

ENVIRONMENTAL RESEARCH
LETTERS

LETTER

OPEN ACCESS

RECEIVED
19 October 2021REVISED
10 May 2022ACCEPTED FOR PUBLICATION
14 July 2022PUBLISHED
18 August 2022

Original content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.

Multi-tier archetypes to characterise British landscapes, farmland
and farming practicesCecily E D Goodwin^{1,*} , Luca Bütikofer² , Jack H Hatfield^{1,3} , Paul M Evans¹ , James M Bullock¹ ,
Jonathan Storkey² , Andrew Mead² , Goetz M Richter² , Peter A Henrys¹ , Richard F Pywell¹
and John W Redhead¹ ¹ UK Centre for Ecology & Hydrology, Maclean Building, Benson Lane, Crowmarsh Gifford, Wallingford, Oxfordshire OX10 8BB, United Kingdom² Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, United Kingdom³ Present address: Leverhulme Centre for Anthropocene Biodiversity and Department of Biology, University of York, York, YO10 5DD, United Kingdom.

* Author to whom any correspondence should be addressed.

E-mail: cecgoo@ceh.ac.uk**Keywords:** landscapes, farming, Great Britain, management, archetypesSupplementary material for this article is available [online](#)**Abstract**

Due to rising demand for both food and environmental services, agriculture is increasingly required to deliver multiple outcomes. Characterising differences, across agricultural landscapes, via the identification of broad archetypal groupings, is an important step in exploring spatial patterns in the capacity of land to deliver these potentially competing functions. Creating characterisations at multiple levels, for landscape and farm management, can allow policy-makers and land managers to harmonise delivery of ecosystem services at different intervention scales. This can identify ways to increase the complementarity of public goods and the sustainability of farmed landscapes. We used data-driven machine learning to create landscape and agricultural management archetypes (1 km resolution) at three levels, defined by opportunities for adaptation. Tier 1 archetypes quantify broad differences in soil, land cover and population across Great Britain, which cannot be readily influenced by the actions of land managers; Tier 2 archetypes capture more nuanced variations within farmland-dominated landscapes of Great Britain, over which land managers may have some degree of influence. Tier 3 archetypes are built at national levels for England and Wales and focus on socioeconomic and agro-ecological characteristics within farmland-dominated landscapes, characterising differences in farm management. By using a non-nested hierarchy, we identified which types of management are restricted to certain landscape settings, and which are applicable across multiple landscape contexts. Understanding variation within and between agricultural landscapes and farming practices has implications for planning environmental sustainability and food security. It can also aid understanding of the scale at which interventions could be most effective, from incentivising changes in farmer behaviour to policy drivers of large-scale land use change.

1. Introduction

Farmland is under great pressure to increase production of food in response to increased and changing demand from growing human populations [1, 2]. Growing awareness of the environmental impacts of intensified farming, the need for agricultural land to deliver multiple functions [3] and competition for other land uses, such as biofuels or carbon storage [4],

puts further pressure on agricultural land to increase production sustainably [5, 6]. Thus, there is a need to holistically assess the potential of different farming systems to deliver multiple public goods and services across landscapes [7]. This may include trade-offs and synergies between the services delivered by different systems [1, 8, 9]. Agricultural policy development can help shape that balance at critical times, such as the current situation in Great Britain, following

the country's withdrawal from the European Union (EU).

Farming systems are formed of the fixed elements of the landscape, and more flexible elements of management practices. Landscapes vary in their soils, climate, land cover and features [10–12], as well as in more anthropogenic elements such as population, infrastructure and land protections. Understanding variation at the landscape scale can allow insight into how different landscapes can constrain management interventions [13, 14]. Set within the landscape, farms themselves are socio-ecological systems [15–17]. For instance, farms can have different crop and livestock management systems, field configurations and employment structures [18–20]. These aspects interact to affect farm productivity and the delivery of ecosystem services [19, 21].

The varied and interdependent nature of farming systems makes it challenging to explore the impacts of potential changes on multiple outcomes. Assessing variation in farming systems at both landscape and farm-scales can allow for an understanding of how these two aspects relate to each other [17, 18]. Through understanding variation in farm management, and how this is influenced and constrained by landscape, more suitable and adaptive management strategies can be devised [22–25]. One route to quantifying this complex variation, and identifying spatial patterns across large extents, is the characterisation of farm management and landscapes into typologies. This can be done through the identification of archetypes, which involves recognising recurrent patterns to create groupings at an intermediate level of abstraction [26]. Archetypes thus represent a balance between generalisation and case-based validity [27, 28].

Archetypes of farming systems enable the contextualisation of local specific cases within larger regional or national frameworks [3, 26, 27]. Relationships between archetypes and their constituent cases help to assess the opportunity space for change [28]. Although requiring care in selection and structuring of data, data-driven methods can detect archetypal patterns in complex multi-dimensional data without the imposition of subjective judgements about the nature of groupings [29–31].

In this study we identify archetypes of land systems at two different tiers for Great Britain (GB), and at a third tier for England and Wales. These three tiers represent gradients of decreasing permanence and increasing intervention potential:

(a) Tier 1, landscape archetypes: highest permanence and lowest intervention capacity, capturing broad differences in land cover, land features and population across GB. These are largely independent of land manager decisions except over long timescales (i.e. 10–100 years).

(b) Tier 2, farmed landscape archetypes: landscape-driven differences within farming dominated GB landscapes, incorporating landscape elements important in farming and potentially modifiable by land managers' strategic decisions over time periods of 1–10 years.

(c) Tier 3, farm management archetypes: management and social differences in farm management across the national level for England and Wales separately, which are largely under land managers' control and thus can potentially change over relatively short timescales given sufficient incentives.

We then assess the relationships between, and distribution of, the three tiers of archetypes and discuss their potential use in assessing current and future environmental impacts.

2. Methods

The three tiers of archetypes were analysed separately and not as a nested structure (i.e. a single Tier 3 archetype can occur in more than one Tier 2 archetype), predominantly to ensure that archetype definitions were easily interpreted across tiers. Tier 1 and 2 archetypes were generated for GB, while Tier 3 archetypes were generated separately for England and Wales, as policy instruments that target land management are determined at this devolved national level. The unavailability of several input variables for agricultural management prevented the generation of Tier 3 archetypes for Scotland.

2.1. Input variables

All spatial variables used to define archetypes (table 1) were processed at a 1 km² resolution, as this reflects the scale at which most data were available, as well as approximating the mean size (0.87 km²) of farms in England [32]. This was done for governmental Ordnance Survey grid cells with over 75% terrestrial land cover [33]. Different data were included in each of the three tiers to achieve good coverage of biogeo-physical, land management and socioeconomic variation. Variables for the three tiers were selected by expert judgement according to their capacity to represent: Tier 1 broad differences in landscape character across the country, modifiable only over long timeframes (10–100 years); Tier 2 differences in the farmed landscape, which landowners decisions may have effects over intermediate timeframes (1–10 years); and Tier 3 elements of the landscape which farmers have the ability to modify, or might influence their decisions over shorter timeframes. Processing and analyses were conducted using ArcGIS v.10.6 (ESRI, Redlands, CA, USA) and R [34].

Table 1. Data used in the derivation of the three archetype tiers. Variable name and shorthand (if used in tables S1–S4) are given. Units are per 1 km².

Variable	Units	Timeframe	Tiers	Source
Land cover	% cover	2015	1, 2	[33]
Land cover (each of: arable, improved grassland, broadleaf woodland, semi-natural grassland, urban)	% cover	2015	1	[33]
Land cover (each of: conifer, coast, freshwater, mountain/heath/bog)	Aggregation Index	2015	2	[33]
Land cover structure (LS) (each of: arable, improved grassland, broadleaf woodland, semi-natural grassland, urban)	Total length (m)	2000	1, 2	[43]
Watercourses (WCL)	Total length (m)	2016	1, 2	[44]
Woody linear features (WLF)	Total length (m)	2016	3	[44]
Inter-field hedges	Relative % content	2013	1, 2	[36, 37]
Soil composition (each of: sand, clay, silt)	pH	2013	1, 2	[36, 37]
Soil pH	% Organic carbon content	2013	1, 2	[36, 37]
Soil organic carbon (SoilOCC)	Mean monthly (mm m ⁻¹)	2005–2015	2	[38]
Soil moisture	Mean monthly deviation	2005–2015	2	[38]
Relative soil dryness (reldry)	Daily mean (K)	2005–2015	1, 2	[35]
Air temperature	Daily mean (kg m ² s ²)	2005–2015	1, 2	[35]
Rainfall	Mean (m)	1990	1	[42]
Elevation (EleMean)	Mean (m)	1990	1	[42]
Elevation variability (EleSD)	Mean tangent (degrees)	1990	2	[42]
Slope	Mean area (km ²)	2018, 2020	3	[54, 56, 57]
Size of each of: farm and field	Spread Index	2018, 2020	3	[54, 56, 57]
Farm spread	Edge:area	2018, 2020	3	[54, 56, 57]
Field shape index	Mean area (km ²)	2015–2020	3	[59]
Mean patch area	Mean NND (km)	2015–2020	3	[59]
Patch isolation	Index	2015–2020	3	[59]
Patch index (each of: Simpson's diversity (Diversity), evenness (Evenness), subdivision (Subdiv), edge contrast (EdgeContr))	Variance of index	2015–2020	3	[59]
Variance (ST) of patch index (each of: mean area, Simpson's diversity and evenness, isolation, subdivision, edge contrast)	Number of heads	2010	3	[61]
Livestock (each of: horses, cattle, dairy cattle, beef cattle, pigs, poultry, sheep)	Total area (km ²)	2010	3	[61]
Crops (each of: beans, fodder, vegetables, old grassland, new grassland)	Total area (km ²)	2015–2020	3	[59]
Crops (each of: maize, oilseed rape, potatoes, spring barley, spring wheat, winter wheat, winter barley, winter wheat, sugar beet, other)	Yearly expenditure (£)	2016	3	[51]
Farm costs (fixed (CostFix) and variable costs (CostVar))	Ethical risk coefficient	2016	3	[51]
Animal welfare (AnimalWel)	Income from AES (£)	2016	3	[51]
AE Scheme Income (AES_FBS)	Proportion index	2016	3	[51]
Economy proportion of agriculture & forestry (PropEcon)	Proportion index	2016	3	[51]
Employment proportion (for each of: agriculture & forestry (PropPop), tourism, woodland management)	Total Income (£)	2016	3	[51]
Income	Total population	2011	1	[45]
Population density (PopDen)	Distance (m)	2019	1	[46]
Distance from minor (Mnroad) and major roads (Mjroad)	% cover	2020	2	[47–50]
Designated area: protected areas & scheduled monuments (Desig_area)	Mean annual (kg km ⁻² yr ⁻¹)	2012–2016	3	[62]
Pesticide application rate				

2.1.1. Land cover and geo-physical variables

Percentage cover of nine aggregate land cover classes was extracted from the UKCEH Land Cover Map 2015 [33]. Mean temperature and precipitation were calculated from daily estimates [35] from January 2006 to December 2015. Soil pH, sand, silt, clay and organic carbon content were obtained from national soil maps [36, 37]. To integrate further measures of soil moisture relevant for farming into Tier 2 archetypes, mean soil moisture was calculated from monthly estimates between January 2006 and December 2015 based on the G2G hydrological model [38]. Mean relative soil dryness was calculated based on the difference of each monthly soil moisture estimate from the per-month per-cell 10 year mean, divided by the difference between the per-cell per-month 10 year minimum and mean.

To describe landscape structure, the Aggregation Index of 25 m pixels of each land cover class was calculated using the landscapemetrics package [39]. The Aggregation Index was chosen as it is relatively independent of the amount of land cover [40, 41]. Landscape structure is less relevant for cells with high or low amounts of land cover as landscape structure possibilities are constrained. We therefore only included structure for cells with between 10% and 70% cover for a given land cover type.

The mean and standard deviation of elevation and slope were calculated based on a Digital Terrain Model [42]. As these variables were inter-correlated, we included elevation variables in deriving Tier 1 landscape archetypes to capture the broad differences in mountainous and lowland landscapes, and slope in deriving Tier 2 landscape archetypes, as it has more relevance for constraining farming practises and types.

Lengths of water courses and woody linear features were calculated based on GB linear features maps [43, 44]. All woody linear features falling within 4 m from a field edge were classified as inter-field hedges.

2.2.2. Socio-economic variables

As the population density of landscapes is related to other features, such as demand for certain land uses, it was included in the generation of Tier 1 archetypes. It is more relevant across these more varied landscapes with a range in population densities, e.g. suburban and urban, and was not included in Tier 2 archetype generation, where the % cover of urban land cover was considered sufficient to capture rural populations. Gridded (1 km²) population density data was based on the 2011 census [45]. Distance from cell centres to the nearest major (motorways, A roads and B roads) and minor road quantified accessibility (cells that contained a road were assigned a value of 0) [46], at Tier 1 level, where there are much broader differences in the isolation of landscapes, e.g. in mountainous areas. Protected area coverage will influence how

the land is managed, and so was included in Tier 2 archetype generation as % cover of two land designations: natural protected areas [47] and scheduled monuments [48–50].

Ten high-level variables were extracted from the Sustainable Intensification Dynamic Typology Tool ([51] see table 1)[63]. These were based on the Farm Business Survey dataset, a questionnaire of farm businesses randomly selected across England and Wales. Absolute values were rescaled to relative measures on a consistent scale. These data were resampled to 1 km² from a resolution of 10 km² by bilinear interpolation.

2.2.3. Farm and field spatial characteristics

The spatial properties of farms and fields were included as they are associated with a range of management characteristics [52, 53]. Farm boundaries were extracted from Countryside Stewardship Scheme [54], Environmental Stewardship Scheme [55] and organic farming scheme [56] data for England; and various Glastir schemes [57] for Wales. These data were interpolated with the ordinary Kriging algorithm from gstat [58] giving predicted farm size. Similarly, a Spread Index was calculated for the fields belonging to each farm (S.I., equation (1)):

$$\text{S.I.} = 1 - \sqrt{F_{\text{size}}} / \sqrt{F_{\text{mcp}}}. \quad (1)$$

Equation (1): Spread Index, where F_{size} is the total farm area and F_{mcp} is the area of the minimum convex polygon encompassing all fields belonging to that farm.

Average field size was computed from the land cover plus: crops map (LCC [59]). To capture the shape of fields, the mean ratio of the perimeter of each field to the perimeter of a square of equal area was calculated (based on [60]).

2.2.4. Crop and livestock-cover

Data on crops and livestock were obtained from LCC [59] and AgCensus data [61]. LCC data was used preferentially, but for crops not in the LCC, the AgCensus dataset was used. The resolution of AgCensus is 5 km for England and 2 km for Wales; so data were resampled to 1 km by bilinear interpolation. As well as cattle (total including calves), a further division into beef and dairy was made due to their different environmental impacts. Average pesticide application rates (2012–2016) were derived from the land cover plus pesticides dataset [62].

Six variables were derived to capture the spatial distribution of crops and grassland, including Simpson's diversity and evenness indices [64, 65]. To represent functional diversity of crop types, the Edge Contrast Index was calculated using a distance matrix (using Gower's method) between pairs of crops, based on functional traits (crop functional type (e.g. cereal vs. legume); mass flowering; narrow/wide row spacing; month of planting and harvest; method

of harvest; height; agrochemical usage intensity). This was then multiplied by the length of perimeter between crop pairs, the weighted lengths summed and divided by the total patch perimeter [65]. The subdivision index quantified the separation of crop patches (a contiguous area covered by the same crop), as one minus the sum of the proportion of area covered by a patch [39]. Because five years of LCC data were available for these crop types, each of these six metrics was calculated as the variance of the metric over five years. The isolation index of crops was calculated as the mean Euclidean nearest neighbour distance between each patch [39].

2.2. Self-organising map (SOM) parameterisation

Clustering and dimensionality reduction were performed using SOMs [66]. SOMs are an artificial neural network method [66, 67], which simplify and visualise complex data by reducing its dimensionality and grouping similar units into clusters, referred to as ‘nodes’ [29, 67]. The method allows for flexibility in the use of input data, and is well suited to land system classification [10, 11, 29, 68].

SOMs work iteratively by competitively mapping each input data vector (in this case, associated with each 1 km² grid cell) to its best matching node, within an $N \times N$ node grid. This in turn influences the variable values (‘codebooks’) for that node and its surrounding nodes through co-operative learning. The SOM’s output nodes represent data clusters, and their codebooks represent the node’s coordinates in the input variable space; thus the (arche-)typical values, or centroids, for cells mapped to that node [30, 69]. For more detail on the SOM methodology, please refer to [29, 66–68].

SOMs were run using the Kohonen R package [70], on normalised (mean-centred and divided by standard deviation) input data. SOMs were set up to accept cells with missing data for up to all but one variable, so that cells would be clustered based on their available data, but missing data did not affect the clustering. SOMs were run using the parallel batch method and data were presented to the grid 500 times.

A challenge in creating archetypes is to maintain a balance between the generalisation and specificity of the archetypes [26]. The configuration of the SOM grid (and therefore the number of nodes, and resultant archetypes) was determined through an assessment of the ‘elbow’ in the graph of the mean distance of cases from archetype centroids for different configurations (figure S1). If there were multiple candidate grid formations which resulted in a sharp decrease in within-cluster distance, including formations with the same number of clusters, the configuration was chosen which increased: accuracy (mean Euclidean distance of cells from their codebook); consistency (the proportion of times cases were assigned to the same node); and the consistency of production of the same archetype set across

multiple iterations (to account for randomness in the SOM [71]). Sensitivity analyses were conducted on correlated and alternative variables, by removing or adding variables, and investigating effects on the archetypes, cluster and classification consistency.

2.3. Deriving archetypes

As the co-ordinates of the SOM nodes are initialised randomly, different outputs can be produced from the same data [71]. To account for this and gain a measure of classification certainty, we ran 1000 iterations of the analysis, holding input parameters constant. To assess consistency, there was a need to recognise similar nodes across iterations. We did this by performing hierarchical clustering using Ward’s method [72] on the codebooks associated with nodes produced from all iterations. To ensure a 1:1 mapping, the iterations which resulted in one node over each cluster (the major archetype groupings) were extracted, along with the mean codebook values for each archetype over these iterations. Each cell was classed as the archetype to which it was most frequently assigned over iterations, as well as recording its Euclidean distance from the central codebook estimates for the archetype. To avoid losing information on the stability of archetypes, we used data across all the iterations to calculate consistency (‘Certainty’) of assignment of each km cell.

Archetype names were derived through extracting and concatenating the names of variables with the strongest negative and positive weightings in each archetype codebook; then simplifying the result to more human-readable, intuitive names. They are therefore assigned for convenience, and do not represent a full description of the archetype characteristics.

2.4. Exploring the spatial configuration of archetypes and their interaction between tiers

We assessed co-occurrence of archetypes of different tiers by calculating their percentage overlap, and thus illustrating the link between landscape types. And landscapes and management systems. Additionally, the configuration of archetypes in the landscape, and the landscape context of individual archetypes, will affect their ability to deliver ecosystem services; for example if there is an aggregation of agricultural archetypes it will likely affect the water quality of freshwater archetypes, or smaller rivers and streams in that area. The dispersion, or aggregation, of archetypes will also impact the ability to deliver a balance portfolio of ecosystem services at different scales. Multifunctional landscapes can be identified which have a high dispersion of different archetypes, and the potential, therefore, to deliver a range of ecosystem services on the local level. To explore the configuration of each archetype of each tier, its aggregation index (0–100) within an 11 × 11 km sliding window was calculated [39]. This window size was large

enough to capture multiple archetypes in each window so that interspersions could be calculated, while also smoothing out the noise in spatial grain (other window sizes were tried and revealed similar patterns). We assessed archetype mixing by calculating an Interspersion Index (The 'Interspersion and Juxtaposition Index' in [65]) within the sliding window for each tier (for windows containing at least three archetypes. Landscapes with fewer than three were assigned a value of zero) [39, 65]. We also calculated the archetype difference index within the sliding window for each tier as the mean of the Euclidean distances between all pairs of the archetype codebooks for cells within the sliding window. Non-farmed areas were ignored for Tiers 2 and 3, and did not contribute to the calculation of these indices. The Archetype difference and interspersions indices were scaled between 0 and 1 and summed to give an archetype diversity index, as functional difference and interspersions represent the variation in composition and configuration respectively [53], and contribute to the overall diversity of land systems.

3. Results

3.1. Landscape and agricultural management archetypes across GB

Examination of elbow plots (figure S1), identification of candidate cluster sets and comparison of consistency led to the formation of 16 (SOM grid of 8×2) and 15 (5×3) Tier 1 and 2 archetypes, respectively, for GB. For Tier 3, 12 (4×3) archetypes were derived for England, and 8 (2×4) for Wales. Across 1000 SOM iterations for the creation of Tier 1 archetypes, 999 runs produced a consistent set of archetypes, where one node per iteration could be assigned to one of 16 groupings. For Tier 2 archetype generation, 969 iterations produced a consistent set of archetypes. For Tier 3 archetypes, 847 and 760 of 1000 iterations were consistent for England and Wales, respectively.

Tiers 1 and 2 captured distinctive patterns of groupings in soils, climate, land cover and features (figures 1, S2 and S3). Some archetypes, especially in Tier 1, are characterised by high values of one or two variables and captured rarer land types (e.g. open coastal landscapes; table S1), but the majority are comprised of more diverse variable combinations, especially in Tier 2 [73].

Farm management archetypes described patterns in land use and management characteristics separately for England and Wales (figure 2). These archetypes were also more likely to be comprised of combinations of many variables (figures S4 and S5).

Euclidean distance (in multi-variable space) values ranged widely (figure 3). The most isolated, rural or built-up areas had the highest distance values from their assigned archetypes (figure 3), indicating unusual landscape types for which the archetype set is a relatively poor descriptor. Tables S1–S4 illustrate

a summary of mean distances, and average certainty of archetype assignment for each archetype within each tier.

3.2. Interactions between farmed landscape and farm management tiers

Broad landscape (Tier 1) archetypes were generally covered by several Tier 2 farmed landscape archetypes (figure S6(A)); with an average of 34% (6.5 SD) of each Tier 1 archetype covered by the Tier 2 archetype with which it most commonly co-occurred. Tier 2 archetypes were often found predominantly within one or two Tier 1 landscapes (figure S6(B)); an average of 63% (20 SD) of each Tier 2 archetype occurred within the Tier 1 archetype with which it most commonly co-occurred. The ten most common co-occurrences of Tier 1 and 2 archetypes covered 53% of agricultural land area.

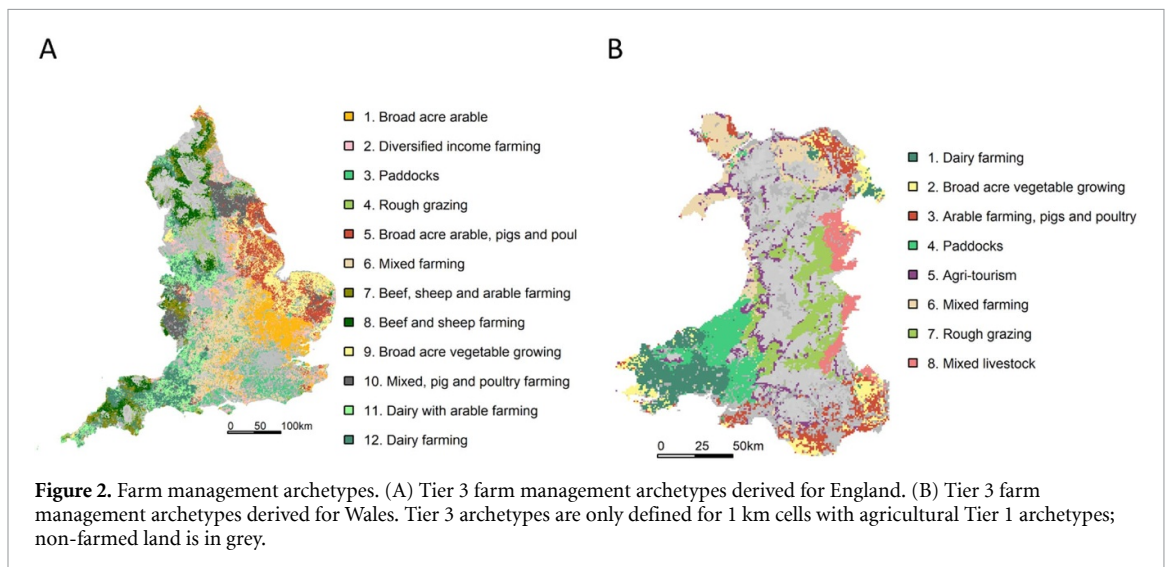
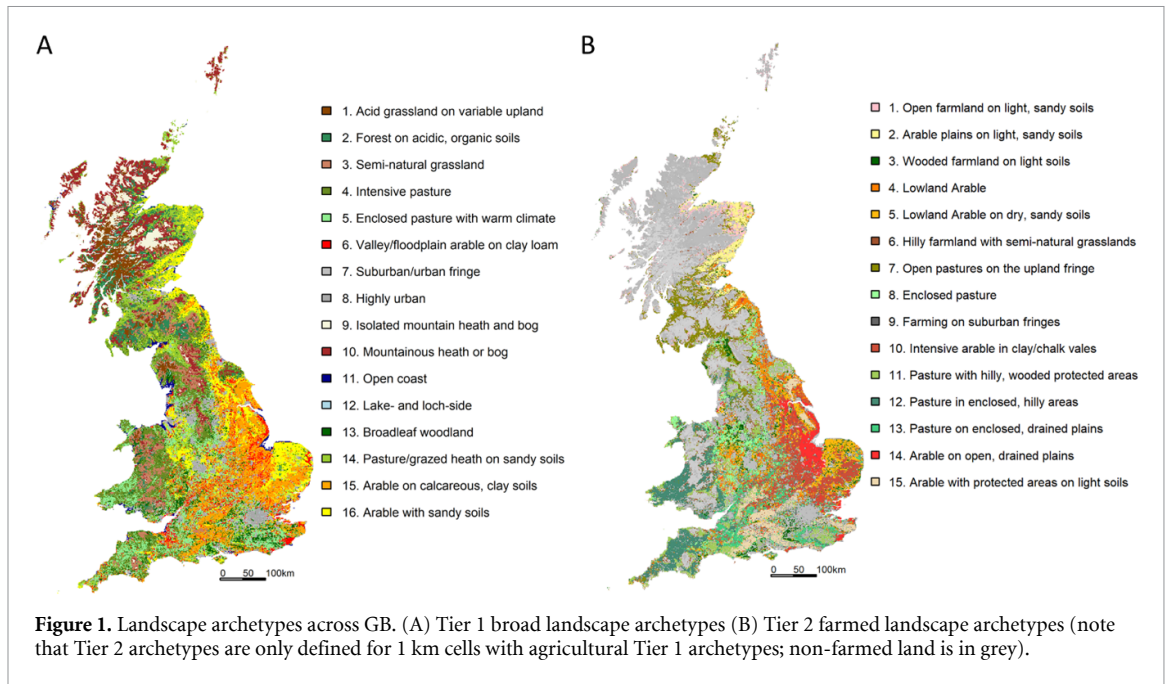
In England, Tier 2 archetypes were covered by a range of Tier 3 archetypes (figures 4(A) and S7). An average of 32% (21%–48%) of each Tier 2 archetype was covered by the Tier 3 archetype with which it most commonly co-occurred. Similarly, different Tier 3 archetypes were spread across each Tier 2 archetype (figure 4(B)); an average of 31% (21%–47%) of each Tier 3 archetype occurred within the Tier 2 archetype where it was most often found. The ten most common co-occurrences of Tier 2 and 3 archetypes covered 31% of agricultural land area (figure S8).

In Wales, Tier 2 landscapes were more associated with particular management archetypes (figures 5(A) and S8); with an average of 54% (18%–100%) of each Tier 2 archetype covered by a single Tier 3 archetype. Management archetypes were also mostly associated with a limited number of farmed landscapes (figure 5(B)); with an average of 53% (32%–88%) of each Tier 3 archetype found within one Tier 2 landscape. In Wales, 62% of the farmed land area was covered by the 10 most dominant combinations (figure S8).

3.3. Spatial patterns in landscape and agricultural management archetypes

Tier 1 archetypes showed varying degrees of spatial aggregation (table S1; figure S8(A)). Tier 2 archetypes were generally less, and less variably, spatially aggregated (table S2; figure S8(B)). Within Tier 3, archetypes generally had less variable aggregation indices in England (figure S8(C); table S3), and higher aggregation indices in Wales (figure S8(D); table S4).

The differing levels of aggregation led to varying archetype diversity across GB. Across Tier 1 and 2, landscapes (11×11 km) surrounding cells on the edges of more semi-natural archetypes (e.g. mountains and coasts) had the highest average difference values (Figures S9(A) and (B)), whereas areas with the highest spatial interspersions of archetypes were in regions of mixed farming, semi-natural areas and around large cities (figures S10(A) and (B)). When



these two measures were combined, the highest spatial diversity of archetypes (figures 6(A) and (B)) were similarly around coasts, edges of large cities and areas where there are large changes in landscape character.

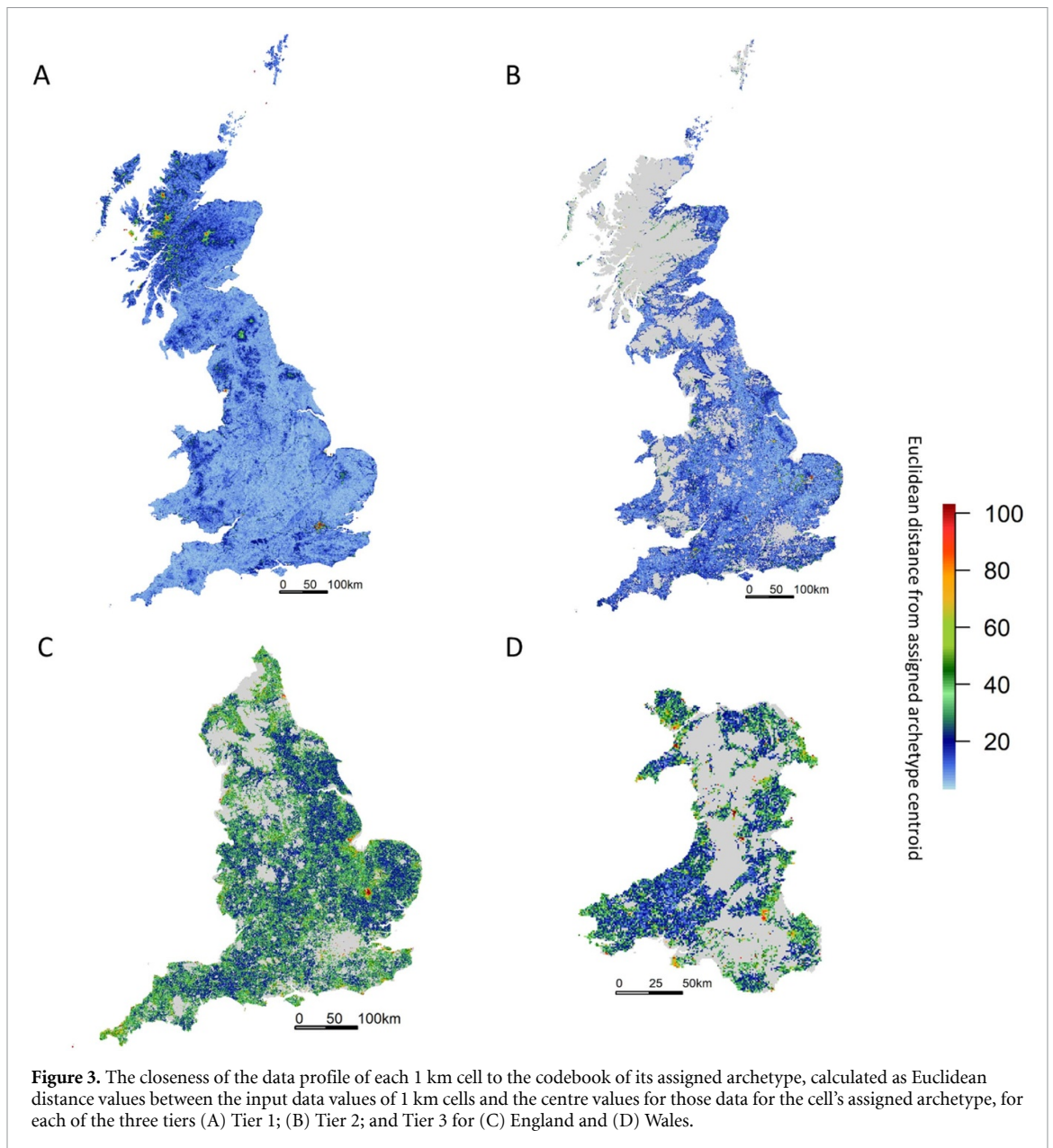
For Tier 3 archetypes in England and Wales, the edges of areas of similar archetype composition had high archetype difference indices (figures S9(C) and (D)), whereas farm management in areas surrounding other types of land use (e.g. surrounding large cities, forests and upland areas) had higher levels of interspersion (figures S10(C) and (D)). The resulting highest levels of diversity in farm management occurred in areas surrounding other land uses and areas of transition between farm management archetypes (figures 6(C) and (D)).

4. Discussion

4.1. Insights and implications for sustainable land use

Archetypes of farming landscapes and practises provide a simple, robust basis for a wide variety of analyses, by reducing multiple complex sources of variation into typologies.

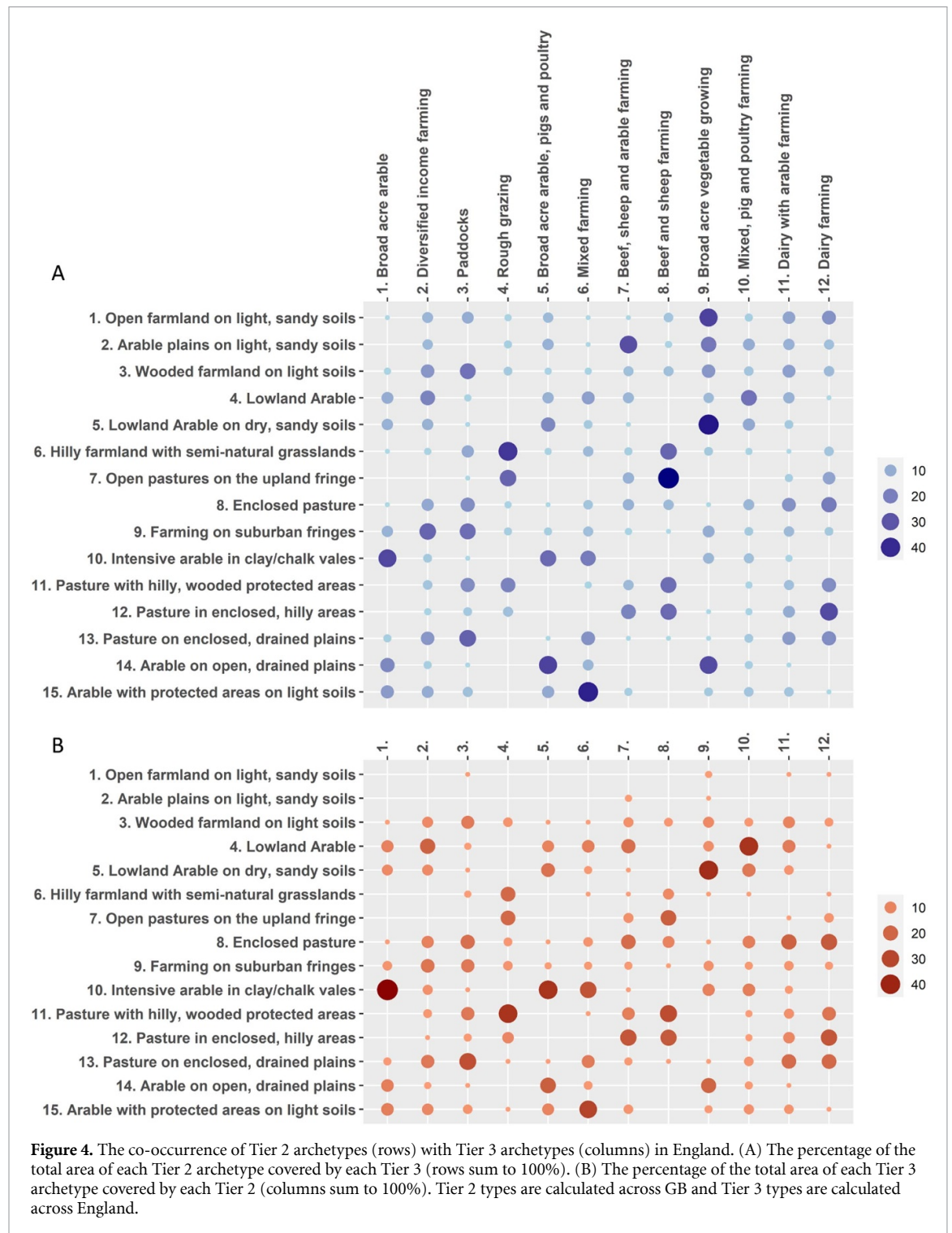
One of the benefits of the non-nested derivation of archetypes is that multiple archetypes in one tier can occur within a single archetype of another tier. The relationships found between archetype tiers can thus be used to distinguish the management strategies associated with particular landscapes. For example, Tier 2 archetypes with more semi-natural habitats predominantly coincide with livestock management



Tier 3 archetypes, which could be targeted for measures to restore these landscapes, a key policy priority in many areas [74–76]. Similarly, some arable and pasture Tier 2 landscapes are characterised by the presence of conservation features, e.g. hedgerows and protected areas [76–78], so improvements to these features could be targeted within the mixed and dairy farming management archetypes with which they most often co-occur. In other cases, the occurrence of a single Tier 3 archetype across multiple Tier 2 archetypes may help indicate widely applicable systems. For example, some more mixed management Tier 3 archetypes occurred across many landscape archetypes. Mixed management styles can be associated with transitions to more sustainable farming [18], so the lack of association with specific Tier 1 or 2 archetypes suggests that these mixed Tier 3 archetypes are less dependent on specific land

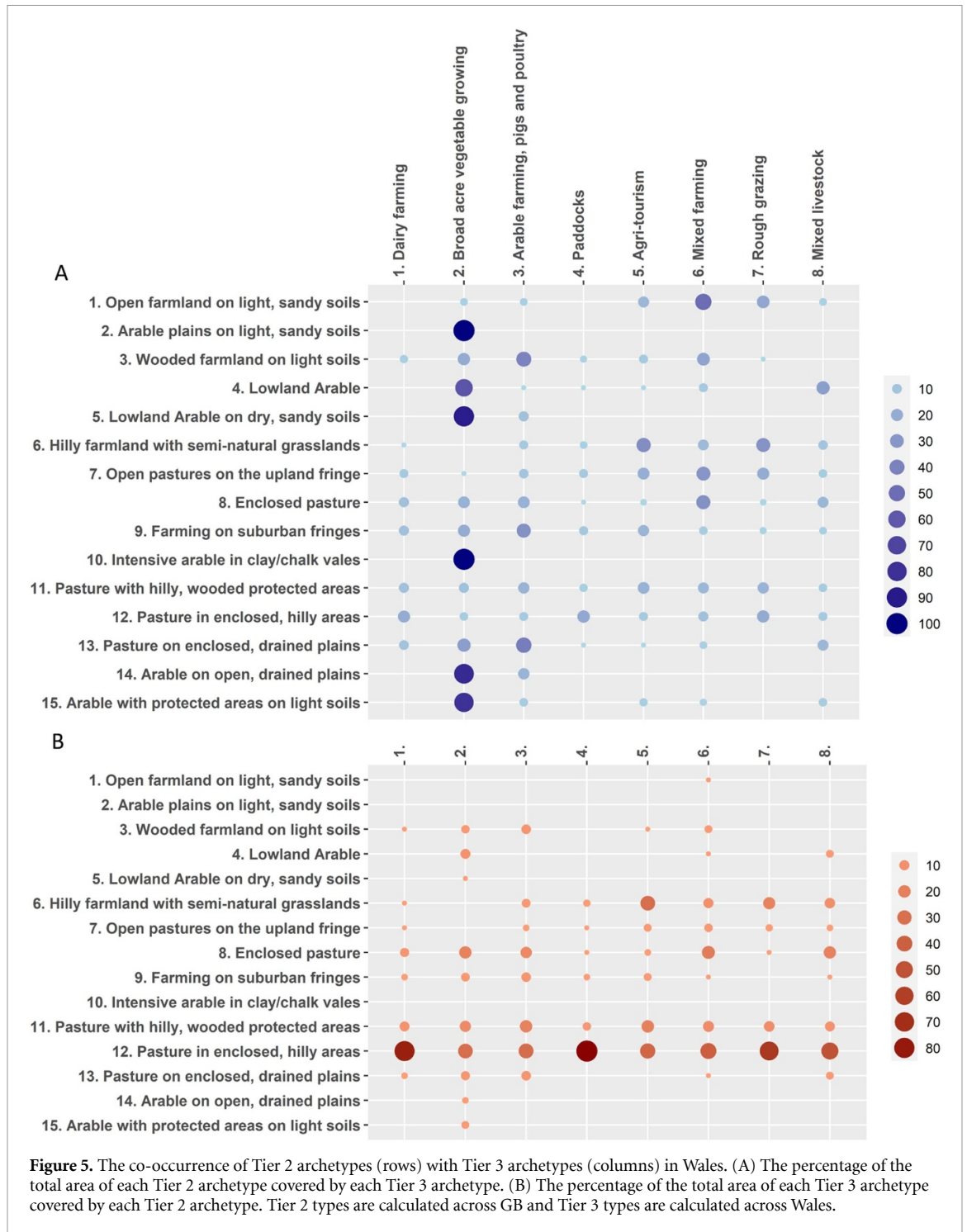
conditions and could be applied in many different landscapes.

The spatial structure of the archetypes is also informative; archetype diversity identifies more varied land systems at the wider scale, which have been associated with beneficial environmental outcomes [17, 79, 80]. Areas with a high diversity of different, interspersed landscapes but a narrow range of management archetypes can be candidates for diversification of management, whilst less diverse, aggregated landscapes with a range of management archetypes could be examined to assess the viability of implementing more diverse farming practise in other parts of the country. Spatially interspersed, diverse wider land systems, for example those north and north west of London, are also more likely to be able to deliver a wider variety of co-ordinated land uses and ecosystem services at the local scale. The non-nested tiers also



mean that one tier effectively defines the opportunity space for transitions between archetypes of another tier. The diversity of Tier 3 archetypes within a given Tier 2 landscape is of particular interest as this represents the existing opportunity space for land managers to adapt within the same constraints to achieve better environmental outcomes. However, it is important to note that archetypes characterise the main axes of existing variation and so more desirable, but currently rare or non-existent, management systems will not be represented. Although this analysis does

not consider socio-economic barriers, it provides a framework for exploring potential trajectories of change. This is particularly relevant in GB for land management policy developments following the UK's exit from the EU. Specifically within the Landscape Recovery scheme of the new Environmental Land Management scheme (ELMs), archetypes could contribute towards understanding how coordination can be achieved across landscapes by defining the range of plausible farm systems (Tier 3) within landscapes of similar environmental constraints (Tier 2) [81]. The



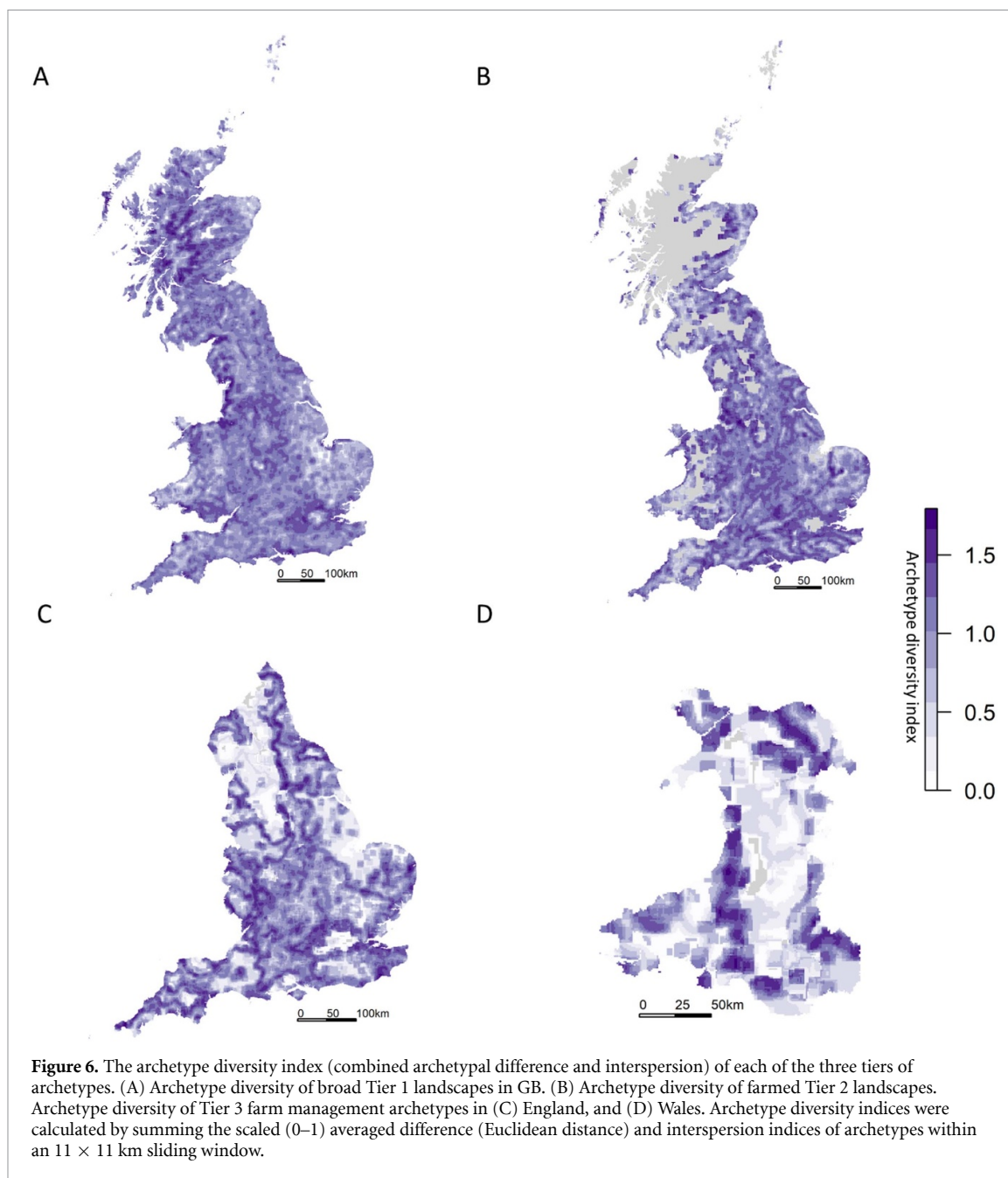
distances of specific 1 km cells from their assigned, and alternate, archetypes could then indicate which cells have greatest potential for transitioning to other landscape and management types, or to identify the opportunities for specific beneficial transitions e.g. afforestation to create more wooded landscapes, or farmland with more areas of woody semi-natural habitat, a major goal of current national policy [7] including habitat creation goals in the UK 25 year environment plan [82].

There is also potential to directly link archetypes with their environmental performance, and

thus demonstrate their validity empirically and for application in decision making [63, 83]. Cells of the same archetype can be compared on their environmental performance, and the reasons for the variation explored. Further research is assessing this by comparing the sustainability and ecosystem service delivery profiles of archetypes, leading to identification of pathways to sustainable farming systems [84].

4.2. Robustness and limitations

The data-driven SOM approach is useful in providing an objective classification of land use systems [85, 86]



based on clusters of co-occurring traits. We have also been able to assess their internal validity through measures of within-archetype variation, and the consistency of their production [83]. An inevitable part of capturing these dominant patterns, however, is that national-level variation may not always fully capture regional and local contexts, which should be considered when interpreting [28] and using the archetypes in practice. Consultation with stakeholders in specific regions would help to elucidate where regional variations differ from the national level, and how the maps generated at a 1 km scale could be downscaled [10, 87].

Areas that were less well described by the archetypes can be identified by their higher Euclidean distance values, indicating that they were in some

way unusual [10, 11]. Therefore, care is needed in interpretation of the archetypes assigned to these areas. However, the ability to highlight the uniqueness of such areas may be advantageous.

The validity of archetype design depends on the attributes selected [83]. We were unavoidably limited by the availability and resolution of input data. Some habitats and features, such as small woods, fall below the spatial resolution of national extent datasets such as LCM2015. Additionally, important components of management practise, such as tillage and fertiliser use, were not available. Some variables included in the third tier of archetypes were only available at the larger 10 km scale, and may not fully capture 1 km variation. This could have affected the spatial patterning of Tier 3 archetypes. However

aggregation indices for Tier 3 archetypes were not especially high, and so while fewer archetypes might occur in some landscapes, there is no indication that interpolation had driven systematically high levels of spatial aggregation. Data collection timescales could also have affected results if shifts in policies or markets caused changes in agriculture across that time. However, as most data were from 2010 to 2015 we do not expect this to have significantly affected the results.

5. Conclusions

We have revealed major drivers, spatial patterns and interrelationships of differences in broad landscapes, farming landscapes and farm management systems for GB. Examining the major groupings of farming systems and how these relate to landscape can help to target different land management strategies, and assess their likely impact on management and landscape elements of farms. The archetypes we have created form a valuable dataset for future research into the delivery of multiple outcomes, and the design and application of sustainable land use planning, from incentivising changes in farmer behaviour to policy drivers of large-scale land use change.

Data availability statement

The archetypes produced by this analysis will be made available as spatial datasets on the Environmental Information Data Centre: <https://doi.org/10.5285/3b44375a-cbe6-468c-9395-41471054d0f3>. R scripts are available on Github: https://github.com/cecilyg/Archetype_SOM.

Acknowledgments

We would like to thank the AgLand project team for co-ordination help and constructive discussion and advice. We would also like to thank the reviewers for their helpful comments. This work was funded by UKRI Grant NE/T000244/2.

ORCID iDs

Cecily E D Goodwin  <https://orcid.org/0000-0003-0093-9838>

Luca Bütikofer  <https://orcid.org/0000-0001-8220-7520>

Jack H Hatfield  <https://orcid.org/0000-0002-6361-0629>

Paul M Evans  <https://orcid.org/0000-0001-6706-420X>

James M Bullock  <https://orcid.org/0000-0003-0529-4020>

Jonathan Storkey  <https://orcid.org/0000-0003-1094-8914>

Andrew Mead  <https://orcid.org/0000-0002-4909-8235>

Goetz M Richter  <https://orcid.org/0000-0003-2180-9514>

Peter A Henrys  <https://orcid.org/0000-0003-4758-1482>

Richard F Pywell  <https://orcid.org/0000-0001-6431-9959>

John W Redhead  <https://orcid.org/0000-0002-2233-3848>

References

- [1] Tilman D, Balzer C, Hill J and Befort B L 2011 Global food demand and the sustainable intensification of agriculture *Proc. Natl Acad. Sci. USA* **108** 20260–4
- [2] Godfray H C J, Beddington J R, Crute I R, Haddad L, Lawrence D, Muir J F, Pretty J, Robinson S, Thomas S M and Toulmin C 2010 Food security: the challenge of feeding 9 billion people *Science* **327** 812–7
- [3] Verburg P H, Mertz O, Erb K H, Haberl H and Wu W 2013 Land system change and food security: towards multi-scale land system solutions *Curr. Opin. Environ. Sustain.* **5** 494–502
- [4] Pretty J and Bharucha Z P 2014 Sustainable intensification in agricultural systems *Ann. Bot.* **114** 1571–96
- [5] Rudel T K et al 2009 Agricultural intensification and changes in cultivated areas, 1970–2005 *Proc. Natl Acad. Sci. USA* **106** 20675–80
- [6] Green R E, Cornell S J, Scharlemann J P W and Balmford A 2005 Farming and the fate of wild nature *Science* **307** 550–6
- [7] Committee on Climate Change 2019 Net zero: the UK's contribution to stopping global warming *Committee on Climate Change* vol 275
- [8] Ekroos J, Olsson O, Rundlöf M, Wätzold F and Smith H G 2014 Optimizing agri-environment schemes for biodiversity, ecosystem services or both? *Biol. Conserv.* **172** 65–71
- [9] Cord A F et al 2017 Towards systematic analyses of ecosystem service trade-offs and synergies: main concepts, methods and the road ahead *Ecosyst. Serv.* **28** 264–72
- [10] Levers C et al 2018 Archetypical patterns and trajectories of land systems in Europe *Reg. Environ. Change* **18** 715–32
- [11] Václavík T, Lautenbach S, Kuemmerle T and Seppelt R 2013 Mapping global land system archetypes *Glob. Environ. Chang.* **23** 1637–47
- [12] Solecka I, Raszka B and Krajewski P 2018 Landscape analysis for sustainable land use policy: a case study in the municipality of Popielów, Poland *Land Use Policy* **75** 116–26
- [13] Tieskens K F, Schulp C J E, Levers C, Lieskovský J, Kuemmerle T, Plieninger T and Verburg P H 2017 Characterizing European cultural landscapes: accounting for structure, management intensity and value of agricultural and forest landscapes *Land Use Policy* **62** 29–39
- [14] van Asselen S and Verburg P H 2012 A land system representation for global assessments and land-use modeling *Glob. Change Biol.* **18** 3125–48
- [15] Malek Ž and Verburg P 2017 Mediterranean land systems: representing diversity and intensity of complex land systems in a dynamic region *Landsch. Urban Plan.* **165** 102–16
- [16] Ostrom E 2009 A general framework for analyzing sustainability of social-ecological systems *Science* **325** 419–22
- [17] Andersen E 2017 The farming system component of European agricultural landscapes *Eur. J. Agron.* **82** 282–91
- [18] Moraine M, Duru M and Therond O 2017 A social-ecological framework for analyzing and designing integrated crop-livestock systems from farm to territory levels *Renew. Agric. Food Syst.* **32** 43–56
- [19] Nair S S, Preston B L, King A W and Mei R 2016 Using landscape typologies to model socioecological systems: application to agriculture of the United States Gulf Coast *Environ. Modelling Softw.* **79** 85–95
- [20] Santos J L, Moreira F, Ribeiro P F, Canadas M J, Novais A and Lomba A 2020 A farming systems approach to linking

- agricultural policies with biodiversity and ecosystem services *Front. Ecol. Environ.* **19** 168–75
- [21] Karrasch L, Klenke T and Kleyer M 2019 Land-use elements and attributed ecosystem services: an archetype approach to land-use evaluation at the German north sea coast *Ecol. Soc.* **24** 13
- [22] Barrantes O, Ferrer C, Reiné R and Broca A 2009 Categorization of grazing systems to aid the development of land use policy in Aragon, Spain *Grass Forage Sci.* **64** 26–41
- [23] Vincent K 2007 Uncertainty in adaptive capacity and the importance of scale *Glob. Environ. Change* **17** 12–24
- [24] Sietz D, Ordoñez J C, Kok M T J, Janssen P, Hilderink H B M, Tittonell P and Van Dijk H 2017 Nested archetypes of vulnerability in african drylands: where lies potential for sustainable agricultural intensification *Environ. Res. Lett.* **12** 095006
- [25] Schirpke U, Leitinger G, Tasser E, Rüdissler J, Fontana V and Tappeiner U 2020 Functional spatial units are fundamental for modelling ecosystem services in mountain regions *Appl. Geogr.* **118** 102200
- [26] Eisenack K, Villamayor-Tomas S, Epstein G, Kimmich C, Magliocca N, Manuel-Navarrete D, Oberlack C, Roggero M and Sietz D 2019 Design and quality criteria for archetype analysis *Ecol. Soc.* **24** 6
- [27] Oberlack C *et al* 2019 Archetype analysis in sustainability research: meanings, motivations, and evidence-based policy making *Ecol. Soc.* **24** 26
- [28] Cullum C, Brierley G, Perry G L W and Witkowski E T F 2017 Landscape archetypes for ecological classification and mapping: the virtue of vagueness *Prog. Phys. Geogr.* **41** 95–123
- [29] Chon T S 2011 Self-organizing maps applied to ecological sciences *Ecol. Inform.* **6** 50–61
- [30] Astel A, Tsakovski S, Barbieri P and Simeonov V 2007 Comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets *Water Res.* **41** 4566–78
- [31] Giraudel J L and Lek S 2001 A comparison of self-organizing map algorithm and some conventional statistical methods for ecological community ordination *Ecol. Modelling* **146** 329–39
- [32] DEFRA 2021 Defra statistics: agricultural facts, England Regional Profiles (New York) (available at: www.gov.uk/government/statistics/agricultural-facts-england-regional-profiles)
- [33] Rowland C S, Morton R D, Carrasco L, McShane G, O'Neil A W and Wood C M 2017 Land cover map 2015 (1km percentage aggregate class, GB) (NERC Environmental Information Data Centre) (<https://doi.org/10.5285/505d1e0c-ab60-4a60-b448-68c5bbae403e>)
- [34] R Core Team 2020 *R: A Language and Environment for Statistical Computing* (Vienna: R Foundation for Statistical Computing)
- [35] Robinson E L, Blyth E, Clark D B, Comyn-Platt E, Finch J and Rudd A C 2017 Climate hydrology and ecology research support system meteorology dataset for Great Britain (1961–2015) [CHESS-met] v1.2 (NERC Environmental Information Data Centre) (<https://doi.org/10.5285/b745e7b1-626c-4ccc-ac27-56582e77b900>)
- [36] Soil Survey of Scotland Staff 1981 Soil maps of Scotland at a scale of 1:250 000 (Aberdeen) (<https://doi.org/10.5281/zenodo.4646891>)
- [37] Cranfield 2020 Soils data © Cranfield (NSRI) 2020 used with permission
- [38] Bell V A, Rudd A C, Kay A L and Davies H N 2018 Grid-to-grid model estimates of monthly mean flow and soil moisture for Great Britain (1891–2015): observed driving data [MaRIUS-G2G-Oudin-monthly] NERC Environmental Information Data Centre (<https://doi.org/10.5285/f52f012d-9f2e-42cc-b628-9cdea4fa3ba0>)
- [39] Hesselbarth M H K, Sciaini M, With K A, Wiegand K and Nowosad J 2019 landscapemetrics: an open-source R tool to calculate landscape metrics *Ecography* **42** 1648–57
- [40] Wang X, Blanchet F G and Koper N 2014 Measuring habitat fragmentation: an evaluation of landscape pattern metrics *Methods Ecol. Evol.* **5** 634–46
- [41] Riitters K 2019 Pattern metrics for a transdisciplinary landscape ecology *Landsc. Ecol.* **34** 2057–63
- [42] Morris D G and Flavin R W 1990 A digital terrain model for hydrology *Proc 4th Int. Symp. on Spatial Data Handling (Zürich, 23–27 July)* vol 1 pp 250–62
- [43] Moore R V, Morris D G and Flavin R W 2000 CEH digital river network of Great Britain (1: 50,000) (NERC Environmental Information Data Centre)
- [44] Scholefield P A, Morton R D, Rowland C S, Henrys P A, Howard D C and Norton L R 2016 Woody linear features framework, Great Britain v.1.0 (NERC Environmental Information Data Centre)
- [45] Reis S *et al* 2017 UK gridded population 2011 based on census 2011 and land cover map 2015 (NERC Environmental Information Data Centre)
- [46] OS Open Roads 2019 Great Britain. Updated: April 2019 (available at: www.ordnancesurvey.co.uk/business-government/products/open-map-roads)
- [47] UNEP, WCMC, and IUCN 2020 Protected planet: the world database on protected areas (WDPA) 04/2020 (Cambridge: UNEP, WCMC, IUCN) (available at: www.protectedplanet.net)
- [48] HE 2020 Scheduled monument data *Historic England. Contains Ordnance Survey Data © Crown Copyright and Database Right 2020, Licensed under the Open Government Licence* (available at: <https://historicengland.org.uk/listing/the-list/data-downloads/>) (Accessed 1 March 2020)
- [49] HES 2020 Scheduled monument data *Contains Historic Environment Scotland and Ordnance Survey data © Historic Environment Scotland—Scottish Charity No. SC045925 © Crown Copyright and Database Right 2020, Licensed under the Open Government Licence v3* (available at: <https://portal.historicenvironment.scot/downloads>) (Accessed 1 March 2020)
- [50] CADW 2020 Designated historic asset GIS data *The Welsh Historic Environment Service (CADW), Licensed under the Open Government Licence v3* (available at: <https://cadw.gov.wales/advice-support/cof-cymru/downloads>) (Accessed 1 May 2020)
- [51] Henrys P *et al* 2016 Dynamic typology tool—initial development and future roadmap *Report for Defra project LM0302 Sustainable Intensification Research Platform Project 2: Opportunities and Risks for Farming and the Environment at Landscape Scales*
- [52] Guiomar N *et al* 2018 Typology and distribution of small farms in Europe: towards a better picture *Land Use Policy* **75** 784–98
- [53] Dufloy R, Ernoult A, Aviron S, Fahrig L and Burel F 2017 Relative effects of landscape composition and configuration on multi-habitat gamma diversity in agricultural landscapes *Agric. Ecosyst. Environ.* **241** 62–69
- [54] Natural England 2020 Countryside Stewardship Scheme agreements (available at: <https://data.gov.uk/dataset/77e84ce1-5f4e-47c6-8bd2-e5ffe1dec9ba/countryside-stewardship-scheme-agreements>) (Accessed 1 February 2021)
- [55] Natural England 2018 Environmental Stewardship Scheme agreements (available at: <https://data.gov.uk/dataset/cff0728e-5179-46e5-bc5a-ab3a7dca8518/environmental-stewardship-scheme-agreements>) (Accessed 1 February 2021)
- [56] Natural England 2020 Organic farming scheme agreements (available at: <https://data.gov.uk/dataset/f79c982a-761f-4c2f-88a7-cc732655add3/organic-farming-scheme-agreements>) (Accessed 1 February 2021)

- [57] lle 2018 Glastir common, glastir organics, glastir gae ga gwc (available at: <http://lle.gov.wales/catalogue/item/Glastir2018Download>) (Accessed 1 February 2021)
- [58] Gräler B, Pebesma E and Heuvelink G 2016 Spatio-temporal interpolation using gstat *R J.* **8** 204–18
- [59] UKCEH 2019 Land cover plus: crops. 2015–2019 *UK Centre for Ecology and Hydrology* (available at: www.ceh.ac.uk/services/ceh-land-cover-plus-crops-2015)
- [60] Patton D R 1975 A diversity index for quantifying habitat 'edge' *Wildl. Soc. Bull.* **3** 171–3
- [61] EDINA Agcensus 2010 *Edina Agcensus* (available at: <http://agcensus.edina.ac.uk/>)
- [62] Jarvis S, Henrys P, Redhead J, Osorio B D S and Pywell R 2019 CEH land cover plus: pesticides 2012–2016 (England and Wales)
- [63] Anthony S *et al* 2016 Landscape typology—a framework for prioritisation of sustainable intensification strategies *Report for Defra Project LM0302 Sustainable Intensification Research*
- [64] Simpson E H 1949 Measurement of diversity *Nature* **163** 688
- [65] Mcgarigal K 2015 Fragstats version 4 Help pp 1–182
- [66] Kohonen T 1990 The self organizing map *Proc. IEEE* **78** 1464–80
- [67] Yin H 2008 The self-organizing maps: background, theories, extensions and applications *Stud. Comput. Intell.* **115** 715–62
- [68] Lek S and Guégan J F 1999 Artificial neural networks as a tool in ecological modelling, an introduction *Ecol. Modelling* **120** 65–73
- [69] Dittrich A, Seppelt R, Václavík T and Cord A F 2017 Integrating ecosystem service bundles and socio-environmental conditions—a national scale analysis from Germany *Ecosyst. Serv.* **28** 273–82
- [70] Wehrens R and Kruisselbrink J 2018 Flexible self-organizing maps in kohonen 3.0 *J. Stat. Softw.* **87** 1–18
- [71] Zhang W, Wang J, Jin D, Oreopoulos L and Zhang Z 2019 A deterministic self-organizing map approach and its application on satellite data based cloud type classification *Proc.—2018 IEEE Int. Conf. Big Data, Big Data 2018* pp 2027–34
- [72] Murtagh F and Legendre P 2014 Ward's hierarchical agglomerative clustering method: which algorithms implement ward's criterion? *J. Classif.* **31** 274–95
- [73] Goodwin C E D, Butikofer L, Hatfield J H, Richter G M and Redhead J W 2022 Multi-tier archetypes to characterise British landscapes, farmland and farming practices *NERC EDS (Environmental Information Data Centre)* (<https://doi.org/10.5285/3b44375a-cbe6-468c-9395-41471054d0f3>)
- [74] Gimona A, Poggio L, Brown I and Castellazzi M 2012 Woodland networks in a changing climate: threats from land use change *Biol. Conserv.* **149** 93–102
- [75] Maxwell S L, Reside A, Trezise J, McAlpine C A and Watson J E 2019 Retention and restoration priorities for climate adaptation in a multi-use landscape *Glob. Ecol. Conserv.* **18** e00649
- [76] Sullivan M J P, Pearce-Higgins J W, Newson S E, Scholefield P, Brereton T and Oliver T H 2017 A national-scale model of linear features improves predictions of farmland biodiversity *J. Appl. Ecol.* **54** 1776–84
- [77] Kremen C and M'Gonigle L K 2015 EDITOR'S CHOICE: small-scale restoration in intensive agricultural landscapes supports more specialized and less mobile pollinator species *J. Appl. Ecol.* **52** 602–10
- [78] Troupin D and Carmel Y 2014 Can agro-ecosystems efficiently complement protected area networks? *Biol. Conserv.* **169** 158–66
- [79] Bennett A F, Radford J Q and Haslem A 2006 Properties of land mosaics: implications for nature conservation in agricultural environments *Biol. Conserv.* **133** 250–64
- [80] Weissteiner C J, García-Feced C and Paracchini M L 2016 A new view on EU agricultural landscapes: quantifying patchiness to assess farmland heterogeneity *Ecol. Indic.* **61** 317–27
- [81] U. K. Government 2022 Environmental land management schemes: overview (available at: www.gov.uk/government/publications/environmental-land-management-schemes-overview/environmental-land-management-scheme-overview)
- [82] Defra 2018 Environment 25 year plan (London) (available at: www.gov.uk/government/publications/25-year-environment-plan)
- [83] Piemontese L *et al* 2022 Validity and validation in archetype analysis: practical assessment framework and guidelines *Environ. Res. Lett.* **17** 025010
- [84] Butikofer L *et al* Pathways to sustainable farming using archetypes and multi-objective optimisation In review
- [85] Sietz D, Frey U, Roggero M, Gong Y, Magliocca N, Tan R, Janssen P and Václavík T 2019 Archetype analysis in sustainability research: methodological portfolio and analytical frontiers *Ecol. Soc.* **24** 34
- [86] Simensen T, Halvorsen R and Erikstad L 2018 Methods for landscape characterisation and mapping: a systematic review *Land Use Policy* **75** 557–69
- [87] van der Zanden E H, Levers C, Verburg P H and Kuemmerle T 2016 Representing composition, spatial structure and management intensity of European agricultural landscapes: a new typology *Landsc. Urban Plan.* **150** 36–49