ECOGRAPHY

Software Note

'ukbioprepr': an R package to support reproducible preparation of environmental data for biodiversity modelling in the UK

Charlotte Rose Rush[1], Joseph Cooper[2], Cecilia Larrosa[3] and Martin Wilkes[1]

[1]University of Essex – Colchester Campus, Life Sciences, Colchester, UK
[2]British Trust for Ornithology, Terrestrial Ecology, The Nunnery, Thetford, Norfolk, UK
[3]Biodiversify Ltd., Sheriff House, 64-66 St James's St, Nottingham, UK

Correspondence: Charlotte Rose Rush (cr23569@essex.ac.uk)

Ecography
2026: e08413
doi: 10.1002/ecog.08413

Subject Editor:
Michael Krabbe Borregaard
Editor-in-Chief:
Christine N. Meynard
Accepted 15 January 2026

Biodiversity modelling is essential for explaining and predicting ecological responses to environmental change and assessing progress towards targets in the Kunming-Montreal Global Biodiversity Framework (CBD 2022). The UK benefits from rich biodiversity time-series data and numerous open-source environmental datasets. However, integrating these into modelling workflows remains challenging – especially for those without considerable data processing expertise. Fragmented sources, spatial and temporal discrepancies and undocumented or unreproducible processing methods often create barriers and hinder coordination. We present 'ukbioprepr', a user-friendly R package developed to address key environmental data preparation challenges in UK biodiversity modelling. It provides functions for downloading, harmonising and extracting site-level environmental variables from open-source datasets on climate, land cover and soil properties. Data products are processed to align spatially and temporarily with UK biodiversity data, allowing consistent covariate generation from 2000 onwards. Substantial data engineering – including reprojection, temporal interpolation and spatial alignment – supports model-ready outputs, whilst limitations (e.g. sparse soil data in urban areas) are transparently documented. 'ukbioprepr' supports both point-based and grid-based survey data and includes methods for aggregating climate variables over biologically relevant time periods (e.g. seasons, custom annual windows). These features enable integration across spatial and temporal scales and support diverse biodiversity modelling approaches. We demonstrate its application in a case study modelling occupancy of the native UK wildflower *Hyacinthoides non-scripta*. Using derived environmental predictors, we show how these data products can inform ecological forecasts under future climate scenarios, predicting a 12.4% reduction in suitable habitat area under the most severe scenario (RCP8.5). By lowering technical barriers and enabling consistent environmental data integration, 'ukbioprepr' supports scalable, reproducible biodiversity modelling across all four nations in

NORDIC SOCIETY OIKOS

www.ecography.org

the UK. The package exemplifies how targeted frameworks can streamline modelling workflows and improve coordination across biodiversity research and policy – principles that can be applied globally.

Keywords: biodiversity, data wrangling, ecological modelling, ecological software, environmental data, R package

Introduction

The world is facing a biodiversity crisis, with an estimated one million species currently at risk of extinction (IPBES 2019). Reversing this trend requires coordinated global efforts to halt biodiversity loss, conserve species and restore ecosystems (Leclère et al. 2020). Biodiversity modelling plays a crucial role in this process by helping researchers and policymakers understand how species respond to environmental change, inform conservation strategies and assess progress towards important targets prescribed by the Kunming-Montreal Global Biodiversity Framework (GBF; CBD 2022).

The availability of environmental data for biodiversity modelling is increasing exponentially, aided by improvements in data collection methods (e.g. remote sensing) and the push towards open-access, online repositories (Zurell et al. 2020). However, integrating these data into modelling workflows remains a considerable challenge. Data preparation methods can be labour intensive, fragmented, poorly documented and unreproducible – slowing progress and hindering collaboration (Munafò et al. 2017). The highly cited ODMAP protocol for reporting biodiversity models (Zurell et al. 2020) highlights the importance of clearly detailing technical data processing steps to enable reproducibility, facilitate peer review and improve model reliability – ultimately leading to more robust biodiversity estimates and predictions (Rapacciuolo 2018).

The UK boasts some of the most comprehensive and well-coordinated biodiversity and environmental monitoring efforts in the world, producing long-term, high-resolution datasets across diverse taxa and habitats (Burns et al. 2018). However, despite this wealth of data, the UK remains one of the most nature-depleted countries globally (Burns et al. 2013). In response, and in alignment with the GBF, the UK nations have committed to halting the decline in species abundance by 2030 under the Environment Act 2021 (c. 30). Meeting this goal, and others of this nature, relies on the ability to translate raw biodiversity data into meaningful policy-relevant outputs that inform conservation action. Robust, transparent methods for preparing and integrating environmental covariates at appropriate spatial and temporal scales are therefore essential to unlock the full potential of UK biodiversity data.

Several R packages support different steps of biodiversity modelling, including data preparation (www.r-project.org; Sillero et al. 2023). However, few are specifically tailored to the geographical and temporal contexts required for detailed national-scale modelling. Of the 60 packages reviewed by Sillero et al. (2023), only three – 'sdmpredictors' (Bosch and Fernandez 2022), 'geodata' (Hijmans et al. 2022) and 'mcera5' (Klinges et al. 2022) – were designed to streamline environmental data compilation. These packages focus primarily on global or continental datasets, often with limitations in spatial or temporal resolution. For example, 'sdmpredictors' includes historical data from WorldClim at a 1 km resolution, but future projections are only available as 20-year averages at ~ 18.5 km resolution (Fick and Hijmans 2017). Similarly, bio-ORACLE, included in both 'sdmpredictors' and 'geodata', is limited to ~ 5.5 km resolution at the equator (Assis et al. 2017). While 'geodata' does incorporate some land cover data, these are derived from ESA WorldCover, which, although available at a fine, 10 m resolution, only spans the years 2020–2021 (Zanaga et al. 2022). The 'mcera5' package provides access to ERA5 climate data at a coarse ~ 31 km resolution (Copernicus Climate Change Service 2017). Furthermore, existing packages are heavily skewed towards climate variables and rarely integrate soil or land cover data to align with the spatial needs of UK biodiversity datasets.

In contrast, 'ukbioprepr' is specifically designed to support UK biodiversity modelling by generating environmental covariates at a 1 km resolution, precisely aligned with monads – 1 × 1 km grid squares defined within a specified coordinate reference system – commonly used in monitoring schemes such as the Breeding Bird Survey (Gregory et al. 2004) and the National Plant Monitoring Scheme (Pescott et al. 2019). Although 1 km may be considered coarse in some ecological contexts, it provides a meaningful resolution for many UK-wide monitoring initiatives. Crucially, 'ukbioprepr' offers harmonised data products for land cover and climate spanning 2000–2023, alongside static soil variables, providing broader temporal coverage and greater breadth than existing tools. By addressing both the spatial and temporal mismatches present in other packages, 'ukbioprepr' fills a key gap for researchers conducting higher resolution, policy relevant biodiversity analyses in the UK.

'ukbioprepr' provides user-friendly functions to download, harmonise, and extract pre-prepared environmental attributes from existing, open-source datasets on land cover, soil properties and climatic variables. Outputs are aligned with UK monads and compatible with both British National Grid (EPSG:27700) and Irish Grid (EPSG:29900), overcoming a common challenge for users working with biodiversity data collected across the UK using different spatial reference systems. To accommodate this, data products and methods were developed to handle both coordinate systems seamlessly. The package streamlines complex spatial and temporal data processing steps – such as reprojection and interpolation – thereby saving substantial time and effort while supporting reproducible, scalable data preparation for a wide range of biodiversity modelling applications.

In this paper, we describe the methods used to generate the environmental data products, outline the core functionality of the package, and demonstrate its application in a case study modeling the occupancy of *Hyacinthoides non-scripta*, an iconic native UK wildflower. The case study highlights how 'ukbioprepr' can not only be used to support biodiversity assessment, but also to forecast species responses under alternative future climate scenarios.

Data products and sources

The environmental data products available in 'ukbioprepr' are derived from a range of open-source datasets on UK soil properties, land cover classes and climate variables. These initial data products originate from diverse sources and vary in format, spatial resolution and coordinate reference systems (Table 1). To ensure consistency and compatibility with UK biodiversity survey data, each original data product underwent a series of processing steps involving projection, alignment, masking, and cropping using methods from the 'terra' package ver. 1.8.42 (Hijmans 2024).

All data products were projected into the British National Grid (EPSG:29903) for the UK and Irish Grid (EPSG:29903) for Northern Ireland using the *project()* function where necessary. This step enables seamless handling of British and Irish grid references or XY coordinates, supporting spatial extraction of environmental values without additional coordinate transformations. Following this, the *align()* function was used to ensure that all rasters were precisely aligned with a 1 km monad grid, ensuring that the environmental data products correspond directly with biodiversity surveying methods, commonly undertaken at the monad level in UK monitoring schemes. Masking was applied using the *mask()* function to remove invalid or placeholder values (e.g. zeros or extreme negative values) that represented areas with no data in the original sources. This ensures that only valid environmental values are retained to avoid erroneous outputs during extraction or analysis. The *crop()* function was also used to spatially subset the data products to either the UK or specifically to Northern Ireland, depending on the coordinate reference system.

These processing steps produce standardised raster data products at a 1 km resolution, fully aligned with UK and Irish grids, and readily usable for modelling and analysis. The harmonised datasets are hosted in dedicated online repositories and can be openly accessed via 'ukbioprepr' functions. In the sections that follow, we detail any specific processing steps and considerations undertaken for each type of environmental data product. Although the current release focuses on long-term, nationally consistent climate, land cover and soil datasets, the same workflow can be extended to incorporate additional products – such as remote-sensing derived metrics – as higher resolution, UK-wide datasets become available.

Soil properties

SoilGrids is a digital mapping system which uses machine learning and environmental covariates to map the spatial distribution of soil properties, using approximately 240 000 soil observations worldwide (Poggio et al. 2021, Turek et al. 2023). SoilGrids was selected for 'ukbioprepr' due to its high spatial resolution, modern predictive methods and comprehensive collection of soil variables. Additionally, predictions available at six standardised depth intervals down to 200 cm offer more options and scope for biodiversity modelling. A further advantage is the inclusion of quantified uncertainty estimates, with the UK possessing a relatively low uncertainty in a global context (Poggio et al. 2021, Turek et al. 2023). While other UK national datasets provide valuable data, they often lack the depth resolution and spatial coverage offered by SoilGrids (Lawley et al. 2014, UKSO 2025).

Data products for 14 soil properties – including *clay proportion*, *organic carbon density* and *pH* – were created using SoilGrids250m ver. 2.0 (Poggio et al. 2021, Turek et al. 2023). Original layers were downloaded and converted to standard units following SoilGrids guidance (Poggio et al. 2021)

Table 1. Sources and original format for all data products available in the 'ukbioprepr' package.

Environmental variable type	Source	Original data product description	Original coordinate reference system
Soil properties	SoilGrids250m ver. 2.0 (Poggio et al. 2021, Turek et al. 2023)	• 250 m resolution • Downloadable as one available soil depth for one soil property at a maximum of a 2 longitudinal × 2 latitudinal degrees • Two derived, eight physical and four chemical soil properties at up six different depths	EPSG:4326
Land cover	UK Centre for Ecology and Hydrology (UKCEH) 1 km target summary rasters (UKCEH 2025)	• 1 km resolution • Separate products for Northern Ireland (NI) and Great Britain (GB) (Isle of Man not included) • Available for years 2000, 2007, 2015, 2017–2023	EPSG:29903 for NI, EPSG:27700 for GB
Climate variables	Met Office HadUK-Grid (Met Office et al. 2018)	• 1 km resolution • Full UK coverage in data product • Monthly measurements for years 1999–2023 for four climate variables: mean, maximum and minimum temperature and precipitation	EPSG:27700

using the *app()* function from the 'terra' ver. 1.8-42 package (Hijmans 2024). Two sets of data products were created: one covering the UK in British National Grid (EPSG:27700) and one for Northern Ireland in Irish Grid (EPSG:29903), following the projection and cropping procedures described previously. A list of all properties at available depths can be found in the Supporting information.

A known limitation of SoilGrids is reduced data availability for urban areas and inland water bodies, which are typically under-represented in soil surveys. These gaps in the training data make model fitting more challenging and result in higher uncertainties. As a result, SoilGrids does not provide data for urban areas and coverage is also often incomplete around coastal and estuarine locations. Users of 'ukbioprepr' are advised to interpret soil estimates near these areas with caution, as the final data products were masked to assign NA values in regions where soil data were sparse. Previous studies have addressed similar challenges using machine learning methods (Tian et al. 2022, Qu et al. 2024, Siqueira et al. 2024, Wang et al. 2024) and spatial interpolation techniques (Madenoglu et al. 2020, Ma et al. 2022, Sanczuk et al. 2022, Ariza-Salamanca et al. 2023).

Land cover

Using the 1 km percentage cover data from the UK Centre for Ecology and Hydrology (UKCEH), we produced four sets of land cover maps: aggregated classes from 2000 onwards and detailed classes from 2015 to 2023, each for the UK and for Northern Ireland separately. The original UKCEH datasets are published separately for Great Britain and Northern Ireland and not available every year until 2017 (Table 1).

To maximise consistency across years, we aggregated land cover classes where classification schemes and methods had changed. For example, the 'Montane' class (present in 2000 and 2007) was removed in 2015 and reclassified as 'Inland rock' or another upland type (UKCEH 2017). To understand how this affected the change in data, we extracted land cover values from upland regions in both 2007 and 2015 using the *extract()* function from the 'terra' package (Hijmans 2024) and compared the outputs. It was not possible to determine whether the differences reflected actual land cover change or reclassification. As a result, we created a composite 'Upland' class to ensure consistency in our secondary data products. Similarly, variation in grassland classifications across years led us to create a general 'Grasses' class (Morton et al. 2011, UKCEH 2017; Fig. 1). These aggregations were implemented using the *app()* function from the 'terra' package (Hijmans 2024).

We note a further methodological change identified by UKCEH, where areas of mixed woodland with over 20% broad-leaved woodland cover may have been assigned to the coniferous woodland class in some years (Rowland et al. 2017a, 2017b, UKCEH 2017). The land cover maps presented here do not correct for this; users particularly interested in woodland classes should interpret these classes with caution.

For years in which land cover maps were unavailable (2001–2006, 2008–2014, 2016), novel data products were generated using a simple linear interpolation method

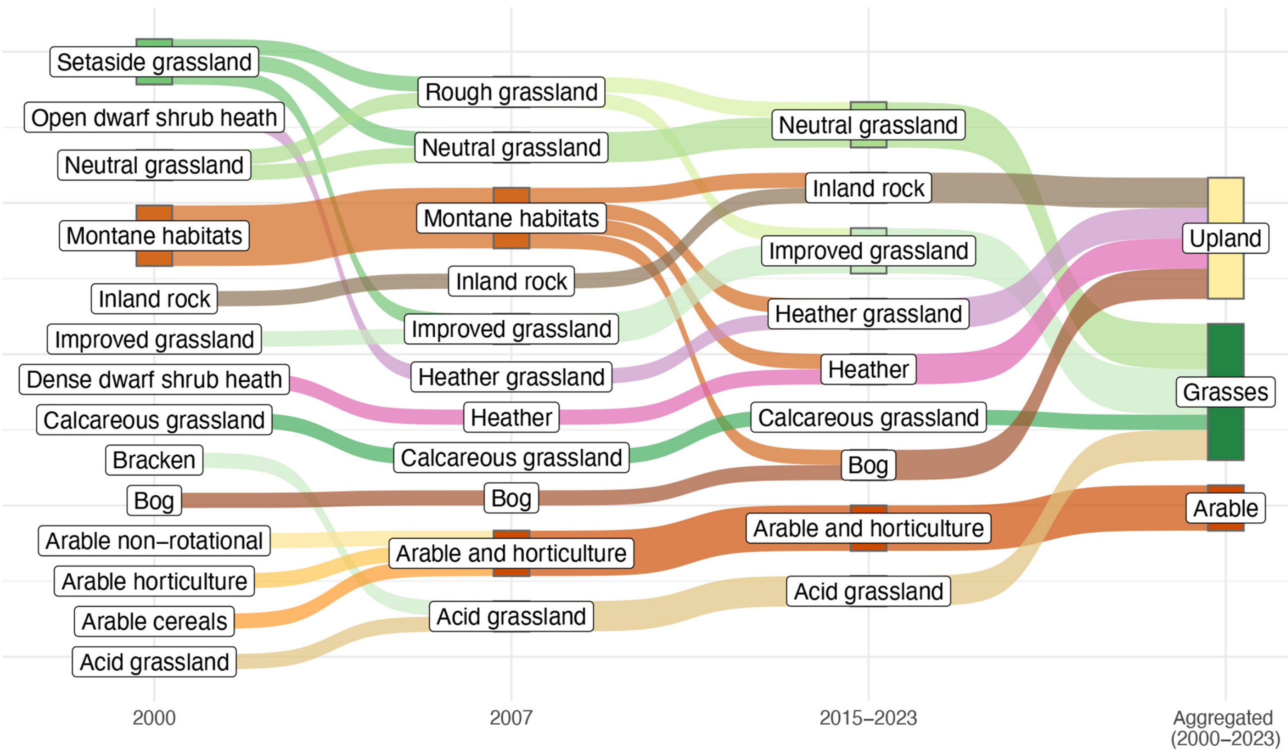

Figure 1. The change between land cover classifications through the available years. Classes which did not undergo any changes have not been included. This figure is for explanatory purposes only and does not show proportional changes.

between the nearest available years across raster cells for each land cover class (Fig. 2). Specifically, values were estimated as:

$$v_{t=}v_1 + \frac{(v_2 - v_1)}{(y_2 - y_1)} \times (t - y_1),$$

Where v_1 and v_2 represent the cell values at years y_1 and y_2, respectively, and v_t is the value for missing year t. This was applied independently to each raster cell and land cover class using the *app()* function from the 'terra' package (Hijmans 2024). As this is a deterministic interpolation, no model-based uncertainty is introduced beyond that inherent in the source data. Northern Ireland products were projected to EPSG:27700 using a bilinear interpolation approach and mosaiced with Great Britain to produce unified UK products.

Some original UKCEH products, particularly in coastal areas, contained cells with minimal or no land cover data. This typically occurred in locations with abrupt environmental transitions, such as cliff edges or shorelines, where land cover classes (e.g. marine, beach, sandy habitats) were not captured by the original classification scheme. In producing our secondary data products, we limited interpolation to areas with consistent data coverage across years and excluded any inconsistent cells by masking them as NA. These areas are therefore not represented in the finalised data products, and

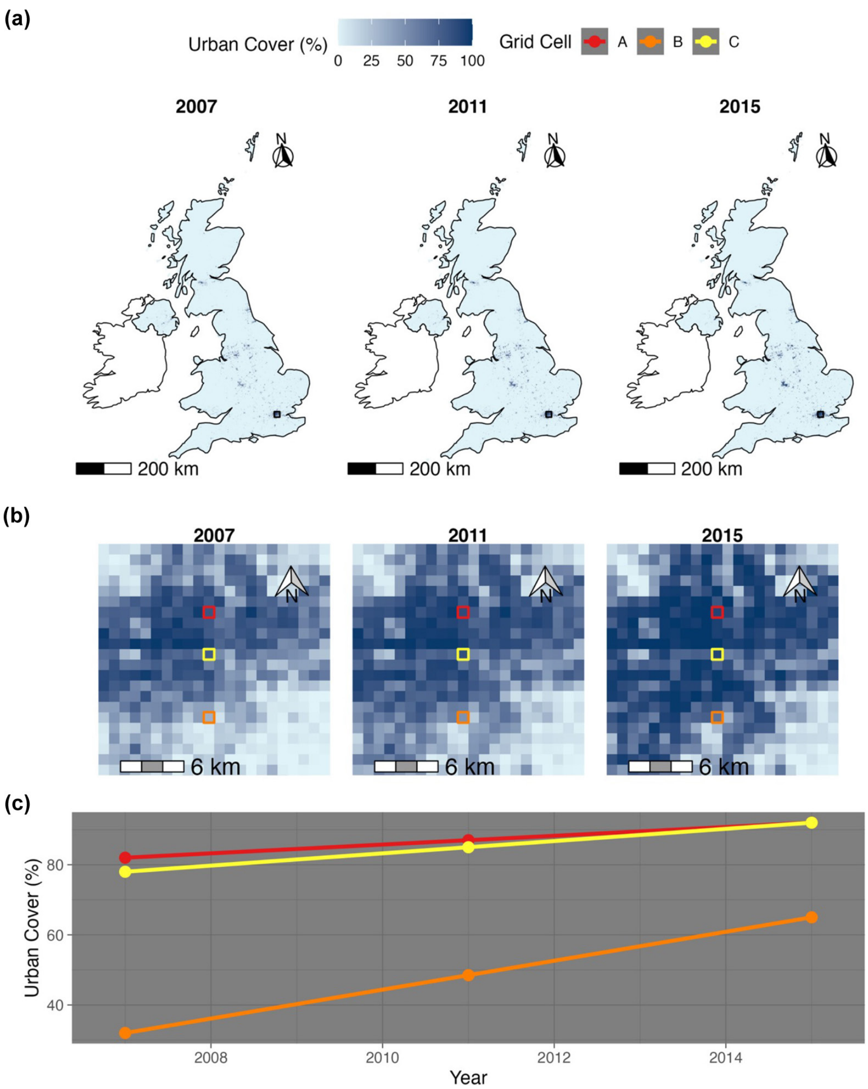

Figure 2. Illustration of the simple linear interpolation method used to estimate missing years of land cover data. (a) UK maps showing the percentage cover of urban land cover for three example years in EPSG:27700 at a 1 km resolution. Greater London is outlined in a black square. (b) Enlarged views of Greater London highlighting the spatial variation of urban cover. Three 1-km grid cells (a–c), shown in different colours, were randomly selected to illustrate temporal changes. (c) For each selected grid cell, changes in urban cover across the example years are plotted, with linear interpolation applied between the years 2007 and 2015, to estimate values for missing years, i.e. 2011, shown by connecting lines.

users should be aware of this limitation when working with data near coastal or inland water boundaries.

The final land cover products integrate and harmonise existing datasets, resolving issues related to missing years, inconsistent classification schemes, and different coordinate reference systems. They offer continuous, annual data from 2000 to 2023, enabling temporally consistent biodiversity modelling. Whilst no additional uncertainty is introduced through the linear interpolation method, errors or misclassifications in the original land cover products will propagate through interpolated layers. Users are encouraged to consider these limitations when conducting downstream analysis. This is an improvement over previous approaches that relied on land cover data from a single year, despite analysing biodiversity data spanning over multiple years (Croft et al. 2019, Boyd et al. 2023, Arsevska et al. 2024). See the Supporting information for class information for both the detailed and aggregated sets of land cover data products, respectively.

Climate variables

Climate data products in 'ukbioprepr' are derived from the UK Met Office HadUK-Grid dataset, which provides high-resolution, gridded climate observations across the UK (Met Office et al. 2018). Four variables were selected: average temperature (tas), maximum temperature (tasmax), minimum temperature (tasmin), and precipitation (rain), measured in degrees Celsius and millimeters, respectively. These variables are provided as monthly values from the years 1999–2023, offering consistent temporal coverage, suitable for long-term biodiversity modelling.

To generate Northern Ireland specific products, the original UK datasets, initially aligned with the British National grid, were cropped to the region and projected to the Irish Grid (EPSG:29903), following the methods described in previous sections. HadUK-Grid was selected over global or continental datasets such as CHELSA (Karger et al. 2017) or Copernicus Climate Change Service (2024) datasets due to its high spatial resolution, greater temporal coverage and use of observed meteorological data interpolated to a uniform grid (Cornes et al. 2018, Chakraborty et al. 2020, Muñoz Sabater et al. 2024). Following projection, the Northern Ireland products were aligned with the Irish Grid (EPSG:29903) monad system to ensure compatibility with monad-based biodiversity monitoring in Northern Ireland. The chosen climate variables align with those commonly used in global modelling frameworks (Karger et al. 2017) facilitating broader compatibility. Methods for aggregating monthly layers into seasonal and annual values are detailed below.

Package and functions

The 'ukbioprepr' package was built using R ver. 4.4.0 (www.r-project.org) and is freely available via GitHub. It provides functions to download (referred to as *fetch()*) and extract environmental variables related to soil properties, land cover, and climate, based on the data products described in section 'Data products and sources'. All required dependencies are installed during the initial package setup. Table 2 summarises the functions and their respective outputs.

Both fetching and extraction functions operate using files hosted in an external online repository and therefore require an internet connection. As some files are large, users are advised to increase their system timeout settings to ensure successful downloads; informative warnings and instructions within the package support this process. Files are downloaded to a temporary directory and remain accessible throughout the user's R session. These files are automatically removed at the end of the session, preventing unnecessary storage use.

fetch_*() functions

The 'ukbioprepr' package includes three *fetch_*()* functions (Table 2) that enable users to download pre-processed raster files of soil properties, land cover classes and climate variables as described in Section 'Data products and sources'. Users can specify whether to download the data for the entirety of the UK (EPSG:27700) or for Northern Ireland (EPSG:29903). These are direct downloads of our harmonised, derived data products stored in the online repositories described above.

The *fetch_climate_raster()* function differs from other *fetch_*()* functions as it offers flexible tools for generating climate rasters from the original monthly data products. Users can aggregate data into seasonal or annual rasters based on standard seasonal groupings (e.g. winter as December–February, spring as March–May, etc.), or define custom annual periods (e.g. June–May) relevant to their ecological research. This is implemented dynamically using the *app()* function from the 'terra' package (Hijmans 2024), streamlining workflows and saving effort otherwise spent on data aggregation and transformation. This function supports both standardised covariates, similar to the BioClim variables (Karger et al. 2017), and tailored metrics relevant to specific biodiversity study, such as flowering response to winter temperatures or species' range activity as a response to summer precipitation.

extract_*() functions

The four *extract_*()* functions in 'ukbioprepr' allow users to extract environmental covariate values from downloaded data products. Extractions can be made using monad grid references in either British National Grid or Irish Grid, reflecting the common formats used in UK biodiversity datasets. The functions automatically detect the coordinate reference system and convert grid references to coordinates of each 1 km square using the 'igr' package (Kennedy 2025) for Irish Grid and the 'rnrfa' package (Vitolo et al. 2018) for British Grid references. These coordinates are then used to extract values via the *extract()* function from the 'terra' package (Hijmans 2024).

As an example, use case, Fig. 3 demonstrates how the *extract_climate_values()* function can be used to generate annual climate summaries (with a custom June–May year) for four UK monads over the period 2000–2023. Code to

Table 2. The functions of 'ukbioprepr' and their outputs. The functions can be split into two categories: *fetch_*()* functions for downloading raster data products and *extract_*()* functions for generating values for chosen variables. More information on the required inputs for these functions can be found on the GitHub for the package.

Function	Description	Output
fetch_soil_raster()	• Downloads prepared soil property raster files for either UK or Northern Ireland (NI) • Users can choose all available soil properties or input a specific list	A list of rasters, one for each property with layers corresponding to soil depths
fetch_landcover_raster()	• Downloads prepared land cover raster files for UK or NI for user specified years • Raster files with aggregated classes are provided for years prior to 2015	A list of rasters, one for each year where layers correspond to each land cover class
fetch_climate_raster()	• Downloads prepared climate raster files for UK or NI for specified start and end date (YYYY_MM) between 1999–2023 of a chosen climate variable • Climate variables include: average temperature (tas), maximum temperature (tasmax), minimum temperature (tasmin) and precipitation (rain) • Original files are in monthly format, however users can choose a time parameter of 'monthly', 'seasonal' or 'annual' for the output. Annual values are calculated from input start month, e.g. an input start month of January would yield annual data consistent with a calendar year • Users can also choose how the output is calculated from monthly data: 'mean', 'min', 'max' and 'sum', e.g. the mean annual maximum temperature or the total annual rainfall using 'sum'	A raster of the selected climate variable, time parameter and date range
extract_soil_values()	• Extracts values of chosen soil properties from raster files for UK and/or NI • Extractions can be performed using X and Y coordinates of a specified coordinate reference system, or grid references in British National Grid (EPSG:27700) and/or Irish Grid (EPSG:29903)	A data frame of locations and extracted values for each soil property at each depth
extract_landcover_values()	• Extracts values of land cover from raster files for UK and/or NI for years between 2000–2023 • Extractions can be performed using X and Y coordinates of a specified coordinate reference system, or monad grid references in British National Grid (EPSG:27700) and/or Irish Grid (EPSG:29903) e.g. J3480 or NT1565, respectively • Based on year, the extractions will be performed using either aggregated land cover classes or full land cover classes	A data frame of locations and extracted values for each land cover class at each year. Land cover measurements are represented as percentage cover
extract_climate_values()	• Extracts values of chosen climate variables across chose time frames between years 1999–2023 of UK and/or NI • Extractions can be performed using X and Y coordinates of a specified coordinate reference system, or grid references in British National Grid (EPSG:27700) and/or Irish Grid (EPSG:29903) • Users can choose which climate variables they wish to extract for as well as the time frames similar to *fetch_climate_raster()* • Users can specific a custom annual window for extractions, e.g. annual mean temperature where a year runs from September–August	A data frame of locations and extracted values of climate variables and time frames. When choosing 'seasonal', the extracted values are the season of the entry and the four seasons previous
extract_all_values()	• A function combining the three extraction functions into one, covering soil properties, land cover and climate • Users can choose all three environmental datasets or use binary operators to choose two as a combination	A data frame containing locations and values for all required variables (as above)

reproduce this example is available in the accompanying repository.

In addition to handling monad grid references, the *extract_*()* functions in 'ukbioprepr' can also extract values using X and Y coordinates, provided with a specified coordinate reference system (CRS). If the input CRS is EPSG:29903, the Northern Ireland data products are used; for all others, the UK-wide products are applied. When the input CRS differs from EPSG:27700, coordinates are automatically projected to British National Grid before

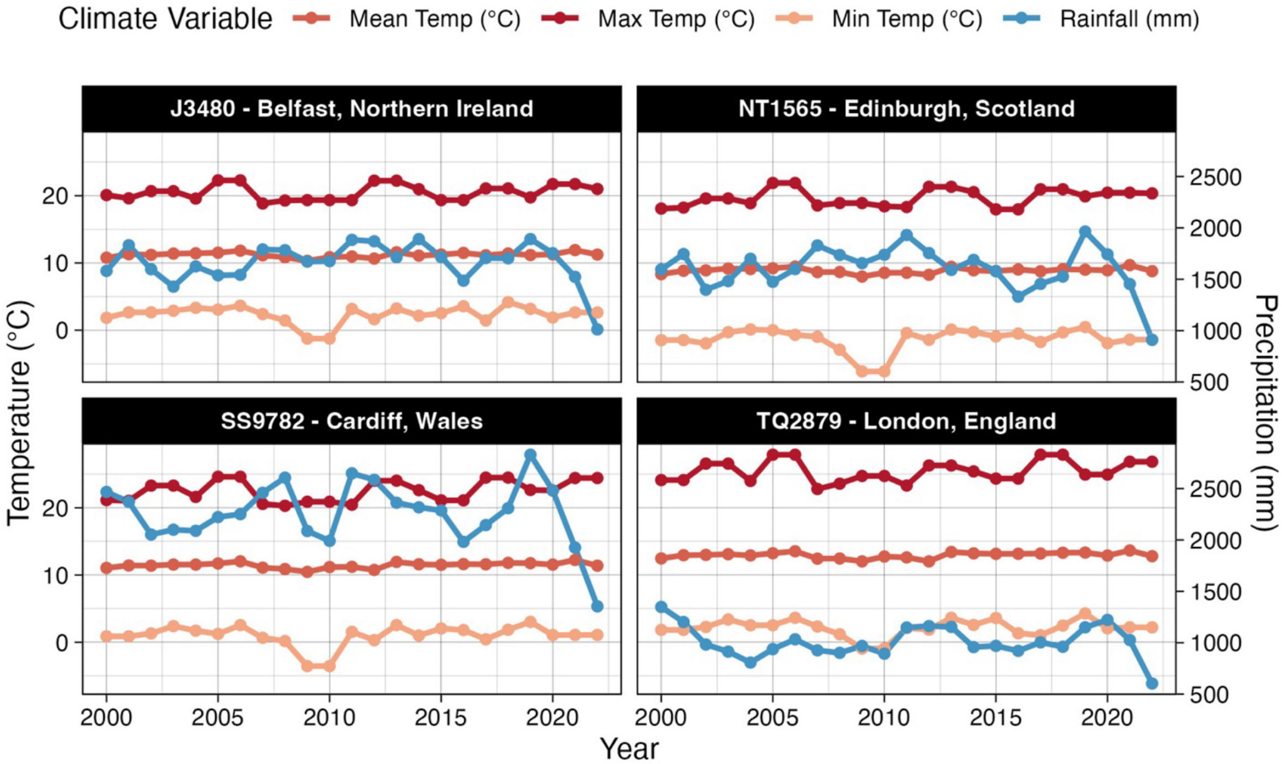

Figure 3. The change in annual climate variables from 2000–2023 using custom annual windows from June–May. Each grid square was chosen from monads used by the National Plant Monitoring Scheme (Pescott et al. 2019) near to the four major cities of the nations of the UK.

extraction. This automated handling of projections enables users to obtain environmental data across the UK regardless of coordinate format, reducing the technical burden for non-expert users.

Case study: modelling occupancy of a native UK wildflower – *Hyacinthoides non-scripta*

Occupancy modelling provides a robust framework for predicting species distributions using detection–nondetection data while accounting for imperfect detection – an inherent challenge in biodiversity survey datasets (MacKenzie et al. 2002, MacKenzie 2005). In this case study, we model the occupancy probability of *Hyacinthoides non-scripta*, a species potentially vulnerable to climate change, under projected climate scenarios. Climate change scenarios were sourced from the UK Climate Projections 2018 (UKCP18), which provides predictions of temperature and precipitation change up to the end of the 21st century (Met Office 2018). Species observations were drawn from the National Plant Monitoring Scheme (NPMS), a UK-wide, monad-based citizen science initiative, that has run from 2015 (Pescott et al. 2019). Volunteers conduct repeated plant surveys twice a year within assigned 1-km grid squares. NPMS data were converted into detection–nondetection records for *H. non-scripta*, and environmental covariates on soil properties, land cover and climate, were extracted using the *extract_all_values()* function from the 'ukbioprepr' package. Grid references, recorded in both Irish Grid and British National Grid formats, were handled automatically during extraction.

The bluebell *Hyacinthoides non-scripta* is an iconic wildflower of the understorey of ancient forests across north-west Europe, with its highest abundance in the UK (Ingrouille 1995, Verheyen et al. 2003, Kohn et al. 2009, Sims et al. 2014, Sanczuk et al. 2022). Native populations have historically been threatened by land-use change, agricultural intensification, and extensive harvesting for trade (Kohn et al. 2009). In 1998, the species was granted legal protection under the Wildlife and Countryside Act 1981 c. 69, prohibiting collection from the wild for domestic or commercial purposes, with substantial fines for non-compliance. In recent times, *H. non-scripta* has faced additional threats from the non-native *Hyacinthoides hispanica* (Kohn et al. 2009), as well as climate change (Sims et al. 2014). The species is a slow coloniser, with seed germination sensitive to seasonal temperature shifts. Additionally, reproductive success depends on successful insect pollination, an ecological interaction increasingly disrupted by climate-related changes to insect abundance (Kohn et al. 2009, IPBES 2019, Sanczuk et al. 2022, Vangansbeke et al. 2022, Mattana et al. 2023). These changes, alongside broader habitat and climatic pressures, could affect seed production and local occupancy persistance.

Table 3. The covariates used in the final occupancy model. All soil properties are extracted at a depth interval of 5–15 cm. Both annual termperature variables were generated from June–May given the impact of seasonal temperatures before flowering (Vangansbeke et al. 2022, Mattana et al. 2023).

Covariate type	Covariate name	Description
Soil property	ocd_D5 to 15 cm	Organic carbon density kg m^{-3}
	phh2o_D5 to 15 cm	pH
	wv0010_D5 to 15 cm	Volumetric water content at −10 kPa (10^{-2} cm^3 cm^{-3}) × 10
Land cover	blw	Broad-leaved woodland (% cover)
	cw	Coniferous woodland (% cover)
	urb	Urban (% cover)
	sub	Suburban (% cover)
Climate	annual_tasmin	Annual min. temperature (°C)
	annual_tasmax	Annual max. temperature (°C)
	winter_rain	Winter precipiation (mm)
	summer_rain	Summer precipitation (mm)

As climate change continues to drive local extinctions and shifts in species distributons (Pereira et al. 2010, Román-Palacios and Wiens 2020), the ability to model and map biodiversity change is essential for informing adaptive conservation management and policy (Peterman et al. 2013, Tulloch et al. 2020, Johnson et al. 2024).

We applied a hierarchical Bayesian occupancy modelling approach using the *stPG0cc()* function from the 'spOccupancy' package, which allows for the incorporation of spatial autocorrelation and imperfect detection, improving model accuracy and inference (Doser et al. 2022). Several models were fitted using different covariate combinations, including a null model, and evaluated using posterior predictive checks. Final model selection was based on five-fold cross validation and comparison using widely applicable information criteria (Watanabe 2010, Hooten and Hobbs 2015, Doser et al. 2022). Additional information on model selection and validation is available in the Supporting information. Table 3 lists the covariates included. Model selection was non-exhaustive therefore results should be interpreted with caution.

A baseline occupancy prediction was generated across the UK using the *predict()* function in 'spOccupancy' (Doser et al. 2022) (Fig. 4). Environmental rasters were downloaded via the *fetch_*()* functions in 'ukbioprepr'. Land cover data was taken from 2023, and baseline climate conditions were determined using means from 2000 to 2023.

To generate climate-based occupancy predictions, climate covariates were adjusted based on projected changes from the UK Climate Projections 2018 (UKCP18; Met Office 2018), which are aligned with Representative Concentration Pathways (RCPs) used in global climate modelling (Moss et al. 2010, Van Vuuren et al. 2011). Three scenarios, broadly corresponding to low (RCP2.6), medium (RCP4.5), and high (RCP8.5) greenhouse gas emission trajectories, were used to represent potential future conditions by 2100. These scenario data were not derived from gridded climate projections but created through uniform adjustments to existing covariates, intended solely to demonstrate potential applications of 'ukbioprepr' data. They are therefore not included as part of the package. Specific adjustments to temperature and precipitation variables under each scenario are detailed in Table 4.

Using the *predict()* function from the 'spOccupancy' package (Doser et al. 2022), the change in occupancy probability from the baseline (Fig. 4) was generated for *Hyacinthoides non-scripta* under all combinations of climate scenarios for every UK grid cell with available covariate data (Fig. 5). To assess habitat availability, we calculated the total weighted suitable habitat area, defined as the sum of predicted occupancy probabilities multiplied by the area of each grid cell. These estimates, in square kilometres (km^2), allow us to quantify how much habitat is climatically suitable under each different scenario. Results are reported with 95% credible intervals to express uncertainty (Fig. 6).

Baseline occupancy probability for *Hyacinthoides non-scripta* varied across the UK, reflecting the species' association with ancient woodlands. Higher occupancy was predicted in well-known woodland areas such as Epping Forest (Greater London), the Wye Valley and South Downs (Fig. 4). Under all future climate scenarios, the change in occupancy probability was largely negative, with losses over 7% in some areas under more severe scenarios (Fig. 5). A corresponding decline was observed in the total weighted area of suitable habitat, with reductions intensifying under higher emissions scenarios; for example, a 5.6% decrease under the low-emissions RCP2.6 and a 12.4% decrease under RCP8.5 (Fig. 6). Notably, the 95% credible intervals also widened with scenario severity, indicating increasing uncertainty about future outcomes under stronger climate shifts.

These results build upon and extend previous projections for *H. non-scripta*, such as those presented by Natural England and RSPB (2019), which modelled species–climate relationships using only climatic variables at a coarser 10 km resolution. This study projected range contractions in southern and central England under a 2°C warming scenario and increases in central Scotland. In contrast, our model incorporates finer scale (1 km) environmental variables generated from 'ukbioprepr', accounting for not only climate but soil properties and land cover – factors known to influence *H. non-scripta* habitat suitability (Kohn et al. 2009, Sims et al. 2014, Vangansbeke et al. 2022). This enables more spatially detailed and ecologically informed predictions. Our results reveal more nuanced patterns of

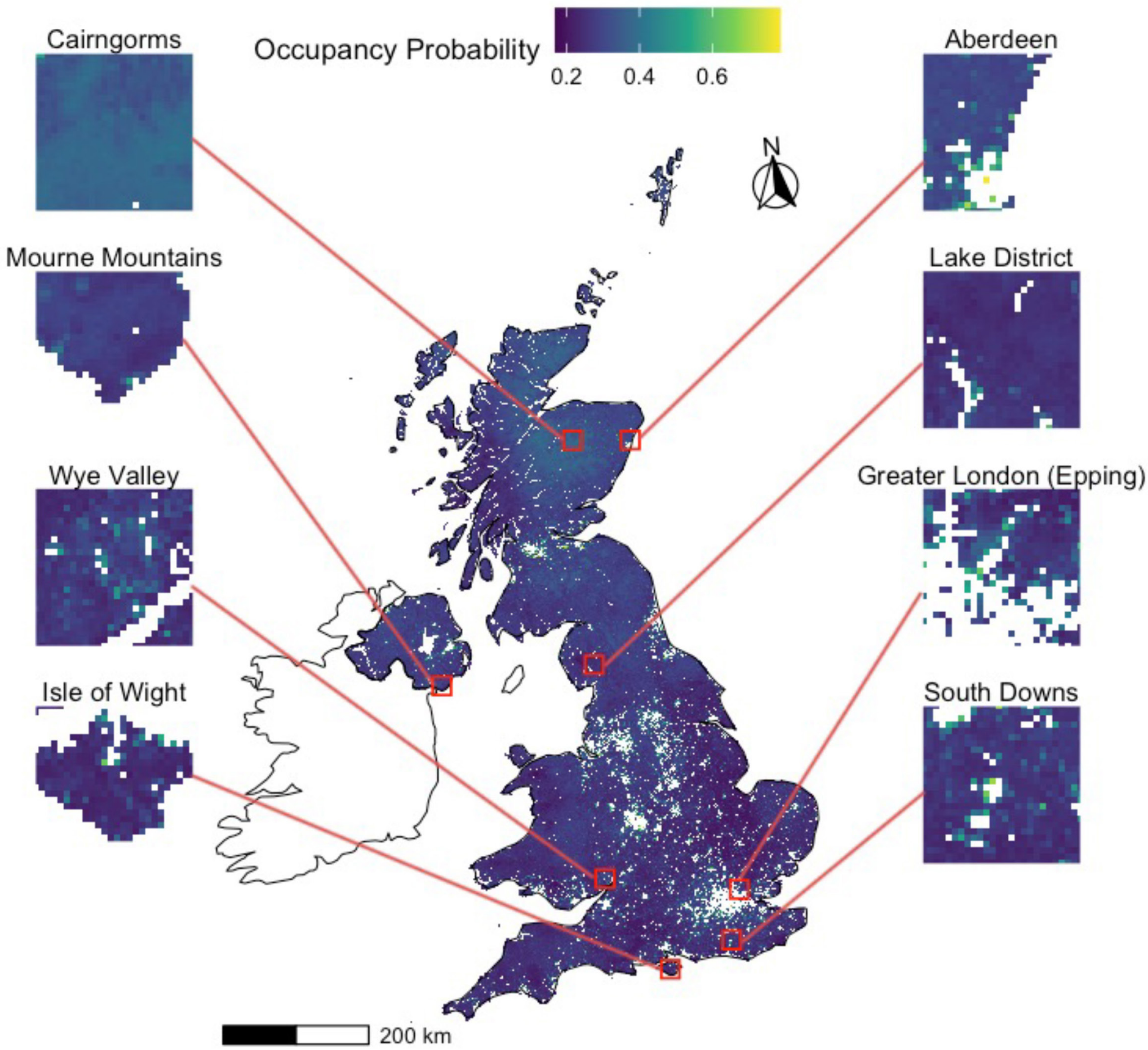

Figure 4. The occupancy probability for *H. non-scripta* in the UK at a 1 km resolution. Eight different locations have been included as zoomed 25 × 25 km squares to highlight the difference in occupancy across the landscape. Areas known for ancient woodlands (Epping Forest, Greater London, South Downs and Wye Valley) were specifically highlighted to show areas of higher occupancy probability. White squares denote areas with missing environmental covariates, and therefore probability predictions could not be produced.

change across nine combinations of temperature and precipitation scenarios. While some areas show relatively stable occupancy probability under low-emission scenarios, occupancy declined more markedly with increasing scenario severity. Areas previously projected as potential refugia in earlier studies, for example in the Scottish Highlands (Natural England and RSPB 2019), exhibited declines from their baseline occupancy in our models under some conditions. These results suggest that *H. non-scripta* may be vulnerable to a wider range of regional climatic shifts than

Table 4. The changes made to climate covariates based on projections from UKCP18, with temperatures increasing and winters becoming drier, and summers becoming wetter.

Relative concentration pathway (RCP)	Scenario	Annual temperature change (°C)	Winter precipitation change (%)	Summer precipitation change (%)
RCP2.6	Best	+1.4	+8	−11
RCP4.5	Mid	+2.0	+13	−20
RCP8.5	Worst	+3.2	+18	−14

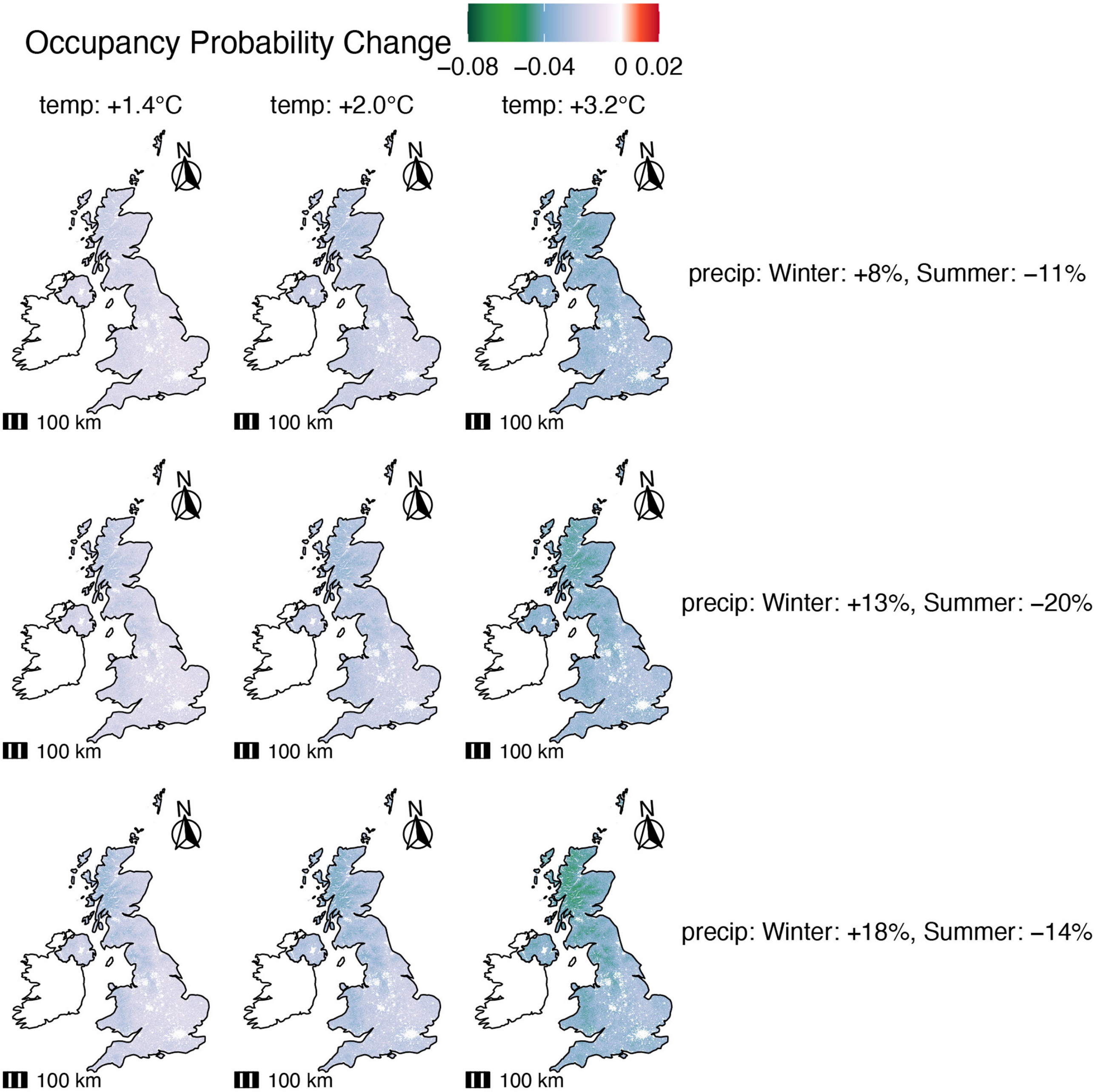

Figure 5. The change from the baseline predicted occupancy probability of *H. non-scripta* under different climate scenarios of temperature (top) and precipitation (right). Light grey cells indicate areas with missing environmental covariates, and therefore probability predictions could not be produced.

previously recognised, highlighting the need for context-specific conservation interventions.

The vulnerability of slow-colonising species such as *H. non-scripta* to climate change underscores the importance of integrating predicitve modelling into conservation planning. Such projections are vital to inform timely interventions and policy measures to mitigate and reverse biodiversity loss, and safeguard culturally and ecologically important, native species.

This case study also illustrates how 'ukbioprepr' can facilitate reproducible modelling using typical UK biodiversity data formats. The package provides accessible functions for downloading and extracting environmental covariates – including soil properties, climate and land cover – tailored for British and Irish grid references. Through streamlined functions, users can seamlessly generate environmental data suitable for spatial analysis and biodiversity forecasting. As demonstrated, the package facilitates users to create

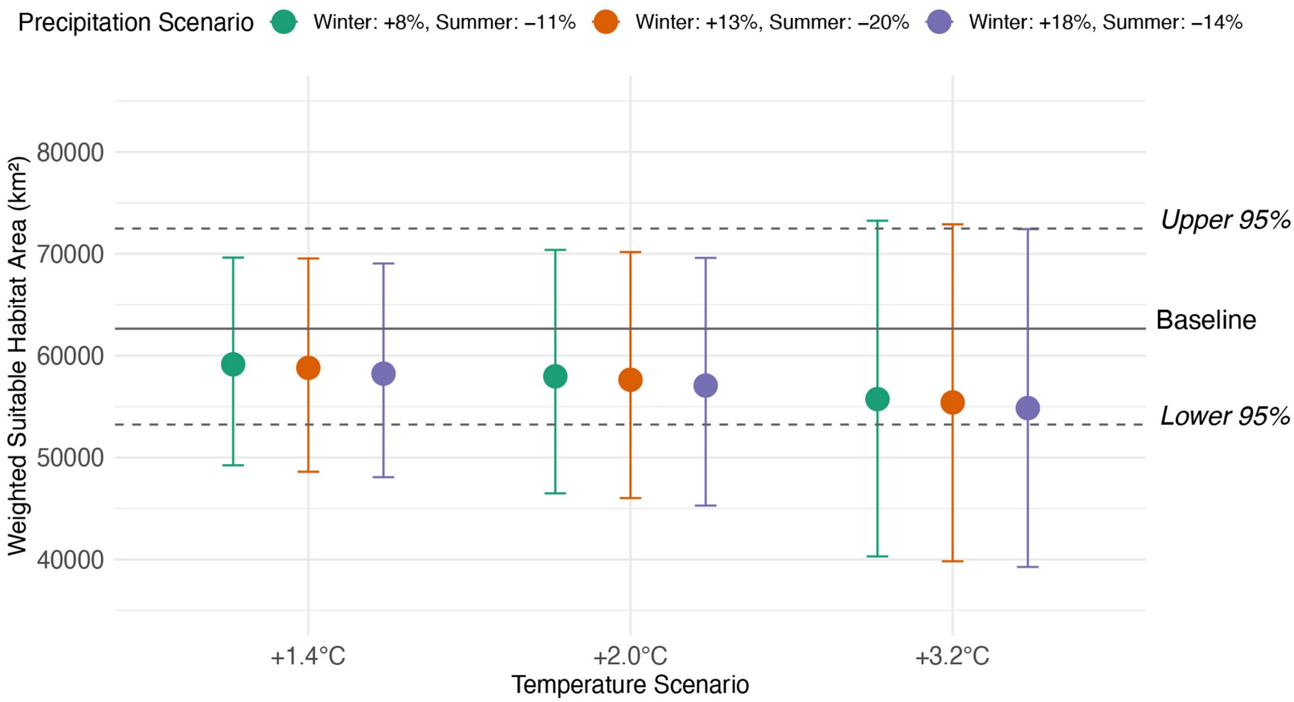

Figure 6. The predicted weighted suitable habitat area (km^2) for *H. non-scripta* across the UK under different temperature and precipiation scenarios along with predicted 95% credible intervals. The habitat area for the baseline prediction is also included.

projections of species under future climate scenarios, supporting wider applications in ecological research, conservation planning and policy implementing.

Conclusion

Biodiversity modelling plays a crucial role in understanding species distributions, forecasting responses to environmental change, and informing effective conservation strategies. However, the process of sourcing, harmonising, and preparing relevant environmental data can be labour intensive, technically challenging, and a barrier to collaboration, particularly to non-experts. 'ukbioprepr' addresses this by providing streamlined functions to seamlessly download and extract soil, land cover and climate covariates tailored for biodiversity data in the UK. In this paper, we have detailed data products, explained how they were processed, and demonstrated the functionality of the package, including any limitations of the data. Through a case study on *Hyacinthoides non-scripta*, we demonstrated how 'ukbioprepr' can be used with typical UK biodiversity datasets to model species occupancy under future climate scenarios. The package offers an accessible tool for researchers, conservation practitioners and decision-makers, supporting a wide range of applications from ecological forecasting to management and policy development. As new, nationally consistent environmental datasets become available, the package framework can be extended easily in the future to incorporate additional covariates.

Whilst 'ukbioprepr' is designed specifically for the UK, its structure and reproducible workflows provide a transferable template for other geographic contexts. This framework could be adapted by substituting equivalent national or regional environmental datasets and building similar data-processing and access pipelines. The underlying principles of harmonising, standardising, and improving access to environmental data products whilst documenting methods transparently, are universally applicable. As countries work towards ambitious targets set out in the Kunming-Montreal Global Biodiversity Framework (CBD 2022), frameworks like 'ukbioprepr' should be considered by other nations to expedite and simplify biodiversity research and action.

Software note:To cite 'ukbioprepr' or acknowledge its use, cite this Software Note as follows, substituting the version of the application that you used for 'version 1.0': Rush, C. R. et al. 2026. 'ukbioprepr': an R package to support reproducible preparation of environmental data for biodiversity modelling in the UK. – Ecography 2026: e08413 (ver. 1.0).

Acknowledgements – We thank the original data providers, the Met Office, UK Centre for Ecology and Hydrology (UKCEH), and SoilGrids, for making their environmental datasets openly available.
Funding – This work was jointly funded by the Natural Environment Research Council and Biodiversify Ltd through the ARIES Doctoral Training Partnership (grant no. NE/S007334/1).
Conflict of interest – The authors declare no conflict of interest.

Author contributions

Charlotte Rose Rush: Conceptualization (lead); Data curation (lead); Methodology (lead); Project administration (lead); Software (lead); Validation (lead); Writing – original draft (lead); Writing – review and editing (lead). **Joseph Cooper**: Supervision (supporting); Writing – review and editing (supporting). **Cecilia Larrosa**: Supervision (supporting). **Martin Wilkes**: Conceptualization (equal); Data curation (equal); Methodology (equal); Software (supporting); Supervision (lead); Validation (supporting); Writing – review and editing (equal).

Transparent peer review

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/ecog.08413.

Data availability statement

The package can be accessed and downloaded through GitHub: https://github.com/crrush7/ukbioprepr.

The following Zenodo repositories contain the data products accessed in the R package – climate: https://doi.org/10.5281/zenodo.14913772, https://doi.org/10.5281/zenodo.14841658; soil properties: https://doi.org/10.5281/zenodo.14973735; and land cover: https://doi.org/10.5281/zenodo.14849882.

The code used to produce the data products and for the case study is available in Zenodo: https://doi.org/10.5281/zenodo.17609019 (Rush et al. 2026).

The biodiversity data used for the case study is accessible at: National Plant Monitoring Scheme (2023). Habitat samples from the National Plant Monitoring Scheme, 2015–2022. NERC EDS Environmental Information Data Centre. https://doi.org/10.5285/a12b5c70-c0e2-48cf-b0b7-3e95791feca9.

Supporting information

The Supporting information associated with this article is available with the online version.

References

Ariza-Salamanca, A. J., Navarro-Cerrillo, R. M., Quero-Pérez, J. L., Gallardo-Armas, B., Crozier, J., Stirling, C., De Sousa, K. and González-Moreno, P. 2023. Vulnerability of cocoa-based agroforestry systems to climate change in West Africa. – Sci. Rep. 13: 10033.

Arsevska, E., Hengl, T., Singleton, D. A., Noble, P. M., Caminade, C., Eneanya, O. A., Jones, P. H., Medlock, J. M., Hansford, K. M., Bonannella, C. and Radford, A. D. 2024. Risk factors for tick attachment in companion animals in Great Britain: a spatiotemporal analysis covering 2014–2021. – Parasit. Vectors 17: 29.

Assis, J., Tyberghein, L., Bosh, S., Verbruggen, H., Serrão, E. A. and De Clerck, O. 2017. Bio-ORACLE v2.0: extending marine data layers for bioclimatic modelling. – Global Ecol. Biogeogr. 26: 1118–1129.

Bosch, S. and Fernandez, S. 2022. sdmpredictors: species distribution modelling predictor datasets, ver. 0.2.14. – R package, https://CRAN.R-project.org/package=sdmpredictors.

Boyd, R. J., Stewart, G. B. and Pescott, O. L. 2023. Descriptive inference using large, unrepresentative nonprobability samples: an introduction for ecologists. – Ecology 105: 2.

Burns, F. et al. 2013. State of nature report. – The State of Nature Partnership, 89pp. (CEH Project No.: C04535).

Burns, F. et al. 2018. An assessment of the state of nature in the United Kingdom: a review of findings, methods and impact. – Ecol. Indic. 94: 226–236.

Chakraborty, D., Dobor, L., Zolles, A., Hlásny, T. and Schueler, S. 2020. High-resolution gridded climate data for Europe based on bias-corrected EURO-CORDEX: the ECLIPS dataset. – Geosci. Data J. 8: 121–131.

Convention on Biological Diversity 2022. Kunming-Montreal Global Biodiversity Framework (Decision 15/4). CBD/COP/15/L.25, www.cbd.int/doc/decisions/cop-15/cop-15-dec-04-en.pdf.

Copernicus Climate Change Service (C3S) 2017. ERA5: fifth generation of ECMWF atmospheric reanalyses of the global climate. – https://cds.climate.copernicus.eu/cdsapp#!/home.

Copernicus Climate Change Service (C3S) 2024. ERA5-land post-processed daily-statistics from 1950 to present. – https://doi.org/10.24381/cds.e9c9c792.

Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M. and Jones, P. D. 2018. An ensemble version of the E-OBS temperature and precipitation datasets. – JGR Atmospheres 123: 9391–9409.

Croft, S., Ward, A. I., Aegerter, J. N. and Smith, G. C. 2019. Modeling current and potential distributions of mammal species using presence-only data: a case study on British deer. – Ecol. Evol. 9: 8724–8735.

Doser, J. W., Finley, A. O., Kéry, M. and Zipkin, E. F. 2022. spOccupancy: an R package for single-species, multi-species, and integrated spatial occupancy models. – Methods Ecol. Evol. 13: 1670–1678.

Environment Act 2021, c. 30 (UK). – www.legislation.gov.uk/ukpga/2021/30/contents.

Fick, S. E. and Hijmans, R. J. 2017. WorldClim 2: new 1 km spatial resolution climate surfaces for global land areas. – Int. J. Climatol. 37: 4302–4315.

Gregory, R. D., Baillie, S. R. and Bashford, R. I. 2004. Monitoring breeding birds in the United Kingdom. – In: Anselin, A. (ed.), Bird No. 1995: Proceedings of the International Conference and 13th Meeting of the European Bird Census Council, Pärnu, Estonia. – Bird Census News 13, pp. 101–112.

Hijmans, R. J. 2024. terra: spatial data analysis. – R package ver. 1.8-0, https://CRAN.R-project.org/package=terra.

Hijmans, R. J., Ghosh, A. and Mandel, A. 2022. geodata: download geographic data. – R package ver. 0.4-11, https://CRAN.R-project.org/package=geodata.

Hooten, M. B. and Hobbs, N. T. 2015. A guide to Bayesian model selection for ecologists. – Ecol. Monogr. 85: 3–28.

Ingrouille, M. 1995. Historical ecology of the British flora. – Chapman/ Hall.

IPBES 2019. Global assessment report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. In: Díaz, S., Settele, J., Brondízio, E. and Ngo, H. T. (eds). – IPBES Secretariat.

Johnson, T. F., Beckerman, A. P., Childs, D. Z., Webb, T. J., Evans, K. L., Griffiths, C. A., Capdevila, P., Clements, C. F., Besson, M., Gregory, R. D., Thomas, G. H., Delmas, E. and Freckleton, R. P. 2024. Revealing uncertainty in the status of biodiversity change. – Nature 628: 788–794.

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P. and Kessler, M. 2017. Climatologies at high resolution for the Earth's land surface areas. – Sci. Data 4: 170122.

Kennedy, J. 2025. igr: Irish grid reference utilities. – R package ver. 1.0.0, https://CRAN.R-project.org/package=igr.

Klinges, D. H., Duffy, J. P., Kearney, M. R. and Maclean, I. M. D. 2022. mcera5: driving microclimate models with ERA5 global gridded climate data. – Methods Ecol. Evol. 13: 1402–1411.

Kohn, D. D., Hulme, P. E., Hollingsworth, P. M. and Butler, A. 2009. Are native bluebells (*Hyacinthoides non-scripta*) at risk from alien congenerics? Evidence from distributions and co-occurrence in Scotland. – Biol. Conserv. 142: 61–74.

Lawley, R., Emmett, B. A. and Robinson, D. A. 2014. Soil observatory lets researchers dig deep. – Nature 509: 427.

Leclère, D. et al. 2020. Bending the curve of terrestrial biodiversity needs an integrated strategy. – Nature 585: 551–556.

Ma, R., Zhu, X., Tian, Z., Qu, L., He, Y. and Liang, Y. 2022. Spatial distribution and scale-specific controls of soil water-stable aggregates in southeastern China. – J. Clean. Prod. 369: 133305.

MacKenzie, D. I. 2005. What are the issues with presence–absence data for wildlife managers? – J. Wildl. Manage. 69: 849–860.

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A. and Langtimm, C. A. 2002. Estimating site occupancy rates when detection probabilities are less than one. – Ecology 83: 2248–2255.

Madenoglu, S., Atalay, F. and Erpul, G. 2020. Uncertainty assessment of soil erodibility by direct sequential Gaussian simulation (DSIM) in semiarid land uses. – Soil Till. Res. 204: 104731.

Mattana, E., Chapman, T., Miles, S., Ulian, T. and Carta, A. 2023. Regeneration from seeds in a temperate native flora: a climate-smart and natural-capital-driven germination risk modelling approach. – Plants People Planet 5: 908–922.

Met Office 2018. UK Climate Projections: UKCP18 overview. – Met Office, www.metoffice.gov.uk/research/approach/collaboration/ukcp/summaries/index.

Met Office, Hollis, D., McCarthy, M., Kendon, M., Legg, T. and Simpson, I. 2018. HadUK-Grid gridded and regional average climate observations for the UK. – Centre for Environmental Data Analysis, http://catalogue.ceda.ac.uk/uuid/4dc8450d889a491ebb20e724debe2dfb.

Morton, D., Rowland, C., Wood, C., Meek, L., Marston, C., Smith, G., Wadsworth, R. and Simpson, I. C. 2011. Final report for LCM2007: the new UK land cover map. – Countryside Survey Technical Report No. 11/07. NERC/Centre for Ecology and Hydrology, 112pp. (CEH Project No.: C03259).

Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., Van Vuuren, D. P., Carter, T. R., Emori, S., Kainuma, M., Kram, T., Meehl, G. A., Mitchell, J. F. B., Nakicenovic, N., Riahi, K., Smith, S. J., Stouffer, R. J., Thomson, A. M., Weyant, J. P. and Wilbanks, T. J. 2010. The next generation of scenarios for climate change research and assessment. – Nature 463: 747–756.

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Du Sert, N. P., Simonsohn, U., Wagenmakers, E. J., Ware, J. J. and Ioannidis, J. P. A. 2017. A manifesto for reproducible science. – Nat. Hum. Behav. 1: 0021.

Muñoz Sabater, J., Comyn-Platt, E., Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thépaut, J.-N., Cagnazo, C. and Cucchi, M. 2024. ERA5-Land post-processed daily statistics from 1950 to present. – Copernicus Climate Change Service (C3S). Climate Data Store (CDS), https://doi.org/10.24381/cds.e9c9c792.

Natural England and RSPB 2019. Climate change adaptation manual: evidence to support nature conservation in a changing climate, 2nd edn. – Natural England.

Pereira, H. M. et al. 2010. Scenarios for global biodiversity in the 21st century. – Science 330: 1496–1501.

Pescott, O. L., Walker, K. J., Harris, F., New, H., Cheffings, C. M., Newton, N., Jitlal, M., Redhead, J., Smart, S. M. and Roy, D. B. 2019. The design, launch and assessment of a new volunteer-based plant monitoring scheme for the United Kingdom. – PLoS One 14: e0215891.

Peterman, W. E., Crawford, J. A. and Kuhns, A. R. 2013. Using species distribution and occupancy modeling to guide survey efforts and assess species status. – J. Nat. Conserv. 21: 114–121.

Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E. and Rossiter, D. 2021. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. – Soil 7: 217–240.

Qu, L., Lu, H., Tian, Z., Schoorl, J. M., Huang, B., Liang, Y., Qiu, D. and Liang, Y. 2024. Spatial prediction of soil sand content at various sampling density based on geostatistical and machine learning algorithms in plain areas. – CATENA 234: 107572.

Rapacciuolo, G. 2018. Strengthening the contribution of macroecological models to conservation practice. – Global Ecol. Biogeogr. 28: 54–60.

Román-Palacios, C. and Wiens, J. J. 2020. Recent responses to climate change reveal the drivers of species extinction and survival. – Proc. Natl Acad. Sci. USA 117: 4211–4217.

Rowland, C. S., Morton, R. D., Carrasco, L., McShane, G., O'Neil, A. W. and Wood, C. M. 2017a. Land cover map 2015 (1 km percentage target class, GB). – NERC Environmental Information Data Centre.

Rowland, C. S., Morton, R. D., Carrasco, L., McShane, G., O'Neil, A. W. and Wood, C. M. 2017b. Land cover map 2015 (1 km percentage target class, N. Ireland). – NERC Environmental Information Data Centre.

Rush, C. R., Cooper, J., Larrosa, C. and Wilkes, M. 2026. Code from: 'ukbioprepr': an R package to support reproducible preparation of environmental data for biodiversity modelling in the UK. – Zenodo, https://doi.org/10.5281/zenodo.17609019.

Sanczuk, P., De Lombaerde, E., Haesen, S., Van Meerbeek, K., Van Der Veken, B., Hermy, M., Verheyen, K., Vangansbeke, P. and De Frenne, P. 2022. Species distribution models and a 60-year-old transplant experiment reveal inhibited forest plant range shifts under climate change. – J. Biogeogr. 49: 537–550.

Sillero, N., Campos, J. C., Arenas-Castro, S. and Barbosa, A. M. 2023. A curated list of R packages for ecological niche modelling. – Ecol. Modell. 476: 110242.

Sims, N. K., John, E. A. and Stewart, A. J. A. 2014. Short-term response and recovery of bluebells (*Hyacinthoides non-scripta*) after rooting by wild boar (*Sus scrofa*). – Plant Ecol. 215: 1409–1416.

Siqueira, R. G., Moquedace, C. M., Fernandes-Filho, E. I., Schaefer, C. E. G. R., Francelino, M. R., Sacramento, I. F. and Michel, R. F. M. 2024. Modelling and prediction of major soil chemical properties with Random Forest: machine learning as tool to understand soil–environment relationships in Antarctica. – CATENA 235: 107677.

Tian, Z., Liu, F., Liang, Y. and Zhu, X. 2022. Mapping soil erodibility in southeast China at 250 m resolution: using environmental variables and random forest regression with limited samples. – Int. Soil Water Conserv. Res. 10: 62–74.

Tulloch, A. I. T., Hagger, V. and Greenville, A. C. 2020. Ecological forecasts to inform near-term management of threats to biodiversity. – Global Change Biol. 26: 5816–5828.
Turek, M. E., Poggio, L., Batjes, N. H., Armindo, R. A., de Jong van Lier, Q., de Sousa, L. M. and Heuvelink, G. B. M. 2023. Global mapping of volumetric water retention at 100, 330 and 15,000 cm suction using the WoSIS database. – Int. Soil Water Conserv. Res. 11: 225–239.
UKCEH 2017. Land Cover Map 2015 Dataset documentation (ver. 1.2, 22 May 2017) – www.ceh.ac.uk/sites/default/files/LCM2015_Dataset_Documentation_22May2017.pdf.
UKCEH 2025. United Kingdom Centre for Ecology and Hydrology Land Cover Map series. – www.ceh.ac.uk/data/ukceh-land-cover-maps.
UKSO 2025. UK Soil Observatory – www.ukso.org.
Van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J. F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S. J. and Rose, S. K. 2011. The representative concentration pathways: an overview. – Clim. Change 109: 5–31.
Vangansbeke, P., Sanczuk, P., Govaert, S., De Lombaerde, E. and De Frenne, P. 2022. Negative effects of winter and spring warming on the regeneration of forest spring geophytes. – Plant Biol. 24: 950–959.
Verheyen, K., Bossuyt, B., Honnay, O. and Hermy, M. 2003. Herbaceous plant community structure of ancient and recent forests in two contrasting forest types. – Basic Appl. Ecol. 4: 537–546.
Vitolo, C., Fry, M., Buytaert, W., Spencer, M. and Gauster, T. 2018. rnrfa: an R package to retrieve, filter and visualize data from the UK National River Flow Archive. – R J. 8: 102–116.
Wang, R., Wan, S., Chen, W., Qin, X., Zhang, G. and Wang, L. 2024. A novel finer soil strength mapping framework based on machine learning and remote sensing images. – Comput. Geosci. 182: 105479.
Watanabe, S. 2010. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. – J. Mach. Learn. Res. 11: 3571–3594.
Wildlife and Countryside Act 1981, c. 69 (UK). – www.legislation.gov.uk/ukpga/1981/69.
Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., Lesiv, M., Herold, M., Tsendbazar, N. E., Xu, P., Ramoino, F. and Arino, O. 2022. ESA WorldCover 10 m 2021 v200 – Zenodo, https://doi.org/10.5281/zenodo.7254221.
Zurell, D. et al. 2020. A standard protocol for reporting species distribution models. – Ecography 43: 1261–1277.