECTIONS, 2013).

Figures 16.1 and 16.2 below show growing season mean temperature and precipitation for the 1961-90 baseline and 2030s and 2060s projections (PROJECTIONS, 2013). These maps imply that areas with < 300 mm precipitation are likely to expand to cover most of lowland England by the 2060s, with mean temperatures increasing to over 15°C. Upland areas of Britain are projected to be less impacted by changes in precipitation but mean growing season temperature increases of around 2°C are quite widespread. It is important to recognise that there is likely to be much annual variability around these middle point projections but changes of this magnitude would clearly have considerable implications for the suitability of different agricultural activities across Britain.

**Figure 16.1 Mean growing season temperature for the 1960-90 baseline, 2030s and 2060s projections.**

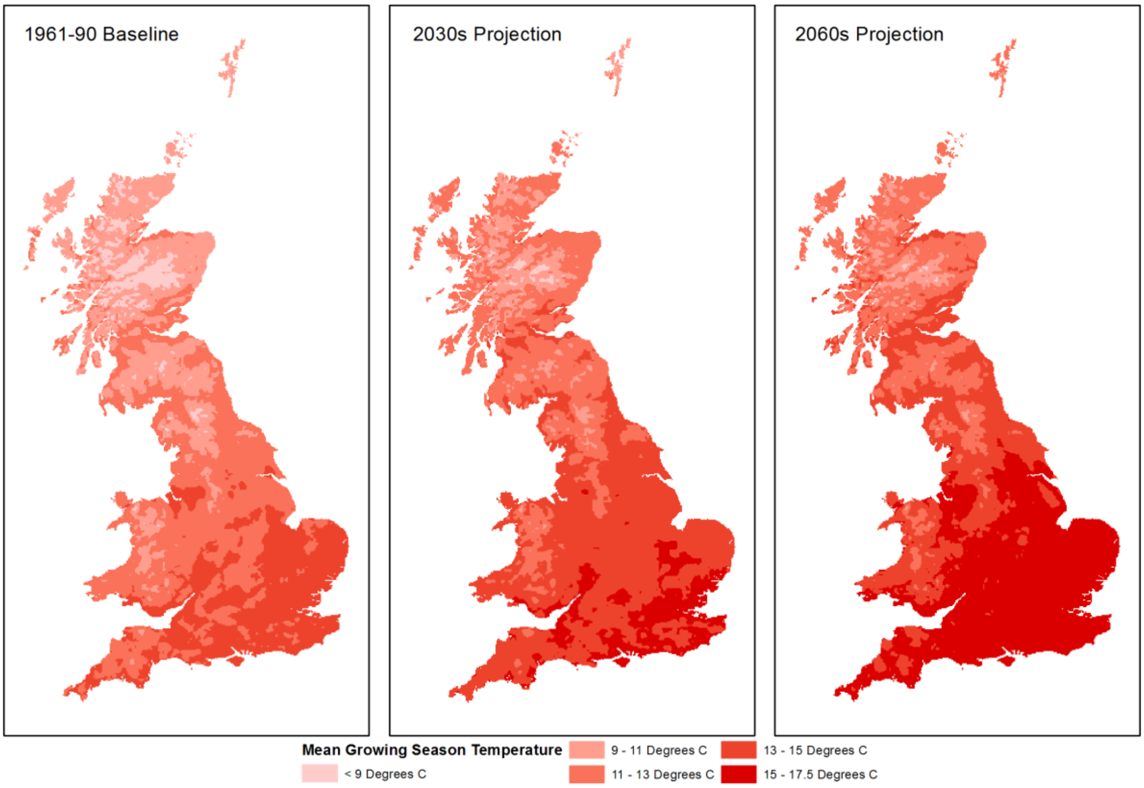
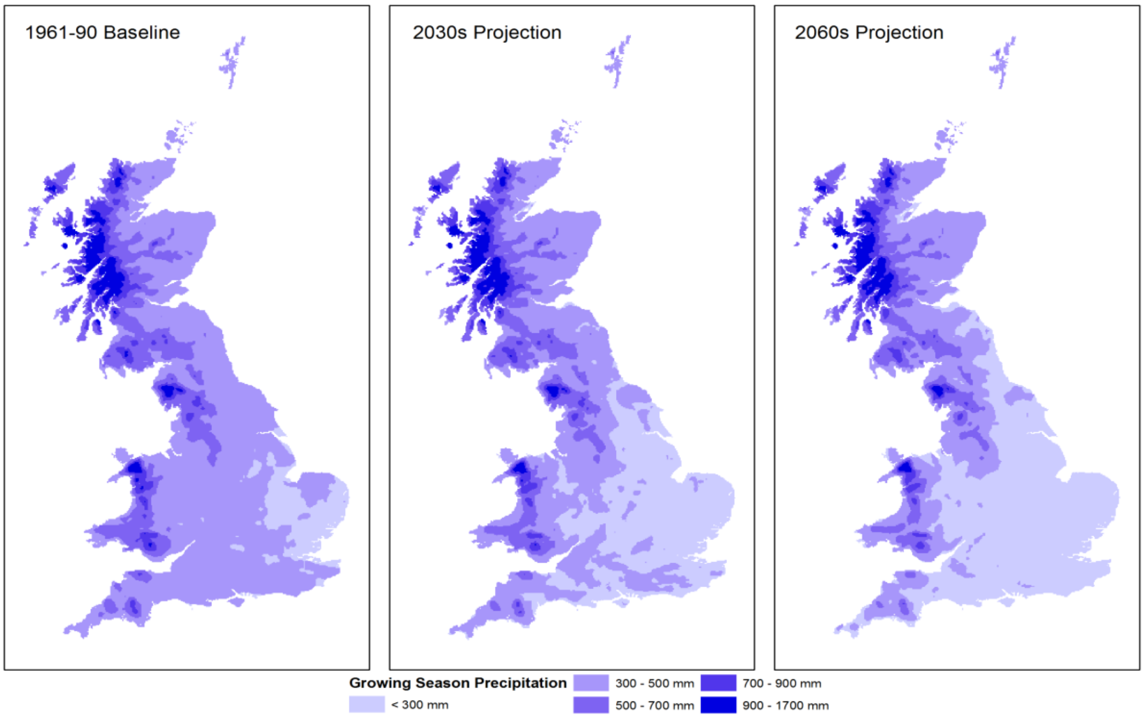


**Figure 16.2 Mean growing season precipitation for the 1960-90 baseline, 2030s and 2060s projections.**




### 16.4.5. Water

The numeric water quality and hydrometric datasets were of a high quality, with fine temporal and spatial resolution. However, they required further sampling and post processing to be fit present purposes. For example, geographic references (sampling locations) for NRFA data and Environment Agency/ SEPA mean flow and water quality samples (described above) were not always consistent with each other or were ambiguous. Processing of water flow and quality observation data is described in the relevant section of this report (Section 10).

Categorical data were selected (sub-sampled) where they coincided with the derived 2km resolution stream network (in the case of GQA data) or within 5km of this network (WFD data). Adequate positioning of the observations on the river network was important; for example, sampling on a tributary must not be assigned to the main river channel at 2 km resolution. Any ambiguous points were removed from the observation dataset. Deviations were ascertained by a manual comparison of individual sample locations with centreline watercourse data (CEH, 2012).

In an initial test phase of The Integrated Model priority woodland was established based on WFD status variables. When these were combined it became apparent that there were many overlaps and sliver polygons on the England-Scotland border which required considerable editing in the ArcGIS software to correct. Another complication was that the WFD status assessment spreadsheets covering all of England, Wales and Scotland did not contain consistent attributes which limited the range of water quality characteristics that could be assessed. Ultimately, by linking the two sets of data using WB-ID codes it was possible to map 8,169 RWBs with a range of WFD status variables.  Other water body polygons such as lakes and coastal locations were not assigned any status variables and coded as -1 (Null values) for the purposes of subsequent analysis (RWBs, 2013).

Additionally, physical response of a watercourse may be influenced by its morphology and this may be proxied by a descriptor of type. Therefore, presence of canal in 2km cell and presence of a surface pipe in a 2km cell were ascertained from CEH watercourse data (CANALS, 2012).

### 16.4.6. Land designations

Welsh greenbelt was digitised to clip to road and county boundaries using information found in Newport Unitary Development Plan (1996-2011)[1]. Scottish greenbelt PDFs were geo-referenced and digitised (scales typically ranging from 1:8,000 to 1:25,000). These national datasets were united with a simple shapefile for England to get total greenbelt in Great Britain. The 2km grid was then overlain and the percentage area of greenbelt in the cell was calculated from the intersection of the two datasets GREENBELT_EW (2012) and GREENBELT_S (2013).

Spatial boundary data for other land designations were available as ESRI shapefiles and the 2km grid was simply overlain.

### 16.4.7. Socio-economic and associated data

Raw demographic statistics, at LSOA-level, were assigned to a LSOA boundary or population weighted centroid (where appropriate, see individual sections for further details). Some statistics have not yet been released for intermediate-level geographies. In these cases, 2001 data were used.

Estimates for the population on mains sewerage, and those using septic tanks, were calculated using DLUA boundaries, LSOA (or Data Zone) boundaries and statistics for the total resident population. First, it was assumed that population was evenly distributed across a LSOA (or Data Zone). Each LSOA (or Data Zone) was given a population density. The DLUAs were then given a 250m buffer and it was assumed that all people within these areas were on mains sewerage, and by default those outside were on septic tank systems. Overlaying the population density surface with the mains sewerage area, and then the 2km grid, allowed an estimate of how the treatment of human effluent is shared in a 2km cell (SEWAGE, 2013). See Section 9 for further details.

### 16.4.8. Beaches, coast and coastal resorts

Beach and bathing data had geographic references and were added as points into ESRI's ArcGIS. Extra attributes were joined by name, where appropriate (e.g. possession of Blue Flag award). Any beaches which were noted in published statistics, i.e. as award winners, but were not otherwise part of spatial datasets were digitised manually.

The spatial extent of coastal resorts was defined by groupings of LSOAs. Where this information was not available (i.e. for Welsh resorts), resort names were matched to OS Meridian Developed Land Use Areas (OS, 2013b).

### 16.4.9. Recreational areas

OSM data were downloaded using an open source software tool called 'Osmosis'. This is a command line Java application which can rapidly process OSM data, and it enables the user to selectively extract data based on elements (nodes, ways and relations) and their tags (keys and values). Data were subsequently converted into an ESRI shapefile format using a two-step importation and conversion

---

[1] Accessible at:
http://www.newport.gov.uk/stellent/groups/public/documents/plans_and_strategies/cont063489.pdf

process with POSTgreSQL (open source object-relational database system) and OpenGeo Suite (open source geospatial software for managing maps and data). Once in shapefile format, data were imported into ESRI's ArcGIS for further processing. First, they were re-projected in to the British National Grid (Projected Coordinate System) and they were then edited and combined with other data sources. Further details follow.

OSM data on parks were edited to remove the following: any areas with access restrictions, including all schools and their recreational grounds; sports clubs; any buildings or parking areas; and areas with a primary land use that would challenge recreational use (e.g. cemetery, allotments and farms). Additionally, very small 'parks' (< 10,000 m2) were removed if they did not contain a playground or were not given a name in OSM. This latter data cleansing process removed small areas that have been classified as a generic 'park' in OSM and are likely to be small community grassland features such as roundabouts or pedestrian areas.

As an intermediary step, the National Trust point file (see Section 16.3.5) was converted into shapefile format and re-projected. The points were overlain (with 150 m tolerance) on English Heritage's Registered Parks and Gardens dataset (see Section 16.3.5). Where selected National Trust-Parks and Gardens were also in the OSM-derived parks dataset, they were removed from the latter.

Next, multiple data sources were merged to obtain a spatial footprint of all major open-access recreation areas. These data sources were: the edited OSM-derived data on parks, National and Local Nature Reserves, Millennium and Doorstep Greens, Woods for People, Country Parks and National Trust properties (see descriptions in Section 16.3.5). Within each of the new recreational areas, the area of land and attributes (e.g. type of wood) under each of these categories were summarised. Additional attributes joined from processed OSM data were: area of pitches, area of playgrounds, length of rivers (inside recreational areas and within 25 m of the boundary), and lake area and perimeter. The amount of land under special types of designation was also calculated (e.g. National Parks, Areas of Outstanding National Beauty).

Finally, the habitat within each park was summarised according to the UK-NEA definition (baseline year 2010; UK-NEA, 2011). The UK-NEA habitat shapefile (1 km resolution) was intersected with the parks layer using tools in Geospatial Modelling Environment (Beyer, 2013).

### 16.4.10. Recreational paths and walks

Connected OSM-derived paths were grouped by a process of applying a small buffer (10 m), dissolving the boundaries of overlapping polygons, assigning a unique ID to the polygons and then joining the polygon ID to each coincident line. Lines were then grouped by polygon ID. Resulting groups that had a total connected line length less than 1000 m were deemed minor places for recreation and were removed.

The terrain (habitat) traversed by each path was summarised by taking an intersection of the path data with the UK-NEA habitat data (UK-NEA, 2011). Spreadsheet-editing software was then used to calculate the length-weighted habitat. A similar intersection was performed to get the total path length in each National Park, in any Area of Outstanding National Beauty and in registered common land (CROW). The length of path beside a river or lake was calculated by taking buffers around these features (10, 50 and 100 m), dissolving their boundaries and taking an intersection with paths.

Additionally, a special category of paths were those along the coast. Buffers of 100 m and 500 m were applied to the coastline and paths were given a TRUE or FALSE indicator if they intersected these.

## 16.5. Discussion and summary

Discussion of supporting data and derived variables can be found in the relevant sections of this report (and references therein). However, general observations are as follows:

- Together these datasets provide a broad set of physical and social descriptors; however, they are not exhaustive.
- For modelling purposes it is necessary to reduce the complexities of ecosystems and care must be taken to not over-simplifying phenomena.
- Natural features cross artificial boundaries and therefore some spatial units are more appropriate than others.
- Due to a lack of alternatives, considerable simplifying assumptions were used for some of the variables (e.g. sewage).