Building a 1911 Historical Land Capacity GIS

Max Satchell and Robert J. Bennett

aems2@cam.ac.uk rjb7@cam.ac.uk

Working Paper 17:
Working paper series from ESRC project ES/M010953:
Drivers of Entrepreneurship and Small Businesses

University of Cambridge, Cambridge Group for the History of Population and Social Structure, Downing Place, Cambridge, CB2 3EN, UK.

May 2019

Comments are welcomed on this paper: contact the authors as above.

> Satchell, Max, and Bennett, Robert J., (2019) *Building a 1911 Historical Land Capacity GIS*, Working Paper 17: ESRC project ES/M010953: 'Drivers of Entrepreneurship and Small Businesses', University of Cambridge, Department of Geography and Cambridge Group for the History of Population and Social Structure.

Keywords: Land capacity, Farms, Land Use, Entrepreneurship, Census

JEL Codes: Q1, Q12, Q15, Q24, L26, D13, D22

Building a 1911 Historical Land Capacity GIS

Max Satchell and Robert J. Bennett

Working Paper 17: ESRC project ES/M010953: Drivers of Entrepreneurship and Small Businesses, University of Cambridge.

1. Introduction

This paper outlines the construction of a historical land capacity GIS for the historical date of c. 1911. The aim of the paper is to construct a database of land capacity that can be used to assess the influence of land on how farmers operated as business proprietors in England and Wales over 1851-1911. This is an important input to the ESRC-supported project ES/M010953 *Drivers of Entrepreneurship and Small Businesses* which has supported the work. A good GIS of land capacity allows investigation of the influence of land/soil capacity on the organisation of farm business operations (as employers or own account), and the size of their business, as well as measuring the potential for the agricultural economy to support wider business development.

It is sought to make the data constructed correspond to the census year 1911. Once created the land capacity GIS can give summary values of type and area for each GIS unit related to census data. The initial output is for census parishes and Registration Sub-Districts. The data mainly derive from combining the GIS files for Agricultural Land Capacity from the 1960s/70s and the Dudley Stamp Land Use Survey data from the 1930s. The GIS output prepares the ground for extending to other data sources in the future, and to other time periods.

The paper is structured as follows. Section 2 summarises the GIS units of census data and the various classes of GIS land use data that potentially pertain to nineteenth century agriculture. Section 3 discusses the Agricultural Land Capacity (ALC) GIS in more detail and explains why the decision was taken to build the historic land use dataset using the ALC GIS datasets as the primary source and how this was achieved. Section 4 discusses modelling historical non-agricultural land-use: woodland, urban and water. Section 5 outlines how the various datasets were assembled into a single shapefile. It also summarises the database.

2. Geographical Units and data sources.

Hitherto historical understanding of the development of farming has been held back by the lack of data on the land capacity at a local level, and estimated for small administrative units such as parishes. Modern digitized data of land suitability / capacity potentially fill this gap, if it is possible to extrapolate backwards to the land capacity of the areas used by farmers in the past. This paper develops a method to give this backwards extrapolation. The outputs sought form a database at the spatial scale of census parishes and Registration Sub-Districts (RSDs). The aim is to link the spatial units adopted for the published census of population and I-CeM digital data of the census, initially for 1911 I-CeM is the digital versions of the census now available at UKDA.[1] For analysis of entrepreneurship a supplement to I-CeM is to be deposited at UKDA as the British Business Census of Entrepreneurs (BBCE).[2] The land capacity data base developed here is aimed at linking with these sources. To facilitate the use with I-CeM/BBCE, in addition to a GIS base for 1911, the data are extrapolated backwards to cover the same units of parishes and RSDs for the whole period 1851-1911. In this initial stage the land capacity data is not adjusted from the 1911 baseline, so that users need to be aware that it does not change, thought the spatial units for which it is constructed do. In future it should be possible to adjust the land capacity data over time to take account of major changes that occurred, such as substantial land drainage.

The outputs are for two spatial scales: (1) parish level units using CONPAR (Continuous parish) polygons for two periods 1851-1891 and 1901-1911; and (2) for Registration Sub-Districts (RSDs) for every census year 1851-1911.[3] Both datasets derive from a common source: the CAMPOP England and Wales Census Parishes GIS. These are referred to as Conpar shapefiles. The Conpars are single parishes or aggregations of them configured so

[1] Higgs, Edward and Schürer, Kevin (University of Essex) (2014) *The Integrated Census Microdata (I-CeM)* UKDA, SN-7481, derived by FindMyPast using a variety of original FMP transcriptions. Version 2 of I-CeM includes a range of valuable additional inputs from colleagues at Campop; see Schürer, K., Higgs, E., Reid, A.M., Garrett, E.M. (2016) *Integrated Census Microdata V.2 (I-CeM.2).*

[2] British Business Census of Entrepreneurs (BBCE) forthcoming.

[3] See data acknowledgements at end of this paper.

that they provide an identical geographical coverage, thus avoiding any spatial discontinuities in boundaries over time, following the method developed by Wrigley (2011).[4] The creation of continuous units is discussed by Satchell *et al*.[5] As continuous units these enable population data to be mapped consistently for two sets of files: 1851-1891, and 1891-1911. About a quarter of census parishes have to be amalgamated to create the continuous data series.

The CONPAR units are based on a dataset created by Kain and Oliver (2001), who derived their boundaries from a variety of maps of various scales most which were at 1:63,360, or lower levels of resolution. This has resulted in a relatively coarse and generalised set of boundaries which means that comparison with some other GIS datasets inherently suffers from a degree of geographical imprecision..

The RSDs of England and Wales number 2,009-2,190 units for each census year from 1851 to 1911. The RSDs were based on poor law unions and were designed to combine where possible urban and rural settlements. Hence, they not only conflate physical regions, but also mix urban and rural. Quantifying the relationship between towns and RSDs is helpful in thinking about the spatial relationships between land use data, the extent of urban land, and the typical size of the unit of analysis. Of the 2,009 RSDs extant in 1911, 1,198 either are encompassed by or intersect with the outline of a town or city. In most areas these RSDs are large and predominantly non-urban places having a mean size of 14,460 acres. They are a little smaller than those RSDs which do not intersect with towns or cities, which have a mean size of 25,599 acres. Sixty-one RSDs have no agricultural land because they are wholly

[4] Wrigley, E.A. (2011) *The Early English Censuses*. Oxford: Oxford University Press.

[5] M. Satchell, G. Newton, E.A. Wrigley, K. Schürer, C. Roughey, M. Anderson, L. Shaw-Taylor, 'Continuous Parish Units of England, Wales and Scotland 1851-1891 shapefile' (2013). A description of the dataset can be found in M. Satchell, 'Continuous Parish Units of England, Wales and Scotland 1851-1891 GIS shapefile documentation' (2017) available at:
https://www.geog.cam.ac.uk/research/projects/occupations/datasets/documentation.html;
M. Satchell, G. Newton, E.A. Wrigley, K. Schürer, C. Roughey, M. Anderson, L. Shaw-Taylor, 'Continuous Parish Units of England, Wales and Scotland 1851-1911 shapefile' (2013). A description of the dataset can be found in M. Satchell, 'Continuous Parish Units of England, Wales and Scotland 1851-1911 GIS shapefile documentation' (2018) available at:
https://www.geog.cam.ac.uk/research/projects/occupations/datasets/documentation.html. The Continuous Parish Units datasets have been created using Satchell, M., Kitson, P.M.K., Newton, G.H., Shaw-Taylor, L., and Wrigley E.A., '1851 England and Wales census parishes, townships and places shapefile' (2016). A description of the dataset can be found in M. Satchell, '1851 England and Wales census parishes, townships and places: documentation' (2016): https://www.geog.cam.ac.uk/research/projects/occupations/datasets/documentation.html. The Satchell *et al* dataset is an enhanced version of Burton, N, Westwood J., and Carter P., GIS of the ancient parishes of England and Wales, 1500-1850. Colchester, Essex: UK Data Archive (May 2004), SN 4828, which is a GIS version of Kain, R.J.P., and Oliver, R.R., Historic parishes of England and Wales: An electronic map of boundaries before 1850 with a gazetteer and metadata. UK Data Archive, May, 2001. SN 4348.

within cities or towns. These are very small having a mean size of 484 acres. Because the mean size of CONPAR parish units is so much smaller than RSDs, any distortion caused by scale problems in the cartographic sources used for land use data is more likely to be significant at parish than RSD level.

The GIS units that are used for mapping and analysing census data, the various classes of GIS data that pertain to the analysis of agriculture, and the degree to which they can be used to construct an agricultural land use GIS are strongly influenced by the graphic scale at which they are available. As the scale of a source paper map increases some features disappear completely and the boundaries of others become more generalised sometimes to such an extent that their areal data is fundamentally distorted. This is not an issue for a GIS with a high resolution - features do not change their attributes whatever the scale they are displayed at. But it is an issue of importance when a shapefile has been digitised at low resolution, or from large scale/ less detailed *paper* maps, which at their worst may have had some features cartographically enlarged for their real size in order to make them more visible in a printed map. In short, the use and comparison of geographic data from paper maps with vastly different source scales can lead to significant errors in geographic data processing. The issue of scale as a source of noise in the various datasets used to generate the ALC GIS will be touched on for each source dataset in the course of this paper.

2.1. Historic Land Use Data

The first historic land use data available at a national level is a GIS of the Dudley Stamp land utilization survey (Environment Agency, 2009; For details see, Webster et al., 2010). This GIS, which was supplied by the Environment Agency, derives from a digitisation of 89 1:63,360 scale land utilisation maps Dudley Stamp, 1933-49). These provide coverage for all of England and Wales and were published between 1933 and 1949, though for rural areas most of the survey work was done between 1931 and 1934. The GIS is very detailed and consists of 1,494,322 polygons representing seven classes of land use. These classes derive from the printed maps and comprise:

(1) Forest and woodland;

(2) Meadowland and permanent grass;

(3) Arable including tilled land, fallow and rotational grass;

(4) Heathland, moorland, commons and rough hill pasture;

(5) Gardens and allotments;

(6) Agriculturally unproductive land such as built-up urban areas and quarries, railways and roads;

(7) Rivers, streams, watercourses and lakes.

The Dudley Stamp survey was largely conducted by volunteers who used letter codes representing different types of land use to mark-up field sheets. These were map sheets of the Ordnance Survey 1:10,560 first and second revision series. These were then collated and the land classification was indicated by colour overprinted on the 1:63,360 "Popular Edition" of the 1 inch mapping of the Ordnance Survey. The scale of the paper mapping is sufficiently detailed that it should not introduce distortion into the ALC shapefile.

Despite the ambitions of Dudley-Stamp (DS), who styled the program as a second Domesday survey, there are several reasons why the land utilization GIS has inaccuracies. One issue was that the original field work was largely conducted by untrained volunteers. The volunteers were mostly school children supervised by their teachers, and some of the distinctions in land use that they were required to make were very difficult to identify. For example, individuals not closely connected with farming would have had problems distinguishing between permanent pasture, rotational grass, spring sown cereals in their early stages, and non-agricultural grassland. In addition, Coppock (1978, p.57) suggested that errors may have occurred when the data were transferred from the 1:10,560 field sheets to the 1:63,360 maps because this was drawn in by hand. Taylor et al. (2010), who conducted a GIS analysis of a block of land of c. 8,800 acres in east Sussex as depicted on the field sheets and printed map of the DS survey, found significant differences in the area of permanent grass and rough grazing. Caution should be exercised before attributing this particular discrepancy to drawing error or assuming the problem was widespread, because the transferring of information from the field sheets was done by professional geographers and cartographers.

Further complications occur because of the periods when the survey was conducted. Though most of the rural part of the survey occurred between 1931 and 1934 some elements are significantly later in date. This is because the editors continued to update their maps to incorporate major changes in land use that they knew about up to the date when the maps were published, the last of which was in 1948-9; urban development, afforestation, and the conversion of agricultural land to airfields were the main changes recorded.

Further issues are also caused by the Environment Agency’s GIS simplifying and distorting the content of the original DS paper maps. The GIS uses a raster colour recognition method to convert the land use colours, signifying different sorts of land use on the paper map, to land-use polygons (Bailey, 2007; Webster et al, 2010). This worked reasonably well for those classes of land use where representation was dependent only on colour, and where the degree of variation locally was small. However, in instances where interpretation of land use was also dependant on context, information was lost. For example, the DS paper maps used the same colour for agriculturally unproductive lands, first class roads, buildings, yards, quarries etc. In addition, land use subdivisions on the paper maps were shown by symbols to distinguish woodland as deciduous, coniferous, mixed and new plantations; these symbols were omitted from the GIS. In addition the raster extraction process widened linear features and enlarged small ones. Also the GIS as supplied by the Environment Agency also did not conform to the standards one would expect of a publicly accessible digital resource in that it had no dataset documentation and a topology check identified overlapping polygons as a significant problem.

The main constraint that arises with using the DS survey, however, is that it covers 1931-4, with lesser information included up to 1949, and we are seeking to construct a baseline of historic land use for 1911. We can overcome this difficulty by noting first that the farming economy did not changed fundamentally from 1911 to 1931, so that many categories of land use did not change. And second, that we limit the initial use of the DS GIS mainly to its information on non-agricultural use and woodland. Hence, whilst the DS GIS in itself cannot be used to indicate land capacity/ use in 1911, its classifications provide a valuable source of supplementary evidence.

3. Modern Land Capacity Data

The modern data used to infer historic land quality is derived from the MAFF land classifications. Before discussing these in detail it is necessary to touch on the issue of whether historic change in the capacity of land to support particular types of farming makes it problematic to use modern data. Change in the capacity of land to support particular types of farming could derive from human agency, such as drainage of wetland, the unintended consequences of human activity, such as the permanent degradation of land due to ill-advised farming techniques, or from climate change, such as prolonged episodes of cooler temperatures. The degree to which historic agricultural land might differ from the present remains largely un-investigated, but for the period 1851-1911 to the present is probably minor for most locations. The major human contribution to land improvement was field drainage, which although well developed by 1851 continued to be undertaken throughout the period. However, climate had not changed in major ways over the period, and while there were examples of the irreversible degradation of agricultural land they were relatively localised.

Agricultural land capacity (ALC) mapping is a framework for classifying land according to the extent to which its physical characteristics impose long-term limitations on agricultural use. The ALC mapping was initiated as part of a national survey by the Agricultural Land Service between 1966 and 1974. This material is now held by the MAFF (see: MAFF, 1988). The ALC classification aims to assess the capacity and limitations of an area of land that affect the range of crops that can be grown, the level of yield, the consistency of yield, and the cost of obtaining it. ALC coding provides a system for classifying land according to the extent its long term potential for agricultural use in terms of physical characteristics. Non-physical factors that influence the type and profitability of land use, such the type and quality of farm management, the availability of capital, distance to markets or road quality, are not taken into account. However, ALC classes are not wholly divorced from human agency. They do acknowledge the influence of long term changes resulting from human influences on land capability, such as large scale drainage schemes.

The grading scheme was first conceived in 1966. This was a simple approach that classified agricultural land into five grades that were intended primarily to assess agricultural land being considered for non-agricultural purposes:

Grade 1 - excellent quality agricultural land

Grade 2 - very good quality agricultural land

Grade 3 - good to moderate quality agricultural land

Grade 4 - poor quality agricultural land. Land with severe limitations which significantly restricted the range of crops and/or level of yields; mainly suited to grass with occasional arable crops the yields of which are variable.

Grade 5 - very poor quality agricultural land. Land with very severe limitations which restricted use to permanent pasture or rough grazing, except for occasional pioneer forage crops.

The grading scheme also had categories for urban land and other land not in agricultural use. This classification was applied to the national survey by staff of the Agricultural Land Service between 1966 and 1974. The output of their work was 113 maps at a scale of 1:63,360 with each sheet accompanied by a report.

As originally conceived the ALC categorisation can be criticised on a number of counts. First, it was found that ALC category 3 land constituted about half the agricultural land, and as such, was too broad. Second, the 1967-1973 fieldwork was conducted with soils and climatic data that is now outdated. Third, the scale of the fieldwork was coarse. ALC grades were generally not assigned to areas smaller than 80 hectares (197.7 acres).

Revised ALC guidelines issued in 1988 remedied many of these faults (MAFF, 1988). ALC class 3 was subdivided into 3A and 3B. Grading was to be based on superior soils data, and superior and more detailed climatic data was to be employed, with a methodology for interpolating it using a distance weighted mean. Unfortunately this superior methodology has not yet been fully implemented. For England it has only been applied piecemeal as particular sections of agricultural land have been considered for development. Hence, most land still remains under the old classification. Wales by contrast conducted a complete national survey in 2017. These developments are represented by three ALC GIS datasets discussed below: the original for 1967-73; new surveys for small parts of England; and a complete new survey for Wales.

3.1 Agricultural Land Class (Provisional) 1967-1973

The original ALC survey GIS has complete national coverage and was digitised from Agricultural Land Classification 1: 250,000 scale maps. These were redrawn from 113

Agricultural Land Classification 1:63,360 scale maps. The fieldwork for the 1:63,360 series was that conducted by MAFF between 1967 and 1973. Not only does the GIS data suffer from the limitations of the original survey, but the digitisation was from the less detailed 1:250,000 series, which means that GIS data derived from this is more generalised than desirable and for those parts of the *1911 Historical Land Capacity GIS* dependent upon it – the great majority of England – it will create problems especially for smaller CONPAR units. The GIS data is available for England from Natural England. For Wales the 1967-73 survey provides similar coarse GIS data. This was supplied by the Wales Land Quality Advisory Service. The dataset comprises 14,513 polygons with a mean size of 2,640.7 acres (1068.6 hectares), though this is because a few of the polygons are extremely large.[6] The 23 largest polygons represent over 50% of the total area of the dataset. There are also issues to do with land that has no ALC grade because it is classed as urban or non-agricultural. These categories of land represent a substantial minority (12.1%) of the dataset. This is a problem because an unknown proportion of this land would have been farmed in the period 1851-1911.

3.2 Agricultural Land Classification (ALC) Grades - Post 1988 Survey (polygons) (England)

This dataset consist of 46,981 polygons and represents ALC classifications for surveys done using the superior 1988 methodology and based on new field surveys. As a consequence, it is much more detailed than the 1:250,000 derived ALC GIS. The average polygon size of the Post 1988 ALC is 14.82 acres (6.0 hectares). The GIS data is available from Natural England.[7] The dataset is not without problems. It has topological issues with 1,051 overlaps and includes 13,313 polygons classified as "other" i.e. non-agricultural land, and 681 polygons classified as "not surveyed". It is also spatially biased to town fringes and highway schemes because it generally represents classifications done prior to development. The dataset is also very small. Once the overlaps are eliminated and the polygons with no ALC grade deleted, what remains comprises 1,411,349 acres (571,153 hectares), and as such, it represents only 1.8% of the English section of the 1967-73 ALC dataset. Since it is based on

[6] The number of polygons given here is the number of discrete polygons generated following geo-processing. As supplied by Wales Land Advisory Service and downloaded from DEFRA each land class was only represented by a single composite polygon.

[7] https://naturalengland-defra.opendata.arcgis.com/datasets/agricultural-land-classification-alc-grades-post-1988-survey-polygons-england/data

detailed surveys in the fields it will not lead to scale problems in the *1911 Historical Land Capacity GIS.*

3.3 ALC Wales 2017

This dataset consists of 949,474 polygons and encompasses all of Wales. The mean polygon size is large being 35,944 acres (14,546 hectares) but if the polygons are ranked in size, 10% of the area of Wales is represented by 13,815 largest polygons. It is derived from raster data with 50m² cells which gives it a graphic scale of 1:100,000. This scale will not create problems in the *1911 Historical Land Capacity GIS*. The Welsh dataset was produced in 2017 by the Cranfield Soils and AgriFood Institute using the superior 1988 methodology and state of the art soil, climatic and gradient data. It has been further refined by the Wales Land Quality Advisory Services with ALC grading being adjusted to show the effect of rocky outcrops, frost exposure, and areas prone to extreme winds (J. Cooke *pers com.*). The dataset was provided by the Wales Land Quality Advisory Service with the permission of the Cranfield Soils and AgriFood Institute (Predictive Agricultural land Classification (ALC) Map, 2017)**.**. It assigns land grades to all of Wales. This means that unlike England, the ALC GIS data for Wales does not have the problem where land farmed in the period 1851-1911 has been subsequently converted to non-agricultural uses and, as a consequence, is not assigned an ALC grade.

To conclude this section the ALC GIS represents a mixed picture. While the ALC grading methodology in many ways fits the needs of a historic land use GIS, in that it is concerned with the agricultural capacity of land as determined by climate and soils, there are problems with two of the three ALC datasets. As already discussed, the most significant problem is the use of outdated methodology for nearly all of England and potential scale problems due to the maps from which it was digitised being significantly larger in scale than those used to create the GIS behind the RSD and CONPAR datasets. To test the robustness of the data to these problems a comparison was made between of the changes between Wales ALC 1967-74 and ALC Wales 2017. Most grades do not change, and only a few have major changes. This suggests that the grading for England that is not covered by the post-1988 polygons is generally either correct or close to being correct under the later ALC scheme. This finding gives weight to the idea that the ALC data from the different sources for England and for Wales are broadly comparable and it is appropriate to use them

3.4 Making ALC into a Useable Historic National Dataset

The next task was to integrate the three ALC GIS datasets into a single shapefile. This was done as follows. (i) Land classed as grades 3a & 3b in the 2017 Wales ALC GIS were amalgamated into a single class grade 3; (ii) The Post-1988 England ALC GIS was cleaned with a topology to identify overlaps which were the rationalised and its grades 3a & 3b amalgamated into a single class grade;[8] (iii) The following elements were deleted from the Post-1988 England ALC GIS because they lacked an ALC grade: 13,313 polygons classified only as "other", 687 polygons classified as "Not surveyed"; one polygon with no attribute data; (iv) The three ALC datasets were then integrated together.

The final stage of dataset manipulation was to reclassify those parts of England which retained their classification as urban and non-agricultural land but would have been farmed in the period 1851-1911. To deal with this, the ALC polygons of urban land and the non-agricultural land were reclassified on the basis of the ALC grade of their adjacent polygons. The dataset was rasterized and then redrawn on the basis of the neighbouring polygons using the Focal Statistics tool of the Spatial Analyst Toolbox using a neighbourhood search of 0.8km. The Focal Statistics tool performs computes an output raster where the value for each output cell is a function of the values of all the input cells that are within the specified neighbourhood around that location. [9] Due to the nature of the rasterization process this generated a very large number of very small polygons slivers - essentially parts of grid cells - and also there remained a much smaller number of larger polygons that had been too large to have values assigned from their neighbours to their entire area using the focal statistics approach. There were also areas of the coastline which extended beyond the limits of ALC GIS for which it was necessary to interpolate values.

---

[8] Almost all overlaps were small slivers around the edges of virtually identical polygons in which it was obvious to which polygon the slivers should be allocated. In the few cases where there was a significant discrepancy in polygon shapes or ALC grades the polygon with the higher grade was selected.

[9] With assistance from Dr Gabriel Amable, Geomatics Officer, Department of Geography, Cambridge.

In these instances their values were interpolated on the basis of their length of boundary with their ALC neighbours. This was straightforward for those polygons whose polygon neighbours all had the same ALC grade. These were given an "INTERP" code of 2. Where a polygon had two or more neighbours with different grades the lengths of the shared boundaries was summed according to grade and whichever grade represented the greatest length was assigned to the polygon. This was given an "INTERP" code of 3. In instances where the two longest summed lengths were equal the grade was assigned randomly but they were also given an "INTERP" code of 2. At the end of this process only some island polygons remained ungraded. These were all assigned ALC classes on the basis of the grade of the nearest landmass to them. They were given an "INTERP" code of 4. The Scilly Isles lacks information in ALC scheme and has been classified as "no data".

It needs to be borne in mind that the various interpolation methods introduce noise into the data. The focal statistics method by its nature will contain fewer errors than the longest neighbour method. Distinguishing interpolated data from the original ALC GIS data with number codes enables the degree to which any GIS census unit consists of interpolated as opposed to original data to be established. The composite graphic scale of the ALC dataset as well as the failure of the original ALC survey to assign grades to areas smaller than 80 hectares may cause issues with the ALC values assigned to GIS census units outside Wales. This is particularly the case for the smaller census units of the CONPAR series. In the subsequent use of these data the source of interpolation can generally be ignored; but the information is retained in the database. For the Drivers of Entrepreneurship project they are aggregated to give one set of ALC measures for each CONPAR or RSD.

4. Estimating Historical Non-Agricultural Land-Use: Urban, Water and Woodland

Having generated a continuous layer of ALC graded land for the whole of England and Wales the next task was to create separate shapefiles of the principal elements of historic non-agricultural land use. The non-agricultural units chosen were woodland, water bodies, and built-up areas of cities and towns. Once created these were then be used to replace those sections of the ALC layer which they overlay. The resultant GIS of ALC graded land, woodland, waterbodies and built-up urban areas provides a crude measure of the potential area of agricultural land available for the various GIS census units for 1911. In addition, the conjunction of some of these historic non-agricultural land use types with modern non-agricultural ones means that substantial areas of the interpolated ALC GIS can be removed from the finished dataset making it more accurate. In what follows each of the historic land-use GIS are discussed first in terms of how they were created, then in terms of how representative they are for 1911 and earlier.

4.1 Woodland

The largest area of non-agricultural land use was woodland which represented about 5% of the area of England and Wales c. 1911. Generating a GIS of woodland for c. 1911 proved the most challenging part of creating the historic land-use GIS not only in terms of sourcing, assessing and editing the woodland GIS data but also in trying to establish the degree to which what was created formed a reasonable representation of woodland c. 1911.

Hitherto there has been no useable historic GIS of the woodland of England and Wales. The only comprehensive historic GIS data is that which can be extracted from the Dudley Stamp GIS of the Land Utilization Survey. This is a valuable as a starting point. But there are a number of problems and issues with the woodland layer in the DS GIS. The method of digitisation used for the DS land use GIS created polygons by land use colour but did not transfer the sub-classifications indicated by symbols for type of woodland. There are also questions about how woodland was classified for the DS project. Woodland is often fluid in nature by being cyclically felled, has capacity for natural regrowth and colonisation, and there can be different interpretations of what constitutes a wood. Also it can have multiple uses, including grazing for livestock blurring the distinction of woodland from farmland. The density of woodland can also make it difficult to distinguish the limits of a wood. However, for the compilers of the DS survey the transfer of woodland detail from the DS field sheets to the 1 inch maps caused few problems because much had not changed and was already present on the 1 inch OS base map. All that had to be done was to apply the correct colour code.

The more important issue is assessment of the degree to which the area of woodland had changed between 1911 and the publication of the Dudley Stamp maps. Some idea of the degree of change comes from looking at information concerning the extent of woodland in England and Wales c.

1911 and comparing it against the DS woodland GIS. To get a control for the extent of change two sources are used: (i) the *Agricultural Statistics* for the area of woodland in England and Wales for 1905 and 1913 (PP, 1905; 1913);[10] and (ii) a Forestry Commission (1952) comprehensive map-based survey of woodland for 1924. The 1913 returns are known to be incomplete but are valuable because they are so close in date to 1911.[11] The percentage difference between the 1913 and 1924 data is used as an adjustment ratio for changes to estimate changes 1913-24, which average 11.9% across England and Wales, though much larger in Wales. The largest losses of woodland over the period were in counties with only modest percentages of woodland in 1913, which suggests that the woodland area data when summarised at the level of the CONPAR or RSD polygon will be most deficient in areas where woodland was least important in 1911.

In assessing the woodland data in the DS GIS it is also important to establish the degree to which it includes areas where woodland had replaced open land after 1911 as well. A small, if unquantifiable, increase in the area of woodland was through colonisation of open land in a variety of settings, such the margins of railway lines, disused quarries, and on land too steep to plough that was no longer grazed. The trend to plant substantial areas of new woodland on the margins of reservoirs for hygiene and amenity also added substantial acreages in a few places But by far the most important in terms of increasing the area of woodland where agriculture was concerned were the new plantations of the Forestry Commission land. The area of extra woodland this added was substantial. The quality of the DS GIS is also variable for woodland, as noted earlier. The only solution is to identify the parts of the DS woodland GIS that need to be edited by checking between the woodland GIS and all the OS 1:10,560 maps of England and Wales closest in date to 1911. This was a substantial task as the mapping encompasses more than 6,500 5km x 5km geo-rectified tiles all of which have to selected, loaded, and inspected before each round of edits. In general, the mapping used was the OS First Revision Series (surveyed 1890-1914), though for the Cheshire, Flintshire, Hampshire, Kent and Sussex the OS Second Revision Series was used as this was closer in date to 1911. However, in instances where the Second Revision had started in a county by

[10] Schedules were distributed by local officers of the Board of Customs and Excise to woodland owners who gave particulars under three categories: coppice, plantations under ten years old, and, other woods.
[11] Work by the Forestry Commission comparing the acreage of woodland measured off 1:10,560 maps mostly of the first and second revision series for England and Scotland.

1911 but was not yet complete, only the older First Revision mapping was used because identifying which map sheets from which series was closest to 1911 was too time consuming.

In terms of editing, a decision had to be made of what to do about scrub - the term used by the Ordnance Survey to categorise and map land with a poor cover of dwarfed oak or birch. Scrub is a significant sub-class of woodland and represented 122,344 acres in 1924. The mapping of woodland by the Dudley Stamp survey may have been compromised because of the difficulty volunteers had distinguishing between scrub and rough grazing. Where there are discrepancies between the OS and DS definitions the OS map definitions were used. The edits to the DS woodland GIS were made as follows: (i) In instances where open land in the OS mapping was woodland in the GIS, it was deleted; but to make the task manageable the GIS was only re-edited where the area misclassified was five acres or more. (ii) Where woods transitioned into areas of open trees and rough grazing, the polygon was cut along the original OS hashed lines and field boundaries used to indicate areas classified as woodland. (iii) Where woods were shown on the OS mapping but were absent from the DS GIS, they were not digitised and added to the GIS because they were generally small fragments that were unimportant. This means that for those counties with substantial areas of woodland that had been felled after 1911 and not replanted at the time each DS map sheet was published, the GIS will under-represent the area of woodland, but this will usually only by a small factor at the level of individual RSDs or ConPar units.

However, the DS GIS woodland is still significantly larger than the estimated figure. This is probably because the Forestry Commission estimate of the amount of woodland omitted from the 1913 returns is too low; and also because the DS digitisation method disproportionately exaggerates the size of small features especially if they are linear, and there are many small woodland polygons in this dataset. There is nothing that can be done about this given the quality and resolution of the original data. However, it must be borne in mind that that when the data are summarised for the RSD or CONPAR, there will be will some exaggeration of the area of woodland for those units which have a large number of small woods. However, whilst the woodland GIS is by no means an accurate representation of the geography of woodland in England and Wales in 1911, its value is indicated by the statistical relationships between the percentage of woodland per county calculated from the 1913 data and the adjusted GIS being very close, with a correlation coefficient of + 0.94.

4.2 Built-up urban places

To categorise built-up urban land 'urban footprints' polygons were used from a previous research project that were drawn to define the built-up area of the cities and towns of England and Wales (Shaw-Taylor et al., 2017). The ones selected where the 929 cities and towns listed by Law (1967) and Robson (1973) (source listing from UKDA: Bennett, 2012). The footprints had been digitised from the OS 1st revision 1:10,560 series (surveyed 1890-1914) because this was the first time in which the Ordinance Survey produced high resolution maps that covered the whole country for a relatively narrow period. For the purpose of the c. 1911 land use/ capacity GIS it might have been better to use OS 2nd revision mapping for the urban footprints of a few counties because their survey dates are closer to 1911 than those of the OS 1st revision, but this would only have minor effect. Outside Wales the urban footprints GIS have the added bonus that they remove a substantial proportion of those areas which are potentially anomalous because they derive their ALC values from interpolation.[12]

4.3 Water

A GIS of major waterbodies was prepared as follows. First, the open water layer was extracted from the modern Ordnance Survey GIS: VectorMap District. This is open data and has a scale of 1:25,000. [13] Second, all waterbodies fifty acres or larger were selected from the layer. These were then overlaid on the OS 1:10,560 OS 1st and second revision mapping closest in date to 1911. All those polygons not shown as water on the historic OS mapping were deleted and those which had substantially changed were re-digitised. Those which, when re-digitised, were smaller than 50 acres were also deleted. In a few instances the historic OS mapping did not provide sufficient information to establish that a reservoir had been filled by 1911. For these water bodies further research was used to establish the date of change to water cover. The edited OS open data used as a source for the waterbodies will scale to RSDs and CONPAR units without difficulty because it has a scale of 1:25,000. Obviously this method does not capture those bodies of water which were filled in before the OS Open data was published in 2017. However, few if any of these missing waterbodies were

[12] For the 'Drivers of entrepreneurship' project, the urban land area and population density derived from the Law-Robson definitions is used as the source to identify the total of urbanised parishes (even though these will in some cases contain small amounts of residual agricultural land); these are generally larger than the footprint data (Smith et al., 2017: WP 6).

[13] https://www.ordnancesurvey.co.uk/business-and-government/products/vectormap-district.html

large, so at worst that there will be only minor inaccuracies in this element of the *1911 Historical Land Capacity GIS*.

5. Assemblage and data base description

The previous stages of geoprocessing created separate shapefiles for ALC, woodland, waterbodies, and built up cities and towns. The final task was to assemble all the datasets together into a single shapefile. This was done by using the “Erase” tool” in ArcMap to remove those parts of the ALC shapefile which were overlain by woodland, water or the built-up cities and towns polygons. The edited ALC shapefile was then merged with the woodland, water or built up cities shapefiles. When finished this comprised 428,217 polygons. Details of its attributes are listed in Appendix 1. Once this was done the shapefile was checked with a topology to confirm there were no overlapping polygons or gaps.

The final shapefile (listed in Appendix) was then processed into ALC classifications for RSD and CONPAR parishes. The Union tool from the Analysis toolbox was used to produce nine hybrid shapefiles in which nine copies of the land use shapefile were each subdivided by the boundaries of one of the seven RSD shapefiles (for each year 1951-1911) and two CONPAR polygons (for the two series 1851-91, and 1901-1911). The area of each land use type was then calculated for each RSD or ConParID number of all nine shapefiles. The resultant tables then define the final dataset used at the RSD and CONPAR levels for the *1911 Historical Land Capacity GIS*.

Acknowledgements

This research has been supported by the ESRC under project grant ES/M010953: **Drivers of Entrepreneurship and Small Businesses.** A special acknowledgement of thanks for advice on some technical issues is recorded to Dr Gabriel Amable, Geomatics Officer, Department of Geography, University of Cambridge.

The Dudley Stamp Land Use Survey digitised files were derived from: Environment Agency, *Dudley Stamp Land Utilisation Survey 1933-1949 Digital Version*, DSTR08286 (2009).

The Land Classification (ALC) GIS was derived from Natural England. The ALC of England for the pre-1988 ALC dataset from:
https://naturalenglanddefra.opendata.arcgis.com/datasets/5d2447d8d04b41d4bbc9a8742f858fd_0

The post-1988 ALC dataset from:
https://naturalengland-defra.opendata.arcgis.com/datasets/agricultural-land-classification-alc-grades-post-1988-survey-polygons-england/data

The ALC data for Wales was provided by James Cooke of the Wales Land Quality Advisory Service with the permission of the Cranfield Soils and AgriFood Institute Wales Land Advisory Service: Predictive Agricultural land Classification (ALC) Map (2017).

Historic OS maps were obtained from Digimap: http://digimap.edina.ac.uk/

Ordnance Survey open data was from:
https://www.ordnancesurvey.co.uk/business-and-government/products/vectormap-district.html

The GIS boundary files for parishes were derived from Satchell, A.E.M., Kitson, P.M.K., Newton, G.H., Shaw-Taylor, L., Wrigley E.A. (2006) *1851 England and Wales census parishes, townships and places*, 2006, ESRC RES-000-23-1579, supported by Leverhulme Trust and the British Academy; Satchell, A.E.M. (2015) *England and Wales census parishes, townships and places*; which is an enhanced and corrected version of Burton, N, Westwood J., and Carter P. (2014) *GIS of the ancient parishes of England and Wales, 1500-1850*, UKDA, SN 4828; which is a GIS version of Kain, R.J.P., and Oliver, R.R. (2001) *Historic parishes of England and Wales: An electronic map of boundaries before 1850 with a gazetteer and metadata*, UKDA, SN 4348.

The GIS boundary files for RSDs were constructed from these parish files by Joe Day for the ESRC fertility project directed by Alice Reid:
http://www.geog.cam.ac.uk/research/projects/victorianfertilitydecline/publications.html

References

Bailey, B. (2007) The extraction of digital vector data from historic land use maps of Great Britain using image processing techniques, *E-perimetron*, 209-23.

Bennett, R. J., (2012) *Urban Population Database, 1801-1911* (2012). [data collection]. UK Data Service. SN 7154, http://dx.doi.org/10.5255/UKDA-SN-7154-1; from C. Law, University of Salford, Department of Geography; B. Robson, University of Manchester, Department of Geography; J. Langton, University of Oxford, School of Geography and the Environment [original data producers].

Coppock, J. T. (1978) Review No. 14. The statistics of land use, pp. 1-101 in W.F. Maunder (ed.) *Reviews of Statistical Sources*, 8, Oxford.

Dudley Stamp, L. (1933-4) T*he Land of Britain: Land Utilisation Survey of Britain, 89 sheets*, London, 1933-1949, 1:63,360 scale.

Environment Agency (2009) Dudley Stamp Land Utilisation Survey 1933-1949 Digital Version, DSTR08286 (2009).

Forestry Commission (1952) *Census of woodlands 1947-1949. Woodlands of five acres and over. Census Report*, London.

Kain, R.J.P., and Oliver, R.R. (2001) *Historic parishes of England and Wales: An electronic map of boundaries before 1850 with a gazetteer and metadata*, UKDA, SN 4348.

Law, C. M. (1967), 'The growth of urban population in England and Wales, 1801-1911', *Transactions of the Institute of British Geographers*, 41, 125-43.

MAFF (1988) *Agricultural land classification of England and Wales. Revised guidelines for grading the quality of agricultural land*, Ministry of Agriculture Fisheries and Food, London.

PP (1905) *Agricultural Statistic Great Britain 1905*, Parliamentary Papers, Cd. 3061.

PP (1913) *Agricultural Statistics England and Wales 1913***,** Parliamentary Papers, Cd. 7325.

Robson, B. (1973), *Urban Growth: An Approach*, London: Methuen.

Satchell, A.E.M., Kitson, P.M.K., Newton, G.H., Shaw-Taylor, L., Wrigley E.A. (2016) *1851 England and Wales census parishes, townships and places*, 2006, ESRC RES-000-23-1579, supported by Leverhulme Trust and the British Academy, available at: https://www.geog.cam.ac.uk/research/projects/occupations/datasets/documentation.html

Satchell, A.E.M., Kitson, P.M.K., Newton, G.H., Shaw-Taylor, L., Wrigley E.A. (2006) *1851 England and Wales census parishes, townships and places*, 2006, ESRC RES-000-23-1579, supported by Leverhulme Trust and the British Academy

Shaw-Taylor, L., M. Satchell, E. Potter, G. Newton, D. Bogart, A. McKenzie, M. Davenport (2017) *Footprints c. 1900 of Candidate Towns of England and Wales 1563-1911*, Cambridge Group for the History of Population and Social Structure.

Smith, H. and Bennett R., J. (2017) *Urban-Rural Classification using Census data, 1851-1911*, Working Paper 6: ESRC project ES/M0010953; https://doi.org/10.17863/CAM.15763

Taylor, K. J., Walford, N. and Short, B**.** (2010) Assessing the land use of interwar Britain: A comparison of the First Land Utilisation Survey field sheets and 1:63,360 scale maps, *Applied Geography*, 30, 50-62.

Webster, N. et al. (2010) *GIS classification of the 1930s-1940s land Utilisation Survey* (LUS). Entec, Final Report June 2010: http://catalogue.ceh.ac.uk/id/864fff73-742d-4cb2-9afb-a0802487ccf7 (accessed 14/01/2019)

Wrigley, E.A. (2011) *The Early English Censuses*. Oxford: Oxford University Press.

APPENDIX: The *1911 Historical Land Capacity GIS* Shapefile

Filename: 1911HistoricLanduse.shp

428,217 polygons

Citation: Satchell, Max (2018) *1911 Historical land capacity shapefile*, ESRC project ES/M0010953: Drivers of Entrepreneurship and Small Businesses', University of Cambridge, Department of Geography and Cambridge Group for the History of Population and Social Structure.

Attribute data

1911HistoricLanduse.dbf

The dbf table contains the following fields.

Field	*Data type*	*Description*
FID	Object ID	Unique ID for each row in the table
Shape	Polygon	Polygons of historic land use
Interp	Long	Interpolation code
HistGrade	Text	Historic Land use
Shape Area	Double	Area in square metres

Interpolation Codes

1	ALC grade interpolated by focal statistics
2	ALC grade interpolated from their length of boundary with the ALC neighbours. All neighbours the same
3	ALC grade interpolated from boundary of ALC neighbours with greatest summed length- neighbours have two or more grades
4	Island with ALC grade assigned from that of nearest landmass
5	Scilly Isles - no assignable ALC grade

Historic Land use Grades

HIstGrade	Description
1	modern ALC grade 1
2	modern ALC grade 2
3	modern ALC grade 3
4	modern ALC grade 4
5	modern ALC grade 5
no data	polygons with a historic land use
Open water	waterbody 50 acres or more c. 1911
Urban	built up area of a city or town listed by Law and Robson
woodland	woodland c. 1911

Co-ordinate system

British National Grid

Projection: Transverse Mercator

False Easting: 400000.000000

False Northing: -100000.000000

Central Meridian: -2.000000

Scale Factor: 0.999601

Latitude of Origin: 49.000000

Linear Unit: Meter

GCS_OSGB_1936

Datum: D_OSGB_1936 James Cooke

Other Working Papers:

Working paper series: ESRC project ES/M010953: *'Drivers of Entrepreneurship and Small Business',* University of Cambridge, Department of Geography and Cambridge Group for the History of Population and Social Structure.

WP 1: Bennett, Robert J., Smith Harry J., van Lieshout, Carry, and Newton, Gill (2017) *Drivers of Entrepreneurship and Small Businesses: Project overview and database design.* https://doi.org/10.17863/CAM.9508

WP 2: Bennett, Robert J., Smith Harry J. and van Lieshout, Carry (2017) *Employers and the self-employed in the censuses 1851-1911: The census as a source for identifying entrepreneurs, business numbers and size distribution.* https://doi.org/10.17863/CAM.9640

WP 3: van Lieshout, Carry, Bennett, Robert J., Smith, Harry J. and Newton, Gill (2017) *Identifying businesses and entrepreneurs in the Censuses 1851-1881*. https://doi.org/10.17863/CAM.9639

WP 4: Smith, Harry J., Bennett, Robert J., and van Lieshout, Carry (2017) *Extracting entrepreneurs from the Censuses, 1891-1911.* https://doi.org/10.17863/CAM.9638

WP 5: Bennett, Robert J., Smith Harry J., van Lieshout, Carry, and Newton, Gill (2017) *Business sectors, occupations and aggregations of census data 1851-1911*. https://doi.org/10.17863/CAM.9874

Data download of classification file: https://doi.org/10.17863/CAM.9874

WP 6: Smith, Harry J. and Bennett, Robert J. (2017) *Urban-Rural Classification using Census data, 1851-1911*. https://doi.org/10.17863/CAM.15763

WP 7: Smith, Harry, Bennett, Robert J., and Radicic, Dragana (2017) *Classification of towns in 1891 using factor analysis.* https://doi.org/10.17863/CAM.15767

WP 8: Bennett, Robert J., Smith, Harry, and Radicic, Dragana (2017) *Classification of occupations for economically active: Factor analysis of Registration Sub-Districts (RSDs) in 1891.* https://doi.org/10.17863/CAM.15764

WP 9: Bennett, Robert, J., Montebruno, Piero, Smith, Harry, and van Lieshout, Carry (2018) *Reconstructing entrepreneurship and business numbers for censuses 1851-81*. https://doi.org/10.17863/CAM.37738

WP 10: Bennett, Robert, J., Smith, Harry and Radicic, Dragana (2018) *Classification of environments of entrepreneurship: Factor analysis of Registration Sub-Districts (RSDs) in 1891.* https://doi.org/10.17863/CAM.26386

WP 11: Montebruno, Piero (2018) *Adjustment Weights 1891-1911: Weights to adjust entrepreneur numbers for non-response and misallocation bias in Censuses 1891-1911*. https://doi.org/10.17863/CAM.26378

Adjustment weights: https://doi.org/10.17863/CAM.26376

WP 12: van Lieshout, Carry, Day, Joseph, Montebruno, Piero and Bennett Robert J. (2018) *Extraction of data on Entrepreneurs from the 1871 Census to supplement I-CeM.* https://doi.org/10.17863/CAM.27488

WP 13: van Lieshout, Carry, Bennett, Robert J. and Smith Harry (2019) *Extracted data on employers and farmers compared with published tables in the Census General Reports, 1851-1881.* https://doi.org/10.17863/CAM.37165

WP 14: van Lieshout, Carry, Bennett Robert J. and Montebruno, Piero (2019) *Company Directors: Directory and Census record linkage.* https://doi.org/10.17863/CAM.37166

WP 15: Bennett, Robert, J., Montebruno, Piero, Smith, Harry and van Lieshout, Carry (2019) *Entrepreneurial discrete choice: Modelling decisions between self-employment, employer and worker status.* https://doi.org/10.17863/CAM.37312

WP 16: Satchell, M., Bennett, Robert J., Bogart, D. and Shaw-Taylor, L. (2019) Constructing Parish-level Data and RSD-level Data on Transport Infrastructure in England and Wales 1851-1911. https://doi.org/10.17863/CAM.37313

WP 17: Satchell, M. and Bennett, Robert J. (2019) *Building a 1911 Historical Land Capacity GIS*.

Full list of all current Working Papers available at:
https://www.campop.geog.cam.ac.uk/research/projects/driversofentrepreneurship