Project Report

Author-formatted document posted on 19/01/2024

Published in a RIO article collection by decision of the collection editors.

DOI: https://doi.org/10.3897/arphapreprints.e119131

D5.5 Short-term ecological forecasts in support of the Bioeconomy Strategy and EU citizens

Ana Ceia-Hasse, Judy Shamoun-Baranes, Néstor Fernández, Henrique Pereira, César Capinha

ARPHA Preprints Author-formatted document posted on 19/01/2024. DOI: https://doi.org/10.3897/arphapreprints.e119131

D5.5 Short-term ecological forecasts in support of the Bioeconomy Strategy and EU citizens

10/01/2024

Lead beneficiary:

Institute of Geography and Spatial Planning (IGOT), University of Lisbon

This project receives funding from the European Commission's Horizon 2020 research and innovation programme, under Grant Agreement n.101003553

Prepared under contract from the European Commission

Grant agreement No. 101003553
EU Horizon 2020 Coordination and Support Action

Project acronym: **EuropaBON**
Project full title: **EUROPA BIODIVERSITY OBSERVATION NETWORK: INTEGRATING DATA STREAMS TO SUPPORT POLICY**

Start of the project: 01.12.2020
Duration: 36 months
Project coordinator: Prof. Henrique Pereira
Martin-Luther Universitaet Halle-Wittenberg (MLU)
www.europabon.org

Type: Coordination and Support Action
Call: The Sc5-33-2020 Call: "Monitoring ecosystems through innovation and technology"

Project ref. no.	**101003553**
Project title	**EUROPA BIODIVERSITY OBSERVATION NETWORK: INTEGRATING DATA STREAMS TO SUPPORT POLICY**

Deliverable title	**Short-term ecological forecasts in support of the Bioeconomy strategy and EU citizens**
Deliverable number	**D5.5**
Contractual date of delivery	**31.01.2024**
Actual date of delivery	**10.01.2024**
Type of deliverable	**Report**
Dissemination level	**Public**
Work package number	**WP5**
Institution leading work package	**Martin Luther University Halle-Wittenberg**
Task number	**T5.5**
Institution leading task	**Institute of Geography and Spatial Planning (IGOT), University of Lisbon.**
Author(s)	**Ana Ceia-Hasse, Judy Shamoun-Baranes, Néstor Fernández, Henrique M. Pereira, César Capinha**
EC project officer	**Laura Palomo-Rios**

Deliverable description	This deliverable describes the workflows developed to produce short-term forecasts of Essential Biodiversity Variables that can support the Bioeconomy Strategy and EU citizens. Two forecasting workflows are described and showcased. The first is applicable to diverse ecological and biological phenomena and is demonstrated here through the fruiting of a wild edible mushroom of recreational interest and the adult life-stage of an invasive species of concern to EU agriculture. The

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

	second workflow focuses on forecasting the aerial biomass of migrating birds.
Keywords	bioeconomy, bird aerial biomass, citizen science, computational workflows, early warning, edible wild mushrooms, invasive pests, iterative modelling, short-term forecasts, weather radar

ARPHA Preprints Author-formatted document posted on 19/01/2024. DOI: https://doi.org/10.3897/arphapreprints.e119131

Contents

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

Executive summary

A relevant number of ecological questions raised by policymakers, managers, and citizens often pertain to the short-term future (e.g., the coming days or weeks). In this sense, short-term ecological and biological forecasts can make substantial and practical contributions to achieving policy objectives and benefit society broadly. Specifically, short-term forecasts of Essential Biodiversity Variables (EBVs) and Essential Ecosystem Service Variables (EESVs) can support decision-making by stakeholders from multiple sectors, enabling to anticipate ecological transformations and support proactive, informed decisions that promote conservation, economic activities, and human well-being.

The aim of this task was to demonstrate how a European Biodiversity Observation Network can support the generation of short-term spatial forecasts of ecological and biological phenomena relevant to the Bioeconomy Strategy and to EU citizens at large. Our specific objectives included showcasing 1) a computational workflow that enables the production of days-ahead forecasts for distinct ecological or biological phenomena and 2) a specialized computational workflow for days-ahead forecasts of bird aerial biomass. The first, ('generic') workflow, is exemplified using two case studies: i) forecasting the fruiting of a wild mushroom of commercial and recreational relevance, and ii) forecasting the life stage of relevance for surveillance of an invasive pest species important for agriculture. These case studies aim to demonstrate specific, tangible contributions that short-term ecological forecasting can make towards the sustainable use of bio-based economy sectors, ecosystem protection, and anticipation of ecological risks. Beyond aligning with the EU Bioeconomy Strategy, our three forecasting targets also offer relevant contributions to a wider range of EU strategies and policies.

We actively involved stakeholders in defining the end-products and in the development of computational modelling approaches of the workflows. This process entailed two distinct approaches. For the generic forecasting workflow, we engaged in a participatory process from the project's start, focusing on stakeholders involved in mushroom foraging and experts in mycology and ecological modelling. For the bird aerial biomass forecasting workflow, we built upon substantial developments that predated the project, with our engagement primarily drawing on insights and input from earlier initiatives. The two workflows serve complementary purposes in terms of the primary data they use. While the first (generic) workflow is based on the growing body of opportunistic biodiversity observation data, particularly from citizen science initiatives, the second workflow requires highly specialized radar data from weather stations. However, both workflows use predictor data from weather observations and forecasts and employ machine learning algorithms to correlate these data with observed variations in the phenomena being forecasted.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

ARPHA Preprints Author-formatted document posted on 19/01/2024. DOI: https://doi.org/10.3897/arphapreprints.e119131

1. Introduction

1.1. State of the art in short term forecasts in support of EU policy and citizens at large

Short-term ecological forecasting refers to the prediction of ecological phenomena in the near future, typically within a timeframe of less than one year, ranging from a few hours to several months (Tulloch et al. 2020). This high temporal resolution and near-term forecasting horizon differentiates it from many ongoing and past ecological modeling efforts, which were predominantly focused on longer time periods, such as multi-decadal scenario-based projections being used in long-term decision support (Pereira et al. 2010).

Short-term ecological forecasts offer significant and tangible contributions towards accomplishing policies objectives and benefiting society at large. Specifically, a relevant number of ecologically related questions being asked by policymakers, managers, and citizens are frequently about the short-term future (e.g., coming days or weeks) (Dietze 2017). These questions encompass a diverse range of ecological or biological phenomena and relate to many distinct aims. For instance, in the USA, real-time tracking and forecasts have been employed to monitor migratory patterns of bird species, aiding monitoring and mortality mitigation efforts (Van Doren and Horton 2018). Similarly, European initiatives have leveraged short-term forecasts for predicting seasonal blooms of harmful species in freshwater ecosystems (Jackson-Blake et al. 2022), the risk of vector-borne disease outbreaks (Semenza 2015), bird migration patterns (Kranstauber et al. 2022; van Gasteren et al. 2019), or tourism-relevant patterns of tree leaf senescence (Tourismus n.d.). Additionally, short-term forecasting can play a crucial role in mitigating damages, be it from invasive species harming crops (Barker et al. 2020) or in the reduction of accidents, like bird or bat collisions with wind farms, aircraft, and urban structures (Van Doren and Horton 2018, Frick et al. 2012, van Gasteren et al. 2019). Crucially, the large potential of these forecasts for supporting environmental and economically oriented decisions, either from experts and decision makers or citizens in general, remains barely touched with many potentially useful applications yet to be tested and operationalised (Dietze et al. 2018).

Short-term ecological forecasts are also relevant within the scope of Essential Biodiversity Variables (EBVs) (Pereira et al. 2013) and of Essential Ecosystem Service Variables (EESVs) (Balnavera et al. 2022). These variables aim at enabling the improved monitoring, researching, and forecasting of patterns and trends in biodiversity and in ecosystem services (Gonzalez et al. 2023). Relevantly, many of the phenomena represented in proposed EBVs or EESVs can show marked intra-annual variation, spanning diverse classes like species abundances, species distributions or ecosystem functioning, and domains such as freshwater, marine, and terrestrial environments. Therefore, the ability to anticipate intra-annual variation for these variables can considerably increase the insights and applied value they provide. The scope of use of the forecasts necessarily varies according to the objectives of end-users; however, most will benefit from an increased ability to respond promptly to anticipated changes in biodiversity or forthcoming shifts in ecosystem services, such as biomass provisioning, pollination, biological control, and recreation-related dynamics. Ultimately, short-term forecasts of EBVs and EESVs further empower stakeholders from multiple sectors (policy, academia, NGOs, citizen science, businesses) to foresee ecological transformations, thereby facilitating proactive monitoring and ensuring timely intervention.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

Despite their confirmed and latent significance in decision-making and societal benefits, the development and operationalization of short-term ecological forecasts still faces important operational challenges. A primary obstacle concerns the pervasive fragmentation of biodiversity observation data across multiple sources. Such sources include real-time *in-situ* monitoring, citizen science contributions, scientific papers, and technical literature, which often originate from disparate entities using varied data reporting and standardization structures (Castro et al. 2023, Weisshaupt et al. 2021). This fragmentation of data sources and standards presents a significant challenge in identifying, harmonizing, and integrating biodiversity information, which is crucial for creating timely and accurate computational models for short-term forecasting, particularly when aimed at wide geographical areas. Given these challenges, the data integration efforts being developed under the EuropaBON project represents a major step forward towards the increase in capacity of implementing short-term ecological forecasts at the scale of the EU.

1.2. Showcase goals

Given the context described above, our aim was to demonstrate how a European Biodiversity Observation Network can support the production of near-real time short-term spatial forecasts of ecological and biological phenomena to support the Bioeconomy Strategy and EU citizens in general. To this end, our specific objectives were to showcase 1) a computational workflow that allows producing short-term forecasts of a diverse range of biological phenomena and 2) a computational workflow for near real-time forecasting of bird aerial biomass. The first ('generic') forecasting workflow is presented here through two case studies: i) the short-term forecasting of the fruiting of wild mushrooms of commercial or recreational value in the EU and ii) the short-term forecasting of the life stages of invasive pest species. The scopes of application of the forecasts specifically aim to demonstrate how short-term ecological forecasts provided to EU citizens at large enable better management of the EU's biological resources and their use as food, as well as in preventing damages from pest and non-pest species. The two workflows also serve a complementary demonstration purpose in terms of the primary data they use. While the first workflow makes use of the growing body of opportunistic biodiversity observation data, especially those from citizen science initiatives (Bonney 2021), the second workflow requires highly specialized data extracted from weather radars (Shamoun-Baranes et al. 2021).

2. Showcase participatory design

2.1. Stakeholder engagement process

Stakeholder engagement took two distinct approaches. For the generic forecasting workflow, we actively involved stakeholders from the beginning of the project. This engagement focused primarily on stakeholders related to mushroom fructification forecasting, the case study that served as the main conceptual and practical basis for the development of the workflow. On the other hand, for the bird aerial biomass forecast workflow, there were substantial developments before the project began; as a result, our engagement with stakeholders for this workflow primarily relied on insights and input from previous initiatives. Stakeholder engagement process is represented in Fig. 1.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

ARPHA Preprints Author-formatted document posted on 19/01/2024. DOI: https://doi.org/10.3897/arphapreprints.e119131

2.1.1. Methodology of stakeholder engagement for the generic forecasting workflow

Regarding the stakeholder engagement process for the generic forecasting workflow, the main stakeholders were initially identified by the task team and consisted of people or institutions with expertise on mushroom phenology, mushroom foraging, and ecological modelling. These diverse backgrounds were chosen to help identify the most suitable characteristics of the forecasts from the point of view of the end users (e.g., foragers and managers), but also to ensure that these attributes can be realistically implemented, considering the knowledge of the species and modelling experts. The engagement process began by inviting the identified stakeholders to take part in workshops organized by the project. We co-organized two workshops, one in November 2021 at the German Centre for Integrative Biodiversity Research (iDiv) - Halle-Jena-Leipzig headquarters in Leipzig, Germany (online and in-person) and one in April 2023 in Troia (Portugal) (in-person). In these workshops, we focused the discussion around three main topics. The first concerned the identification of the desirable properties of forecasts from the end-user's point of view and included discussing subjects such as the temporal and spatial resolution of forecasts and how to present the forecast products to end-users. The second topic concerned the identification of relevant drivers of mushroom seasonality and how these could be converted into spatial predictors to use in the forecast models. Finally, the third topic was centered around the modelling approaches and the operational assembly of the workflow and included subjects such as the relevant data sources available and discussing pros and cons of distinct modelling choices. In addition to the joint discussions in the workshops, we had follow-up meetings, in person and online, with several of the participants focusing on specific sections of the workflow.

2.1.2. Stakeholder representativeness and gaps for the generic forecasting workflow

The scientific representation of stakeholders in the participatory process was comprehensive, including eight 'core' experts, who assisted in the co-development of the workflow in a consistent and continuous manner and covering the full range of research-related issues involved. On the end-user side, counting the project participants, we received feedback from more than ten individuals with experience in mushroom collecting in various European countries (Portugal, Spain, Poland, the UK, Denmark, Italy, and Germany). Although we expect that this good level of individual participation guarantees the representativeness of the intended end users, we note the absence of representatives from mushroom collecting associations, to which a significant number of invitations were sent, but from which we received no response.

2.1.3. Methodology of stakeholder engagement for monitoring and forecasting aerial biomass of birds

Concerning the workflow of the forecasts of aerial biomass of birds, stakeholder engagement was done through discussions with stakeholders as part of previous national and international initiatives (e.g., user committees) and through discussions for developing future tools with new stakeholders, contacted sometimes directly by stakeholders (e.g., different offices within the ministry of infrastructure, ministry of economic affairs, provincial governments) and through international collaborative projects (e.g. BioDivERsa GloBAM). Short term forecasts of biomass flows had already been developed for military aviation and are in operational use in several countries. Discussions with these users focused on improving biomass flow monitoring techniques (for near real-time warnings

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

and as input for forecast model development), access to and integration of biomass flow data from different sensors and forecast accuracy. Stakeholders for the wind energy sector included policy makers, energy companies, energy grid operators and nature conservation. This group is particularly interested in the development of forecast models to provide early warnings for wind energy curtailment. Topics discussed with this group include challenges related to forecasting migration (particularly over sea, which is novel), transitioning from research and development to operational systems, access to weather forecast data, and realistic forecast windows.

2.1.4. Stakeholder representativeness and gaps for monitoring and forecasting aerial biomass of birds

Several research groups are active in further developing tools for monitoring and forecasting aerial biomass of birds and discussions with researchers from the Netherlands, Switzerland, Sweden, Belgium, Finland, Germany, Norway, and the USA have been incorporated. The workflow for automated processing of weather radar data to extract aerial biomass flows was developed and updated through various collaborative projects, resulting in the ALOFT repository (Table 1). Short term forecasts were developed for aviation safety by researchers in the Netherlands and Finland and spatial predictions developed by researchers in Switzerland, resulting in the aerial biomass flow EBV. Official agreements for collaboration with OPERA were established during the cost action ENRAM (2013-2017) and while these agreements are periodically renewed, stronger collaboration would be beneficial in the future and represents a current gap in the stakeholder representation. Currently individual meteorological institutes are involved in discussions, but official representation from the leadership of OPERA would have been an improvement. Forecast models have been designed at a regional scale in Europe where long term and high-quality aerial biomass data is available, with examples in the Netherlands and Finland. However, many other European countries could benefit from short term forecasts for a broad range of stakeholders. Wind energy stakeholders were represented in discussions as were representatives for military aviation. Civil aviation may be potential stakeholders and have periodically shown interest in such developments but were not actively involved in discussions at this stage.

2.2. Key inputs from stakeholders

2.2.1. Key inputs from stakeholders for the generic forecasting workflow

The co-design participatory process employed for the generic forecasting workflow provided prolific and highly valuable input from stakeholders. Key input concerning the desirable properties of the product (i.e., the forecast maps) included a spatial resolution in the range of 10×10 km to 50×50 km, a daily or multi-daily temporal resolution and a forecast horizon of at least one week (8 days). Based on species experts, we also identified a set of potential target species to test the workflow and the identification of most likely relevant environmental factors driving the seasonality of fructification. Together with ecological modelling experts, species experts also provided highly valuable input on the identification of available data sources of mushroom observation data and on how to best encode environmental predictors for maximizing predictive performance (e.g., which features of temperature or precipitation variation to use as predictors). Finally, experts in ecological modelling were instrumental for defining the full modelling cycle and implementation, providing advice on key aspects,

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

such as likely successful modelling techniques and optimal validation approaches (e.g., machine learning models, validation with independent data, etc.). Another key contribution referred to the joint vision among participants of the capacity of the workflow to adapt seemingly to other ecological phenomena, prompting us to test it for a second case study (life-stages of invasive pests; mentioned above). However, many other potential applications were envisioned, including the early warning of harmful algal blooms, forecasting life-stages of disease vectors, forecasts of levels of activity for game species or of species of relevance for touristic activity (birdwatching, firefly tourism, etc.).

2.2.2. Key inputs from stakeholders for the workflow of the aerial biomass of birds

An important issue for bird forecasts is the tradeoff between improved bird forecasts and the need for a certain forecast window to take appropriate mitigation action. The energy grid operators for example may require a longer forecast window to balance the energy market and address supply and demand issues. Yet, the longer the forecast window the higher the uncertainty in migration forecasts. Energy companies, on the other hand, want to avoid unnecessary wind curtailment. Input data for forecasts was also mentioned regarding availability, and the need for a license or open access, e.g. from the European Centre for Medium-Range Weather Forecasts (ECMWF). For example, reanalysis of weather data (ERA5, Hersbach et al. 2020) is available open access at an hourly temporal resolution. Such data is used to develop predictive models of avian migration (Kranstauber et al. 2022). Yet, to run a predictive model to forecast migration, weather forecasts are needed but the ECMWF medium range forecasts are not freely available. The propagation of errors in migration forecasts due to uncertainty of weather forecasts is a relatively unexplored issue and requires further research. Another issue relates to the need to go from scientific code to operational software, who would be responsible and how this can be implemented. An issue of concern raised by scientists developing predictive models and data users is access to long term biomass flow data of sufficient quality. Data flows from operational weather radars are currently at risk of being cleaned of biological information before being made available for biodiversity related research and model development (Shamoun-Baranes et al. 2022).

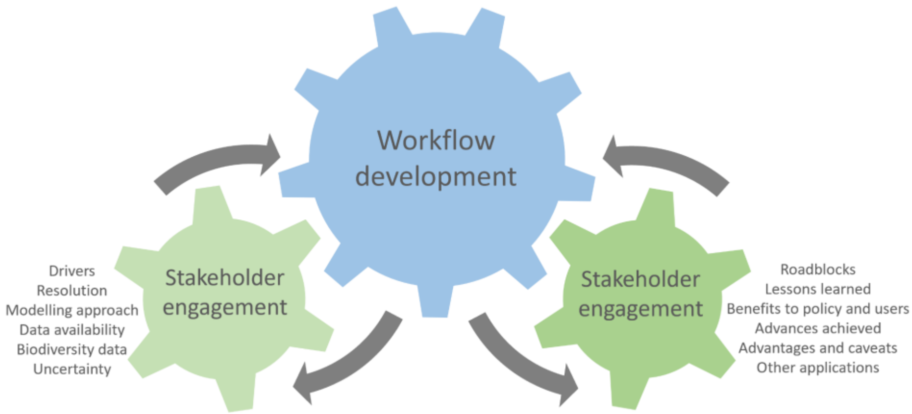

Figure 1. Schematic representation of stakeholder engagement and contributions to workflow development. Workflow development involves stakeholder engagement at different steps of the process, represented by the different wheels of stakeholder engagement in the figure. At a more initial stage (left wheel in the figure), stakeholders are involved for input on workflow characteristics and technical details such as: which drivers of the phenomena being forecasted should be considered (e.g.,

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

ARPHA Preprints Author-formatted document posted on 19/01/2024. DOI: https://doi.org/10.3897/arphapreprints.e119131

weather), at which spatial and temporal resolution should the forecasts be produced, or which is the best modelling approach for the intended objective? At a later stage (right wheel in the figure), issues such as advances achieved, roadblocks, or advantages and caveats, are key factors to be discussed with stakeholders.

3. Contributions of the short-term forecasts to policy and to end users

As mentioned, the two forecasting workflows are described and demonstrated by providing days-ahead insight into 1) the likelihood of fruiting for wild edible mushrooms, 2) the likelihood of the occurrence of invasive pest species in specific life stages and 3) the aerial biomass of birds. These case studies were selected having in mind EU citizens as end users and aim primarily at contributing to the EU Bioeconomy Strategy, which intends to foster sustainable and circular bio-based economy sectors, reduce industries environmental footprint, diversify sources of renewable biological resources, and protect ecosystems (EC-DGRI 2018).

By providing EU citizens with timely information on when to forage for edible wild mushrooms, we expect to significantly promote the adoption of this naturally renewable food source, which is still an overlooked forestry by-product in many European countries (De Cianni et al. 2023, Huber et al. 2023). Furthermore, mushroom foraging can also contribute significantly to ecotourism (Latorre et al. 2021). Activities such as workshops, guided tours, and foraging events can draw tourists and catalyze economic growth in rural areas (Latorre et al. 2021), aligning with the bioeconomy strategy aspirations for rural development and job creation. Having *a priori* information on the timing of occurrence of fruiting bodies of species of foraging interest can contribute to maximizing the success of such initiatives. Moreover, for the EU Bioeconomy Strategy to truly harness the benefits of mushroom foraging, adherence to sustainable practices is critical. In this context, if managers of natural habitats, farmlands, or forestry areas adopt these forecasts, they could also be better informed about the optimal periods to implement measures countering possible negative effects of harvesting activities, ensuring mushroom foraging aligns with both ecological and economic objectives.

The contribution of forecasts for pest species can also be substantial. Specifically, by knowing when specific life stages are occurring, citizens (e.g., under citizen-science initiatives) and experts can better program the timing for pest monitoring and early detection initiatives (Latombe et al. 2016), as well as to implement any impact minimization or eradication efforts (Barker et al. 2020, Dietze et al. 2018, Tulloch et al. 2020). This, in turn, contributes to a more informed and proactive approach to pest management, aligning with the EU Bioeconomy Strategy's goal of safeguarding ecosystems and preventing significant economic losses in agriculture and forestry sectors, two crucial pillars of the bioeconomy. The timely eradication or control of invasive pests also minimizes the use of potentially harmful chemicals, further reducing the environmental footprint of industries.

The bird biomass flow forecasts, in turn, is highly valuable to inform civil and military aviation about periods when large numbers of birds are expected in a particular air space, and subsequent flight planning can improve aviation safety and reduce economic losses due to damage to aircraft or delayed flights (Kranstauber et al. 2022, Metz et al. 2020, Nilsson et al. 2021, van Gasteren et al. 2019). Finally, migration is a natural phenomenon of high touristic potential and is appreciated by many nature enthusiasts (see, for example, BirdCast in North America; Horton et al. 2019). Similarly, to the

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

ARPHA Preprints Author-formatted document posted on 19/01/2024. DOI: https://doi.org/10.3897/arphapreprints.e119131

mushroom case study, forecasts can help improve the success of birdwatching-related touristic activities by informing on the optimal timing for observation activities.

Beyond our primary alignment with the EU Bioeconomy Strategy, the three forecasting targets also provide relevant contributions to a broader array of EU strategies and policies. For example, these forecasts can support the EU Biodiversity Strategy for 2030 (EC-DGRI 2018), contributing to conservation, by helping support the definition of prohibition dates for sites with habitats sensitive to mushroom foraging activities or to inform wind turbine curtailment (e.g., bird biomass in the North Sea) and temporarily turning off urban lighting to reduce avian mortality ("lights out" programs, Horton et al. 2019). Bird migration forecasts can also contribute to targets of the Convention for migratory species, the Ramsar convention and potentially the Birds Directive. For example, they could be used to identify future areas being targeted for restoration or conservation to support changing migratory patterns in face of climate change (Barrett 2011, Howard et al. 2020). The EU Farm to Fork (F2F) Strategy is also supported, as with deeper insights into invasive pest life cycles, one can significantly optimize pest management, reducing the dependency on chemical pesticides. This aligns with the strategy's goal of minimizing chemical usage, ensuring both healthier ecosystems and safer food chains (EC 2020). Relevantly, better knowledge on the stages of development of invasive pest species also provides clear support to the EU efforts to fight invasive alien species (IAS), namely under the EU IAS regulation (EU 2014), by enabling the optimization of measures for preventing their further spread in the EU territory. Finally, the European Green Deal is also supported, since in addition to supporting conservation, our forecasts also help promote sustainable and environmentally friendly economic practices.

4. Essential Biodiversity Variables design and outputs

4.1. EBV design characteristics

4.1.1. EBV description

The EBVs produced in the workflows developed belong to the Species populations and Community composition EBV classes and to the terrestrial environmental domain. They represent phenology and biomass flows and are provided as daily probabilities (phenology of mushrooms or invasive pest) or hourly averages and annual timing of arrival and departure in the case of bird density (Table 1).

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

Table 1. Characterization of the EBVs produced by the two forecasting workflows described.

			EBV		
Forecast	**Class**	**Name**	**Description**	**Metric**	**Domain**
Phenology of mushrooms	Species Populations	Phenology of fructification of mushrooms and wild fruits	Short-term forecasting of the fructification of wild mushroom species within contiguous spatial units (grid cells) across the EU over time	Daily probability of observing fruiting bodies of selected wild mushroom species of commercial and recreational interest across the EU	Terrestrial
Phenology of invasive pests	Species Populations	Phenology of life-stages of selected terrestrial invasive pest species	Short-term forecasting of the phenology of invasive pest species within contiguous spatial units (grid cells) across the EU over time	Daily probability of observing selected invasive pests in specific life stages	Terrestrial
Bird biomass	Community composition	Aerial biomass of migrating birds	Biomass flows of aerial migrants (birds) across Europe within contiguous spatial units (grid cells) over time	Summary statistics of migration densities of birds derived from vertical profile time series of weather radar data (e.g. hourly averages of bird density and speed)	Terrestrial
	Species Populations	Phenology of migration of terrestrial birds	The annual timing of arrival and departure of European terrestrial migratory bird species at breeding, staging and wintering sites over time	Migration phenology metrics	Terrestrial

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

4.1.2. EBV spatial and temporal resolution and extent

The EBVs produced in the forecast workflows have a daily temporal resolution and temporal extent going from days to one year. The spatial resolution is 0.25° (c. 28 x 28 km at the Equator) and the intended uses include supporting mushroom foraging and management activities, pest monitoring and management, and informing aviation and wind turbine activities, while contributing to bird conservation (Table 2).

Table 2. Spatial and temporal characteristics of the EBVs produced in the forecast workflows.

<table>
<tr><th>Forecast</th><th>Spatial extent</th><th>Spatial resolution</th><th>Temporal extent</th><th>Temporal resolution</th><th>Intended uses</th></tr>
<tr><td>Phenology of wild mushrooms</td><td rowspan="2">Countries within Europe</td><td rowspan="2">0.25° (~28x28km at the Equator)</td><td rowspan="2">up to nine days</td><td rowspan="3">daily</td><td>Support wild mushroom foraging and management activities, by indicating the expected timing of occurrence of fruiting bodies of selected species.</td></tr>
<tr><td>Phenology of invasive pests</td><td>Support invasive pest monitoring and management activities, by indicating the expected timing of occurrence of specific life stages of selected species.</td></tr>
<tr><td>Bird biomass</td><td>Countries within western Europe</td><td>0.25°</td><td>one year</td><td>Support aviation safety, wind turbine curtailment, and bird migration conservation, by informing about levels of aerial bird biomass.</td></tr>
</table>

4.2. Input biodiversity data

4.2.1 Input biodiversity data for the generic forecasting workflow

The generic forecasting workflow is based on temporally discrete biodiversity observation records. To be usable within the scope of the workflow, these records need to 1) provide the full date of the observation (i.e., day, month, and year), 2) have geographical coordinates with a spatial resolution equal to or better than that of the final product, and 3) be supported by visual media, such as photographs, or textual descriptions confirming the representation of the phenomenon of interest (e.g., the fruiting body of a selected wild mushroom species or the adult life stage of an invasive pest). These records can come from a large variety of sources, particularly citizen science projects and research-based initiatives, and are commonly available from online databases (e.g., inaturalist.org; mushroomobserver.org, observado.org), research institutions (e.g., natural history collections), the scientific literature and social media (Marcenò et al. 2021). The Global Biodiversity Information Facility (GBIF; gbif.org; Table 3) is currently a reference meta-repository for these observations and, for many ecological phenomena, can already provide several thousands of usable records with sufficient spatial and temporal coverage (Capinha et al. 2023). However, in the case of a perceived limited sampling based solely on observations records available from GBIF, the other mentioned sources could be considered.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

ARPHA Preprints Author-formatted document posted on 19/01/2024. DOI: https://doi.org/10.3897/arphapreprints.e119131

Apart from the three above-mentioned criteria, the observation records need to undergo a strict quality control before being used in the workflow. First, they need to have been made within a temporal period of reference (e.g., from 2010 to 2020), which is used to characterize each record in terms of predictor variables. Records made outside this period should be excluded. Next, retained records need also to undergo a fine cross-checking of validity of observation dates. One common error in this attribute concerns the automated assignment of the first day of the month for records that are missing the exact day number (Belitz et al. 2023). In records coming from GBIF, these cases could be identified by the simultaneous attribution of a null timing (i.e., 00:00:00). Records having this or any other inconsistencies in terms of the attribution of the date of observation do not qualify for use.

Additionally, potential georeferencing issues must also be assessed. As previously mentioned, only records offering a spatial resolution compatible with the desired end product (0.25 degrees) are suitable. In terms of coordinate uncertainty, this means retaining only those records with a level of uncertainty up to 0.125 degrees (i.e., half of the spatial resolution of reference), to ensure that the record accurately matches the corresponding grid cell rather than neighboring ones. It is also crucial to examine common indicators of positional error in records for which the positional uncertainty is unknown (e.g., when only the coordinates are given). These indicators include coordinates that match the centroids of countries or provinces, capital cities, urban areas, or having identical values of longitude and latitude, or coordinates located at sea for terrestrial phenomena (or on land for marine phenomena). Flagged observation records, for which there is no additional information allowing to confirm sufficient positional accuracy, should be excluded.

It is also imperative to confirm that the records effectively represent the ecological or biological phenomenon of interest (i.e., the one being forecasted). For instance, records of a mushroom fruiting body should be meticulously checked for taxonomic inaccuracies, such as the misattribution of species names. The most common form of evidence supporting the representation of the phenomenon of interest, particularly in citizen science records, is photographic documentation. Therefore, each record and its supporting media should be individually scrutinized to confirm they specifically depict the event of interest. This means, for example, ensuring that records of a mushroom species accurately represent the fruiting body of the species rather than another distinguishing feature (e.g., spores). Furthermore, even if the images correctly depict the expected phenomenon, it is crucial to verify that they represent its occurence at the date of observation and not from a previous period. This includes, for example, checking for signs of significant deterioration in a fruiting body of a mushroom, which, if confirmed, should lead to the exclusion of the record, since there was an apparent time lag between the occurrence of the phenomenon and its recording.

4.2.2 Input biodiversity data for the aerial bird biomass forecasting workflow

The aerial bird biomass workflow uses data on the numbers of birds in the air during nocturnal migration. Aerial biomass data is extracted from operational weather radars in western Europe using the bioRAD package (Dokter et al. 2019) running at the Swedish Meteorological Institute. Data is processed into 5 - 15 minute vertical profiles of bird migration and stored in an online repository (https://aloftdata.eu). Continuous, high resolution spatio-temporal maps of nocturnal bird migration densities, flight speed in east/west direction and flight speed in north/south direction across western Europe were interpolated from vertical profile time series datasets measured by 37 weather radars in

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

ARPHA Preprints Author-formatted document posted on 19/01/2024. DOI: https://doi.org/10.3897/arphapreprints.e119131

France, Germany, The Netherlands and Belgium, operating between 13 February 2018 and 1 January 2019. To go from point data (from each radar) to gridded data, bird flow (i.e., average bird movement) of long- and short-distant nocturnal migrants is modeled on a nightly time scale and a spatial resolution of 0.25°. The modeled flow (density, speed and direction) represents average fluxes of all birds moving through the area and used to create the EBVs (Nussbaumer et al. 2021). Measured or modeled bird flows can be used to develop predictive models of migration.

Ensemble models of spring and autumn bird migration were developed to forecast bird migration for military aviation safety in the Netherlands. Weather radar data was collected from the KNMI (Royal Netherlands Meteorological Institute) and vertically integrated bird densities were extracted at a temporal resolution of 5 minutes using the bioRAD (Dokter et al. 2019) package to create a 10 year time series. Phenological predictors were derived from the migration time series and meteorological predictors from ERA5 data. Migration was predicted using general additive models; for more details regarding the forecast modelling workflow see Kranstauber et al. (2022).

Table 3. Raw data availability and access for the two workflows developed.

DATASET TITLE	GBIF	Mushroom Observer	Aloft
Raw data collection design	The Global Biodiversity Information Facility (GBIF; https://www.gbif.org) is a leading aggregator of biodiversity observation records, including those from opportunistic observations - citizen science, research, monitoring programs, museum collections .	Mushroom Observer (https://mushroomobserver.org) is a citizen science platform (not included in GBIF) providing opportunistic observation records of mushroom species	The Aloft data repository contains bird movement data extracted from European weather radar data which is sent to the central OPERA repository (https://aloftdata.eu).
Monitoring programs	NA	NA	OPERA (https://www.eumetnet.eu/activities/observations-programme/current-activities/opera/).
Types of data access	Open access (license: CC BY-NC, CC BY).	Open access upon registration (license: CC BY-NC, CC BY).	Vertical profile data of bird movement publicly available data (license: CC BY).
Data repositories	https://www.gbif.org	https://mushroomobserver.org	https://aloftdata.eu
Persistent identifier(s)	*Craterellus tubaeformis*: https://www.gbif.org/occurrence/search?taxon_key=2554536 *Popillia japonica*: https://www.gbif.org/occurrence/search?taxon_key=4425774	*Craterellus tubaeformis*: https://mushroomobserver.org/observer/observation_search?pattern=Craterellus+tubaeformis	https://aloftdata.eu
Metadata description	Metadata discoverability: metadata is indexed in a searchable resource and accessible with standard protocols.	Metadata discoverability: metadata description is provided as a static resource in the project website. Metadata standards: rules and guidelines used to describe and	Metadata discoverability: https://github.com/enram/data-repository Provenance: https://github.com/enram/data-repository/tree/master/file_transfer

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

ARPHA Preprints Author-formatted document posted on 19/01/2024. DOI: https://doi.org/10.3897/arphapreprints.e119131

	Metadata standards: Darwin Core	catalog information is specific to the project.	
Other provenance information	-	-	Data originally extracted from OPERA and processed using the bioRAD toolbox (https://adriaandokter.com/bioRad/)

4.3. The EBV model

4.3.1. Generic forecasting workflow

4.3.1.1. Workflow description

The full 'generic' workflow consists of seven main data processing and modelling steps (see Capinha et al. 2023) for the full details of the modelling approach and Fig. 2 for a representation of the workflow).

The first step of the workflow consists in compiling biodiversity observation records and is described above in section "4.2.1 Input biodiversity data for the generic forecasting workflow". For example, in developing a forecasting workflow for the edible winter Chanterelle (*Craterellus tubaeformis*), we compiled fruiting body observation data from GBIF and Mushroom Observer for the period 2015 to 2021. Following the described procedures, we obtained about 1,700 suitable observation records. We undertook a similar process for our second case study, which aims to forecast the occurrence of the adult stage of the Japanese beetle (*Popillia japonica*). This invasive alien species, recently established in continental Europe (notably in Italy; EFSA 2019), has the capacity to consume hundreds of plant species, causing significant economic damage. Forecasts of the adult phase of this species are relevant to inform monitoring and invasion surveillance efforts, as this is when the species has a higher visual detectability (EFSA 2019). For this species, we gathered data from GBIF from 2015 to 2021. After implementing the previously mentioned quality control criteria, we obtained a total of c. 10, 000 records of the species in its adult stage.

The second step involves identifying and preparing environmental predictor data. The selection of environmental factors in this phase is primarily influenced by the nature of the phenomena being forecasted. For many such phenomena, a detailed representation of meteorological variation is essential, as it is a primary driver of intra-annual variation for many ecological and biological phenomena (Diez 2013, EFSA 2019). Additional factors that may be considered include photoperiod (Buonaiuto et al. 2023), along with 'static' factors like soil properties, topographical features, and land use classes, which are not expected to change over the forecast period.

When selecting temporal predictors, it is crucial to source them from providers that offer forecasts covering the intended forecasting horizon. For example, both the European Centre for Medium-Range Weather Forecasts (ECMWF) and the Global Forecast System (GFS) deliver global weather forecasts that stretch beyond 15 days, covering an extensive range of variables. Our general workflow is contingent upon these forecasted predictor variables being available. Therefore, our choice of temporally varying predictors is bound by the pool of variables offered by these forecasting services. For our case studies - the fruiting bodies of the winter chanterelle and the adult phase of the Japanese beetle - we used weather predictor data from the GFS. This includes time series data on daily minimum, maximum, and average temperatures, as well as snow depth and average wind speed. Unlike ECMWF,

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

GFS offers these datasets and real-time forecasts freely. Specifically, for model calibration (see below), we compiled spatial time series of these meteorological variables at the global scale using a 0.25° spatial resolution, from 2014 to 2021.

The third step consists of addressing potential spatial or temporal bias in the observation data of the phenomena of interest. Typical spatial biases in these data correspond to having records highly concentrated in a few areas, which if used for model calibration may dominate the overall patterns of the data and make model fitting unrepresentative of other locations (i.e., reduce their spatial transferability). There are distinct procedures that could be used to mitigate this issue (Isaac and Pocock 2015), most of them involving the subsampling of the observation data and reducing the number of records in oversampled regions (e.g., Zhang et al. 2018). A typical way to approach this is to overlay a predefined grid of equal-sized cells and identify those that are upper outliers in terms of the number of records they have (Capinha 2019). Then for each of these cells only a subset of records is randomly selected (e.g., the support threshold value or the average number of records across all cells).

Temporal biases are less straightforward to address. Clearly, some phenomena will have more observation records in some periods than others, simply because there was higher recording activity (e.g., more activity from citizen scientists). In other words, the temporal patterns in the frequency of observation records results from the joint variation in the actual timing of occurrence of the phenomenon being recorded and in the levels of activity of the recorders. One possibility to address these biases is to 'remove' the effect of unequal levels of recording activity. This can be achieved, to some extent, by using proxies of the level of recorder's activity and resampling the observation records so that those made in periods of higher activity will have lower chances of being selected, and vice versa. Precise proxies can be hard to obtain, however the temporal variation in observations for certain 'benchmark' taxa can give a useful indication (Capinha et al. 2023). Specifically, taxa that show little or no phenological variation along the year could be good indicators of recording levels. The rationale is that the appeal for recording these species is constant, implying that any temporal fluctuations in their records are likely indicative of levels of recording activity rather than the species variability. Taxa that could be considered as possible proxies include some conifer species, which are evergreen and have inconspicuous or visually unassuming reproductive structures, making their levels of recording appealing similar along the year.

For the two case studies, we dealt with spatial bias first by using a grid of 250x250 km cells to identify upper outlier regions in terms of number of observation records (i.e., above the upper boundary of Q3 +1.5 IQR, with Q3 corresponding to the third quartile and IQR to the interquartile range). Then for each outlier region, we randomly selected a number of records matching the outlierness threshold. Finally, we also only kept one record per 0.25° grid cell and date of observation.

We also tested the benchmark taxa approach to deal with temporal bias in observation data. Specifically, we used records of observation of pine trees (*Pinus* spp.). We collected data of observation of these taxa from 2015 to 2021 from GBIF, and performed the geographical and temporal quality filtering procedures, as described above for the phenomena of interest.

Next, we generated an equal number of observations, having the same geographical coordinates as *Pinus* spp. records, but with dates generated at random within the selected years (i.e., 2015 to 2021). This allowed us to obtain a distribution of records that would be expected if observations were made

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

randomly over time. For both types of records (i.e., observations and randomly generated dates), we then extracted the day of the week, month, average temperature of the day, total precipitation of the day, and average wind speed of the day.

To avoid collinearity issues among the predictors, we computed the variance inflation factor (VIF) using the *car* package (Fox and Weisberg 2019) within R (R Core Team 2023). The results indicated that all predictors were within acceptable collinearity limits (i.e., squared scaled generalized VIF ≤ 1.4). Subsequently, we used a generalized linear model (GLM) with a binomial error distribution to correlate the record classes with calendar and weather predictors. The model yielded plausible results, effectively reflecting significative positive association between the availability of *Pinus* spp. records and warmer temperatures, absence of precipitation, lower wind intensities, weekends, and months associated with summer holidays.

Next, we applied this model to estimate the sampling effort level associated with each record of the fruiting body of winter chanterelle and of adults of the Japanese beetle. The predicted values reflect the likelihood of increased number of records due to observer-friendly conditions (such as preferred days, months, and weather conditions). To compensate for this bias in the datasets of both the Japanese beetle and the winter chanterelle, we used inverse probability weighting (Mansournia et al. 2016). Specifically, for each case study we created an alternate dataset of observation records, where the inclusion probability of each original record was inversely proportional to the recording effort predicted by the model. Consequently, records made under more favorable conditions for observers were less likely to be chosen, and vice versa.

Despite the advantages of data-processing techniques aimed at reducing spatial and temporal biases in observational data, it is important to note that this step is entirely optional in our workflow. Specifically, one may choose to directly calibrate models using data gathered and cleaned in the initial workflow stage (Fig. 2). In our case studies, we observed that models using temporally debiased data and those without such processing yielded comparable predictive performances (Capinha et al. 2023). However, in the absence of prior knowledge regarding the extent of data biases, we recommend employing both approaches – with and without debiasing attempts – and comparing the model predictions from each.

The next step in the workflow involves calibrating models to perform the forecasts. This can include a variety of techniques, ranging from traditional statistical-based methods such as logistic regression (Shanubhogue and Gore 1987) to supervised 'classical' machine learning algorithms such as random forests or boosted regression trees (Cutler et al. 2007, Elith et al. 2008, Zhang et al. 2020), and more recent deep-learning approaches such as convolutional or recurrent neural networks (Capinha et al. 2021; Ceia-Hasse et al. 2023). In essence, these models are trained by contrasting the set of temporal environmental conditions associated with the phenomenon of interest against the temporal environmental conditions that are available in the places of their occurrence. To represent the latter, one may use temporal pseudo-absences (Capinha et al. 2023), which are records that share the geographical location of the observed phenomenon but have randomly selected dates within the temporal range of the data. Both types of records are then characterized by a set of predictor variables, representing conditions such as preceding temperature, accumulated precipitation, snow depth or wind speed. Prior to training the models, it is also crucial to evaluate and eliminate high levels of collinearity among the predictors, e.g., by selecting a subset of non-collinear ones.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

ARPHA Preprints Author-formatted document posted on 19/01/2024. DOI: https://doi.org/10.3897/arphapreprints.e119131

Specifically for the two case studies, we trained two classical machine learning algorithms, random forests (Cutler et al. 2007) and boosted regression trees (Elith et al. 2008), using a total of 67 predictor variables representing multiple aspects of preceding conditions of temperature, precipitation, accumulated snow, and wind speed (Capinha et al. 2023). To address potential issues of collinearity, we calculated the Pearson correlation coefficient among pairs of predictors and selected only predictors having absolute correlation values below 0.8.

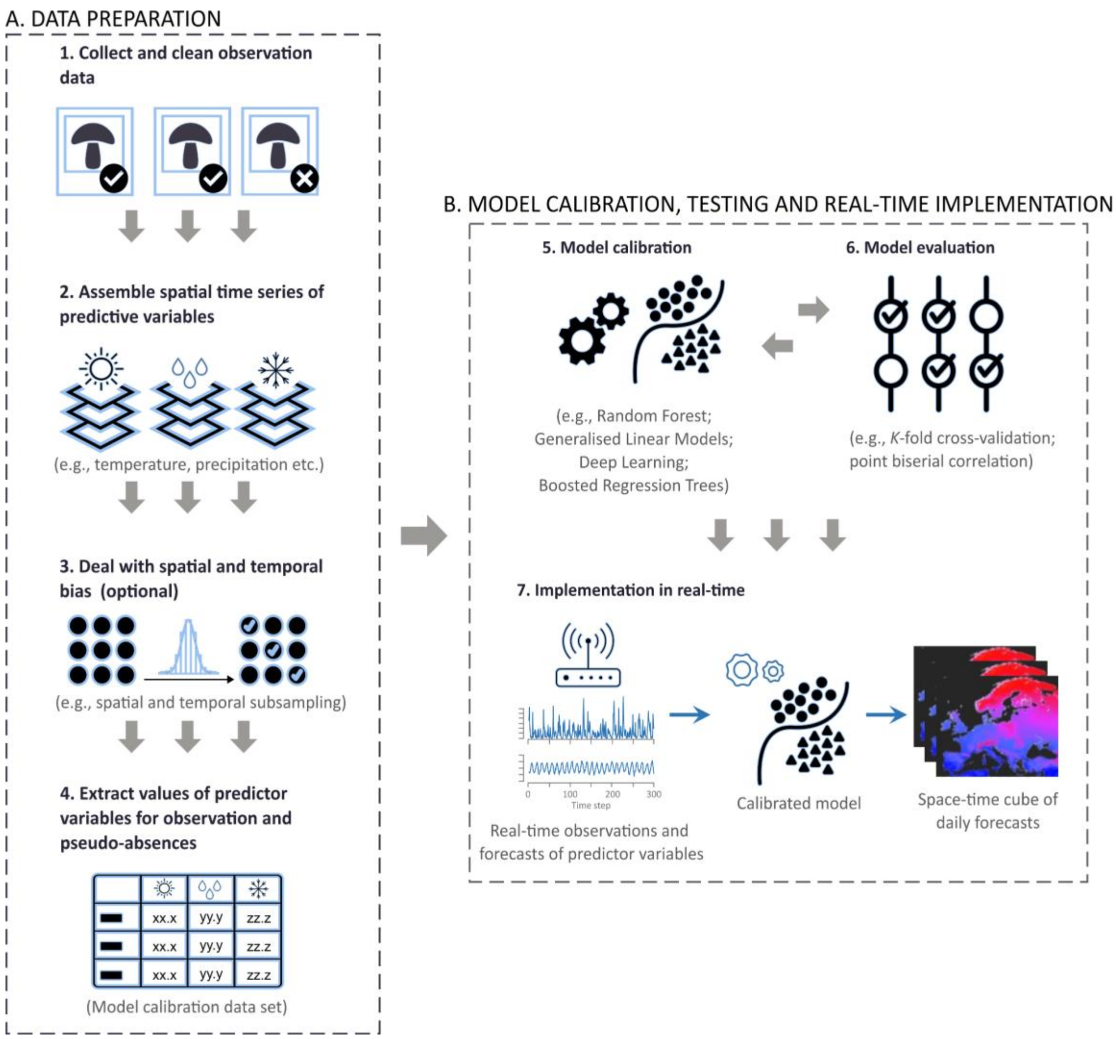

Figure 2. Representation of the data processing and modeling procedures implemented in the generic forecasting workflow.

Next, it is crucial to measure the predictive performance of the models. This can be done in various ways, with the fundamental principle being that the evaluation should not be performed with data used in model calibration (e.g., *k*-fold cross validation). One approach could be to use data from all years except one, reserving the excluded year for comparison with the predictions of the model. This procedure can be repeated, each time using a different year as the evaluation dataset (i.e., a 'leave-one-year' out approach). This ensures that model testing is made for time periods independent of those used in the model calibration. In addition, one could also use data from regions or periods where high-quality observational data is available, such as those obtained through systematic observation

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

efforts. The congruence between model predictions and actual observations can be measured using a range of complementary metrics (e.g., the area under the receiver operating characteristic curve (AUC), point bi-serial correlation, or the F-score). This assessment should also cover the entire forecast horizon, meaning that it should evaluate predictions made for all future days-ahead periods covered by the forecast.

For the two case studies, we used a 'leave-one-year' out approach, and also compared predictions of models with observation data in regions having a high recording density of the phenomena (i.e., in Denmark for the winter chanterelle and in northwest Italy for the Japanese beetle). In the first case, we used the AUC as the performance metric, while in the second we measured the point bi-serial correlation between the probabilities forecasted and the observed presence or absence of records of the phenomenon. In both cases, all predictions achieved AUC and point bi-serial correlation values representative of a good to very good predictive performances (minimum AUC for the winter chanterelle = 0.81 and minimum AUC for the adult stage of the Japanese beetle = 0.88; minimum point bi-serial correlation for the winter chanterelle = 0.70 and minimum point bi-serial correlation for the adult stage of the Japanese beetle = 0.81).

Finally, once the models achieve a desired level of predictive accuracy, it is necessary to implement them for running in near-real-time. This implementation involves the preprogramming of a series of routines that automatically trigger task execution at predetermined intervals or after the completion of preceding processes. A key routine to run concerns the recurrent updating of predictive variables (e.g., by collecting daily updated data from GFS and processing it into the pre-defined predictor variables, for each future day within the forecasting horizon). The calibrated model(s) then generate(s) forecasts based on these updated variables. To guarantee a consistent and reliable near-real time operational implementation, it is essential to use a robust and fail-safe computational infra-structure. This includes having built-in redundancy features to safeguard against potential interruptions, such as unforeseen system downtimes, and continuous backups of data.

4.3.1.2. Representation of the resulting EBV

This workflow allowed obtaining up-to-date short-term forecasts of the target phenomena. Specifically, it enabled producing well-performing forecasts of the probability of occurrence of fruiting bodies of the winter chanterelle and of the adult stage of the Japanese beetle within a time frame of 9 days, e.g., for the date of production of the forecast and two days after that date, for four to six days after the date of production of the forecast, or for seven to nine days after the date of production (see Figs. 3 and 4 for an example forecast of each of the target phenomena considered).

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

ARPHA Preprints Author-formatted document posted on 19/01/2024. DOI: https://doi.org/10.3897/arphapreprints.e119131

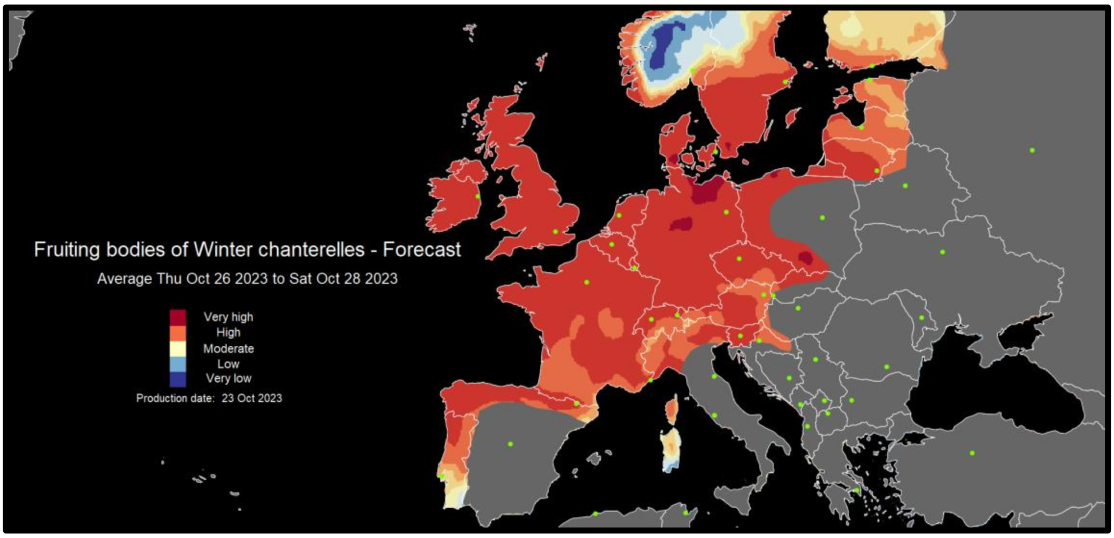

Figure 3. Output map from the generic forecasting workflow applied to the winter chanterelle. Probability of occurrence of fruiting bodies of *Craterellus tubaeformis* for four to six days from the day of production of the forecast (date of production: 23rd October 2023; forecast for 26-28 October 2023). Areas in gray are deemed unsuitable for the species or are outside its recorded range of distribution.

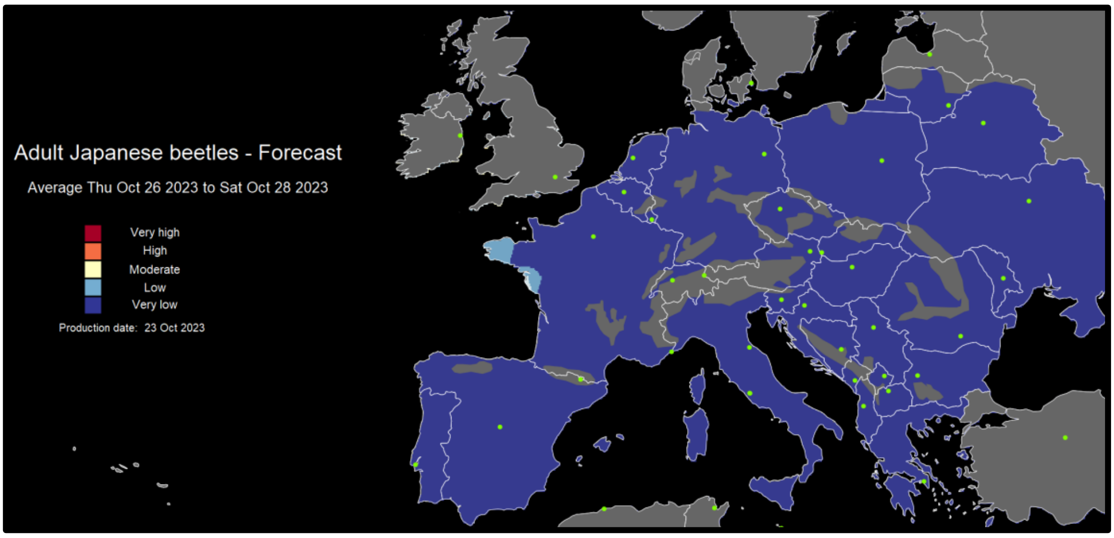

Figure 4. Output map from the generic forecasting workflow applied to the Japanese beetle. Probability of occurrence of the adult stage of *Popillia japonica* for four to six days from the day of production of the forecast (date of production: 23rd October 2023; forecast for 26-28 October 2023). Areas in gray are deemed unsuitable for the species or are outside its recorded range of distribution.

4.3.1.3. Uncertainty assessment

In the generic workflow, predictive uncertainty can arise from multiple sources, such as variability in the input data, the inherent stochastic nature of the models used, and potential limitations in the representativeness of the observational data. To address this, one can incorporate uncertainty assessment as a critical component of our workflow. This can be achieved through the implementation of bootstrapping and sensitivity analysis, enabling to quantify variability in model outputs due to changes in input data. Additionally, ensemble modeling methods can also be used, which involve

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

running multiple iterations of the models with slightly varied parameters to understand the range of potential outcomes. This allows for a more robust interpretation of the forecasts, providing a clearer picture of the likely bounds of the forecasted phenomena.

4.3.2. Aerial biomass of birds forecasting workflow

4.3.2.1. Workflow description

This workflow forecasts bird migration - including moments of peak migration - using bird migration intensity data extracted from operational weather data (Fig. 5). Kranstauber et al. (2022) detail the predictive modelling approach to obtain bird migration forecasts using ensemble models and differentiating between spring and autumn (see also van Belle et al. 2007 and van Gasteren et al. 2019).

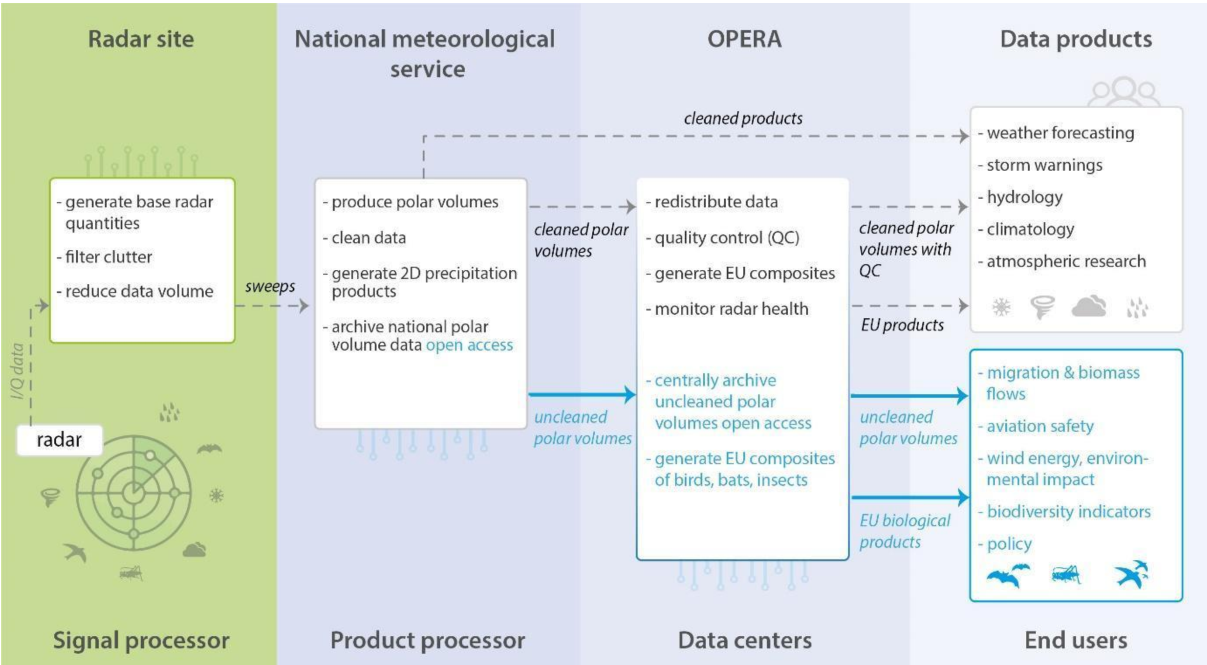

Figure 5. Schematic representation of the forecast workflow for the aerial biomass of birds (from Shamoun-Baranes et al. 2022).

Migration intensity was modeled with estimates of vertically integrated bird density (birds/km2), which is derived from weather radar data. Bird density was calculated with the software package *vol2bird* within a range from 5 to 25 km from the radar (Dokter et al. 2011). Data quality was further ensured by visually inspecting the vertical profile time series of peak nights, where periods of rain and other non-bird reflections not filtered out by *vol2bird* were then excluded. Additionally, the lowest altitudinal bin (0-200m) was not used since at this height birds were not consistently identified for vertical integration.

Regarding the spatiotemporal covariates, frequently used environmental variables correlating with migrant numbers are wind conditions, occurrence and abundance of rain, surface pressure and air temperature (e.g., van Belle et al. 2007, Van Doren and Horton 2018). Every 5-min radar measurement of bird density was related to environmental conditions belonging to three classes: local weather (e.g., wind and temperature at several heights, cloud cover, precipitation rate), remote weather (e.g., pressure and surface wind at departure locations), and an index related to the accumulation of birds not migrating due to poor weather conditions (see Table 1 in Kranstauber et al. 2022 for the full list of

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

variables). Weather conditions were linearly interpolated in time from the ECMWF ERA5 dataset (Hersbach et al. 2020). Data was extracted from the location of the radar site (local weather) and from the conditions at locations where birds could depart from (remote weather).

The modelling approach consisted in two steps. A phenological model was first developed, representing seasonal and diurnal long-term migration dynamics, and then the value of including weather variables in the model was evaluated. Additionally, differences between spring and fall were assessed, and top models were averaged to create ensemble models, with the aim of increasing predictive performance (Dormann et al. 2018). Migration density was modeled using generalized additive models with a quasi-Poisson distribution family, which do not assume a particular relationship between the predictor variables and the dependent variable (Wood et al. 2015). The general phenological trends in migration were first captured in a model including three variables: day of the year (to capture seasonal trends), solar elevation (to capture circadian effects) and time derivative of solar elevation (to differentiate between sunset and sunrise). Models including environmental variables were then fitted using the phenological model as a basis. Each model represents the expected seasonal migration and one or a pairwise combination of two environmental variables. The best model was identified using cross-validations. In Kranstauber et al. (2022), the dataset was split 10 times for cross-validation datasets with a 70:30 division and the deviance by the sum of squares was calculated to the excluded 30% of data, for each environmental model.

4.3.2.2. Representation of the resulting EBV

This workflow allows obtaining point location specific bird migration forecasts. The approach can be applied on a continental scale using weather radars across Europe (Fig. 6). In addition, point based migration intensity (forecasted or measured) can be converted into a spatially explicit representation of migration intensity (biomass EBV) following the methodology developed by Nussbaumer et al. (2021) (Fig. 7).

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

ARPHA Preprints Author-formatted document posted on 19/01/2024. DOI: https://doi.org/10.3897/arphapreprints.e119131

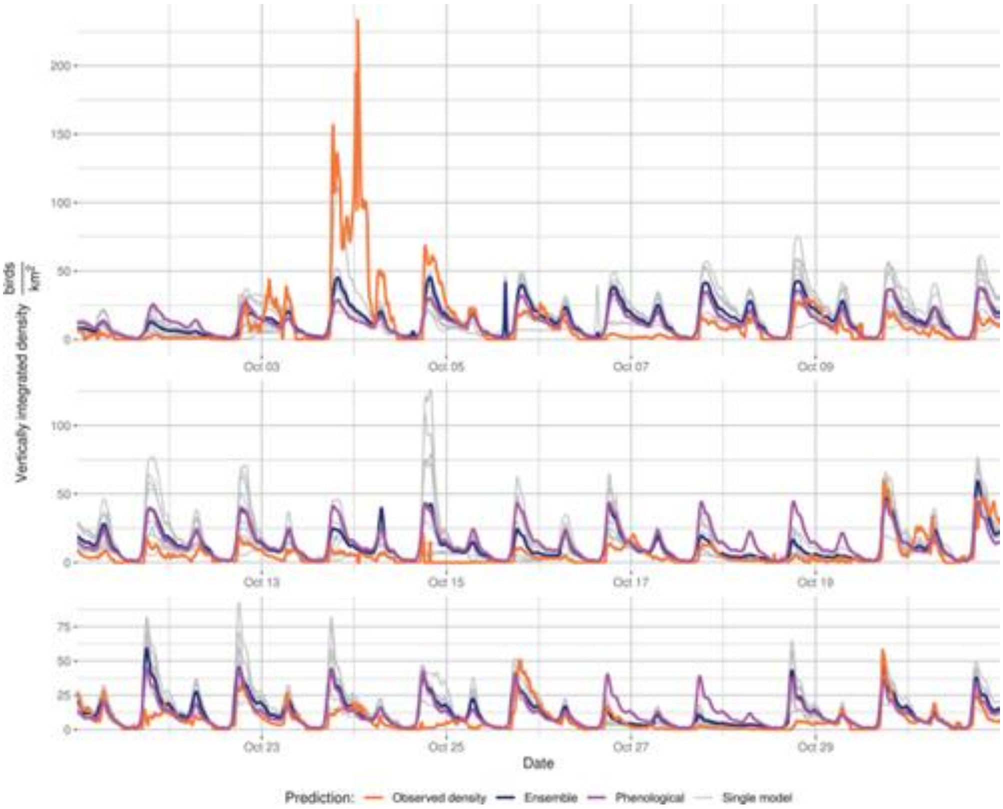

Figure 6. Observed and predicted bird density (birds/km2) in the Netherlands for October 2016. The orange line corresponds to the observed bird density, while the dark blue line corresponds to the ensemble forecast model. Other predictions are also represented: predictions from single environmental models are represented by the gray lines, and predictions from a phenological model are represented by the purple line (from Kranstauber et al. 2022).

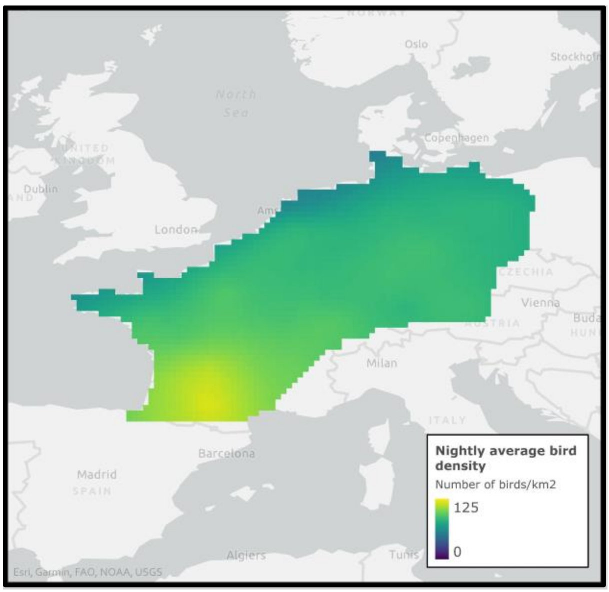

Figure 7. Map of nightly average bird density (birds/km2) on 18 October 2018, a night of intense migration in many regions in western Europe.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

4.3.2.3. Uncertainty assessment

Model performance was assessed by iteratively omitting single years from the dataset for which predictions are calculated. The predictive capacity of phenological models was compared to that of the models including environmental variables, and single models were compared with ensemble model predictions for evaluating the advantages of ensemble modelling. Additionally, receiver operating characteristic (ROC) curves were computed to visualize the predictive performance for BirdTAM (Bird-notice-to AirMen) density thresholds (10, 20, 40 birds/km2) of the phenological and ensemble model (Fawcett 2006, van Gasteren et al. 2019). Furthermore, the importance of the different environmental predictors was evaluated by analyzing how frequently they were selected in the cross-validations.

In general, the ensemble models considering local and remote weather conditions performed better than the baseline phenological models and than single environmental models. The use of such ensemble models is thus recommended in operational settings - such as providing warnings for aviation flight planning, to reduce bird aircraft collisions during intense bird migration (Kranstauber et al. 2022). When forecasts were produced separately for spring and fall (Kranstauber et al. 2022) there were differences in the magnitude of peak migration, in the best performing modelling approach and in the weather conditions selected as most important in each season, which should be further investigated and taken into account in future implementations of this workflow. Furthermore, the models can fail to adequately capture extreme (migration peaks) and rare events (Kranstauber et al. 2022), for which the inclusion of additional variables (such as the accumulation of easterly winds) can be useful, and which shows the importance of long training datasets and suitable data archives (Shamoun-Baranes et al. 2021). Similar models can also be valuable for civil aviation to support decisions (Nilsson et al. 2021) and to support shutdown or curtailment decisions for wind turbines (Marques et al. 2014).

5. Discussion

5.1. Advantages and breakthroughs of the forecasts

The workflows described here can significantly empower stakeholders from multiple sectors (policy, academia, NGOs, citizen science, businesses) by enabling them to anticipate ecological changes in the short term, thus facilitating proactive and timely decision-making. They offer significant, tangible contributions towards achieving multiple EU policy objectives and benefiting society at large.

The large diversity and relevance of potential gains from short-term ecological forecasting are clearly exemplified by the three test cases demonstrated. For example, mushroom fructification forecasts contribute to the bioeconomy and benefit EU citizens at large, also informing habitat and species conservation, and ecotourism. The invasive beetle forecasts can aid in pest surveillance and the timely implementation of control actions, thereby reducing crop damage and agricultural and economic losses. The aerial biomass of birds can inform aviation safety and wind farm curtailment, contributing to conservation, the economy, and citizen safety. It is noteworthy that these are just a limited number of possible applications, and the full potential of applications is much vaster. In this regard, it is relevant to highlight the demonstrated ability of the 'generic' workflow described to handle multiple, distinct

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

ecological phenomena, hence potentially allowing the wider potential of ecological short-term forecasts to be harnessed.

5.2. Caveats, outstanding challenges, and proposed solutions

However, there are still some caveats to the described workflows. The forecasts produced have relatively coarse spatial resolution and limited forecast-horizon extent, which largely reflect constraints in the available input data, particularly weather forecasts. Both workflows would also benefit from increased automation of certain procedures (e.g., in data collection and preprocessing steps), especially if aiming at an operational setting. In this regard, a relevant workflow component concerns the updating of model calibration (e.g., with data becoming available as time progresses), which is not currently automated in the workflows. Thus, future recalibration of models as new observational data becomes available still relies entirely on manual implementation. Additionally, we did not consider some of the weather data being produced and made available for Europe by ECMWF since these are not open access, being freely delivered for time-limited research projects only. The quality of the weather data provided by ECMWF is well recognized and to consider these data fully in the workflows would most likely result in further improvements. The forecasting workflows should also continuously seek improvements in other fronts. Specifically, improved understanding of the modeled systems would also be beneficial by help guiding model refinement, for example in deciding which predictors to include or whether more spatially general or locally specialized models are preferred.

Several "upstream" challenges also hinder the widespread development and operationalisation of short-term ecological forecasts in the EU. One significant issue is the widespread scattering and lack of harmonization of the ecological and biological data needed for developing forecasting workflows. Efforts of data inventorying and integration, as performed by the EuropaBON Project are thus crucial to lessen this obstacle. Moreover, ensuring the long-term sustainability and operational functionality of these workflows is challenging, depending on the availability of dedicated human personnel and IT infrastructure.

Despite the identified limitations of the described workflows and the upstream challenges in implementation and operationalization, we demonstrate the substantial potential of short-term iterative ecological forecasts for informing policy and citizens. The increasing availability of vast volumes of biological data across wide geographical scales, along with ongoing efforts in data mobilization and harmonization, indicates that it is an opportune time for increased and robust research investment in short-term ecological forecasting.

6. References

Balvanera P, Brauman KA, Cord AF, Drakou EG, Geijzendorffer IR, Karp DS, Martín-López B, Mwampamba TH, Schröter M (2022). Essential ecosystem service variables for monitoring progress towards sustainability. *Current Opinion in Environmental Sustainability* 54: 101152. https://doi.org/10.1016/j.cosust.2022.101152

Barker BS, Coop L, Wepprich T, Grevstad F, Cook G (2020). DDRP: real-time phenology and climatic suitability modeling of invasive insects. *PLoS One* 15: e0244005. https://doi.org/10.1371/journal.pone.0244005

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

Barrett RT (2011) Recent response to climate change among migrant birds in northern Norway. *Ringing & Migration* 26: 83-93. https://doi.org/10.1080/03078698.2011.587242

Belitz MW, Larsen EA, Shirey V, Li D, Guralnick RP (2023). Phenological research based on natural history collections: Practical guidelines and a lepidopteran case study. *Functional Ecology* 37: 234–247. https://doi.org/10.1111/1365-2435.14173

Bonney R (2021). Expanding the impact of citizen science. *BioScience* 71: 448–451. https://doi.org/10.1093/biosci/biab041

Buonaiuto DM, Donahue MJ, Wolkovich EM (2023). Experimental designs for testing the interactive effects of temperature and light in ecology: The problem of periodicity. *Functional Ecology* 37: 1747–1756. https://doi.org/10.1111/1365-2435.14329

Capinha C (2019). Predicting the timing of ecological phenomena using dates of species occurrence records: a methodological approach and test case with mushrooms. *International Journal of Biometeorology* 63: 1015-1024. https://doi.org/10.1007/s00484-019-01714-0

Capinha C, Ceia-Hasse A, Kramer AM, Meijer C (2021). Deep learning for supervised classification of temporal data in ecology. *Ecological Informatics* 61: 101252. https://doi.org/10.1016/j.ecoinf.2021.101252

Capinha C, Ceia-Hasse A, de-Miguel S, Vila-Viçosa C, Porto M, Jaric I, Tiago P, Fernandez N, Valdez J, McCallum I, Pereira HM (2023) Predicting the timing of ecological phenomena across regions using citizen science data. *bioRxiv* 2023.05.05.539567. https://doi.org/10.1101/2023.05.05.539567.

Castro A, Ribeiro J, Reino L, Capinha C (2023). Who is reporting non-native species and how? A cross-expert assessment of practices and drivers of non-native biodiversity reporting in species regional listing. *Ecology and Evolution* 13: e10148. https://doi.org/10.1002/ece3.10148

Ceia-Hasse A, Sousa CA, Gouveia BR, Capinha C (2023). Forecasting the abundance of disease vectors with deep learning. *Ecological Informatics* 78: 102272. https://doi.org/10.1016/j.ecoinf.2023.102272

Cutler DR, Edwards Jr TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ (2007). Random forests for classification in ecology. *Ecology* 88: 2783–2792. https://doi.org/10.1890/07-0539.1

De Cianni R, Varese GC, Mancuso T. (2023) A Further Step toward Sustainable Development: The Case of the Edible Mushroom Supply Chain. *Foods* 12: 3433. https://doi.org/10.3390/foods12183433

Dietze MC (2017). Ecological forecasting. Princeton, USA: Princeton University Press.

Dietze MC, Fox A, Beck-Johnson LM, Betancourt JL, Hooten MB, Jarnevich CS, Keitt TH, Kenney MA, Laney CM, Larsen LG, Loescher HW, Lunch CK, Pijanowski BC, Randerson JT, Read EK, Tredennick AT, Vargas R, Weathers KC, White EP (2018). Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proceedings of the National Academy of Sciences* 115: 1424-1432. https://doi.org/10.1073/pnas.1710231115

Diez JM, James TY, McMunn M, Ibáñez I (2013) Predicting species-specific responses of fungi to climatic variation using historical records. *Global Change Biology* 19: 3145-54. https://doi.org/10.1111/gcb.12278

Dokter AM, Liechti F, Stark H, Delobbe L, Tabary P, Holleman I (2011). Bird migration flight altitudes studied by a network of operational weather radars. *Journal of the Royal Society Interface* 8: 30–43. https://doi.org/10.1098/rsif.2010.0116

Dokter AM, Desmet P, Spaaks JH, van Hoey S, Veen L, Verlinden L, Nilsson C, Haase G, Leijnse H, Farnsworth A, Bouten W, Shamoun-Baranes J (2019). bioRad: biological analysis and visualization of weather radar data. *Ecography* 42: 852-860. https://doi.org/10.1111/ecog.04028

Dormann CF, Calabrese JM, Guillera-Arroita G, Matechou E, Bahn V, Bartoń K, Beale CM, Ciuti S, Elith J, Gerstner K, Guelat J, Keil P, Lahoz-Monfort JJ, Pollock LJ, Reineking B, Roberts DR, Schröder B, Thuiller W, Warton DI, Wintle BA, Wood SN, Wüest RO, Hartig F (2018). Model averaging in ecology: A review of Bayesian, information-theoretic, and tactical approaches for predictive inference. *Ecological Monographs* 88: 485–504. https://doi.org/10.1002/ecm.1309

EFSA (2019). Pest survey card on *Popillia japonica*. *EFSA Supporting Publications* 16: 1568E.

Elith J, Leathwick JR, Hastie T (2008). A working guide to boosted regression trees. *Journal of Animal Ecology* 77: 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

European Commission - Directorate-General for Research and Innovation (2018) *A sustainable bioeconomy for Europe – Strengthening the connection between economy, society and the environment – Updated bioeconomy strategy.* Publications Office. https://data.europa.eu/doi/10.2777/792130

European Commission (2020) A Farm to Fork Strategy for a Fair, Healthy and Environmentally-Friendly Food System. COM(2020) 381 Final. Brussels: European Commission. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0381

EU Regulation (2014) No 1143/2014 of the European Parliament and of the Council of 22 October 2014 on the prevention and management of the introduction and spread of invasive alien species. *Official Journal of the European Union* L 317: 35–55.

Fawcett T (2006). An introduction to ROC analysis. *Pattern Recognition Letters* 27: 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Fox J, Weisberg S (2019). An R Companion to Applied Regression. Third edition. Sage, Thousand Oaks CA. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Frick WF, Stepanian PM, Kelly JF, Howard KW, Kuster CM, Kunz TH, Chilson PB (2012) Climate and Weather Impact Timing of Emergence of Bats. *PLOS ONE* 7: e42737. https://doi.org/10.1371/journal.pone.0042737

Gonzalez A, Vihervaara P, Balvanera P, Bates AE, Bayraktarov E, Bellingham PJ, Bruder A, Campbell J, Catchen MD, Cavender-Bares J, Chase J, Coops N, Costello MJ, Dornelas M, Dubois G, Duffy EJ, Eggermont H, Fernandez N, Ferrier S, Geller GN, Gill M, Gravel D, Guerra CA, Guralnick R, Harfoot M, Hirsch T, Hoban S, Hughes AC, Hunter ME, Isbell F, Jetz W, Juergens N, Kissling WD, Krug CB,

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

Le Bras Y, Leung B, Londoño-Murcia MC, Lord J-M, Loreau M, Luers A, Ma K, MacDonald AJ, McGeoch M, Millette KL, Molnar Z, Mori AS, Muller-Karger FE, Muraoka H, Navarro L, Newbold T, Niamir A, Obura D, O'Connor M, Paganini M, Pereira H, Poisot T, Pollock LJ, Purvis A, Radulovici A, Rocchini D, Schaepman M, Schaepman-Strub G, Schmeller DS, Schmiedel U, Schneider FD, Shakya MM, Skidmore A, Skowno AL, Takeuchi Y, Tuanmu M-N, Turak E, Turner W, Urban MC, Urbina-Cardona N, Valbuena R, van Havre B, Wright E (2023). A global biodiversity observing system to unite monitoring and guide action. *Nature Ecology & Evolution*. https://doi.org/10.1038/s41559-023-02171-0

Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J, Nicolas J, Peubey C, Radu R, Schepers D, Simmons A, Soci C, Abdalla S, Abellan X, Balsamo G, Bechtold P, Biavati G, Bidlot J, Bonavita M, De Chiara G, Dahlgren P, Dee D, Diamantakis M, Dragani R, Flemming J, Forbes R, Fuentes M, Geer A, Haimberger L, Healy S, Hogan RJ, Hólm E, Janisková M, Keeley S, Laloyaux P, Lopez P, Lupu C, Radnoti G, de Rosnay P, Rozum I, Vamborg F, Villaume S, Thépaut J-N (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* 146: 1999-2049. https://doi.org/10.1002/qj.3803

Horton KG, Nilsson C, Van Doren BM, La Sorte FA, Dokter AM, Farnsworth A (2019). Bright lights in the big cities: migratory birds' exposure to artificial light. *Frontiers in Ecology and the Environment* 17: 209-214. https://doi.org/10.1002/fee.2029

Howard C, Stephens PA, Pearce-Higgins JW, Gregory RD, Butchart SHM, Willis SG (2020). Disentangling the relative roles of climate and land cover change in driving the long-term population trends of European migratory birds. *Diversity and Distributions* 26: 1442–1455. https://doi.org/10.1111/ddi.13144

Huber P, Kurttila M, Hujala T, Wolfslehner B, Sanchez-Gonzalez M, Pasalodos-Tato M, de-Miguel S, Bonet JA, Marques M, Borges JG, Enescu CM, Dinca L, Vacik H (2023). Expert-Based Assessment of the Potential of Non-Wood Forest Products to Diversify Forest Bioeconomy in Six European Regions. *Forests* 14: 420. https://doi.org/10.3390/f14020420

Isaac NJ, Pocock MJ (2015). Bias and information in biological records. *Biological Journal of the Linnean Society* 115: 522–531. https://doi.org/10.1111/bij.12532

Jackson-Blake LA, Clayer F, Haande S, Sample JE, Moe SJ (2022). Seasonal forecasting of lake water quality and algal bloom risk using a continuous Gaussian Bayesian network. *Hydrology and Earth System Sciences* 26: 3103-3124. https://doi.org/10.5194/hess-26-3103-2022

Kranstauber B, Bouten W, van Gasteren H, Shamoun-Baranes J (2022). Ensemble predictions are essential for accurate bird migration forecasts for conservation and flight safety. *Ecological Solutions and Evidence* 3: e12158. https://doi.org/10.1002/2688-8319.12158

Latombe G, Pyšek P, Jeschke JM, Blackburn TM, Bacher S, Capinha C,Costello MJ, Fernández M, Gregory RD, Hobern D, Hui C, Jetz W, Kumschick S, McGrannachan C, Pergl J, Roy HE, Scalera R, Squires ZE, Wilson JRU, Winter M, McGeoch MA (2017). A vision for global monitoring of biological invasions. *Biological Conservation* 213: 295-308. https://doi.org/10.1016/j.biocon.2016.06.013

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

Latorre J, de Frutos P, de-Magistris T, Martinez-Peña F (2021). Segmenting tourists by their motivation for an innovative tourism product: mycotourism, *Journal of Ecotourism* 20: 311-340. https://doi.org/10.1080/14724049.2021.1892123

Mansournia MA, Altman DG (2016). Inverse probability weighting. *BMJ* 352 :i189. https://doi.org/10.1136/bmj.i189

Marcenò C, Padullés Cubino J, Chytrý M, Genduso E, Salemi D, La Rosa A, Gristina AS, Agrillo E, Bonari G, del Galdo GG, Ilardi V, Landucci F, Guarino R (2021). Facebook groups as citizen science tools for plant species monitoring. *Journal of Applied Ecology* 58: 2018-2028. https://doi.org/10.1111/1365-2664.13896

Marques AT, Batalha H, Rodrigues S, Costa H, Pereira MJR, Fonseca C, Mascarenhas M, Bernardino J (2014). Understanding bird collisions at wind farms: An updated review on the causes and possible mitigation strategies. *Biological Conservation* 179: 40–52. https://doi.org/10.1016/j.biocon.2014.08.017

Metz IC, Ellerbroek J, Mühlhausen T, Kügler D, Hoekstra JM (2020). The Bird Strike Challenge. *Aerospace* 7: 26. https://doi.org/10.3390/aerospace7030026

Nilsson C, La Sorte FA, Dokter A, Horton K, Van Doren BM, Kolodzinski JJ, Shamoun-Baranes J, Farnsworth A (2021). Bird strikes at commercial airports explained by citizen science and weather radar data. *Journal of Applied Ecology* 58: 2029-2039. https://doi.org/10.1111/1365-2664.13971

Nussbaumer R, Bauer S, Benoit L, Mariethoz G, Liechti F, Schmid B (2021). Quantifying year-round nocturnal bird migration with a fluid dynamics model. *Journal of the Royal Society Interface* 18: 20210194. https://doi.org/10.1098/rsif.2021.0194

Pereira HM, Leadley PW, Proença V, Alkemade R, Scharlemann JP, Fernandez-Manjarrés JF, Araújo MB, Balvanera P, Biggs R, Cheung WW, Chini L, Cooper HD, Gilman EL, Guénette S, Hurtt GC, Huntington HP, Mace GM, Oberdorff T, Revenga C, Rodrigues P, Scholes RJ, Sumaila UR, Walpole M (2010). Scenarios for global biodiversity in the 21st century. *Science* 330: 1496-501. https://doi.org/10.1126/science.1196624

Pereira HM, Ferrier S, Walters M, Geller G, Jongman RHG, Scholes RJ, Bruford MW, Brummitt N, Butchart SHM, Cardoso AC, Coops NC, Dulloo E, Faith DP, Freyhof J, Gregory RD, Heip C, Höft R, Hurtt G, Jetz W, Karp D, McGeoch MA, Obura D, Onoda Y, Pettorelli N, Reyers B, Sayre R, Scharlemann JPW, Stuart SN, Turak E, Walpole M, Wegmann M (2013). Essential Biodiversity Variables. *Science* 339: 277-78. https://doi.org/10.1126/science.1229931

R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org

Semenza JC (2015). Prototype early warning systems for vector-borne diseases in Europe. *International journal of environmental research and public health* 12: 6333-6351. https://doi.org/10.3390/ijerph120606333

Shamoun-Baranes J, Bauer S, Chapman JW, Desmet P, Dokter AM, Farnsworth A, Haest B, Koistinen J, Kranstauber B, Liechti F, Mason THE, Nilsson C, Nussbaumer R, Schmid B, Weisshaupt N, Leijnse

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

H (2021). Weather radars' role in biodiversity monitoring. *Science* 372: 248–248. https://doi.org/10.1126/science.abi4680

Shamoun-Baranes J, Bauer S, Chapman JW, Desmet P, Dokter AM, Farnsworth A, van Gasteren H, Haest B, Koistinen J, Kranstauber B, Liechti F, Mason THE, Nilsson C, Nussbaumer R, Schmid B, Weisshaupt N, Leijnse H (2022). Meteorological Data Policies Needed to Support Biodiversity Monitoring with Weather Radar. *Bulletin of the American Meteorological Society* 103: E1234-E1242, doi: https://doi.org/10.1175/BAMS-D-21-0196.1

Shanubhogue A, Gore AP (1987). Using logistic regression in ecology. *Current Science* 56: 933–935. http://www.jstor.org/stable/24091356

Tourismus S (n.d.). Foliage map. Switzerland Tourism. Retrieved 20 October 2023, from https://www.myswitzerland.com/en/experiences/summer-autumn/autumn/foliage-map/

Tulloch AIT, Hagger V, Greenville AC (2020). Ecological forecasts to inform near-term management of threats to biodiversity. *Global Change Biology* 26: 5816-5828. https://doi.org/10.1111/gcb.15272

van Belle J, Shamoun-Baranes J, van Loon E, Bouten W (2007). An operational model predicting autumn bird migration intensities for flight safety. *Journal of Applied Ecology* 44: 864-874. https://doi.org/10.1111/j.1365-2664.2007.01322.x

Van Doren BM, Horton KG (2018). A continental system for forecasting bird migration. *Science* 361: 1115-1118. https://doi.org/10.1126/science.aat7526

van Gasteren H, Krijgsveld KL, Klauke N, Leshem Y, Metz IC, Skakuj M, Sorbi S, Schekler I, Shamoun-Baranes J (2019). Aeroecology meets aviation safety: early warning systems in Europe and the Middle East prevent collisions between birds and aircraft. *Ecography* 42: 899-911. https://doi.org/10.1111/ecog.04125

Weisshaupt N, Lehikoinen A, Mäkinen T, Koistinen J (2021). Challenges and benefits of using unstructured citizen science data to estimate seasonal timing of bird migration across large scales. *PLoS One* 16: e0246572. https://doi.org/10.1371/journal.pone.0246572

Wood SN, Goude Y, Shaw S (2015). Generalized additive models for large data sets. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 64: 139–155. https://doi.org/10.1111/rssc.12068

Zhang Z, Capinha C, Karger DN, Turon X, MacIsaac HJ, Zhan A (2020). Impacts of climate change on geographical distributions of invasive ascidians. *Marine Environmental Research* 159: 104993. https://doi.org/10.1016/j.marenvres.2020.104993

Zhang Z, Capinha C, Weterings R, McLay CL, Xi D, Lü H, Yu L (2019). Ensemble forecasting of the global potential distribution of the invasive Chinese mitten crab, *Eriocheir sinensis*. *Hydrobiologia* 826: 367-377. https://doi.org/10.1007/s10750-018-3749-y

Acknowledgments

The figures in this report incorporate graphical elements that are freely available for use from Project Noun, courtesy of the following individuals and entities: Tahsin Tahil, PEBIAN, Marc Torrada,

Dinosoftlabs, Alex Burte, Douglas Machado, Stefanie Peschel, M. Oki Orlando, Formidable Design Studio, Boris, and Fabiana Antonioli.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.