Agricultural and Environmental Modelling 8: e167439 (2026)
DOI: 10.3897/aem.8.167439

Methods

Development of a spatio-temporal representation of agricultural landscapes as the modelling environment for spatially explicit agent-based models in the Animal Landscape and Man Simulation System (ALMaSS)

Elżbieta Ziółkowska[1,2,3], Barbara Jaśkowiec[1], Geoffrey Brian Groom[3], Christopher John Topping[3]

1 Institute of Environmental Sciences, Faculty of Biology, Jagiellonian University, Kraków, Poland
2 Institute of Geography and Spatial Management, Faculty of Geography and Geology, Jagiellonian University, Kraków, Poland
3 Social-Ecological Systems Simulation Centre, Department of Agroecology, Aarhus University, Aarhus, Denmark
Corresponding author: Elżbieta Ziółkowska (e.ziolkowska@uj.edu.pl)

OPEN ACCESS

This article is part of:
PollinERA – Understanding pesticide-Pollinator interactions to support EU Environmental Risk Assessment and policy
Edited by Christopher John Topping, Xiaodong Duan, Fabio Sgolastra, Andreas Focks, James Henty Williams

Academic editor:
James Henty Williams
Received: 1 August 2025
Accepted: 9 March 2026
Published: 29 April 2026

Citation: Ziółkowska E, Jaśkowiec B, Groom GB, Topping CJ (2026) Development of a spatio-temporal representation of agricultural landscapes as the modelling environment for spatially explicit agent-based models in the Animal Landscape and Man Simulation System (ALMaSS). Agricultural and Environmental Modelling 8: e167439. https://doi.org/10.3897/aem.8.167439

Abstract

The model environment constitutes the core component of the Animal, Landscape, and Man Simulation System (ALMaSS). ALMaSS was developed to evaluate the effects of changes in landscape structure and management on key animal species within agro-ecosystems. Consequently, it is designed to work with representations of actual agricultural landscape areas, capturing both spatial and temporal landscape heterogeneity to fulfil the specific requirements of the ALMaSS species models.

This article presents the methodology for describing the land-use/land-cover of agricultural landscapes and generating detailed landscape representations for use in ALMaSS landscape simulations. We outline the external data requirements and data handling procedures necessary to prepare the final set of input files for an ALMaSS run. We provide a mapping algorithm to generate a landscape (land-use/land-cover) raster map and describe the methods for classifying and defining ALMaSS farm types and crop rotations. We present exemplary results and discuss potential applications beyond the ALMaSS modelling framework. Finally, we examine the ALMaSS landscape model generation process in the context of input data quality, accessibility, and data processing challenges and offer a perspective on future developments.

Key words: Agricultural data, farming structure, Integrated Administration and Control System (IACS), land-use/land-cover data, Land Parcel Identification System (LPIS), landscape heterogeneity, landscape simulation

Introduction

The proper definition of the model environment, which includes all necessary properties of the system that may affect the behaviour of modelled entities, is a critical component of spatially explicit agent-based models (ABMs) of complex socio-ecological systems (Tang and Bennett 2010; McLane et al. 2011). The model environment provides the spatio-temporal context for agents, allowing them to obtain information, make decisions, and behave accordingly. It can be designed at different levels of detail, depending on the level of spatial complexity included, the most basic being the definition of the spatial bounds for the agents. The model environment may represent a real-world situation or be artificially created, integrating biotic or abiotic factors. It may utilise different data structures, i.e. discrete object-based, continuous grid-based, or network/graph-based structures, and may be implemented as either static or dynamic (Tang and Bennett 2010; McLane et al. 2011).

Spatial data and Geographic Information System (GIS) techniques and functionalities may form part of the external data-handling procedures for ABMs or may be directly integrated into the simulation/modelling systems used to build ABMs, such as (Geo)MASON (Luke et al. 2005; Sullivan et al. 2010), Repast (North et al. 2006), or GAMA (Taillandier et al. 2019). Within such simulation/modelling systems, the model environment must be designed separately for each ABM, depending on its purpose and the research question under investigation. In terms of modelling species population dynamics with spatially explicit ABMs, this often means adopting a species-specific perspective and viewing the environment from the perspective of the resources available, i.e. including layers of information representing different biotic variables, e.g. food resources and human density as disturbance, or abiotic environmental variables, e.g. elevation and topography, related to species movement or foraging activities. For example, in the SEARCH (Spatially Explicit Animal Response to Composition of Habitat) programme (Pauli et al. 2013), the environment is represented as four polygon maps, defined separately for movement, food, risk, and suitability, and one point map assigning the location and number of released individuals. Similarly, in the BEEHAVE and Bumble-BEEHAVE models, for honeybee and bumblebee colonies, respectively (Becher et al. 2014, 2018), the landscape is viewed from the perspective of pollination activities, with flowering patches providing nectar and pollen. These models illustrate how the species-specific environmental perspective translates into computational representation: the map of flowering resources becomes a discrete habitat agent network in which individual agents, i.e. flowering patches, are characterised by state variables reflecting their ecological relevance, e.g. type, distance from the colony, size, and daily nectar and pollen production. This approach to modelling resource landscapes as agent networks has proven transferable, as demonstrated by its application in the SyrFitSource model of the hoverfly *Episyrphus balteatus*, whose adult stage feeds on nectar and pollen (App et al. 2025).

The Animal, Landscape and Man Simulation System (ALMaSS) (Topping et al. 2003; Topping 2022) is designed as a C++ system of spatially explicit ABMs for predicting the effects of human landscape management on a range of key animal species. Unlike other simulation/modelling systems, the model environment in ALMaSS has been designed to serve all other species models in the system and thus provide all the information needed by any other model in space and time. The range of species for which ABMs already exist in the ALMaSS system is broad, both in terms of size, from ground beetles to hares and roe deer; movement and dispersal capabilities, from a few metres to a few kilometres per day; and resource use, e.g. pollinators relying on floral resources, pest-control species using open agricultural land, and hares and skylarks using meadows and grasslands. This range of models has required the model environment to be sufficiently detailed and flexible to reflect accurately the complexity of the real landscape, which is crucial for all modelled species. In addition, because agricultural landscapes have been an important focus ecosystem for ALMaSS since the beginning of its development, the model environment has had to represent the specific spatial and temporal heterogeneity resulting from the organisation of cropping systems by farmers (Vasseur et al. 2013).

The emphasis on realistic landscape representation in ALMaSS reflects its operational role in European policy assessment. ALMaSS is regularly employed to assess the impacts of agricultural and environmental policies on biodiversity and ecosystem services across European landscapes (e.g. Ziółkowska et al. 2022, 2024). The use of real agricultural landscapes in these assessments, rather than simplified or theoretical representations, provides crucial advantages for integrated policy support. By working with actual landscape structures that capture the full complexity of farming systems, land ownership patterns, and ecological features, ALMaSS enables the evaluation of multiple policy instruments simultaneously within the same landscape context. This capability allows policymakers to identify synergies between various measures, such as agri-environmental schemes, biodiversity conservation programmes, and pesticide regulations, while revealing potential conflicts that might otherwise remain hidden (Nilsson et al. 2012). Such integrated assessments are particularly valuable given the increasing complexity of agricultural and environmental policy frameworks in the EU, where multiple directives and regulations must work coherently to achieve sustainable agricultural landscapes (Lefebvre et al. 2015).

The model environment in ALMaSS has been designed as a landscape simulation that captures both spatial and temporal heterogeneity in the landscape. Spatial heterogeneity is represented by the ALMaSS landscape map, a detailed land-use/land-cover raster with 1 m^2 resolution, typically covering 10 × 10 km. Each cell of the ALMaSS landscape map references a homogeneous landscape polygon characterised by its land-use/land-cover type and associated attributes, including soil type, elevation, and farm type. The temporal component allows polygon properties such as vegetation height, biomass, floral resources, and crop management to change on a daily time step, capturing the dynamic character of agricultural landscapes throughout the year.

The software implementation of the ALMaSS landscape model, including its class architecture and internal data structures, is described in Topping and Duan (2024). The present paper addresses a complementary and independently applicable problem: how to acquire, integrate, and process the heterogeneous geospatial data required to generate a high-resolution agricultural landscape map suitable for spatially explicit ABM simulation. This is a non-trivial challenge because input datasets differ in format, coordinate reference system, spatial resolution, thematic scope, and provenance, and their combination inevitably produces artefacts and data gaps that must be resolved before a consistent, complete raster can be produced. The core methodological contribution of this paper is a multi-stage data-integration and quality-assurance algorithm for agricultural landscape map generation, including an improved procedure for the systematic removal of sliver artefacts and unclassified gaps, representing a significant advance on the approach first described in Topping et al. (2016). Although the methodology was developed in the context of ALMaSS and the format of the output files conforms to its input specifications, the data-processing pipeline is general, and the resulting land-use/land-cover rasters are able to support any spatially explicit model requiring a detailed representation of agricultural landscape structure. We demonstrate this by discussing uses of the generated maps beyond ALMaSS simulation.

Methods

Input files for landscape simulation in ALMaSS

The data-processing methodology described in this paper, focusing on the spatial component of the ALMaSS landscape model (Fig. 1), enables the generation of a raster landscape map and additional data outputs necessary to produce three mandatory external input files required to run a landscape simulation in ALMaSS: a binary raster map (.lsb format), a polygon reference text file (PolyRef.txt), and a farm reference text file (FarmRef.txt). The full specification of these file formats and their implementation within the ALMaSS software architecture is described in Topping and Duan (2024); here we provide only the detail necessary to understand the data requirements that drive the processing pipeline.

The .lsb raster map defines the spatial extent and resolution of the landscape and assigns each raster cell a reference number linking it to a specific landscape polygon. The PolyRef file describes the attributes of each polygon using that reference number as a unique identifier. The current list of polygon attributes is shown in Table 1; it is this set of attributes that the external data processing must ultimately populate. The FarmRef file lists all farm units within the landscape, each assigned a farm type, and links them to individual field polygons via the PolyRef file. Each farm can be further linked with a specific crop rotation scheme described in a .rot text file. Crop management plans, weather inputs, vegetation growth curves, and floral resource curves, which drive the temporal component of the landscape simulation, are generated separately and are not covered here; details are provided in Topping and Duan (2024) and Topping and Olesen (2005).

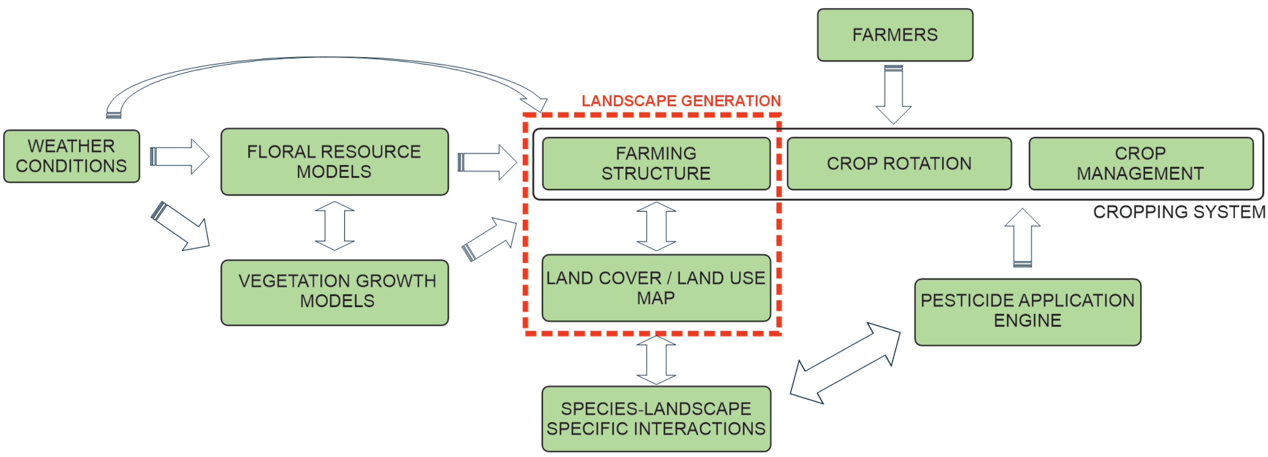

Figure 1. Simplified overview of the elements of the ALMaSS landscape model. Elements forming the spatial component addressed in this paper are marked in red.

Table 1. Polygon attributes referenced in the polygon reference (PolyRef) input file. The data category provides the basic indication of the spatial data required to define a given attribute. The data categories are further described in the 'Main spatial data categories and applied data sources'. A data category set to 'Output of external data processing' indicates that an attribute is automatically calculated during data-processing steps (see 'Overview of data-processing steps' for details).

Attribute	Description	Domain	Data category
PolyType	Polygon type.	An integer number according to the ALMaSS classification of Type of Landscape Elements (TOLEs).	Topography Land use / land cover Habitat and vegetation Agriculture
PolyRefNum [key]	Unique identifier of a polygon on the .lsb raster landscape map.	An integer number from 0 to one less than the number of polygons within a landscape map.	Output of external data processing
Area	Polygon's area in m^2.	Integer numbers.	Output of external data processing
FarmRef	Reference number of a farm to which the polygon belongs and is managed.	An integer number from 0 to one less than the number of farm units within a landscape map. −1 if a polygon is not part of a farm.	Agriculture
UnSprayedMarginRef	PolyRefNum of the associated unsprayed margin, if applicable. It indicates whether the polygon (if it is a field) has a field margin not treated by pesticides.	An integer number in the range of PolyRefNum values.	Output of external data processing
SoilType	Code of the majority soil type, if known.	Integer number from 0 to 14 denoting the soil type as referenced inside ALMaSS code with their FAO soil texture classification and agricultural usage class of crop management (see Topping and Duan 2024).	Soil
Openness	Measure of distance to the nearest tall or large artificial object from the polygon's centre point (CentroidX and CentroidY).	Integer number.	Calculated within ALMaSS after an initial load and run of the landscape map
CentroidX	Raster-relative X coordinate of the centroid of the polygon.	Integer number, limited by the size of the landscape raster map.	Calculated within ALMaSS after an initial load and run of the landscape map
CentroidY	Raster-relative Y coordinate of the centroid of the polygon.	Integer number, limited by the size of the landscape raster map.	Calculated within ALMaSS after an initial load and run of the landscape map
Elevation [optional]	Mean elevation of a polygon.	Integer number denoting the meters a.s.l.	Surface
Slope [optional]	Mean slope of a polygon.	Integer number denoting degree.	Surface
Aspect [optional]	Mean aspect of a polygon.	Integer number denoting the compass direction the surface faces at that location. It is measured clockwise in degrees from 0 (due north) to 360 (again due north), coming full circle. Flat areas are given a value of −1.	Surface
PollenNectarCurve [optional]	Reference number of a habitat/flora type associated with a certain plant composition, allowing modelling of the production of floral resources (pollen, nectar and sugar).	Integer number according to the ALMaSS habitat/flora look-up table.	Habitat and vegetation

Main spatial data categories and applied data sources

Land-cover/land-use and topography

This category includes spatial data representing topographic features related to land cover and land use, such as the water network, the transport network (including roads and railways), buildings, and other infrastructure elements. All important natural or semi-natural landscape features (e.g. rivers or streams, lakes, bare sand or rock) and forms of semi-natural vegetation (e.g. hedgerows, open vegetation types, forest, isolated trees) should be adequately represented.

Such information typically originates from various data sources, making data integration a crucial component of landscape data processing. In most cases, the primary source of topography and land-cover/land-use information is national or regional topographic object databases, which provide a level of detail corresponding to topographic maps at scales of 1:5,000 to 1:25,000. Various satellite-based products often provide important additional layers of information. For example, the Copernicus Land Monitoring Service (https://land.copernicus.eu/en) offers several thematic layers with data at a spatial resolution of 10 m or higher (e.g. high-resolution forest-type layer or high-resolution small woody features layer). In addition, various national or regional thematic mapping products are available in many EU Member States.

Habitats and vegetation

Maps of habitats or vegetation types (and their associated vegetation) serve as an additional layer of information used by ALMaSS to detail land-cover classes and assign habitat or flora types for floral resource modelling. One such example is the Biological Valuation Map (BVM), a comprehensive database of land and vegetation cover for the Flemish Region of Belgium.

Agriculture

As noted earlier, accurately representing agricultural areas and their heterogeneity is particularly important for the ALMaSS landscape model. ALMaSS requires that polygons in the landscape map representing agriculturally managed areas be assigned to specific farm units, with each unit characterised by its farm type. Such a structure requires highly precise spatial and temporal agricultural data. For EU Member States, these are provided by the Integrated Administration and Control System (IACS), a control mechanism for Common Agricultural Policy (CAP) interventions. One of the elements of IACS is the Land Parcel Identification System (LPIS), which identifies and stores geographical information on reference parcels (geographically delimited areas with unique identification codes). The IACS system also maintains direct payment registers, which provide information on the type of crops grown on reference parcels; agricultural holding registers, which allow individual reference parcels to be grouped into farm units; and registers of animals kept for agricultural purposes. The system is managed at the national or regional level (depending on the Member State) and is usually updated annually. By combining different components of the IACS database, it is possible to subdivide agricultural land into individual fields and farm units and to classify farm units into types, principally by combining information on crops grown and animals reared per farm unit.

The LPIS reference parcel layer is integral to the EU subsidy payment system, as farmers applying for subsidies must delineate the parcel or parcels. However, the level of detail varies depending on the country or region. For landscapes where reference parcel information is scarce, incomplete, or difficult to obtain, a possible strategy is to use fine spatial resolution imagery (e.g. Sentinel-2 time series or orthophoto maps) and pixel- and object-based image processing and analysis methods to determine field boundaries and crop types. However, the use of IACS/LPIS data is recommended because imagery-based methods do not allow the delineation of farm units.

Additional information layers: soil and elevation

The ALMaSS landscape simulator modulates crop management on each field as a function of the dominant soil type. Soil maps are therefore required to define the dominant soil type for each polygon on a landscape map. ALMaSS employs a 15-class classification of soils based on the FAO soil texture classification, which is linked to agricultural use and soil management types (Topping and Duan 2024). Soil classes from national or regional soil maps must be mapped to those used in ALMaSS. Alternatively, soil data at the EU level can be obtained from the European Soil Database (https://esdac.jrc.ec.europa.eu/resource-type/datasets) provided by the Joint Research Centre.

Elevation, slope, and aspect are additional characteristics that can be defined at the landscape polygon level. These variables can be relevant for ALMaSS simulations because they may influence local environmental conditions affecting habitat suitability for modelled species and may also affect crop growth processes represented in the model. Digital elevation models (DEMs) can be used to provide elevation information directly and can be further processed using GIS tools to produce slope and aspect maps. Various national, European, or global DEM products with different spatial resolutions are commonly available, including those provided by the Copernicus Land programme (i.e. Copernicus DEM – Global and European Digital Elevation Model). At the national or regional level, high-resolution (e.g. 1 m) digital elevation models are often available as products of airborne laser scanning (ALS).

Overview of data processing steps

Here, we describe all processing steps applied to the input data described above that are required to generate the full set of mandatory external input files (a '.lsb' binary raster map file with an associated PolyRef.txt file describing the landscape structure and a FarmRef.txt file associating the farming structure with farm types) needed to run a landscape simulation in ALMaSS.

The ALMaSS landscape methodology was first developed for Danish landscapes, based on topographic, land-cover, and agricultural data for 2012–2013 (Topping et al. 2016). The mapping algorithm used to generate a landscape raster map was coded using a Python script and the Python library ArcPy to access geographic information system (GIS) functions available in ArcGIS. R scripts were used to prepare the PolyRef and FarmRef reference files (https://github.com/flemmingskov/python-landscapegen). Since then, the methodology has undergone several changes, one of the most important being the shift from filling data gaps in the landscape map with more general land-cover data to the development of a multi-stage algorithm that removes data inconsistencies (sliver polygons) while simultaneously filling data gaps. The new procedure was first tested and applied to Polish landscapes (Ziółkowska et al. 2021) and subsequently to Dutch landscapes (Ziółkowska et al. 2022). Furthermore, slight modifications were made during the implementation of the methodology in several other countries within the Horizon 2020 EcoStack project. Here, we present the most recent version, which is entirely coded in Python scripts (see section 'Workflow automation'), resulting from experience gained over several years of work with ALMaSS landscape simulations.

This section is divided into three parts. First, we present a mapping algorithm to generate a landscape raster map ('ALMaSS landscape map'). Second, we describe the method for classification and definition of farm types and crop rotations ('Farming structure and cropping system'). Third, we present the algorithm used to generate the ALMaSS input PolyRef and FarmRef reference files and to transform the raster ALMaSS map into the .lsb format for ALMaSS ('Producing ALMaSS inputs').

ALMaSS landscape map

The raster ALMaSS landscape map is a single-layer representation of reality, meaning that each landscape polygon corresponds to a distinct land-cover or land-use element, such as agricultural fields, forests, roads, or water bodies. As such, the ALMaSS landscape map does not incorporate vertical stratification of landscape elements and lacks a multi-layered representation of vegetation or structural complexity. However, two important aspects enrich this simplified structure. First, properties assigned to landscape polygons via the PolyRef text file allow the incorporation of additional "layers of information" related to structural complexity. For example, the PollenNectarCurve property enables the association of a given polygon with a specific plant composition. Second, the temporal component of the ALMaSS landscape model allows polygon properties to change over time, further enhancing the informational depth of the landscape representation.

The ALMaSS landscape map results from a multi-stage process of spatial data processing using GIS functions. The aim of this process is to obtain a raster map of landscape elements covering areas in a complete and disjoint manner, i.e. such that each pixel of the raster map is unambiguously assigned to a specific polygon representing a given type of landscape element. The whole process can be divided into several steps, as described below.

Identification of data sources and mapping of spatial information layers to TOLEs

ALMaSS uses its own typology of landscape elements (TOLEs), designed to support ABMs of the species present in the system (Suppl. material 1: table S1). The typology is therefore habitat-oriented, linking land-cover information with land use, vegetation, and topography. ALMaSS also utilises the classificatory attributes of the PolyRef text file, which hold information about specific characteristics of TOLEs. Each TOLE can then be individually interpreted from the perspective of modelled species, for example, by assigning quality values describing breeding habitat or food resources.

Identifying geodata that allow the mapping of ALMaSS TOLEs for a given study area is the first important step in the ALMaSS landscape model generation process. Reviewing the metadata, geographical scope, credibility, and thematic scope of potential input data is crucial for proper understanding of the data content and context. For example, within the ALMaSS TOLE typology, it is necessary to distinguish between managed and semi-natural grassland habitats. The first group includes permanent grasslands under agricultural management that may be mown, grazed, or fertilised. They are further subdivided into specific TOLEs according to the type and intensity of management (e.g. permanent pasture, low-yield permanent pasture) and linked to specific management plans. The second group includes natural areas of grasses and herbaceous plants, further subdivided into dry grasslands and wet grasslands. This distinction is important because vegetation type determines the availability of food resources, shelter, and other essential elements, while management type modifies these conditions. Unfortunately, grasslands are defined and mapped differently in national topographic databases across Europe, and even layers with the same thematic name can have different meanings (Table 2). This example demonstrates that thorough understanding of the input data is necessary to map ALMaSS TOLEs accurately.

The output of this step is a landscape data description table (LandDataDescriptionTable.csv) and a database organising the input geodata (in the form of a single geodatabase or folder). The landscape data description file lists all geospatial layers used in the landscape generation process, together with information on their definition or thematic scope and additional attributes required for subsequent processing steps. Depending on the structure of the input data, a particular information layer may be represented by a single geospatial layer or by a subset of a layer's features. In the latter case, a query expression must be defined on the source geospatial layer to obtain the features required for the information layer. Thus, the landscape data description file may distinguish between the information layer and the source geospatial layer, where the latter forms part of the input database and the former undergoes further processing steps. The file includes query expressions where necessary. Other information contained in the landscape data description file relates to the data type (vector or raster), spatial resolution (in the case of raster data), geometry type (in the case of vector data; point, line, or polygon), applied buffer (in the case of point and line vector data, used to convert zero- or one-dimensional objects into two-dimensional ones), and the corresponding TOLE code.

In addition, each information layer is assigned to one of the following themes: natural (including natural or semi-natural land-cover types), cultivable (including agriculturally managed areas: fields in rotation and permanent crops, such as orchards or pastures), built-up (including built-up areas of different types and individual buildings), cultural (complexes of land uses or other objects constituting important habitats, such as cemeteries or hedges), water (water network and related infrastructure), and transport (road and railway networks and related infrastructure). The information layers in each theme are assigned three-digit codes. The first digit of the code identifies the theme, while the other two form a number indicating the order of the layers in terms of their spatial priority, which is highly significant for the next step of processing the layers and their combination within each theme (see below).

A database organising the input geodata contains datasets harmonised to a common coordinate reference system (CRS). If necessary, input geodata are reprojected to the official, legally recognised projected CRS of the country or region for which the ALMaSS landscape map is generated. This ensures metric consistency and avoids distortions in distance-based operations (e.g. buffering and Euclidean distance calculations). The final ALMaSS landscape raster products are provided in this projected CRS.

Table 2. Comparison of grassland definitions in selected national topographic databases in Europe. The definitions differ in terms of thematic scope, inclusion criteria, and minimum mapping unit applied. The example demonstrates that the mapping between the objects represented in the topographic databases and the ALMaSS TOLEs is not necessarily one-to-one and may sometimes require the application of additional datasets for specification. In some cases, due to the lack of sufficiently detailed data, a certain degree of generalisation must be applied. For example, in the German topographic database, it is possible to distinguish between semi-natural grassland and grassland under agricultural management. In the Finnish topographic database only natural grassland is included, whereas in the Netherlands the topographic database does not allow a distinction between managed and unmanaged grassland.

Country	Topographic database name	Scale represented	Reference layer name with indication of the target group, object class and attribute*	Definition	Minimum mapping unit	Inclusion criteria
Finland	Maanmittauslaitoksen maastotietokohteet	1:10 000	Maasto/1: Niitty	A natural area of grasses and herbaceous plants.	5000 m^2 (0.5 ha)	Inclusions: • uncultivated fields that are in their natural state and are unlikely to be cultivated again Exclusions: • afforested or forested areas (these are set to 'forest land') • uncultivated field edges and yards (these are set to 'unclassified background') • manicured lawns (these are set to 'parks' or 'sports and recreational areas') • open mineral waterlogged areas along the coast (these are set to either 'lake' or 'open aquatic reservoirs' depending on the amount of hay, grass, reed and reed canary grass)
Netherlands	Basisregistratie Topografie (BRT) TOP10NL	1:5 000–1:25 000	Terrein/Type landgebruik/Grasland	Land predominantly covered with grassy vegetation.	For sites bounded only by hard topography (such as road sections, rail sections, water sections and some landscaping elements), no minimum size applies. For sites not entirely bounded by hard topography, the minimum area is 1000 m^2	Inclusions: • grass seed nurseries and sod farms • narrow strips of grassy areas bounded by topography features (if located between a watercourse and a forest edge, then mapped if wider than 2 m; if permanent and not along a watercourse, then mapped if wider than 6 m Exclusions: • (ornamental) lawns within yards (these are set to 'other')
Germany	Digital Basic-Landscape model (Basic-DLM)	Not set	Vegetation/ Landwirtschaft/ Grünland	Area of grass that is mown or grazed (under the agricultural use)	1 ha	Not provided
			Vegetation/Unland/ Vegetationslose Fläche/Naturnahe Fläche	Area not used for the cultivation of crops, which is overgrown with grass, wild herbs and other plants	1 ha	Not provided
			Besondere Vegetationsmerkmale / Vegetationsmerkmal / Gras	Area covered with slender, herbaceous monocotyledonous flowering plants	1 ha	Not provided

Clipping, querying and converting data

ALMaSS landscape model generation requires high-accuracy and high-resolution input data, the processing of which for large areas (e.g. on a regional or national scale) would be computationally very demanding and time-consuming, possibly requiring high-performance computers. Therefore, the mapping algorithm is designed to process data at the study-area level, which for ALMaSS typically means a 10 km × 10 km processing window.

Prior to further processing, vector geometries are validated to ensure topological correctness. Invalid geometries (e.g. self-intersections, sliver polygons, duplicate vertices, or gaps) are repaired using standard geometry validation and cleaning procedures. This step ensures robust rasterisation and prevents artefacts during layer merging. All input datasets are then clipped to the spatial extent of the study area. To minimise boundary effects during spatial operations, an additional buffer (typically 1 km) is added around the study-area extent. All subsequent spatial analyses are performed within this buffered processing window.

A base raster layer, referred to as 'land', is generated for the full buffered extent. This raster defines the target CRS, spatial resolution (1 m^2), spatial extent, and grid alignment of the final landscape map. The land raster serves as the snap raster for all subsequent rasterisation procedures, ensuring that all derived raster layers share identical cell size, grid alignment, spatial extent, and CRS. This guarantees spatial consistency during layer merging.

The next step is to process the input data according to the information provided in the landscape data description file. This step involves generating individual information layers (using query expressions where necessary) and then converting the vector layers into raster layers. Polygon features are directly rasterised using the predefined 1 m^2 grid and snap raster settings. Point and line features (i.e. zero- and one-dimensional geometries such as roads, streams, or individual trees) require preprocessing to ensure appropriate nominal spatial representation in raster format. First, Euclidean distance raster surfaces are calculated for these features. A predefined cut-off value (specified in the landscape data description file) is then applied to add a buffer to the features, effectively transforming them into two-dimensional raster representations. This procedure preserves the continuity of linear elements and ensures that narrow features are not fragmented or omitted during conversion from vector to raster.

If input datasets include raster data with spatial resolutions different from 1 m^2, they are resampled to the target 1 m^2 grid using a nearest-neighbour assignment method. This approach preserves categorical data integrity and avoids the creation of artificial intermediate classes.

During conversion to raster, all pixels representing objects from a given information layer are assigned a value according to the three-digit code provided in the landscape data description file, and all background pixels are assigned a value of 1. The information layer for the agricultural fields is an exception, as pixels representing each field are assigned separate values (codes) starting at 1000.

Combining layers into thematic maps and stacking them into a 'raw' landscape map

Individual raster layers within each theme are then combined into thematic layers representing natural areas (Natural theme), cultivable areas (Cultivable theme, excluding agricultural fields), built-up areas (Built-up theme, excluding individual buildings), cultural features (Cultural theme), water features (Water theme), and transport infrastructure (Transport theme). Agricultural fields and buildings are treated as separate layers. The layers within each theme are overlaid based on their spatial priority, as defined by the three-digit code assigned in the previous step according to the landscape data description file. Overlaying is achieved by applying a 'maximum' cell statistic tool, which for each pixel returns the highest spatial priority value from the input raster layers. This procedure enables objects with a higher priority (those assigned a higher three-digit code value) to overlay objects with a lower priority (those assigned a lower three-digit code value) (Fig. 2).

The set of resulting single-theme rasters is then stacked in sequence so that the final raw landscape map displays the ecologically meaningful layers at the top (Fig. 3). The ordering of the themes depends on data characteristics and data quality (so that more reliable data are prioritised); therefore, this part of

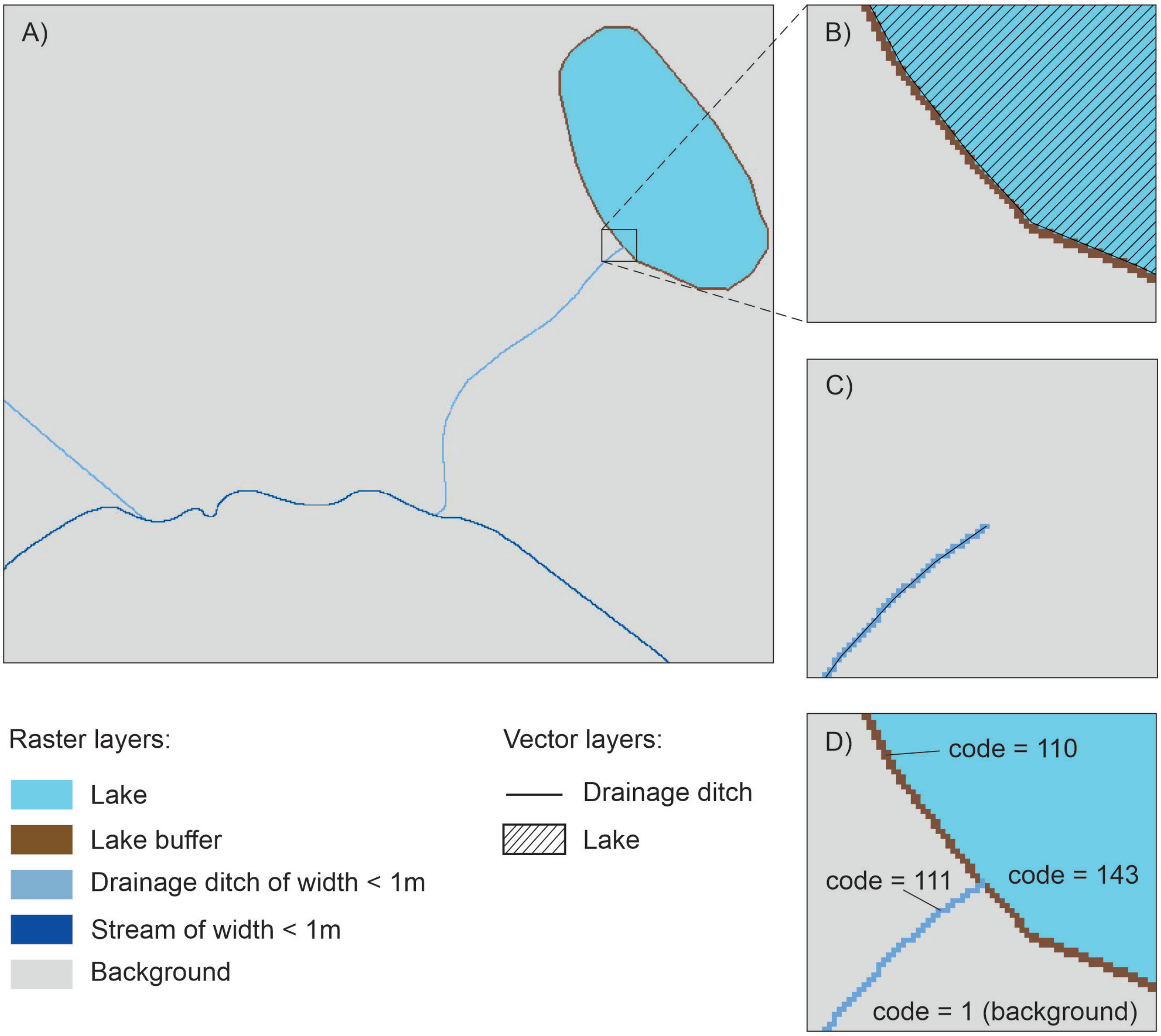

Figure 2. An example of combining layers within the Water theme. **A.** A section of the area showing a raster model of two drainage ditches (width < 1 m) draining into a stream (width < 1 m), with one of the ditches connecting a small lake with the stream. (**B**), (**C**), and (**D**) show a zoom-in on the area where the drainage ditch connects to the lake, with vector layers in (**B**) and (**C**) shown for reference. Within each theme, individual layers are processed separately. Here, (**B**) shows a raster layer of lakes created from a polygon vector layer (shown with a dashed signature). Each lake was surrounded by a 1 m buffer of riparian vegetation as part of the data processing, as detailed riparian vegetation data were not available for this area. (**C**) shows the drainage ditch raster layer created from a line vector layer (black line). This particular drainage ditch object is < 1 m wide, so after conversion to raster, it is represented by a collection of individual pixels (at 1 m resolution) forming a line. (**D**) shows how individual raster layers from (**B**) and (**C**) are overlaid within the water theme. As objects from the lake layer have the highest three-digit code (code = 143), the lake overlays the drainage ditch (code = 111) at the connection point. Because the drainage ditch layer objects have a higher three-digit code (code = 111) than the lake buffer layer objects (code = 110), the drainage ditch overlays the riparian vegetation at the connection point.

the mapping algorithm needs to be adapted each time a new region or country is processed. The resulting raw landscape map may still contain some unclassified areas ('background'), especially if the input data come from different sources and are of varying quality.

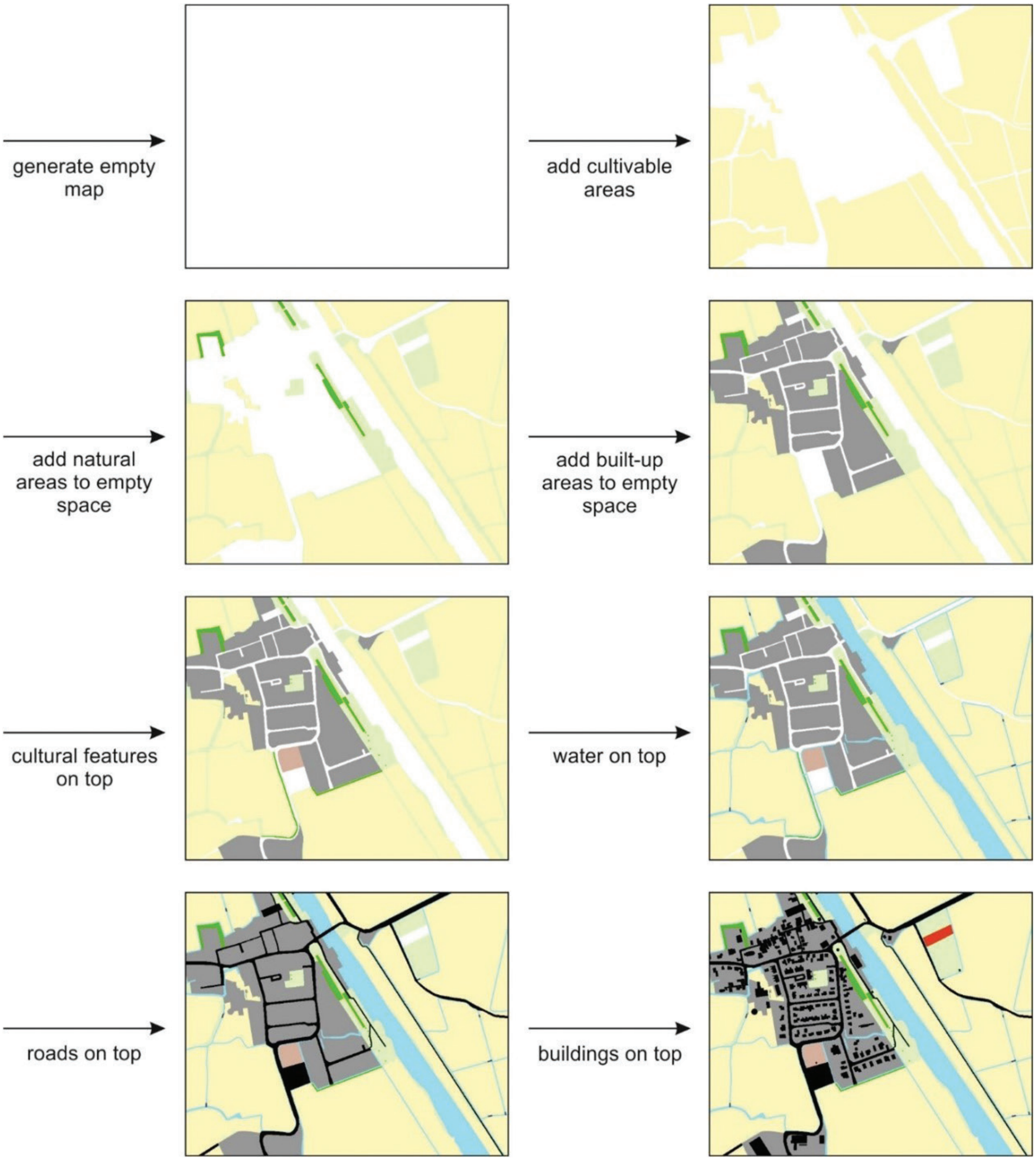

Figure 3. An example of the stacking process. The stacking process starts with cultivated areas (fields in both rotational and permanent cropping). The Natural theme parts (such as forest patches or heaths) and then the Built-up theme parts (such as residential or commercial areas) are added to the empty space not filled with cultivated areas, indicating that information on cultivated areas is treated as more important. Next, the Cultural theme parts, such as individual trees or hedges, are added on top. The spaces occupied by such features may occur within fields or built-up areas. Next, the Water theme parts and then the Transport theme parts are added on top, as features from these themes can cross agricultural areas, built-up areas, and semi-natural habitats. The Transport theme overlays the Water theme, indicating the possibility of bridges crossing watercourses. Finally, all buildings are added on top. The result of the stacking process is a raw landscape map, which may still contain some unclassified areas ('background'), shown here in red.

Multi-stage process of removing artefacts and data gaps in the "raw" landscape map

In addition to unclassified ('background') areas, the raw landscape map may also contain artefacts related to the spatial alignment between input geodata features (i.e. gap or overlap slivers between those features) (Fig. 4). The extent to which these issues are present will again depend on the quality and consistency of the input data. Information layers from a single topographic database should, by design, be consistent with each other. However, their combination with LPIS agricultural field data or other thematic data from different sources may introduce various artefacts related to input data inconsistencies rather than reality. We have developed a multi-stage procedure to resolve instances of unclassified areas (hereafter referred to as 'background') and input data inconsistency artefacts. Table 3 presents the main stages of this process in the order in which they are usually applied, indicating whether they relate to the removal of data gaps or error artefacts (slivers). The order of the stages presented has been chosen to provide the best output (based on our experience in generating ALMaSS landscape models for different European countries and regions) but can be rear-

Figure 4. An example of data gaps (unclassified 'background' areas) and inconsistencies related to the spatial alignment of features. The number in brackets indicates the type of fix applied, as detailed in Table 4.

Table 3. Description of the main processing steps of the raw landscape map. 'Type of fix' indicates whether the processing stage relates to the removal of 'background' pixels (gaps) or the removal of 'sliver' areas (artefacts of errors). 'Target areas (objects)' describe the type of areas or objects targeted, and 'Conditions' list any geometry, land cover, and neighbourhood rules applied in a particular processing stage. 'Workflow' provides details of the implementation, and 'Reclassification target' defines the target landscape element type and the corresponding ALMaSS TOLE code ('ALMaSS_TOLE'). Finally, 'Processing order' indicates the order in which the processing steps are usually applied.

Type of fix	Target areas (objects)	Conditions	Workflow	Reclassification target	ALMaSS_ TOLE	Processing order
Gaps removal	Large 'background' polygons of regular geometry (i.e. rectangular-like shape) with predominantly arable land cover. These represent agricultural fields not present in current LPIS data (i.e. 'missing' fields) as not all fields may be registered for the subsidies. As a data source, LPIS data from previous years and/or topographical data could be used. However, if available, LPIS data are prioritised.	Geometry rules *A* > minimum threshold defined based on field size structure in the processed area (as field size structure could differ substantially between countries or even between regions within one country) $\frac{A}{P} > 2.9$ Land cover rule Arable land > 75%	Applying a minimum area threshold and 'regular' geometry rule to all 'background' areas. For those fulfilling conditions, calculating the share of arable land according to land cover data is possible from the topographical database, but other, less detailed, data could also be used.	Agricultural fields	20	1
Gaps removal	Elongated 'background' areas.	Geometry rules *A* divided by the area of the minimum bounding circle < 0.2 $\frac{A}{P} < 2$ Neighbourhood rule Neighbouring an agricultural field	Applying 'elongated' geometry rules to all 'background' areas. For those fulfilling conditions, checking the type of neighbouring areas.	Field margins	160	2
Sliver removal	Small part of multi-divided fields.	Geometry rule *A* < minimum threshold defined based on field size structure in the processed area	Applying the geometry rule to all agricultural fields consisting of more than one part.	Agricultural fields/ Wasteland	20/209	3
Gaps removal	Remained 'background' areas next to water areas	Neighbourhood rule Neighbouring a watercourse or lake	Checking the type of neighbouring areas	Riparian vegetation	98	4
Gaps removal	Remained 'background' areas – small artefacts	Geometry rule $A < 10\ \mathrm{m}^2$	Applying a minimum area threshold to all 'background' areas.	Non, objects are nibbled out	-	5
Gaps removal	Remained 'background' areas next to built-up areas	Neighbourhood rule Neighbouring built-up areas	Checking the type of neighbouring areas	Yards/ Wasteland	11 /209	6
Silver removal	Elongated objects next to agricultural fields (field margins)	Geometry rule Polsby-Popper test of shape compactness $\frac{4\pi A}{P^2} > 0.12$	Checking the 'elongated' geometry rule for all field margins.	Field margins	160	7
Silver removal	Small artefacts	Geometry rule $A < 10\ \mathrm{m}^2$	Applying geometry rule to all objects besides those from water and transportation themes, individual trees, tree lines and hedges, and individual buildings	Non, objects are nibbled out	-	8
Silver removal	Elongated parts of permanent crops	Geometry rules $\frac{A}{P} < 1.1$	Applying the geometry rule to permanent crops	Non, objects are nibbled out	-	9

A – area, P – perimeter.

ranged as deemed appropriate. In addition, the conditions applied in some stages (e.g. geometry rules) often need to be adapted to geographic context-specific landscape and agricultural structure characteristics. For example, a minimum area threshold applied to identify agricultural fields not present in the LPIS data (i.e. 'missing' fields) may vary considerably depending on the cropping system and field size structure. Furthermore, in some specific cases, additional processing steps may be necessary due to the thematic resolution and quality of the input data (see 'Operational implementation in Polish agricultural landscapes' section). For example, subclassification of grassland may be necessary if the input data does not distinguish between natural and cultivated grassland (Table 2).

Once this processing step is complete, a landscape map without any unclassified pixels or erroneous landscape feature artefacts has been produced, with each pixel of the resulting map having specific associated land-cover/land-use information.

Reclassification, regionalisation and export of output files for further processing

The landscape map produced from the preceding processing stage typically exhibits a higher thematic resolution than that specified by the TOLE classification in ALMaSS (Suppl. material 1: table S1). Consequently, a reclassification step is required to reduce the thematic resolution of the landscape map, ensuring alignment with the TOLE typology as defined in the data description text file. Executing this processing step at this stage ensures that the input data processing remains independent of the TOLE typology within ALMaSS. Consequently, if modifications are made to this typology or its structure (e.g. transitioning to a hierarchical classification system), it is unnecessary to repeat the entire landscape map preparation process. Instead, only the reclassification step and the subsequent steps need to be revisited.

The final raster landscape map for ALMaSS must delineate polygons of homogeneous land-use/land-cover types (according to the TOLE) by assigning them unique reference numbers ('PolyRefNum'). This is achieved through the application of a regionalisation function, which records the identity of the connected region for each cell in the output map (Fig. 5). In practice, the algorithm scans the raster grid and records the identity of each connected region of cells representing the same values. Connectivity between cells is evaluated using an eight-neighbourhood rule, meaning that cells sharing either an edge or a corner are considered connected. In this context, the delineation of regions is determined by the allocation of cells to a specific TOLE. Eight neighbouring cells are used when evaluating connectivity between cells that define a region. The resulting raster representation of polygons (homogeneous land-use/land-cover types) is exported in ASCII format for further processing (ASCII_{study_area}.txt).

In addition, two other mandatory and up to five optional outputs are used to produce the final ALMaSS landscape external inputs (see section 'Producing ALMaSS landscape external inputs') (Fig. 5). One of the two mandatory outputs is the attribute table (in .csv format, ATTR_{study_area}.csv) of the polygon raster map, which must include a 'link' field storing the TOLE value of the zone to which the cells of each region in the output belong. The other mandatory output is the attribute table (in .csv format, FIELDS_{study_area}.csv) of the vector layer with agricultural fields, which is generated at the end of the multi-stage process of removing inconsistencies and data gaps in the "raw" landscape map. This attribute table should include a "link" field storing the field ID number from the agricultural register. The field ID number from the agricultural register can

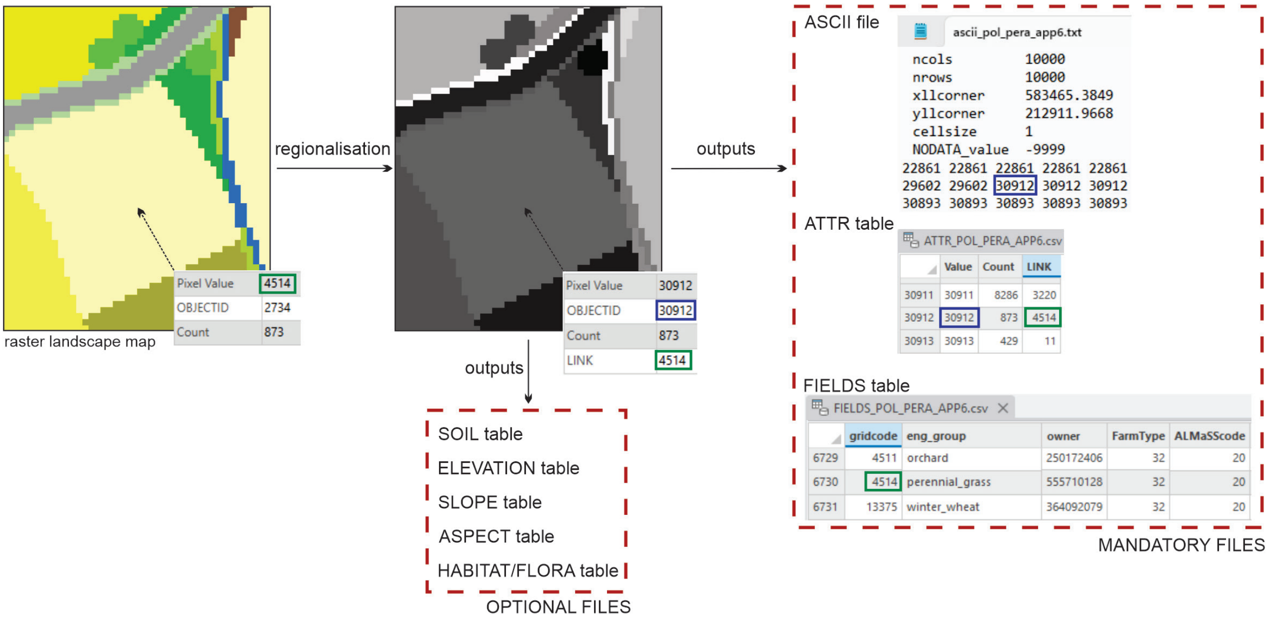

Figure 5. The final raster landscape map, the map after regionalisation, and all possible outputs (mandatory and optional) used to produce the final ALMaSS landscape inputs. The map with homogeneous land-use/land-cover regions is exported to ASCII format. The region ID (marked blue) and field ID (marked green) numbers enable the connection of each landscape parcel to all outputs.

be directly linked to other field-related data, such as the farm unit ID. The five optional outputs include tables (in .csv format) linking each polygon (region) ID ('PolyRefNum') of the raster landscape map with information on: (1) soil type (SOIL_{study_area}.csv), (2) elevation (ELEV_{study_area}.csv), (3) slope (SLOPE_{study_area}.csv), (4) aspect (ASPECT_{study_area}.csv), and (5) habitat/flora type (HAB_{study_area}.csv). Using sourced raster maps of these variables, a zonal statistics function can be applied to calculate the majority (in the case of soil type and habitat/flora type) or mean (in the case of elevation, slope, and aspect) values of each of these variables within each polygon of the raster landscape map and to store these results as .csv tables.

Farming structure and cropping system

In ALMaSS, agricultural management (including the allocation of crops to individual fields according to rotation plans) is applied at the farm level, necessitating the grouping of individual agricultural parcels (fields) into farm units of various farm types. Typically, farm type assignment is not included in the agricultural register and must be generated separately for each country or region based on available crop and farm animal data. Unless specified otherwise, the standard farm types considered in ALMaSS include farms specialised in field crops (arable), horticulture (vegetables and flowers), potato, beet, permanent crops (orchards – fruits, wine, and olives), grazing livestock, granivores, and mixed stock, with distinctions between conventional and organic practices. Additionally, hobby farms, which are smallholdings or small farms maintained without the expectation of being a primary source of income, are considered separately. This typology generally aligns with the types of farming recognised in the Farm Sustainability Data Network (FSDN), formerly known as the Farm Accountancy Data Network (FADN), the EU's farm-accounting data collection system. We, however, also distinguish between farms specialising in field crops and those specialising in potato and beet production due to the particularly large amount of agrotechnical treatments, including fertilisers and pesticides, applied to these crops. Alternative sets of ALMaSS farm types may be applied for different regions or countries, depending on the types of farming present. For example, for Denmark the farm type set does not include permanent crops, and grazing livestock is considered either a pig or cattle farm type.

The initial step in ALMaSS farm type classification involves compiling a list of individual farms within a country or region, including information on their field acreage, total crop acreage by specific categories, and the number of animals of different types owned, expressed in terms of livestock units (LSUs). Specific crop categories include cereals (common wheat and spelt, durum wheat, rye, barley, oats, summer cereal mixes, grain maize, and other cereals for grain production); other field crops (dry pulses, potatoes, sugar beet, herbaceous oilseed and fibre crops including seed, hops, tobacco, and other industrial crops); energy crops; vegetables and flowers (fresh vegetables, melons, strawberries, flowers, and ornamental plants); vineyards; total permanent crops (fruit and berry orchards, citrus fruit orchards, olive groves, nurseries, and other permanent crops) broken down into orchards, olive groves, and other permanent crops; forage crops (fodder roots and brassicas, other fodder plants, temporary grass, meadows and permanent pastures, and rough grazing); agricultural fallows; and set-aside land. The number of animals of different types is recalculated to LSUs according to Eurostat guidelines (https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Livestock_unit_(LSU)). Livestock types include, where possible, dairy cows, other cattle, sheep and goats, pigs, and poultry. The crop categories and livestock types correspond to the statistics used by FADN. Consequently, crop and animal FADN statistics for different types of farming determined for the country or region under investigation can be compared with similar crop and animal statistics calculated for farm units based on the agricultural register. This comparison allows the determination of rules for classifying farm units into specific farm types (Table 4).

Following farm type classification, each farm unit in the agricultural register is assigned a farm type, and summary statistics on the proportions of crops cultivated by farms of different types are calculated. Based on these proportions, crop rotation schemes can be generated individually for each farm type. Typically, only crops with more than a 1% share of the area of a given farm type are considered. It is therefore assumed that the rotation could be represented by 100 crops (1 crop for each 1%). The sequence of crops adheres to typical agronomic practices, and the built-in ALMaSS farm code addresses issues such as late harvests, which can lead to impossible sowing conditions. This prevents unrealistic agronomic crop sequences. The result is a pattern of changing crops on a field that precisely matches the overall crop distribution pattern for that farm type over 100 seasons. For instance, if a specific crop, such as maize for silage, appears 13 times out of 100 in the rotation, it will, on average, occupy 13% of all fields covered by that rotation at any given time. Additionally, predefined crop rotation schemes can be used.

Table 4. Classification rules for farm units in Poland. Defined based on crop and animal FSDN statistics for types of farming in Poland in 2023 (Suppl. material 1: table S2).

Farm type	Minimum acreage [ha]	Maximum acreage [ha]	Rules related to crops	Rules related to animals	Comment
Permanent crops	-	-	≥ 50% permanent crops	-	Mainly farms specialised in fruit production.
Horticulture	6	-	≥ 10% vegetables and flowers	-	
Beet	16	-	≥ 25% beet	-	
Potato	10	-	≥ 20% potato	-	
Grazing livestock	6	-	≥ 40% grazing and fodder plants OR	≥ 20 LSU total > 70% cattle, sheep and goats	Crop OR animal rule.
Granivores	6	-	-	≥ 20 LSU total ≥ 80% pigs*	
Mixed stock	6	-	20–40% grazing and fodder plants OR	≥ 20 LSU total and not classified as grazing livestock and granivores	Crop OR animal rule.
Hobby	-	6	-	< 20 LSU	
Other	-	6	≥ 10% vegetables and flowers	-	-
Arable	6	-	-	-	All other farms are not classified yet; i.e. those with large acreage specialised in field crops but not beet and potato, with few or no animals and little or no grazing.

*A granivore farm type usually indicates a high proportion of pigs and poultry together. However, animal data available for Poland from ARMA (Agriculture Restructuring and Modernisation Agency) contain data only for pigs and not for poultry. The share of pigs in the total livestock unit calculated without poultry is about 99%. Therefore, the threshold for the granivore farm type differs from the pigs' FADN statistics presented in Suppl. material 1: table S2.

Producing ALMaSS landscape external inputs

The production of ALMaSS landscape inputs begins with the preprocessing of input data related to agricultural fields. The attribute table of the vector layer with agricultural fields (FIELDS_{study_area}.csv) is merged with the agricultural register based on the ID numbers of agricultural fields. This integration ensures that each field in the analysed landscape window is assigned information on ALMaSS crop type, farm unit ID, and farm type. There may be several agricultural fields missing this information. These fields, generated during the multi-stage process of removing inconsistencies and data gaps in the "raw" landscape map, represent agricultural fields not present in the LPIS data (i.e. 'missing' fields), as not all fields may be registered for subsidies. A special procedure is required for these fields, as agricultural fields without assignment to a farm unit are not permitted in the final ALMaSS landscape input files. Consequently, these fields are randomly divided into groups forming farm units. The size of each artificially generated farm unit is derived from a uniform distribution with minimum and maximum values based on the size statistics of farm units present in the analysed landscape window. All new farm units are assigned the most common farm type in the analysed landscape window. Next, the crop type for each agricultural field is verified. If an agricultural field is managed as a field in rotation, it receives an ALMaSS TOLE number '20'. Conversely, fields with permanent crops are assigned a specific ALMaSS TOLE number corresponding to the given permanent crop (Suppl. material 1: table S1). For example, conventionally managed orchards receive

TOLE number '56', while conventionally managed permanent pastures receive TOLE number '35'. All agricultural fields generated during the multi-stage process of removing inconsistencies and data gaps in the "raw" landscape map are treated as fields in rotation.

The subsequent step involves processing the attribute table of the polygon landscape raster map (ATTR_{study_area}.csv). In this step, ATTR_{study_area}.csv is merged with the attribute table of the vector layer with agricultural fields (FIELDS_{study_area}.csv) and tables providing information on additional attributes of landscape polygons (soil type, elevation, slope, aspect, and habitat/flora type). Farm unit ID numbers and landscape polygon attributes are assigned to specific attributes of the polygon reference (PolyRef_{study_area}.txt) file (Table 1). If any attribute tables are missing, the corresponding attributes in the polygon reference file are assigned a value of −1 for all landscape polygons. If a given landscape polygon is not agriculturally managed, it receives a value of −1 for the 'FarmRef' attribute. Based on the polygon reference file, a list of all farm units and their farm types is generated and saved as the FarmRef_{study_area}.txt file. In this file, farm unit ID numbers are renumbered starting from 0, and farm types are recoded to numbers starting from 32, as required by ALMaSS. Finally, the polygon reference file is cleaned to include only the necessary attributes (Table 1), and the landscape raster binary file {study_area}.lsb is generated from the polygon landscape raster map ASCII_{study_area}.txt.

Workflow automation

The entire workflow is implemented in three Python 3.x scripts, utilising the arcpy library to access ArcGIS geoprocessing functions and the Pandas library for tabular data handling. The scripts and associated running directories are publicly available at the GitLab repository (https://gitlab.com/ALMaSS/almass_methodology/-/tree/main/landscape_generation), with the overall processing flow summarised in Fig. 6. The protocol supports batch processing across multiple landscape windows, enabling scaling to regional or national extents on mainframe computers or clusters.

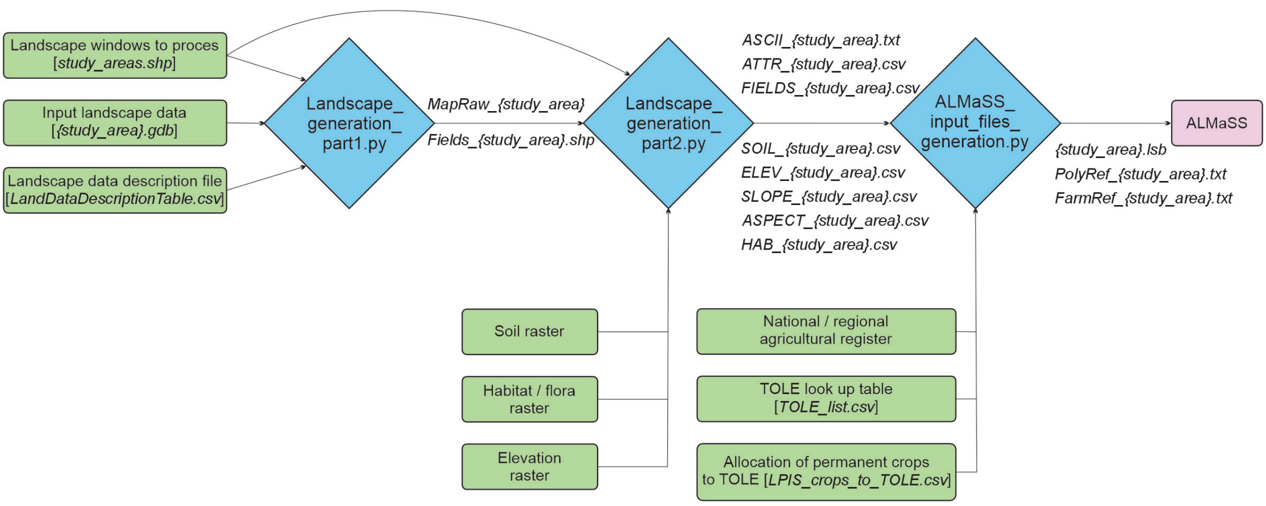

Figure 6. A workflow diagram for the generation of ALMaSS landscape input files.

Operational implementation in Polish agricultural landscapes

The operational implementation of the presented methodology depends on the availability, structure, and thematic scope of spatial input data. Consequently, minor adaptations are typically required when generating ALMaSS landscape models for different countries or regions. Here, we present an example implementation for Polish agricultural landscapes.

The source code and running directory used for this implementation are publicly available in the ALMaSS methodology repository (folder landscape_generation): https://gitlab.com/ALMaSS/almass_methodology/-/tree/main/landscape_generation. The repository contains scripts adapted to Polish conditions as well as the files necessary to execute the workflow, including example geospatial input data for a small 2 × 2 km study area in Poland (Fig. 7). Running the workflow generates the complete set of external input files required for ALMaSS simulations, including the binary landscape file (.lsb), the associated PolyRef file describing landscape elements, and the FarmRef file linking farms to farm types.

The generation of ALMaSS landscapes in Poland relies on several national geospatial datasets describing land cover, land use, infrastructure, and agricultural management. A summary of all datasets used in the workflow, including their spatial resolution, currency, coordinate reference system, and data sources, is provided in Table 5. Within the workflow, datasets are organised through a configuration table (PL_LandDataDescriptionTable.csv, located in the Ex_workspace subfolder), which lists all geospatial layers used in the landscape generation process along with their thematic definitions and additional attributes required for processing. Table 6 presents an excerpt of this structure for water-related layers.

While most processing steps follow the general landscape generation procedure described earlier, some additional adjustments were necessary to address characteristics of the Polish input datasets. In the Polish topographic database, grasslands are included in a general land-cover layer without distinction between managed and unmanaged types. To avoid overestimation of natural grasslands, supplementary information on managed grasslands from the LPIS dataset was combined with geometry- and neighbourhood-based rules to reclassify grassland polygons into more specific categories, such as pastures, low-yield pastures, yards, amenity grass, field margins, and road verges. Another issue was the creation of small "wasteland" polygons during earlier cleaning stages when areas could not be assigned to any defined land-use class. Because such objects lack ecological characteristics relevant for ALMaSS simulations, additional geometry- and neighbourhood-based rules were implemented to reassign these areas to appropriate land-use categories where possible (Table 7).

Fig. 7 presents the process of integrating and stacking the input datasets to produce the final ALMaSS landscape map for the exemplary 2 × 2 km agricultural area in Poland.

Outcomes

The primary objective of the landscape model generation process described here is to provide the necessary inputs for simulating landscapes in ALMaSS, thereby supporting other simulation components (see Topping and Duan

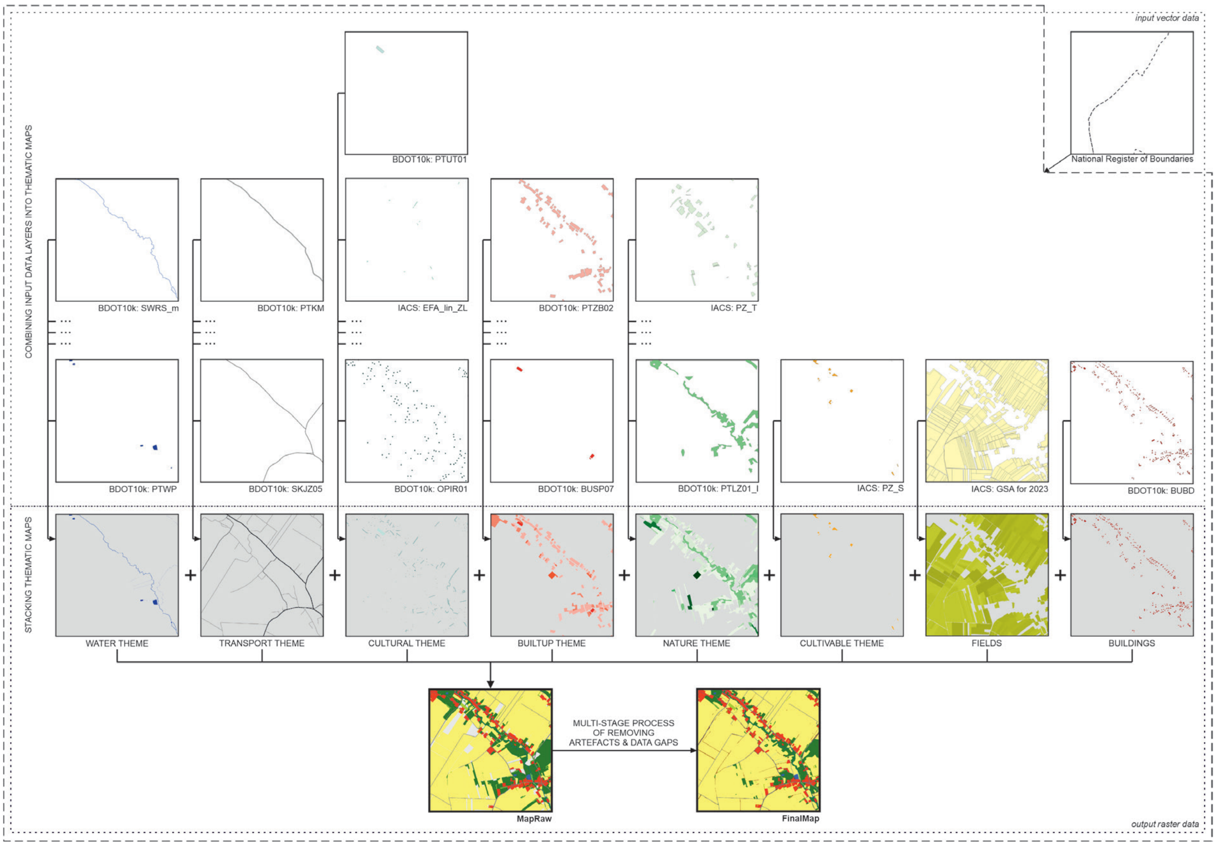

Figure 7. An example ALMaSS landscape generated for a 2 × 2 km agricultural area in Poland after integration and processing of the input datasets. Descriptions (BDOT10k and IACS) refer to the input datasets listed in Table 5, while the layer names (e.g. PZ_T and SKJZ05) originate from the landscape data description table (PL_LandDataDescriptionTable.csv). MapRaw – "raw" landscape raster map (output of the stacking of thematic maps); FinalMap – output of the multi-stage process of removing artefacts and data gaps in the "raw" landscape map. MapRaw (MapRaw.tif) and FinalMap (map_final.tif) are available in the ALMaSS methodology repository (in the subfolder Ex_results of the landscape_generation folder).

Table 5. Input datasets for the generation of the ALMaSS landscape model in Poland.

Dataset	Layers	Data type	Coordinate Reference System	Spatial resolution	Date currency (year)	Source
Topographic Objects Database (BDOT10k)	Network of watercourses (SW)	vector	EPSG: 2180	-	2024–2025	GUGiK, Geoportal (http://www.geoportal.gov.pl)
	Network of roads and railways (SK)					
	Land cover (PT)					
	Protected areas (TC)					
	Buildings, structures and equipment (BU)					
	Land development complexes (KU)					
	Other objects (OI)					
IACS data	LPIS data	vector	EPSG: 2180	-	2024	ARMA Geoportal (http://www.geoportal.arimr.gov.pl; available only after verification and approval by ARMA)
	GSA (Geospatial Application) data				2023	
	Land cover (PZ)				2024	
	EFA (Ecological Focus Areas) data					
European Soil Database	TXSRFDO	raster	EPSG: 3035	1 km	2006	European Soil Data Centre (ESDAC; https://esdac.jrc.ec.europa.eu/content/european-soil-database-v2-raster-library-1kmx1km; after registration)
National Register of Boundaries		vector	EPSG: 4258	-	2024	GUGiK, Geoportal (http://www.geoportal.gov.pl)

Table 6. Example, for the theme Water, of a landscape data description file (LandDataDescriptionTable.csv) from the ALMaSS landscape generation for Poland. Information_layer – the name of the information layer containing features of certain types and characteristics used in the subsequent processing steps. Information_layer is a result of the processing of the Source_layer contained in the input database by applying, if necessary, a query expression (provided in the Query column). The information layers are grouped into themes (Theme column); here only the theme 'Water' is shown. Other included data properties are the data type (Data_type: vector or raster), the spatial resolution (Spatial_resolution; only for raster data), the geometry type (Geometry_type: point, line, or polygon; only for vector data), the applied buffer (Buffer; only for point and line vector data; used to convert zero- or one-dimensional objects into two-dimensional ones), the three-digit code (Code column) defining the order of the layers in terms of their spatial priority (necessary for the combination of layers within each theme), and the corresponding ALMaSS TOLE code (ALMaSS_TOLE). The example below illustrates how a single source layer can be utilised to generate multiple information layers. For example, the layer 'SWRM_L' is a source for three information layers ('SWRM_ss', 'SWRM_s', and 'SWRM_m'), which distinguish three classes of drainage ditches based on their width using corresponding SQL query expressions. These information layers are further processed by converting them from vector to raster layers and applying a buffer that allows one-dimensional line objects to be converted into two-dimensional ones. The output raster layers are indicated in the Output column. The process of generating information layers from source layers and converting vector to raster layers is described in the section 'Clipping, querying and converting data').

Information_layer	Definition	Theme	Source_layer	Data_type	Spatial_resolution	Geometry_type	Query	Code	Buffer	Output	ALMaSS_TOLE
SWRM_ss	Drainage ditch of width < 1 m	Water	SWRM_L	Vector	−1	Line	"SZEROKOSC" <= 1	111	0.5	SWRM111	222
SWRM_s	Drainage ditch of width between 1 and 2 m	Water	SWRM_L	Vector	−1	Line	"SZEROKOSC" >1) AND "SZEROKOSC" <= 2	112	1.0	SWRM112	222
SWRM_m	Drainage ditch of width > 2 m	Water	SWRM_L	Vector	−1	Line	"SZEROKOSC" > 2	113	2.0	SWRM113	222
SWKN_ss	Channel of width < 1 m	Water	SWKN_L	Vector	−1	Line	"SZEROKOSC" <= 1	121	0.5	SWKN121	223
SWKN_s	Channel of width between 1 and 2 m	Water	SWKN_L	Vector	−1	Line	"SZEROKOSC" >1 AND "SZEROKOSC" <= 2	122	1.0	SWKN122	223
SWKN_m	Channel of width > 2 m	Water	SWKN_L	Vector	−1	Line	"SZEROKOSC" > 2	123	2.0	SWKN123	223
SWRS_ss	Stream of width < 1 m	Water	SWRS_L	Vector	−1	Line	"SZEROKOSC" <= 1	131	0.5	SWRS131	207
SWRS_s	Stream of width between 1 and 2 m	Water	SWRS_L	Vector	−1	Line	"SZEROKOSC" >1 AND "SZEROKOSC" <= 2	132	1.0	SWRS132	207
SWRS_m	Stream of width between 2 and 5 m	Water	SWRS_L	Vector	−1	Line	"SZEROKOSC" >2	133	2.0	SWRS133	207
PTWP02	River of width > 5 m	Water	PTWP_A	Vector	−1	Polygon	RODZAJ = 'woda płynąca'	141	−1	PTWP141	96
PTWP03	Lake	Water	PTWP_A	Vector	−1	Polygon	RODZAJ = 'woda stojąca'	143	−1	PTWP143	90
BUZT02	Fire ponds in the area under roads	Water	BUZT_A	Vector	−1	Polygon	RODZAJ = 'zbiornik'	150	−1	BUZT150	219
OISZ_A	Reedbed	Water	OISZ_A	Vector	−1	Polygon	none	161	−1	OISZ161	98

2024). Leveraging various European and national projects, we have successfully generated ALMaSS landscape models for a wide range of European agricultural landscapes characterised by diverse farming systems with varying degrees of landscape and farming heterogeneity (Fig. 8). This capability enables us to use ALMaSS to simulate species population-level responses to different current and future landscape management practices across various agroecosystems. Consequently, we can better understand species-specific responses and their variability in relation to various spatial dynamic factors, such as the spatial distribution of source and sink habitats or underlying habitat suitability.

Table 7. Description of the additional processing steps of the raw landscape map applied in the Polish case. 'Target areas (objects)' describe the type of areas or objects targeted, and 'Conditions' list any geometry, land cover, and neighbourhood rules applied in a particular processing stage. 'Workflow' provides details of the implementation, and 'Reclassification target' defines the target landscape element type and the corresponding ALMaSS TOLE code ('ALMaSS_TOLE'). Finally, 'Processing order' indicates the order in which the processing steps were applied (after the main processing steps presented in Table 3).

Type of fix	Target areas (objects)	Conditions	Workflow	Reclassification target	ALMaSS_ TOLE	Processing order
Wasteland issue	Wastelands next to buildings	Neighbourhood rule Neighbouring built-up areas Geometry rule $A > 10\ \mathrm{m}^2$, $\frac{A}{P} > 1.5$, $\frac{P^2}{A} \leq 100$ (shape compactness ratio)	Checking the type of neighbouring areas. For those fulfilling conditions, applying geometry rules.	Yards	11	10
	Large, field-like-shaped wastelands	Geometry rule • $A > 1000\ \mathrm{m}^2$, $\frac{P^2}{A} \leq 250$ or $\frac{A}{P} > 2.8$ (in the south Poland case, where the fields are more elongated) • $A > 500\ \mathrm{m}^2$, $\frac{A}{P} > 3.5$	Applying geometry rules to all wastelands.	Agricultural fields (permanent)	26 33	
	Elongated wastelands next to watercourses	Neighbourhood rule Neighboured to watercourses Geometry rule $A > 5\ \mathrm{m}^2$, $\frac{4\pi A}{P^2} \leq 0.3$, $\frac{A}{P} < 2.8$	Checking the type of neighbouring areas. For those fulfilling conditions, apply geometry rules.	Natural grasslands along watercourses	205	
	Elongated wastelands next to agricultural fields	Neighbourhood rule Neighbouring agricultural fields Geometry rule $A > 5\ \mathrm{m}^2$, $\frac{4\pi A}{P^2} \leq 0.29$, $\frac{A}{P} < 2$	Checking the type of neighbouring areas. For those fulfilling conditions, applying geometry rules.	Field margins	160	
	Elongated wastelands next to roads	Neighbourhood rule Neighbouring roads Geometry rule $A > 5\ \mathrm{m}^2$, $\frac{4\pi A}{P^2} \leq 0.3$, $\frac{A}{P} < 1.6$	Checking the type of neighbouring areas. For those fulfilling conditions, applying geometry rules.	Road verges	13	
Grasslands issue	Grasslands next to buildings	Neighbourhood rule Neighbouring built-up areas Geometry rule $A > 10\ \mathrm{m}^2$ and $A < 10\ 000\ \mathrm{m}^2$, $\frac{A}{P} > 1.5$, $\frac{P^2}{A} \leq 100$ (shape compactness ratio)	Checking the type of neighbouring areas. For those fulfilling conditions, applying geometry rules.	Yards	11	11
	Large, field-like-shaped grasslands	Geometry rule • $A > 1000\ \mathrm{m}^2$, $\frac{P^2}{A} \leq 250$ or $\frac{A}{P} > 2.8$ (in the south Poland case, where the fields are more elongated) • $A > 500\ \mathrm{m}^2$, $\frac{A}{P} > 3.5$	Applying geometry rules to all grasslands.	Agricultural fields (permanent)	35 / 26 33	
	Elongated grasslands next to watercourses	Neighbourhood rule Neighbouring watercourses Geometry rule $A > 5\ \mathrm{m}^2$, $\frac{4\pi A}{P^2} \leq 0.35$, $\frac{A}{P} < 2.8$	Checking the type of neighbouring areas. For those fulfilling conditions, applying geometry rules.	Natural grasslands along watercourses	205	
	Elongated grasslands next to agricultural fields	Neighbourhood rule Neighbouring agricultural fields Geometry rule $A > 5\ \mathrm{m}^2$, $\frac{4\pi A}{P^2} \leq 0.29$, $\frac{A}{P} < 2$	Checking the type of neighbouring areas. For those fulfilling conditions, applying geometry rules.	Field margins	160	

Type of fix	Target areas (objects)	Conditions	Workflow	Reclassification target	ALMaSS_TOLE	Processing order
Grasslands issue	Elongated grasslands next to roads	Neighbourhood rule Neighboured to roads Geometry rule $A > 5\ \mathrm{m}^2, \frac{4\pi A}{P^2} \leq 0.3, \frac{A}{P} < 1.6$	Checking the type of neighbouring areas. For those fulfilling conditions, apply the geometry rules.	Road verges	13	11
Yards issue	Elongated yards next to agricultural fields	Geometry rule $\frac{A}{P} < 0.72$ Neighbourhood rule Neighbouring agricultural fields	Applying the geometry rule for yards. For those fulfilling the condition, checking the type of neighbouring areas.	Field margins	160	12

A – area, P – perimeter.

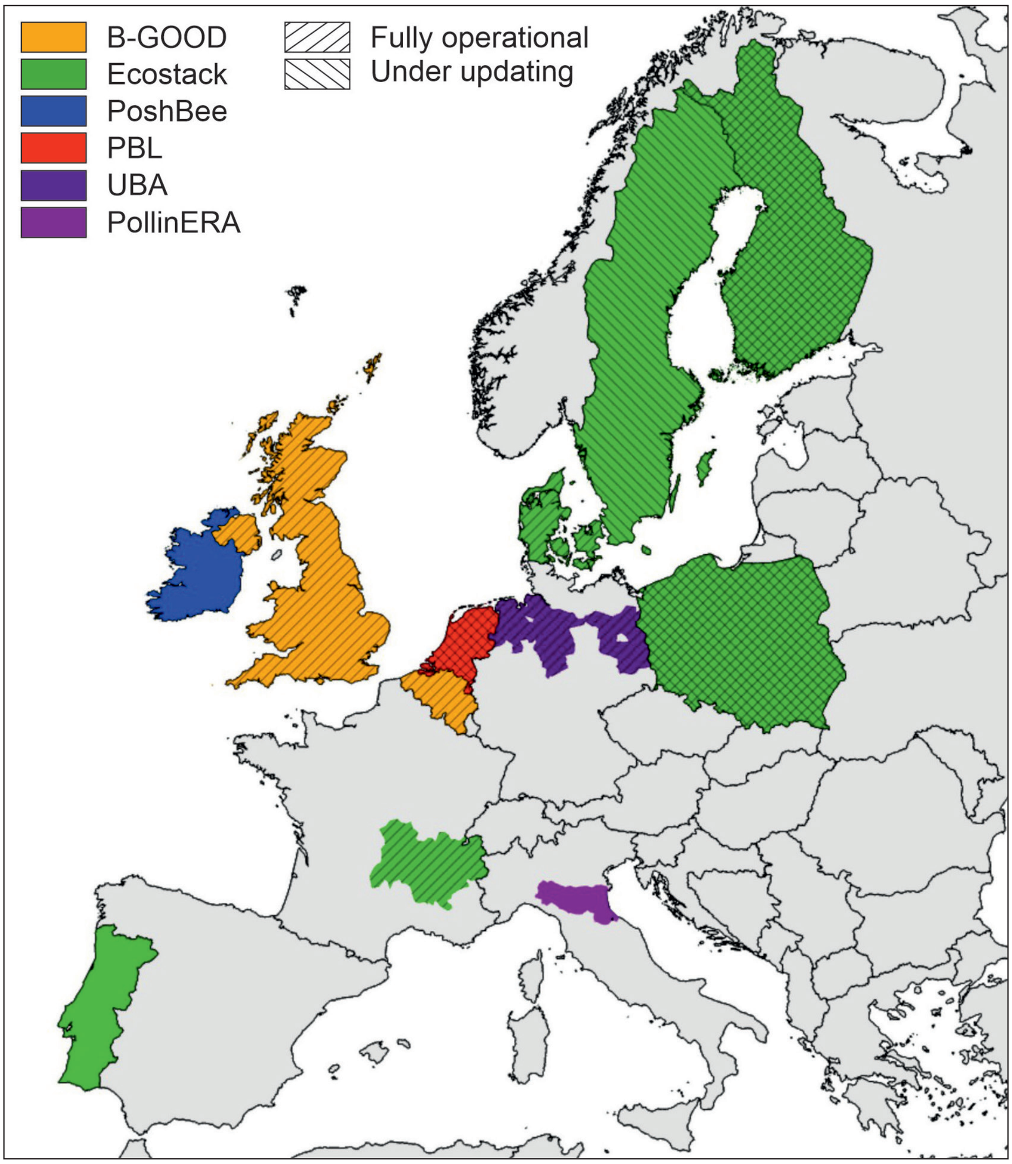

Figure 8. ALMaSS landscapes' coverage in Europe. ALMaSS landscape models were developed or are being developed under various Horizon 2020 projects (B-GOOD, EcoStack, PoshBee), the Horizon Europe project (PollinERA), and national projects (PBL – Netherlands Environmental Assessment Agency; UBA – German Environmental Agency). Fully operational – regions or countries for which a fully operational ALMaSS landscape model exists and can be used in simulation runs. Under updating – regions or countries that are currently being updated under the Horizon Europe PollinERA project.

Beyond ALMaSS applications

Detailed land-use/land-cover maps generated as part of the ALMaSS landscape models can be utilised to describe structural and farmland heterogeneity through a set of landscape- and class-level metrics (Fig. 9). This type of analysis serves multiple purposes, including assessing the impact of landscape and

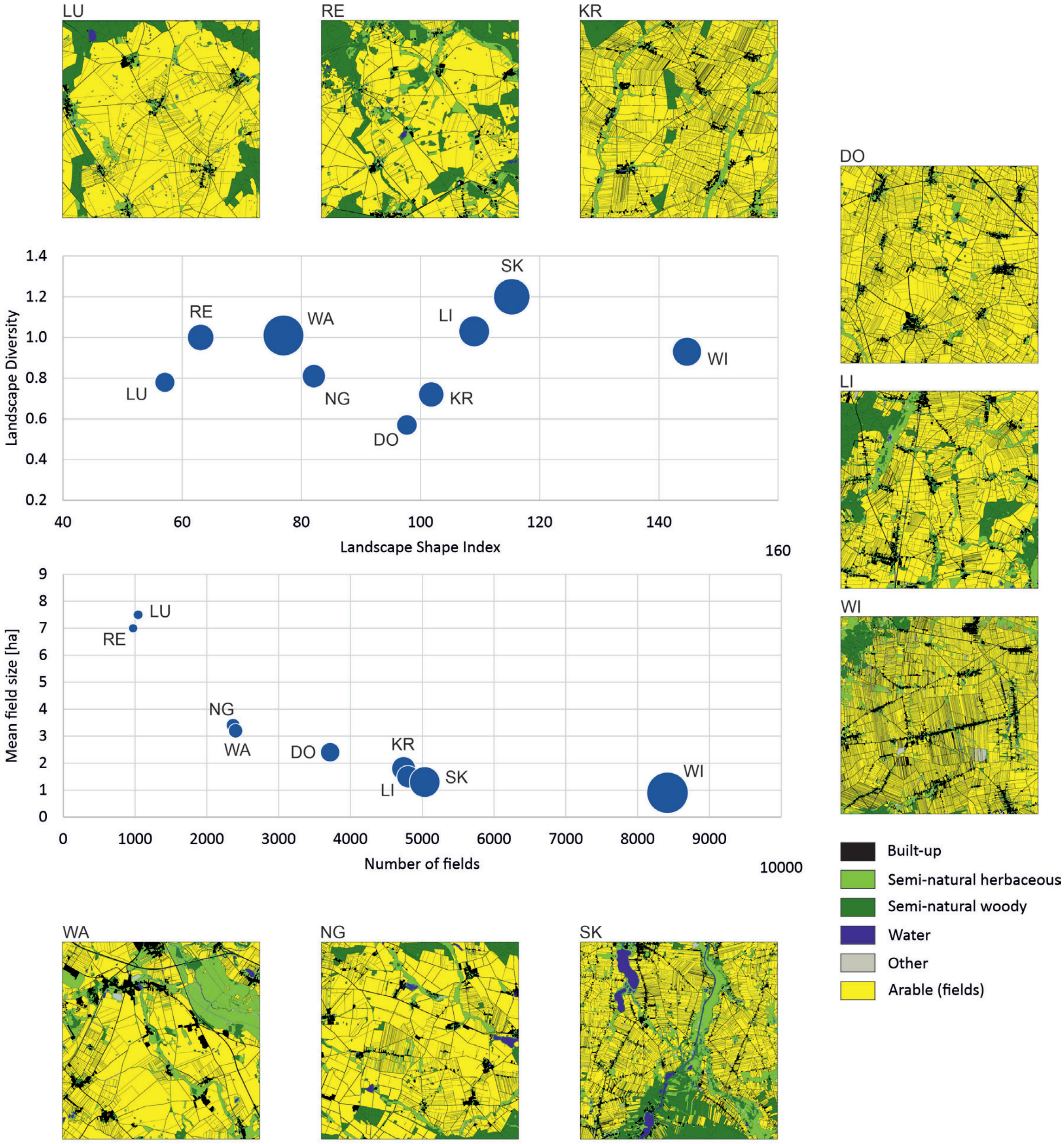

Figure 9. Exemplary landscapes 10 × 10 km from Poland mapped using landscape model generation algorithms for ALMaSS. The presented landscapes exhibit variations in structural and farming heterogeneity. These differences are quantified through various landscape- and class-level metrics, providing insights into the complexity and diversity of the mapped areas. ALMaSS types of landscape elements (TOLEs) were generalised to six main classes: built-up areas (with transport infrastructure); semi-natural herbaceous areas, including meadows, extensively managed grasslands, field margins, and road verges; semi-natural woody areas, including forests, coppices, shrubs, individual trees, tree lines, and hedgerows; water; arable (with individual fields); and other. Each analysed landscape was coded with the first two letters of the main town within or on the border of its area (LU – Lubicz, RE – Redło, KR – Krotoszyn, DO – Domaniów, LI – Lipno, WI – Wieluń, WA – Warnice, NG – Nowy Gołębin, SK – Skulsk). Details about the landscapes are provided in Ziółkowska et al. (2021).

farmland heterogeneity on changes in species population dynamics simulated with ALMaSS. For instance, Ziółkowska et al. (2021, 2022) employed ALMaSS to simulate the population dynamics of the ground beetle *Bembidion lampros* in landscapes with varying degrees of landscape and farmland heterogeneity and representing different farming systems. Both the composition (e.g. coverage of herbaceous and woody semi-natural habitats) and configuration (e.g. landscape and farming diversity, landscape shape index, number of fields, mean field size, and field boundary density) of landscape elements were analysed in the studied landscapes of 10 × 10 km. The authors found that the observed effects on beetle populations were highly context-specific, with beetle populations being larger and distributed over larger areas in more diverse landscapes with a higher proportion of herbaceous semi-natural areas.

Given that ALMaSS simulations are highly CPU-intensive, it is often necessary to make choices regarding simulation scenarios. This decision-making process can be supported by analyses of structural and farmland heterogeneity conducted on a set of potential landscape windows before running ALMaSS simulations. This ensures that simulation scenarios include landscapes with desired properties (e.g. a gradient from small to large fields) or landscapes representing specific types, as demonstrated in Ziółkowska et al. (2024).

Detailed land-use/land-cover maps generated using the methodology presented here can also be employed to select study sites for fieldwork, ensuring that they meet specific conditions related to landscape characteristics. For instance, this methodology was applied in the study by Sowa et al. (2022) to generate detailed land-use/land-cover raster maps for two distinct landscapes, each 12 km × 16 km. One landscape was dominated by large fields, while the other was characterised by small fields and family farming. Based on these maps, field study sites for sampling the ground beetle *Poecilus cupreus* were delineated within each landscape by analysing the types of crops cultivated within a 500 m radius around the midpoints where the beetle traps were located. Similarly, in the study by Bednarska et al. (2021), the selection of study sites and the location of red mason bee (*Osmia bicornis*) nests in the field were based on a detailed analysis of landscape structure and crop characteristics derived from land-use/land-cover maps generated using the presented methodology. Importantly, these detailed land-use/land-cover maps enable the characterisation of landscapes from a species perspective by analysing agricultural landscape properties tailored to the species of interest. For example, Mikołajczyk et al. (2021) evaluated landscape composition and configuration at different spatial scales from the perspective of *O. bicornis*.

Discussion

The simulation of landscapes in ALMaSS necessitates a specific ALMaSS external input data structure (Topping and Duan 2024). There are no predefined requirements concerning the quality, spatial, or thematic resolution of the input data. Even simple geometric maps can be utilised as inputs, provided they meet the structural data requirements, for instance, to address more theoretical questions (Hoye et al. 2012; Topping and Lagisz 2012). Nevertheless, ALMaSS was designed to assess the impacts of changes in landscape structure and management on beneficial organisms within agroecosystems. As such, it is intended to operate with representations of actual agricultural landscape areas. These representations must encapsulate the landscape structure details crucial for the diverse species modelled in ALMaSS, encompassing not only land-cover information but also land use, which determines the management actions applied at the landscape element level. Herein, we present a methodology for describing the land use/land cover of agricultural landscapes and generating detailed landscape representations for use in ALMaSS landscape simulation.

Compared to the workflow described by Topping et al. (2016), the present methodology introduces an improved algorithm for resolving spatial conflicts and integrating multiple thematic layers into a coherent polygon structure. This modification enhances the internal consistency of the resulting landscape map and reduces manual post-processing requirements. The improvement is particularly relevant when integrating heterogeneous national datasets, where overlapping or inconsistently classified polygons are common. By formalising these processing steps, the current workflow moves beyond a purely technical description of data requirements and provides a more robust and transferable landscape generation framework.

The presented methodology is highly data-intensive, requiring detailed agricultural and topographical data. The representation of agricultural areas in the ALMaSS landscape model is unique, with no comparable representation found in other existing spatially explicit individual-based models or systems. Utilising detailed information from the IACS, agricultural areas are subdivided into individual fields, which collectively form farms of various types. By utilising data on crops cultivated in these fields, insights into crop production characteristics are obtained at both the farm and landscape levels. This approach enables the simulation of changes in the spatio-temporal patterns of crops.

Access to IACS data is therefore crucial for developing ALMaSS landscape models for new areas. However, experience indicates that obtaining these data can be challenging. While spatial data delineating field boundaries and crop information are readily available in many European countries (e.g. downloadable services from webpages or available upon request), field ownership information, which is necessary for defining farm units, is typically provided only upon request and often requires additional agreements outlining the terms of use and access with data providers. Furthermore, these data may be managed at different administrative levels, either nationally (as in Poland or Denmark) or regionally (as in Germany or Italy).

An alternative approach involves using satellite-based products to identify crop types. For instance, a 10-metre resolution map of crops for the EU and Ukraine for 2022 was generated using LUCAS Copernicus polygons, Sentinel-1 and Sentinel-2 satellite imagery, land surface temperature, and a digital elevation model (Ghassemi et al. 2024). Similar national products include digital maps of cropping in Great Britain (UK Centre for Ecology & Hydrology Land Cover® Plus: Crops), available annually since 2015. However, this product for Great Britain utilises Sentinel-1 and Sentinel-2 for categorising land parcels with predefined boundaries.

While satellite-based crop classification products can provide unified crop information over extensive areas, they often exhibit relatively low thematic resolution, as evidenced by the 19 crop types identified in Ghassemi et al. (2024). Their accuracy depends on the heterogeneity of agricultural areas, with small and narrow fields being more challenging or even impossible to detect and classify. Additionally, these products do not facilitate the delineation of farm units, which poses a challenge for the ALMaSS landscape model. Artificial delineation can be achieved by applying assignment algorithms based on regional farm statistics, potentially enabling the generation of landscape models that represent typical farm units and farm types.

The requisite detail levels of land use/land cover information for the ALMaSS landscape model strike a balance between functionality (supporting multiple species models) and logistical considerations. One might propose transitioning from national or regional topographic data to European-wide products, such as those offered by the Copernicus Land Monitoring Service, at resolutions of 5 or 10 metres. Although these products provide consistent information across large areas, they are designed to map land cover rather than land use. This distinction is crucial. For example, natural meadows, managed mown grasslands, and amenity grass differ significantly in the resources they provide but are not differentiated within the High Resolution Layer Grassland of the Copernicus programme. Furthermore, these products may lack important details, such as small woody or grassy landscape elements (e.g. field margins, individual trees), which are vital refuge or source habitats in agricultural landscapes. The impact of reduced resolution on ALMaSS model species has not been comprehensively studied. However, simulation scenarios testing mitigation measures involving changes in small grassy elements for various invertebrates (Jepsen et al. 2005; Ziółkowska et al. 2021, 2022) suggest a substantial impact, although species-specific. European-wide products provided by, for example, the Copernicus Land Monitoring Service may serve as supplementary data, especially when national or regional data lack the thematic resolution required for some information layers. For example, the High Resolution Layer Dominant Leaf Type can be used to subdivide forest patches into coniferous, deciduous, and mixed ones.

The extension of the ALMaSS landscape model to other countries has also highlighted differences in the thematic scope of topographic databases across countries, as well as discrepancies in the definitions of land-use and land-cover types within these databases. While this may not pose a significant issue when testing different landscape scenarios within a single country, it requires careful investigation and harmonisation for comparisons between countries.

The ALMaSS landscape model generation methodology presented herein exhibits flexibility by decoupling land type or cover information from supplementary attributes that pertain to specific characteristics of landscape elements (polygons), such as habitats, altitude, soil type, elevation, slope, aspect, and habitat or flora type. These attributes enable model species to classify landscape features as suitable habitats for reproduction, shelter, or food sources. New attributes are under development, with soil moisture planned for addition in the near future. As these supplementary attributes are defined in a separate text file (polygon reference file), their list can be easily expanded as ALMaSS incorporates new model species. Since these attributes are specified at the level of individual landscape elements rather than landscape element types, they introduce additional variability into the landscape model.

With this structure of input data, it is relatively easy to modify the represented landscape structure for simulation scenarios testing the impacts of landscape management and modifications on species population dynamics. For example, one may test the impact of introducing new or expanding existing field margins or flower strips. Such landscape alterations may be handled both outside (using GIS functionalities) and within the ALMaSS modelling framework.

Limitations and future perspectives

One limitation of the presented workflow is its reliance on the ArcPy library within the ESRI ArcGIS environment. ArcPy is not open source and requires a licensed GIS platform, which may affect reproducibility, accessibility, and transparency. While the algorithmic logic is fully documented, execution depends on proprietary software. From a computational perspective, ArcPy provides robust geoprocessing tools, but processing speed depends on landscape complexity and extent, local hardware, and the internal optimisation of ESRI tools. Many geoprocessing steps, including raster operations such as Nibble (which assigns values from nearest neighbours to fill missing or erroneous raster cells) and vector spatial joins, become computationally intensive at high resolution (1 m), representing the main performance bottlenecks. Although some tools support parallel processing, not all benefit equally, and performance can vary considerably across computing environments. Equivalent operations implemented in modern, in-memory Python libraries such as GeoPandas or GDAL/OGR may execute faster.

To address these limitations, efforts are underway to develop an open-source implementation of ALMaSS landscape generation utilising Python packages from the osgeo family, specifically GDAL and OGR. This transition would improve accessibility, transparency, and customisation possibilities and would enable execution on high-performance computing (HPC) systems based on Linux. Such capability is crucial for handling large-scale simulations and complex data processing tasks, thereby enhancing efficiency and scalability.

Another source of uncertainty arises from integrating spatial datasets of differing resolution and coordinate systems. In the workflow, heterogeneous vector and raster datasets are harmonised into a 1 m raster to capture narrow landscape elements such as field margins and hedgerows. Downscaling coarser datasets may introduce artificial boundary sharpness, while reprojection can cause minor geometric distortions along complex boundaries. These effects may influence ecologically important small landscape elements, potentially affecting habitat availability, edge density, and connectivity. Sensitivity analyses could help evaluate how spatial resolution impacts species modelling and whether slightly coarser input data could produce similar outcomes with more efficient processing. A formal sensitivity analysis evaluating the influence of spatial resolution and dataset harmonisation on model outcomes would therefore be a valuable direction for future research.

The landscape simulation in ALMaSS is dynamic in terms of crop rotations, agro-technical treatments, and vegetation growth in response to weather conditions. However, the underlying landscape polygon structure remains static, and changes in land use or land cover are not modelled outside arable fields. A crucial step in future development is incorporating the ability to modify the landscape structure during simulation to account for land-use or land-cover changes. ALMaSS could be integrated with agent-based models of land-use and land-cover change in agricultural landscapes, considering processes such as farm cessation, expansion, and diversification that shape landscape structure (Valbuena et al. 2010; Bakker et al. 2015; Beckers et al. 2018). Potential trajectories of change could be associated with each landscape element or applied at the landscape scale based on a probability matrix of changes for defined land-use or land-cover types.

The long-term objective is to develop a Data Linkage and Integration module facilitating automated generation of landscape inputs for any selected location within the EU, drawing on the best available national and European datasets (Fig. 10). The primary challenge lies in maintaining dynamically updated data flows and developing algorithms for flexible, quality-aware data selection across heterogeneous sources, ensuring seamless connections between input data and the processing tools described here.

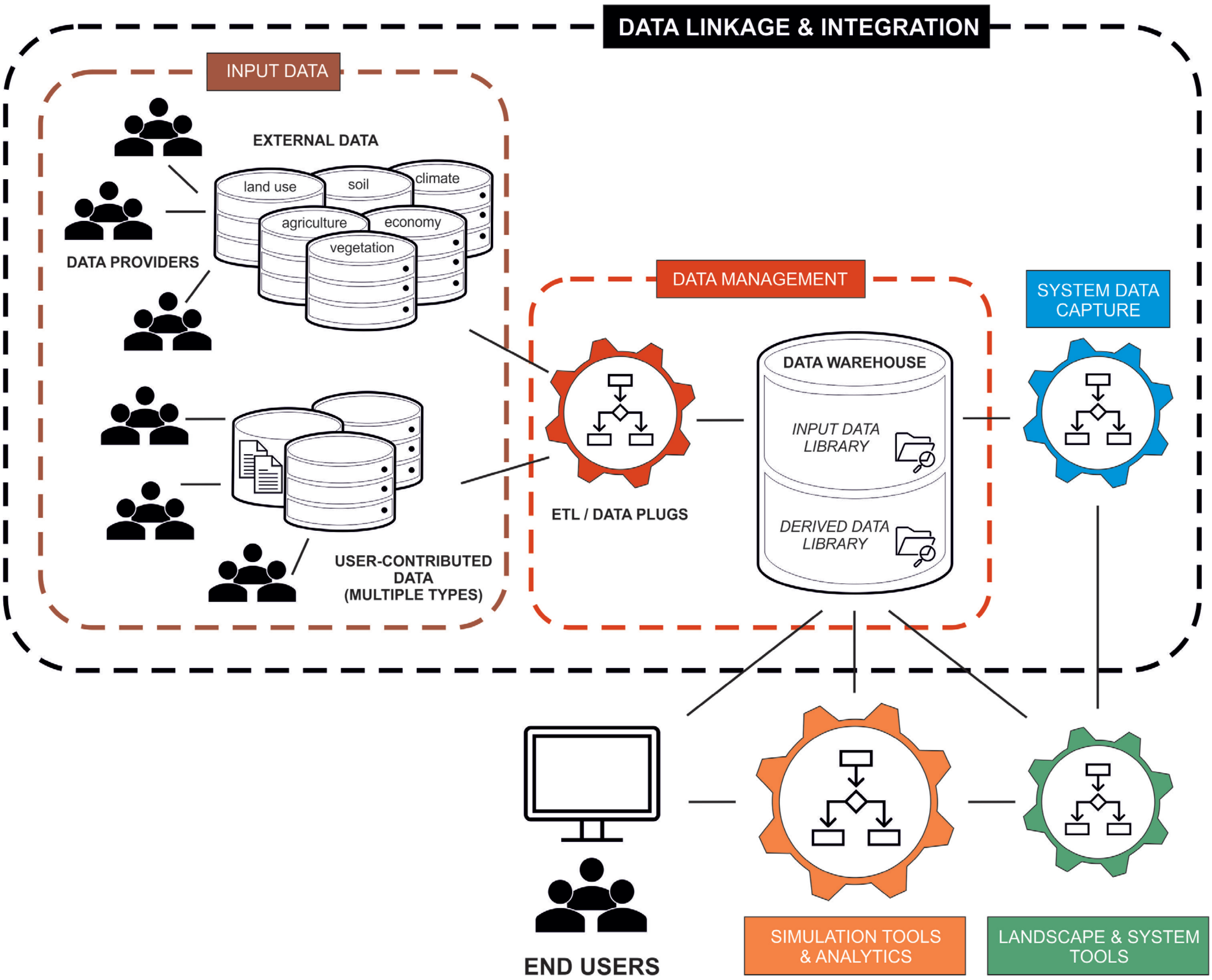

Figure 10. Suggested design of the Data Linkage and Integration module for landscape model generation. External and user-contributed datasets will be accessed or gathered and fed into the Input Data Library. This will be supported by designing dedicated input Data Plugs including necessary ETL (Extract, Transform, and Load) processes to meet the internal data formats and standards defined to allow seamless connections between data and the tools using it. Input Data Library and Landscape and Systems Tools will be linked via System Data Capture tools. Here, an adaptive hierarchical spatial data selection algorithm could be applied to select, for a certain geo-location, a set of "best available" and up-to-date landscape and systems input data. This set of data will be used for the generation of detailed, dynamic landscape models to feed simulations. These models and simulation outputs will be stored in the Derived Data Library.

Acknowledgements

Many people have contributed to the ALMaSS landscape generation over the years, either by providing the necessary input data or adjusting the methodology to be used with specific national or regional landscape data. Lars Dalby and Flemming Skov developed the first version of the ALMaSS landscape models for Denmark. Jamie Alison made a significant contribution to the development of a prototype version of the landscape generation scripts, which leverage open-source GIS tools for spatial data processing.

References

App M, Hellwig N, Schneider-Hohenbrink A-K, Burmeister J, Schröder B, Thiele J (2025) SyrFitSources: An agent-based model to investigate the effects of land use on the population dynamics of aphidophagous hoverflies. Ecological Modelling 507: 111098. https://doi.org/10.1016/j.ecolmodel.2025.111098

Bakker MM, Alam SJ, van Dijk J, Rounsevell MDA (2015) Land-use change arising from rural land exchange: an agent-based simulation model. Landscape Ecology 30: 273–286. https://doi.org/10.1007/s10980-014-0116-x

Becher MA, Grimm V, Thorbek P, Horn J, Kennedy PJ, Osborne JL (2014) Beehave: a systems model of honeybee colony dynamics and foraging to explore multifactorial causes of colony failure. Journal of Applied Ecology 51: 470–482. https://doi.org/10.1111/1365-2664.12222

Becher MA, Twiston-Davies G, Penny TD, Goulson D, Rotheray EL, Osborne JL (2018) Bumble-BEEHAVE: a systems model for exploring multifactorial causes of bumblebee decline at individual, colony, population and community level. Journal of Applied Ecology 55: 2790–2801. https://doi.org/10.1111/1365-2664.13165

Beckers V, Beckers J, Vanmaercke M, Van Hecke E, Van Rompaey A, Dendoncker N (2018) Modelling farm growth and its impact on agricultural land use: A country scale application of an agent-based model. Land 7(3): 109. https://doi.org/10.3390/land7030109

Bednarska AJ, Mikołajczyk Ł, Ziółkowska E, Kocjan K, Wnęk A, Mokkapati JS, Teper D, Kaczyński P, Łozowicka B, Śliwińska R, Laskowski R (2021) Effects of agricultural landscape structure, insecticide residues, and pollen diversity on the life-history traits of the red mason bee *Osmia bicornis*. The Science of the Total Environment 809: 151142. https://doi.org/10.1016/j.scitotenv.2021.151142

Ghassemi B, Izquierdo-Verdiguier E, Verhegghen A, Yordanov M, Lemoine G, Martínez ÁM, De Marchi D, van der Velde M, Vuolo F, d'Andrimont R (2024) European Union crop map 2022: Earth observation's 10-meter dive into Europe's crop tapestry. Scientific Data 11: 1048. https://doi.org/10.1038/s41597-024-03884-y

Hoye TT, Skov F, Topping CJ (2012) Interpreting outputs of agent-based models using abundance-occupancy relationships. Ecological Indicators 20: 221–227. https://doi.org/10.1016/j.ecolind.2012.01.017

Jepsen JU, Topping CJ, Odderskær P, Andersen PN (2005) Evaluating consequences of land-use strategies on wildlife populations using multiple-species predictive scenarios. Agriculture, Ecosystems & Environment 105: 581–594. https://doi.org/10.1016/j.agee.2004.08.007

Lefebvre M, Espinosa M, Gomez y Paloma S, Paracchini ML, Piorr A, Zasada I (2015) Agricultural landscapes as multi-scale public good and the role of the Common Agricultural Policy. Journal of Environmental Planning and Management 58: 2088–2112. https://doi.org/10.1080/09640568.2014.891975

Luke S, Cioffi-Revilla C, Panait L, Sullivan K, Balan G (2005) MASON: A multiagent simulation environment. Simulation 81(7): 517–527. https://doi.org/10.1177/0037549705058073

McLane AJ, Semeniuk C, McDermid GJ, Marceau DJ (2011) The role of agent-based models in wildlife ecology and management. Ecological Modelling 222: 1544–1556. https://doi.org/10.1016/j.ecolmodel.2011.01.020

Mikołajczyk Ł, Laskowski R, Bednarska AJ, Ziółkowska E (2021) Species-specific landscape characterisation method in agro-ecosystems. Ecological Indicators 129: 107894. https://doi.org/10.1016/j.ecolind.2021.107894

Nilsson M, Zamparutti T, Petersen JE, Nykvist B, Rudberg P, McGuinn J (2012) Understanding policy coherence: Analytical framework and examples of sector–environment policy interactions in the EU. Environmental Policy and Governance 22: 395–423. https://doi.org/10.1002/eet.1589

North MJ, Collier NT, Vos JR (2006) Experiences creating three implementations of the Repast agent modeling toolkit. ACM Transactions on Modeling and Computer Simulation 16(1): 1–25. https://doi.org/10.1145/1122012.1122013

Pauli BP, McCann NP, Zollner PA, Cummings R, Gilbert JH, Gustafson EJ (2013) SEARCH: Spatially Explicit Animal Response to Composition of Habitat. PLOS ONE 8(5): e64656. https://doi.org/10.1371/journal.pone.0064656

Sowa G, Bednarska AJ, Ziółkowska E, Laskowski R (2022) Homogeneity of agriculture landscape promotes insecticide resistance in the ground beetle *Poecilus cupreus*. PLOS ONE 17(4): e0266453. https://doi.org/10.1371/journal.pone.0266453

Sullivan K, Coletti M, Luke S (2010) GeoMason: Geospatial support for MASON. Technical Report. Dept. of CS, George Mason University.

Taillandier P, Gaudou B, Grignard A, Huynh Q-N, Marilleau N, Caillou P, Philippon D, Drogoul A (2019) Building, composing and experimenting complex spatial models with the GAMA platform. Geoinformatica 23: 299–322. https://doi.org/10.1007/s10707-018-00339-6

Tang W, Bennett DA (2010) The explicit representation of context in agent-based models of complex adaptive spatial Systems. Annals of the Association of American Geographers 100(5): 1128–1155. https://doi.org/10.1080/00045608.2010.517739

Topping CJ (2022) The Animal Landscape and Man Simulation System (ALMaSS): a history, design, and philosophy. Research Ideas and Outcomes 8: e89919. https://doi.org/10.3897/rio.8.e89919

Topping CJ, Dalby L, Skov F (2016) Landscape structure and management alter the outcome of a pesticide ERA: Evaluating impacts of endocrine disruption using the ALMaSS European Brown Hare model. Science of The Total Environment 541: 1477–1488. https://doi.org/10.1016/j.scitotenv.2015.10.042

Topping CJ, Duan X (2024) ALMaSS Landscape and Farming Simulation: software classes and methods. Food and ecological Systems Modelling Journal 5: e121215. https://doi.org/10.3897/fmj.5.121215

Topping CJ, Hansen T (2003) ALMaSS, an agent-based model for animals in temperate European landscapes. Ecological Modelling 167: 65–82. https://doi.org/10.1016/S0304-3800(03)00173-X

Topping CJ, Lagisz M (2012) Spatial dynamic factors affecting population-level risk assessment for a terrestrial arthropod: An agent-based modeling approach. Human and Ecological Risk Assessment 18: 168–180. https://doi.org/10.1080/10807039.2012.632292

Valbuena D, Verburg PH, Bregt AK, Ligtenberg A (2010) An agent-based approach to model land-use change at a regional scale. Landscape Ecology 25: 185–199. https://doi.org/10.1007/s10980-009-9380-6

Vasseur C, Joannon A, Aviron S, Burel F, Meynard J-M, Baudry J (2013) The cropping systems mosaic: How does the hidden heterogeneity of agricultural landscapes drive arthropod populations? Agriculture, Ecosystems & Environment 166: 3014. https://doi.org/10.1016/j.agee.2012.08.013

Ziółkowska E, Laskowski R, Dominic AR, Stein S, Wenzel B, Kehlenbeck H, Riddervold M, Topping CJ (2024) Evaluating the impact of landscape structure and source-sink dynamics on non-target arthropod pesticide risk assessments in Germany. German Environment Agency, Texte | 58/2024, 150 pp. https://doi.org/10.60810/openumwelt-6012

Ziółkowska E, Tiktak A, Topping CJ (2022) Is the effectiveness of policy-driven mitigation measures on carabid populations driven by landscape and farmland heterogeneity? Applying a modelling approach in the Dutch agroecosystems. PLOS ONE 17(12): e0279639. https://doi.org/10.1371/journal.pone.0279639

Ziółkowska E, Topping CJ, Bednarska AJ, Laskowski R (2021) Supporting non-target arthropods in agroecosystems: Modelling effects of insecticides and landscape structure on carabids in agricultural landscapes. Science of the Total Environment 774: 145746. https://doi.org/10.1016/j.scitotenv.2021.145746

Supplementary material 1

Supplementary tables S1, S2

Authors: Elżbieta Ziółkowska, Barbara Jaśkowiec, Geoffrey Brian Groom, Christopher John Topping

Data type: pdf

Explanation note: **table S1.** Types of landscape elements (TOLEs) defined in ALMaSS. **table S2.** Selected information on production on Polish farms according to the type of farming, based on the standard results of the FADN Public Database for 2023.

Link: https://doi.org/10.3897/aem.8.167439.suppl1

Additional information

Conflict of interest

The authors have declared that no competing interests exist.

Ethical statement

No ethical statement was reported.

Artificial Intelligence (AI) use

The authors accept full responsibility for the content of the manuscript, including the disclosure of any use of AI.
No AI tools were used in the preparation of this manuscript.

Funding

This work was partially supported by the EcoStack project funded by the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 773554 and by the PollinERA project funded by the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101135005.

Author contributions

Conceptualization: CJT, EMZ. Funding acquisition: CJT. Methodology: EMZ, CJT, BJ, GBG. Software: CJT, EMZ, GBG, BJ. Supervision: CJT. Visualization: EMZ. Writing – original draft: EMZ. Writing – review and editing: BJ, CJT, GBG.

Author ORCIDs

E. Ziółkowska https://orcid.org/0000-0002-7213-2200
B. Jaśkowiec https://orcid.org/0000-0002-6737-1346
G.B. Groom https://orcid.org/0000-0003-3820-6621
C.J. Topping https://orcid.org/0000-0003-0874-7603

Data availability

All data that support the findings of this study are available in the main text or Supplementary Materials.

The source code and the running directory illustrating the operational implementation of the presented methodology in Polish agricultural landscapes are publicly available at the GitLab repository: https://gitlab.com/ALMaSS/almass_methodology/-/tree/main/landscape_generation.

The ALMaSS landscape models developed for various countries (as part of different projects mentioned in the section "Outcomes") are available from the authors upon request.