

# DuneFront

Deliverable 4.2

April 2025





# Biological boundary conditions

DuneFront – D4.2

Deliverable information

Title	Biological boundary conditions
Deliverable number	D4.2
WP number	4
Author(s)	Maxime Dahirel, Dries Bonte
Lead beneficiary	UGent
Contributors	
Type	Report
Dissemination level	Public
How to cite	Dahirel, M. and Bonte, D. (2025). Biological boundary conditions, version 1.0, DuneFront project deliverable 4.2, Ghent University.
Copyright license*	© Authors and DuneFront consortium, 2024–2027. <i>This work is openly licensed via <a href="#">CC-BY</a>. For the separate license on the data layers associated with the report, follow links under Data Availability Statement at the end of this report.</i>

Versioning and contribution history

Version	Date	Authors (Institution)	Notes
Version 0.1	2025-03-03	UGent	Draft created by Beneficiary A
Version 1.0	2025-04-08	UGent	Final version approved by all Beneficiaries

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.





## Cover page

*The goal of the DuneFront project is to develop research on Dune-Dike hybrid (DD-hybrid) Nature-based Solutions (NbS), to improve coastal protection in a context of climate change and rising sea levels. Coastal dune dynamics are intrinsically driven by aeolian processes that interact with vegetation. The design and development of dunes as Nature-based Solutions therefore needs an understanding of how their vegetation changes under various natural and anthropogenic pressures. This deliverable (4.2) aims to predict and map the coastal suitability for key dune vegetation species at the European level, providing a broad-scale picture that may serve as foundation for further or more local studies.*

*The study domain was chosen in collaboration with DuneFront work package 4.1 (WP4.1), and spans coastlines beyond the strict European region, in order to increase the transferability of model predictions: it includes the Mediterranean, Black Sea, Baltic Sea, and European Atlantic, Channel, and North Sea regions. We used the resources curated in WP4.1 to obtain geophysical, hydrodynamic and climate predictors used to fit our models, as well as to determine which sites to select. Given the focus on dune habitats, we selected only grid cells containing sandy coasts, leading to a study extent of 39150 1 km<sup>2</sup> sites. We collected biodiversity records from open repositories, focusing on about 20 species including major dune-building species, species of conservation interest and non-native introduced species. We created our Species Distribution Models using Bayesian Additive Regression Trees, a versatile method that allows for complex non-linear relationships between predictor and species, and carries over uncertainty to predictions in a principled Bayesian manner. Given the importance of coastal protection for urbanized coastal zones, we also developed separate occupancy models specifically to tease out the expected effect or urbanization from those of sampling bias in a way that is not possible otherwise from classical Species Distribution Models.*

*The models recover accurately geographical patterns broadly known from the literature, and are able to refine them and then link them to key predictors. In particular, we show that even for the species with the broadest ranges, some climatic variables always constrain suitability to some extent. Second, we show that the degree by which the coastline is open versus embayed is a major driver of dune species distribution, being recovered in most species. This likely reflects the physical dynamics that drives the development of dunes. We show nonetheless that responses to predictors vary widely between species, in ways that show no clear distinction between major dune-building species, species of conservation interest and non-native introduced species. Importantly, the use of a Bayesian method allows us not only to provide maps of suitability across Europe for all these species, but to associate each site-level value with an uncertainty. We show that this uncertainty can be quite high and is often non-randomly distributed: for many species the models are quite certain of where they are not present, but predictions are much more uncertain in areas of generally high suitability. Part of this uncertainty may reflect irreducible biological variability, but a substantial part is likely due to missing local predictors, which may be changed in future updates. With some caveats, which we remind the reader of, predictions from Species Distribution Models can be transferred to predicted future climates to make hypotheses about changes in future suitabilities. If key assumptions hold, we can expect that our focal species will generally respond negatively to climate change, with no clear patterns between dune-building species, species of conservation interest and introduced species. Finally, our use of occupancy models allows us to pinpoint precisely how urbanization affects our focal species. Interestingly here, we show a divide between native and introduced species: while the former show generally negative responses to urbanization, the latter show neutral to positive responses. This aligns with expectations on the role of cities as hubs for introduced species, and suggests a potential for conflicts between the coastal protection and biodiversity goals of dune Nature-based Solutions in cities, which need to be investigated further. Despite the limitations we discuss throughout this report, the delivered maps provide a first broad-scale overview of the suitability of sandy coasts to select dune plants, which is key information for evaluating the impact of DD-hybrids at the European level.*



# Table of Content

1.	Introduction.....	10
1.1	Work package 4 overview.....	10
1.1.1	General objectives.....	10
1.1.2	Work Package sub-tasks, milestones and deliverables.....	10
1.2	Goals of WP4.2.....	10
2.	Data sources.....	11
2.1	Delimitation of the study area.....	12
2.1.1	Region of interest.....	12
2.1.2	Reference coastline and reference grid.....	13
2.2	Environmental data.....	15
2.2.1	Geophysical and hydrodynamic variables.....	15
2.2.2	Climate.....	17
2.2.3	Urbanization.....	17
2.2.4	Ecoregions.....	18
2.2.5	Data filtering.....	18
2.3	Plant records.....	19
2.3.1	Defining the list of target species.....	19
2.3.2	Main data sources.....	21
2.3.3	Data cleaning and pre-processing.....	23
3.	Models.....	24
3.1	Environmental suitability maps: Bayesian Additive Regression Trees.....	24
3.2	Separating the effect of urbanization from the effect of sampling bias: occupancy models.....	26
4.	Results and Discussion.....	27
4.1	BART models.....	27
4.1.1	Models performance.....	27
4.1.2	Present-day species distribution patterns.....	28
4.1.3	Expected future shifts in suitability.....	33
4.2	Occupancy models.....	36

4.2.1	Differences in detection between species .....	36
4.2.2	Effect of urbanization.....	36
4.2.3	Important note .....	38
5.	Conclusion and perspectives .....	39
6.	Data availability statement.....	40
7.	References.....	41

## List of abbreviations

Abbreviation	Explanation
WP	Work Package
DD-hybrid	Dune-Dike hybrid
NbS	Nature-based Solution(s)
EUNIS	European Nature Information System
SDM	Species Distribution Model
EU, EEA	European Union, European Economic Area
GHSL	Global Human Settlement Layer
CHELSA	Climatologies at High resolution for the Earth's Land Surface Areas
SSP	Shared Socio-economic Pathways
SMOD	Settlement MODel
UC, DUC	Urban Centre(s), Dense Urban Cluster(s)
GBIF	Global Biodiversity Information Facility
CART	Classification And Regression Trees
BART	Bayesian Additive Regression Trees
PCA	Principal Component Analysis

# 1. Introduction

## 1.1 Work package 4 overview

This section is taken, with modifications, from the equivalent section in the WP4.1 report (Castelle & Dahirel, 2024).

### 1.1.1 General objectives

The primary goal of the DuneFront Work Package (WP) 4 is to quantify and map at the level of Europe, large scale physical, biological, and socioeconomic boundary conditions that are expected to influence the effectiveness of Dune-Dike-hybrid (DD-hybrid) Nature-based Solutions (NbS). The spatial data compiled and curated within WP4 can then be used to support the other DuneFront tasks, including but not limited to numerical modelling initiatives (WP7, WP11, WP13), ecological (WP10) and physical experimental campaigns (WP12), as well as data-driven assessments for up-scaling efforts (WP14).

### 1.1.2 Work Package sub-tasks, milestones and deliverables

WP4 is subdivided into three sub-WPs listed below, each associated with a deliverable, with milestone (M4.1) involving the compilation of all boundary conditions for transfer to other WPs:

WP4.1 - Physical Boundary conditions: with D4.1 (December 2024), a database of physical coastal boundary conditions for subsequent numerical/physical investigations, and an associated report (Castelle & Dahirel, 2024).

W4.2 - Distribution of important species: with D4.2 (April 2025, this report), a database of the suitability of European coasts for important species based on Species Distribution Models, providing essential data for data driven analysis and subsequent modelling.

W4.3 - Socio-economic/admin boundaries: with D4.3 (October 2024), a catalogue/database of socio-economic and administrative boundaries for subsequent data driven analysis and upscaling analysis (Lojek et al., 2024).

## 1.2 Goals of WP4.2

The stated goal of WP4.2 is to deliver maps of environmental suitability for a set of species key to coastal dune development, at high resolution across Europe under both current conditions and expected future scenarios. This differs from existing European-scale efforts from e.g. EUNIS which are focused on predicting *habitats*, not species (e.g. Hennekens, 2019). This can be divided into three steps:

- From the large set of plant species inhabiting European coasts (Van Der Maarel & Van Der Maarel-Versluys, 1996), **delineate a smaller, manageable set of species of conservation concern and/or functional importance** in the coastal dune context, and **collect publicly available occurrence data** for these species. This includes species known to contribute to dune growth, species listed on various conservation lists, and alien introduced species which have or might have a negative impact.
- Use the physical boundary conditions maps compiled in WP4.1, as well as other data sources, to compile an **ecologically relevant set of potential predictors** of species distributions. These should include where possible (e.g. climate) expected values under future scenarios.
- **Develop Species Distribution Models (SDMs)** that link each species' occurrence to these environmental predictors. The performance and shortcomings of these models should be critically assessed, before their output is projected spatially as maps of current and expected future suitability of the entire European coast for each species.

An overarching goal of DuneFront is to develop and maintain biodiversity in Dune-Dike hybrids as Nature-based solutions for **coastal protection**. As such, we may need to pay specific attention to urbanized coastal areas, where protection needs may be more pressing, and where the risks of sandy coast erosion may be exacerbated by hard “grey” coastal infrastructure (Nawarat et al., 2024). Urbanization is expected to have major disruptive impacts on biodiversity (see e.g. Aronson, 2023; Carlon & Dominoni, 2024; Knapp et al., 2021; McKinney, 2008; Szulkin et al., 2020), including potentially our plants of interest. While urbanization data can be added directly into classical Species Distribution Models, its effect there often cannot be separated from spatial biases in sampling (locations with more people and/or easier access are usually more likely to be sampled, e.g. Geurts et al., 2023). To better inform about the ways urbanization might alter the suitability of coasts to key plant species, we therefore added a supplementary step to this sub-WP:

- Develop models that can identify **the effects of urbanization** and separate them from those of sampling bias and imperfect detection. While these models may not translate to high resolution maps as the SDMs do, their outputs will help adjust expectations regarding plant distributions in urban vs. non-urban coasts.

## 2. Data sources

We detail here the various data sources we accessed to achieve WP4.2's goals, as well as the way they were processed and combined prior to analysis. In line with the open access goals of the European Union, we only used openly available data sources.

Building broad-scale distribution models of coastal species brings additional challenges, linked to the linear and moving nature of the coastline. As a result, several SDM methods explicitly based on 2D space cannot be used directly, and many coastal-specific data sources of interest rely on different underlying delimitations of the coastline, which requires careful data reconciliation and compromise before proceeding to the analyses proper. We explicitly detail below the choices we make in this respect.

## 2.1 Delimitation of the study area

### 2.1.1 Region of interest

The area of interest in this report is defined the same way as in [Castelle and Dahirel \(2024\)](#) (**Fig. 1**) and can be summarized as: the European Union “mainland” coastline, as well as biogeographically relevant adjacent areas. In more details, it includes the entire Mediterranean, Black Sea and Baltic Sea coastlines *sensu lato* (i.e. including marginal seas such as the Adriatic), and the European Atlantic (including marginal seas such as e.g. the Irish Sea), Channel and North Sea coasts from Gibraltar up to the approximate latitude of the Arctic circle in Norway (matching the latitudes of the northernmost coastal points in mainland EU territory, in the Baltic Sea). Islands that fall within these boundaries were included (though very small islands may be excluded during the data filtering process). Extra-European territories of EU and EEA member states (Outermost Regions and Overseas Countries and Territories) were excluded as they are biologically and geographically distinct from the European mainland (this biological criterion for exclusion extends to even the nearest Outermost Regions, the Canaries, Azores and Madeira, see [Hernández-Cordero et al. \(2015\)](#))



*Figure 1: Coastlines (in bold black) studied in WP4.1 and WP4.2. Note that coastlines are highlighted here independently of sediment type, but only sandy sites were used in WP4.2 analyses.*

### 2.1.2 Reference coastline and reference grid

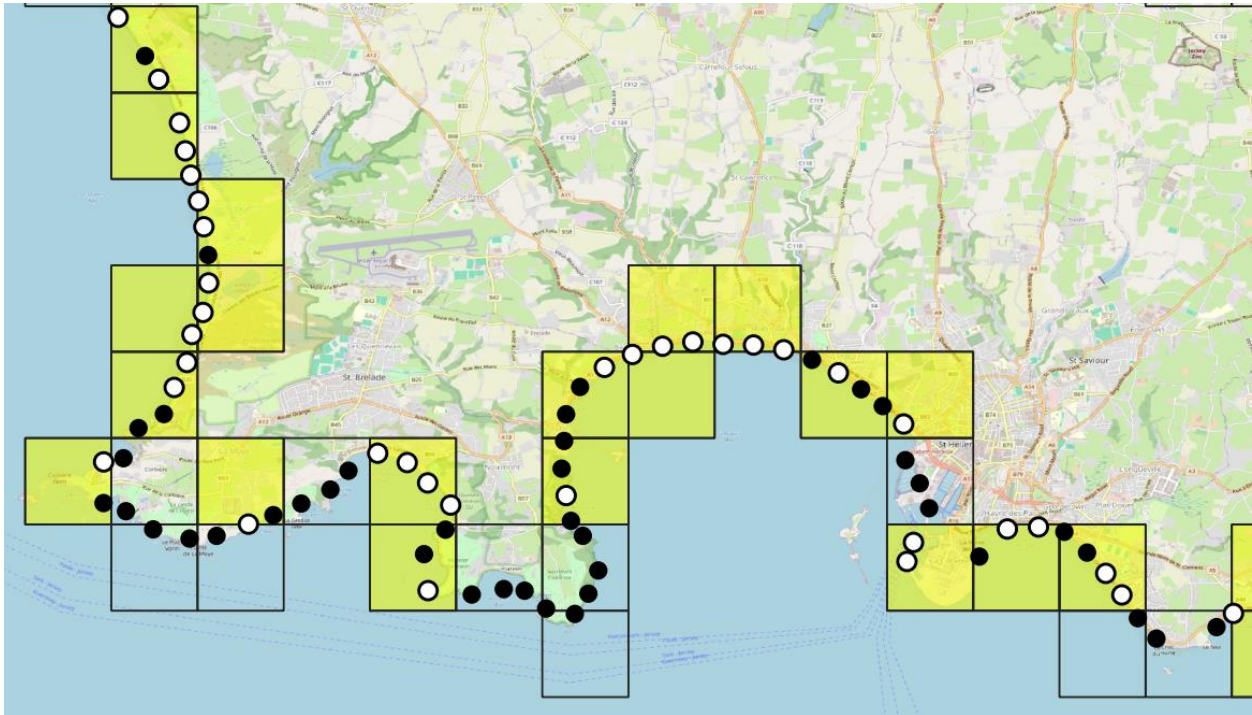
Publicly available data sources vary in where they place the coastline, both due to resolution and definition choices, with deviations sometimes reaching hundreds of metres. Following others (e.g. Athanasiou et al., 2024; Hulskamp et al., 2023; Lujendijk et al., 2018), we use OpenStreetMap as our reference here (OpenStreetMap contributors (2025), data as of 2025-01-05, downloaded from <https://osmdata.openstreetmap.de/>). The OpenStreetMap coastline is generally drawn at a much finer grain than most other Europe-wide sources, and is explicitly aiming to be based on mean high water springs (<https://wiki.openstreetmap.org/wiki/Coastline>). Contrary to e.g. Athanasiou et al. (2024), we use the highest resolution of this coastline, and do not smooth out local variations: as we will need to filter plant observations that are precisely georeferenced, this “anchor” coastline needs to be as precise as possible.

The statistical methods used in occupancy or distribution models are generally based on a discretization of the area of interest, with suitability/probability of occurrence defined at the level of the site or of the grid cell. The choice of the target resolution is thus a key parameter to consider

prior to the analysis. Especially because observation records are geographically biased (Isaac & Pocock, 2015), there is a fundamental trade-off between accuracy and usefulness tied to spatial resolution (Boyd et al., 2024). On the one hand, studies at very fine resolutions can more directly link species responses to the environment, but will often have poor accuracy, because of sampling error, and because many sites will simply not have been visited. On the other hand, studies at coarser resolutions will be more accurate (at the limit, to directly cite Boyd et al. (2024), “we can be surer a species occupies planet Earth than a set of small plots on its surface”); but at the cost of understanding how the target species responds to key local environmental drivers.

Europe-wide maps at  $\leq 50$  km resolution (or even  $\approx 10$  km locally) have long been available for many dune plant species (see for instance the Biological Flora of the British Isles series, which despite its name also includes European maps in all articles; Davy et al. (2006); Huiskes (1979); Bond (1952); Sanyal & Decocq (2015)). This provides an upper limit on the target resolution. At the other end of the scale, resolutions finer than 1 km are probably not workable (at least not without making very strong assumptions), as (i) most environmental predictors of interest are available at European scale only at grid resolutions of a few hundreds metres to a few km, (ii) the overwhelming majority of sites would be unvisited and contain no plant records *of any species*, let alone focal ones, (iii) location errors when recording plants may become non-negligible at the smallest grid resolutions.

Therefore, we chose for this deliverable a reference grid of 1 km resolution, as in e.g. Outhwaite et al. (2019). As we use the Global Human Settlement Layer project (Joint Research Centre (European Commission), 2023) as a source of environmental data related to urbanization in our analyses (see below), and they provide ready-made raster layers with nominal resolution 1 km (using the Mollweide projection, ESRI:54009), we used the grid underlying these layers as our reference grid. The GHSL 1 km grid cells that intersect with the OpenStreetMap coastline therefore form the initial set of potential “sites”, which was further filtered out based on the presence of sandy coasts (**Fig. 2**, and see next).



**Figure 2:** *Interplay between the GHSL reference grid used in analyses in this report and the shoreline data from [Castelle and Dahirel \(2024\)](#), using the western and southern coasts of Jersey as an example. Shoreline transects midpoints are marked as white if their sediment type was classified as sandy, black otherwise. 1 km<sup>2</sup> grid cells that intersect with the coastline are shown. Grid cells that contain at least one sandy point are highlighted in yellow; those are the ones used in further analyses. Map background is © [OpenStreetMap contributors](#), under the [Open Database License \(ODbL\)](#).*

## 2.2 Environmental data

We provide an overview of the environmental layers used in our analyses [Table 1](#). Further details are given below.

### 2.2.1 Geophysical and hydrodynamic variables

The layers produced in DuneFront WP4.1 ([Castelle & Dahirel, 2024](#)) contain the midpoints of a series of reference transects, spaced every 300-500 m along the entire coastline, taken from previous studies ([Castelle et al., 2024](#); [Hulskamp et al., 2023](#); [Luijendijk et al., 2018](#)). As the transects were made by alignment to a smoothed, lower resolution, of the OpenStreetMap coastline, we first realigned these midpoints to our reference coastline by a nearest neighbour approach. Each transect contains information about whether the coastline it intersected with was sandy, rocky, vegetated, muddy or “other”. After realignment, we classified as “sandy” all grid cells that contained at least one “sandy” transect ([Fig. 2](#)).

**Table 1:** Variables used in WP4.2 analyses, with their data sources.

	Variable name	Source
<b>Geophysical variables</b>		
	Sediment type (sandy/not sandy)*	Castelle and Dahirel (2024) (compilation and curation of various primary sources, see there for full list)
	Coastal orientation (angle relative to North)	
	Shoreline trend (accretion/erosion rates)	
	Orthogonal distance to nearest shore	
<b>Hydrodynamic variables</b>		
	100-year return storm surge	Castelle and Dahirel (2024) (compilation and curation of various primary sources, see there for full list)
	Wave height	
	Wave period	
	Highest astronomic tide	
	Tide range	
<b>Climate variables (averages over 30-year climatological period)</b>		
	Annual daily temperature**	CHELSA ( <a href="https://chelsa-climate.org/">https://chelsa-climate.org/</a> ) (Karger et al. 2017, 2019, 2020, 2021)
	Average daily minimal temperature of coldest month**	
	Average daily maximal temperature of warmest month**	
	Annual precipitation**	
	Wind speed	
	Average wind speed of windiest month	
<b>Bias-control covariate</b>		
	Distance to nearest city***	GHSL-SMOD (Schiavina et al. 2023)

\*not included as predictor in models, used to define which sites are modelled (**Fig. 2**)

\*\*The only variables in the dataset for which future projections under various climate change scenarios are also available

\*\*\*Distance to the nearest DUC or UC, whichever is closest for each site.

These transects also contain information about other geophysical parameters, such as the rate of coastal accretion or erosion, the coastline orientation, or the orthogonal distance to the nearest shoreline, a measure of whether a site was exposed to the open sea (high distances) or embayed (low distances). They also contain hydrodynamic information about wave parameters (height, period), tides (range, highest astronomic tide) and storm surges. For each of these except distance to the nearest shoreline, we calculated for each grid cell the mean value of its transects, using the circular mean for coastal orientation as it is an angle. For the distance to the nearest shoreline,

values in the WP4.1 dataset were capped after 200 km. Because of this, summarizing transects using the mean would lead here to biases, so we used the minimum value in each grid cell.

### 2.2.2 Climate

As in [Castelle and Dahirel \(2024\)](#), high resolution climate data for the “present”/recent past (1989-2018) were obtained from the CHELSA database (Climatologies at High resolution for the Earth’s Land Surface Areas, <https://chelsa-climate.org/>, version 2.1, [Karger et al., 2017, 2019, 2020, 2021](#)). Here however, rather than simply retaking the point values associated with each transect, we recalculated the average values of each climate variable within each grid cell from the source raster layers developed during WP4.1. With the exception of wind speed (for which the data was not available), 2041-2070 and 2071-2100 predictions were also obtained from CHELSA under three Shared Socio-economic Pathways (IPCC, 2023; SSPs, [Riahi et al., 2017](#)):

- SSP1-RCP2.6 (low greenhouse gas emissions, “Sustainability – Taking the Green Road” narrative),
- SSP3-RCP7 (high greenhouse gas emissions, “Regional Rivalry – A Rocky Road” narrative),
- SSP5-RCP8.5 (very high greenhouse gas emissions, “Fossil-fueled Development – Taking the Highway” narrative).

### 2.2.3 Urbanization

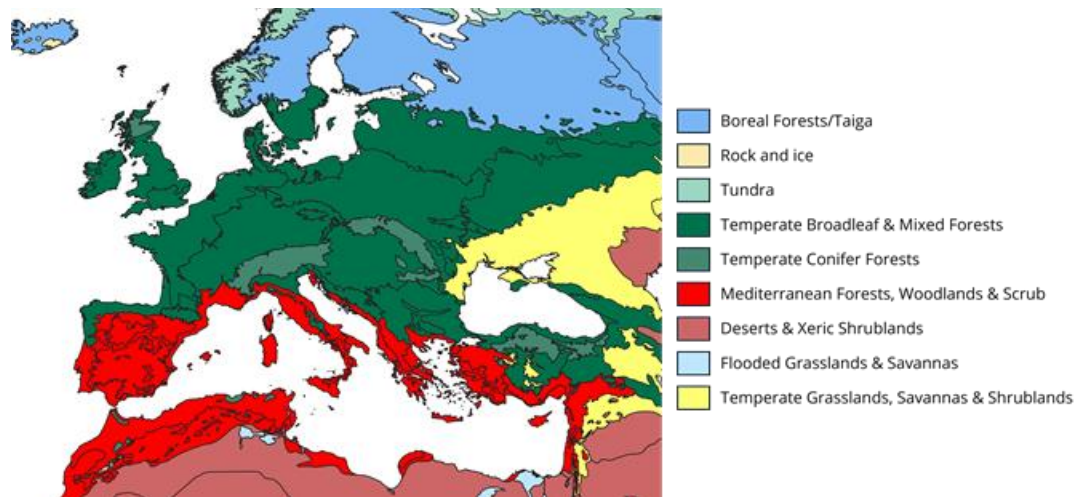
The Global Human Settlement Layer project (GHSL, Joint Research Centre (European Commission), 2023) is an EU project providing open and globally standardized data on human presence across the planet, including population density, built-up area and the structure of the urban fabric. The GHSL has rapidly become a key resource in urban studies and urban ecology ([McCallum et al., 2022](#); e.g. [Pacifici et al., 2020](#); [Rezaei & Millard-Ball, 2023](#); [Richardson et al., 2023](#); [Venter et al., 2021](#)). We used the Degree of Urbanization as recorded in the Settlement Model layers (GHS-SMOD, Eurostat (European Commission), 2021; [Schiavina et al., 2023](#)). The Degree of Urbanization methodology is a hierarchical classification based on population density, built-up area and contiguity rules. At the highest level, grid cells are classified as either part of continuous high-density Urban Centres (UC), as smaller Dense Urban Clusters (DUC) or as low-density “rural” cells. For each coastal site, we determined whether it belonged to an Urban Centre or Dense Urban Cluster, and measured the distances to the nearest UC and to the nearest DUC (from centroid of focal grid cell to perimeter of nearest UC/DUC; the value was 0 if the cell was part of a UC/DUC).

## 2.2.4 Ecoregions

For the models aiming to separate the effect of urbanization on occurrence from sampling biases (see below), preliminary tests showed it was not computationally practical to include all the relevant environmental layers, and especially their possible non-linear effects. Nonetheless, one should still attempt to control for the potential effect of these variables on baseline occupancy. As climate and other environmental variables vary across biomes and ecoregions (nested in biomes), including these as random effects in models offers a way to implicitly account for this biogeographical variation without worrying about a large number of parameters or their functional forms. We used the Ecoregions 2017 classification (Dinerstein et al., 2017), an update of the widely used Olson et al. (2001) classification (**Fig. 3**). Again, the coastlines do not systematically align between layers, resulting in some coastal grid cells in our reference system falling into the sea in the Ecoregions 2017 database. We used a nearest neighbour approach to assign biome and ecoregion values to each site.

## 2.2.5 Data filtering

We filtered the set of coastal grid cells to only keep the 39,350 “sandy” sites, which we then combined with the various other environmental datasets described above. A small number of sites have missing data for some of the variables coming from Castelle and Dahirel (2024) (see there for details). After removing those from the dataset, the extent on which we attempted to predict suitability consists of 39,150 sandy grid cells.



**Figure 3:** Biomes and ecoregions of Europe and the surrounding regions, according to Dinerstein et al. (2017). Colour code follows Dinerstein et al. (2017) and distinguishes between biomes, while lines within the same biome delineate ecoregions.

## 2.3 Plant records

### 2.3.1 Defining the list of target species

At least 2400 vascular plant species or subspecies can be found on European coastal dunes (Marcenò et al., 2018), and of these, around 300 could be considered as typical dune taxa (i.e. not usually found outside of dunes, Van Der Maarel & Van Der Maarel-Versluys, 1996). To fulfill the goals of this deliverable in a timely fashion, a much smaller set of key or reference species had to be defined. With respect to the present work package, the DuneFront project proposal mentions the need to investigate “species of conservation concern, invasion risk and dune-building capacity”. From the lists in Van Der Maarel & Van Der Maarel-Versluys (1996) and Marcenò et al. (2018), and the list of alien dune species in Giulio et al. (2020), we aimed to refine a final shorter list of about 20 species of interest covering all three of these categories, using a multi-step approach. First, we elicited expert suggestions from the ecologists involved in the DuneFront project, which combined with a qualitative review of the literature led to a list of about 50 species. Then, for the “dune-building” and “conservation concern” categories, we examined the official definitions of Annex I dune habitats under the EU Habitats directive, as well as interpretation documents extending and detailing these definitions for practitioners (Bensettiti et al., 2004; Directorate-General for Environment (European Commission), 2013). Within the list of expert suggestions, we prioritized species explicitly mentioned in these habitat definitions. Given the focus of DuneFront, we prioritized species of embryonic and mobile dunes (2110, 2120) and Baltic vegetated sandy beaches (1640), but also considered mentions in driftline vegetation (1210) and fixed dunes (other 21xx habitats, 22xx) where deemed relevant by consulted experts. One species (the hybrid species *Calamagrostis × baltica*) was removed from the set later at the modelling stage, as there were not enough records in the selected datasets for analyses (< 50 sites with records). For “invasion risk” species, we relied on the expert suggestions, the official EU list of Species of Invasive Concern (Commission Implementing Regulation (EU) 2022/1203 of 12 July 2022 Amending Implementing Regulation (EU) 2016/1141 to Update the List of Invasive Alien Species of Union Concern, 2022) and the list in Giulio et al. (2020) to define a shortlist of up to five taxa. The final list of modelled species is presented **Table 2**.

**Table 2:** Final list of retained plant species, by category.

Current accepted name (GBIF Backbone Taxonomy)	Number of 1 km <sup>2</sup> sites with records (2015-2024)
<b>A - Major dune-building species*</b>	
<i>Calamagrostis arenaria</i> (L.) Roth	751
" <i>Elytrigia juncea</i> " species complex ( <i>Thinopyrum</i> spp)**	130
<i>Leymus arenarius</i> (L.) Hochst.	448
<b>B - Species of conservation importance</b>	
<i>Achillea maritima</i> (L.) Ehrend. & Y.P.Guo	516
<i>Anthemis maritima</i> L.	197
<i>Cakile maritima</i> Scop.	3789
<i>Cyperus capitatus</i> Vand.	212
<i>Echinophora spinosa</i> L.	389
<i>Eryngium maritimum</i> L.	2825
<i>Euphorbia paralias</i> L.	1446
<i>Helichrysum stoechas</i> (L.) Moench***	457
<i>Honckenya peploides</i> (L.) Ehrh.	1593
<i>Medicago marina</i> L.	653
<i>Pancratium maritimum</i> L.	1919
<i>Polygonum maritimum</i> L.	371
<i>Salsola kali</i> L.	734
<b>C - Introduced species non-native in the study area</b>	
<i>Baccharis halimifolia</i> L.	112
<i>Carpobrotus acinaciformis</i> (L.) L.Bolus****	334
<i>Carpobrotus edulis</i> (L.) N.E.Br.****	1078
<i>Rosa rugosa</i> Thunb.	1353
<i>Senecio inaequidens</i> DC.	369
<b>D - Special case</b>	
<i>Hippophae rhamnoides</i> L.*****	787

\*While some of the species listed in the "Conservation concern" category can also contribute to initial sand deposition in embryonic dunes (e.g. Davy et al., 2006), these grass species are widely acknowledged as the major dune builders along European coasts (Doody, 1991; Maun, 2009).

\*\*The dune grass referred to as *Elytrigia juncea* or *Elymus farctus* in older (or even recent) publications (Bensettiti et al., 2004; Directorate-General for Environment (European Commission), 2013; e.g. Doody, 1991; Marcenò et al., 2018; Maun, 2009; Van Der Maarel & Van Der Maarel-Versluys, 1996) is now split into two accepted species by most reference databases' taxonomies, including GBIF: *Thinopyrum junceiforme* (Á.Löve & D.Löve) Á.Löve and *Thinopyrum junceum* (L.) Á.Löve, which were previously considered as two subspecies (*boreoatlantica* and *juncea*, respectively, Tison & de Foucault, 2014; Banfi, 2018). However, while they are generally described as having distinct ranges (Atlantic/North Sea/Baltic and Mediterranean, respectively), there is actually substantial geographic overlap in GBIF between records assigned to the two new species. This may result (from our examination of records and literature) from a

*combination of genuine sympatry, identification errors, and ambiguous reassignments of older data that antedate the species split. Rather than attempting to sort these, we decided for the present report to treat and model this taxon as one species complex. New data and further literature investigation might allow us to revise this choice in potential updates.*

*\*\*\*Helichrysum stoechas is mainly a species of grey dunes and dune slacks but can be found on the sheltered inland side of white dunes (e.g. Demichele et al., 2025).*

*\*\*\*\*Carpobrotus acinaciformis and C. edulis can sometimes be confused with one another (the former has pink-purple flowers, but the latter's yellow flowers can fade to pink as they age) and additionally can hybridize (Campoy et al. 2018). In the present report, we present results based on data from the two species as they were recorded. However, we additionally ran models, and propose optional maps, treating the complex of the two species as a whole.*

*\*\*\*\*\*Hippophae rhamnoides is a special case in this list in that it is native from the overall study area, and the definitional species of one coastal Annex I habitat (2160 - Dunes with H. rhamnoides). However, its growth can negatively impact other dune habitats, especially but not only in areas of the UK and Ireland where it is not native (Isermann et al., 2007; Richards & Burningham, 2011).*

### 2.3.2 Main data sources

The Global Biodiversity Information Facility (GBIF) is a massive “international network and data infrastructure funded by the world’s governments and aimed at providing anyone, anywhere, open access to data about all types of life on Earth” (<https://www.gbif.org/what-is-gbif>). The GBIF website provides a central hub from which researchers can access open biodiversity data from a myriad of small and large data providers, including research groups, museums or crowdsourced biodiversity data collection apps, all under a common standardised format (including a standardized species taxonomy). It is a *de facto* one-stop shop for many Species Distribution Modelling applications.

As mentioned above, the GBIF compiles information from a variety of data sources, which can vary in temporal and spatial extent, but most importantly in spatial resolution. Some data sources provide exact coordinates, while some only provide open information at administrative unit level or on a coarser resolution grid, with exact locations either unavailable or withheld. While many coarser-grained SDM applications often download all observations of the relevant species irrespective of this (with some post-hoc filtering of edge cases), our goal of (relatively) finer-grained SDMs means we have to be more selective *a priori*.

We explored the GBIF website for datasets that fulfilled the following criteria:

- Provides point coordinates for each observation, i.e. excluding gridded data.
- Covers the whole extent of the study area (though see below for precisions on that). We aimed to avoid national level datasets, on the basis that not all focal countries had one in GBIF. Their inclusion could accentuate potential issues linked to sampling biases.

Moreover, many national-level datasets also only provided gridded data, failing the previous criterion. This filter may be relaxed in future updates.

- Provides large numbers of vascular plant (Trachaeophyta) observations.
- Some degree of human input in the species identification. This excludes datasets containing solely machine learning based identification suggestions. While computer-vision identification of species can often be quite accurate (e.g. <https://www.inaturalist.org/blog/104015-new-computer-vision-model-v2-18-with-nearly-1-500-new-species>), the error rate in the absence of human validation remains non-negligible, and cannot be accounted for by the modelling methods we use here.

This resulted in a primary set of three datasets, though others may be added in future updates. All three datasets come from portals dedicated to opportunistic crowdsourced biodiversity observation, which have become a major source of biodiversity data in recent years:

- iNaturalist Research-Grade Observations (<https://www.inaturalist.org/>; <https://doi.org/10.15468/ab3s5x>): Georeferenced records uploaded to iNaturalist are subject to crowdsourced validation by other users based on provided proofs of observation (typically photographs). Observations reach “Research-grade” status only if at least 2/3 of the contributing users agree on an identification; that status is dynamic and can be reversed if new identifications disagree. While iNaturalist provides well-regarded computer vision suggestions, the final decisions always come from human users.
- Observation.org data (<https://observation.org/>; <https://doi.org/10.15468/5nilie>): observations are subjected to expert validation either manually based on photographic proof, or automatically through predefined criteria based on species life histories, identification difficulty and known broad-scale distributions. While image recognition tools can be used, the final decision always also relies on the above expert validation systems. We note that while most of global Observation.org data are made available with point coordinates, data from Belgium and the Netherlands are under different management regimes and are only openly available at coarse (> 4 km) resolution, which is unsuitable for our purposes. These were excluded before further analyses.
- Pl@ntNet observations (<https://plantnet.org/>; <https://doi.org/10.15468/gtebaa>): as with iNaturalist, identifications are based on photographic proofs, with here a combination of human input and automated identification used. Pl@ntNet also provides a dataset containing records that have been confidently identified only by the deep learning algorithm, without human input (<https://doi.org/10.15468/mma2ec>); despite being one order of magnitude larger, we did not use it, in accordance with our selection criteria.

For each of these three datasets, we downloaded on January 16, 2025 all observations of vascular plants (Trachaeophyta) within the study region for the 2015-2024 decade (see **Data availability statement**).

### 2.3.3 Data cleaning and pre-processing

We filtered plant observations, of any species, to keep only those occurring within the extent defined in **Environmental data**. We further removed observations that were too imprecise and might have their true locations outside the focal grid cells (documented precision > 250 m; we include in these the Belgian and Dutch Observation.org data, as mentioned above). We then formatted the data in two ways, for each modelling method.

For the environmental suitability maps, we thinned the data to keep only one observation per grid cell, for each of the focal species **Table 2**. That is, was the species recorded at least once at this site, or not, during the 2015-2024 decade?

For the models investigating the effects of urbanization, we divided the data in recorded visits, with one visit containing all the observations made in one grid cell by any observer on one day (a visit may aggregate observations from multiple observers). During the 2015-2024 decade, 149,658 visits where at least one vascular plant was recorded were made to 19,376 out of the 39,350 sandy sites. For each visit we again determined whether the focal species **Table 2** were recorded at least once, or not. The definition of which taxa to use to define a visit is a key aspect of applying occupancy models to opportunistic data (see below), as it determines the set of sites that were visited but had no detections of the focal species. Our choice here implies that users that recorded any vascular plant within a grid cell could plausibly have also recorded one of our focal species. This is not an unreasonable assumption given the small size of the grid cells, and the fact that plant observations mostly do not require specialized equipment. We considered using narrower reference groups, but issues specific to sandy coastal ecosystems appeared:

- We could have set a target group based only on the set of species found strictly on sandy coasts (based on e.g. Van Der Maarel & Van Der Maarel-Versluys, 1996). However, many urban sandy coasts are heavily managed beaches, where there may be no plant at all to be found on the coastline itself but recorder activity may exist immediately inland in the city. Using a baseline based strictly on sandy coast plants would mean these sites would be excluded as having received no visits. We wanted here to include beach management as a key component of the effect of urbanization on sandy habitats, so needed to include these sites. Different choices may be made depending on the question of interest.
- Another possibility would have been, rather than using the same reference group for all species, to set taxonomically narrower, species-specific references, for instance using detections of grasses (Poaceae) as the reference group for dune grasses. However, we run into the issue that dunes are often dominated by one or a few very common and drab grass species. We can expect that many dune visitors would record non-grass, visually

distinct dune species, while ignoring the dominant grass (especially in areas where they assume, or know, the dune grass was planted rather than spontaneous). Using taxonomically narrower reference groups may remove such visits from the data pool, which we do not consider acceptable as they are clearly informative about detection.

## 3. Models

All analyses were done in R, version 4.4.2 (R Core Team, 2024).

### 3.1 Environmental suitability maps: Bayesian Additive Regression Trees

Classification and regression trees (CARTs) are popular tools in Species Distribution Modelling, in large part due to their ability to handle the expected non-linear relationships between environmental predictors and species occurrences without having to specify explicit functional forms, thanks to their reliance on nested binary splits. Popular CART methods such as random forests or boosted regression trees generally perform well in SDM benchmarks, but have known shortcomings, such as tendencies to overfit in the absence of careful tuning, or the absence of principled ways to describe uncertainty in model outputs (Carlson, 2020).

Bayesian Additive Regression Trees (BARTs) have been used in other disciplines since their inception by Chipman et al. (2010), but have been more recently popularized for species distribution modelling by Carlson (2020). BART differs from other CART methods in that it is Bayesian, meaning that outputs are explicitly samples from a posterior distribution, carrying over uncertainty to predictions in a principled manner. As BART relies on a sum-of-many weak trees approach, the probability that a variable is retained among the set of trees at each iteration can provide a measure of relative variable importance. It relies only on a small set of priors, for which suitable ranges of default values and distributions have been established, facilitating its out-of-the-box use. These priors are also typically set in a way that penalises tree depth, limiting overfitting. We send the reader to Carlson (2020) for a more thorough overview suitable for ecologists, from which the details in this paragraph are largely drawn.

When using presence-only data as we do here, many SDM methods, including BART, need the user to define a sample of pseudo-absences points from the set of sites in which the species has not been detected (which does not mean it wasn't present, hence "pseudo-"absences). There are *many* ways to define pseudo-absences, including purely random sampling or sampling that accounts for the distance to presences in geographic or environmental space, with extensive discussion of the advantages and shortcomings of each method in the literature, and ongoing development of new methods. For each species, we sampled pseudo-absences using the uniform method defined in Da Re et al. (2023), which was shown to perform as well or better as common alternatives in simulations. Briefly, it is based on sampling the environmental space (defined using a PCA on the environmental predictors) uniformly rather than randomly (to better sample rare

vs. overrepresented environments in the study area), while avoiding sampling the parts of the environmental space where presence records are. We excluded distance to cities from the set of variables used in defining the environmental space for this, as it is meant to be a “control” variable rather than a “true” environmental variable of interest. As BART is sensitive to class imbalance in the training data, we set for each species to sample a number of pseudo-absences equal to the number of presences (Carlson, 2020; Carlson et al., 2022) (the specifics of Da Re et al. (2023) mean that an exact 50-50 split was not always reached, but deviations were minor).

For each species, we fit a BART model to the presence and pseudo-absence data, using the *embarcadero* package (Carlson, 2020) as a wrapper around the *dbarts* package (Dorie et al., 2024), using 4 chains with 1000 iterations each post burn-in and otherwise default settings. We initially included all geophysical, hydrodynamic and present-day climate variables mentioned in **Table 1** as predictors, as well as distance to the nearest city (whether UC or DUC) as a bias control covariate. We then used the model selection procedure embedded in *embarcadero* (based on variable importance) to remove potentially uninformative variables. While this tends to perform well for BARTs specifically (Carlson, 2020), the shortcomings of stepwise selection in general are well known (Whittingham et al., 2006). Therefore, while we only review the reduced models (post-model selection) in the rest of this report, we make available map layers based on both the full and reduced models (**Data availability**).

We used a random 80%/20% training/test split, splitting presences and pseudo-absences separately. Using the posterior mean predictions, we collected multiple metrics of model’s performance, as usually recommended (e.g. Da Re et al., 2023): the area under the receiver-operator curve (ROC-AUC), the continuous Boyce index (Hirzel et al., 2006), the maximized True Skill Statistic (Allouche et al., 2006) and the corresponding sensitivity and specificity (Kuhn & Wickham, 2020; Leonardi et al., 2024). We estimated and provide all these metrics for both the training and test sets. In this report, we focus especially on the Boyce index, as it is the only one of these metrics specifically designed for “presence-only + pseudo-absence” workflows. We examined partial dependence plots to look at the shape of the relationships between predictors and response, and their uncertainty (Carlson, 2020).

Finally, we used our BART models to estimate the mean suitability at every sandy grid cell (rather than just those included in the models), as well as the lower and upper bounds of the corresponding 95% credible intervals. The difference between these bounds is used as a measure of local prediction uncertainty, with the (un)suitability of cells with wider intervals being more uncertain. To control for the effect of sampling bias, we fixed the distance to cities to its average value when making predictions for all sites. Map layers containing posterior means and credible interval information for each species are made available along with this report (**Data availability**). We also generated maps of expected near-future suitability by using the climate variables from the SSP scenarios in predictions. Transforming maps of relative suitability into binary presence/absence maps using thresholds is a common practice for SDM interpretation. We do

agree with others showing that for presence-only data it can introduce biases and a false sense of certainty in estimates, especially in the context of forecasting (Santini et al., 2021; Smith & Levine, 2025). We do acknowledge, however, that it may facilitate between-species comparisons, and the demand for such maps in applications. We therefore make available along the continuous probabilities the discretization thresholds used by *embarcadero*, which are based on maximizing the True Skill Statistic over the average posterior predictions for training data (Carlson, 2020).

### **3.2 Separating the effect of urbanization from the effect of sampling bias: occupancy models**

Occupancy models are powerful tools to study species occupancy and its variation in response to environmental drivers in the face of imperfect detection (Kéry & Royle, 2015; MacKenzie et al., 2002). While they can be quite complex and computationally intensive in some cases, the basic underlying principle is simple:

The probability of observing a species during a site visit depends, intuitively, both on its probability of being present and on the probability of detecting/recording it if present. Detection probabilities from a single visit are almost never = 1, even for plants (Perret et al., 2023). Assuming periods (so-called “closure periods”) in which the true occupancy of a site does not change, repeated visits to sites within such intervals allow us to separate detection from occupancy. Both occupancy and detection submodels can include predictors, allowing for a finer understanding of both the true ecological state and the observation process. Because occupancy models focus only on sites that have been visited, the spatial bias in overall sampling effort shared by all species is removed from consideration in the model.

Occupancy models have originally been developed for classical surveys where ecologists visit a number of sites repeatedly in a standardized way. However, Kéry et al. (2010), van Strien et al. (2010) and van Strien et al. (2013) showed that they could be adapted to opportunistic data (or mixes of standardized and opportunistic data) with only minor changes, and Isaac et al. (2014) confirmed that they were one of the most efficient methods to separate ecological signal from biases in such data. Importantly given the number of visits per site is often power-law distributed in opportunistic schemes (with a plurality or slight majority of sites visited only once), occupancy models are robust to a substantial proportion of sites receiving only one visit per closure period (von Hirschheydt et al., 2023).

We follow here an approach inspired by e.g. Outhwaite et al. (2019), with differences due to different goals and software choices. For each species:

- the occupancy submodel (on the logit scale) included fixed effects of urbanization (as a three-level factor: rural, DUC, UC), as well as random intercept effects of year, to account in fluctuations in occupancy between years. Because sites were visited over multiple

years, the model also included random effects of site identity. Finally, to account for broad-scaled spatial patterns in occupancy, the model included random effects of biome and ecoregion.

- the detection submodel (also on the logit scale) included a fixed effect of the so-called “list length”, i.e. the number of species seen during a specific visit, as a proxy of sampling effort. Visits that recorded one or a few species are unlikely to have been exhaustive, so we can generally expect detectability to increase with list length. In practice, we follow others and include list length as a three-level factor: 1 species recorded per visit, 2-3 species, 4 or more (Outhwaite et al., 2019; van Strien et al., 2013). This is because (i) detectability might in some cases not increase with list length, as visually striking species for instance may be overrepresented in 1-species visits; (ii) using a continuous list length variable, differences in regional species richness could lead to wrongly assume higher detectability in more species-rich regions (such as the Mediterranean). We considered also including non-linear effects of the time of year, to account for the impact of phenology on detection. We ended up not doing so because we found in preliminary tests, as in Outhwaite et al. (2019), that this greatly increases the computational burden with limited impacts on the resulting estimates of occupancy.

We fit our models in a Bayesian framework using the *brms* (Bürkner, 2018) and *flocker* (Socolar & Mills, 2023) R packages as interfaces to the Stan language (Gabry et al., 2024; Stan Development Team, 2024). We use weakly informative priors based on McElreath (2020) for proportional responses, and ran our models with 4 chains for 2000 iterations, with the first half used as warm-up.

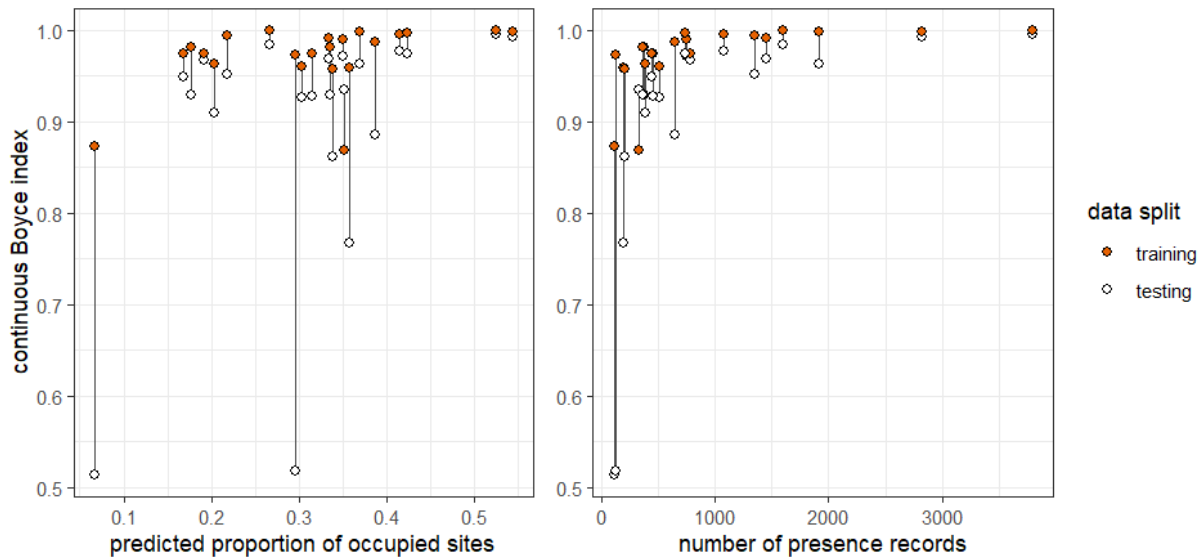
## 4. Results and Discussion

### 4.1 BART models

#### 4.1.1 Models performance

We find that model performance varied between species, but that models have generally good to very good performance on testing data, based on the continuous Boyce index (as a correlation coefficient, values  $> 0$  are better than random). Whether a species was predicted to be rare or common (based on the predicted proportion of sites with presence, using the binary discrimination threshold) was not a clear predictor of model performance (**Fig. 4**, left). By contrast, the number of sites with presence records is strongly linked with model performance: models for species with smaller amounts of available data tended to perform worse (**Fig. 4**, right) (species predicted to be more common tend to be those with more presence records in the database, but the correlation is relatively low,  $r = 0.61$ ). We note that the relationship between Boyce index and number of records is much steeper on testing vs. training data; this is because the models overfit much more on species with lower numbers of records (see the testing-training

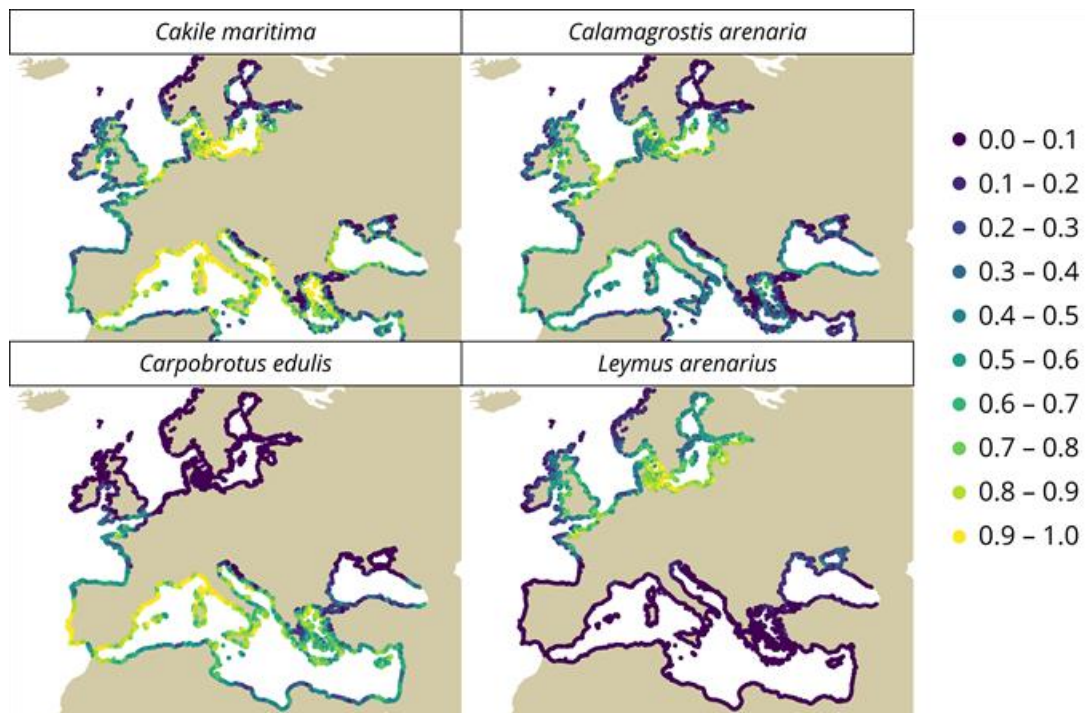
differences **Fig. 4**). We also note that some of the species with the lowest numbers of records (and therefore where the models perform worse) include the key dune-building species, as seen **Table 2**. Other metrics showed similarly good performance, but tended instead to perform slightly *worse* for more common species, or species with more records (not shown, but see **Data availability**). This may reflect the fact that they were not designed for presence-only data, and place similar weight on presences and pseudo-absences in their estimation.



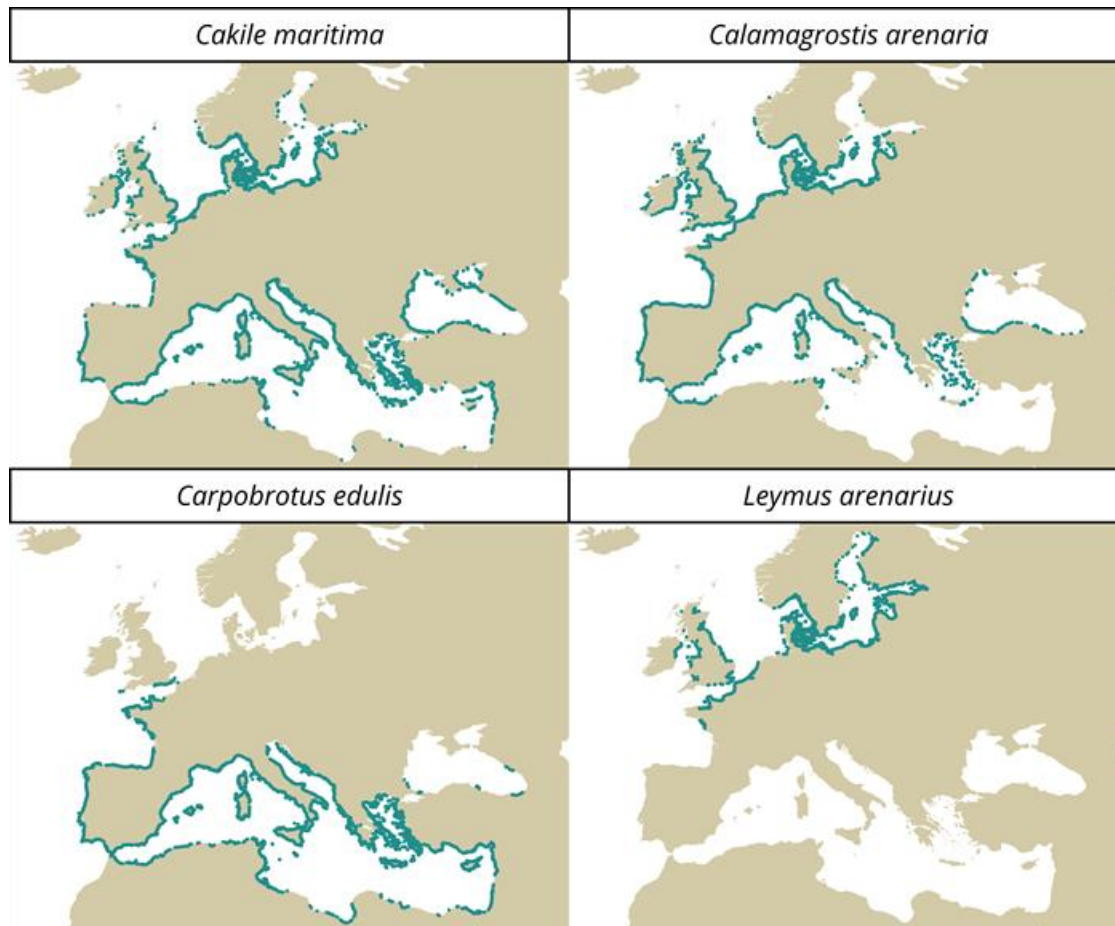
**Figure 4.** Relationship across species between the continuous Boyce index and the predicted occupancy (based on discretized probabilities, left) or the number of sites with observed presence records (right)

#### 4.1.2 Present-day species distribution patterns

In line with the overall good performance of our models based on continuous Boyce indices, we generally recover expected broad scale distribution patterns from the literature, whether it is plant species restricted to southern (Campoy et al., 2018) or northern coasts (Bond, 1952), or widely distributed across Europe (Davy et al., 2006; Huiskes, 1979) (**Figs 5, 6**). Even for the widely distributed marram grass *Calamagrostis arenaria*, the model accurately recovers the few regions of Europe where it is expected to be absent (northern Baltic; Huiskes (1979); Doody (1991)).

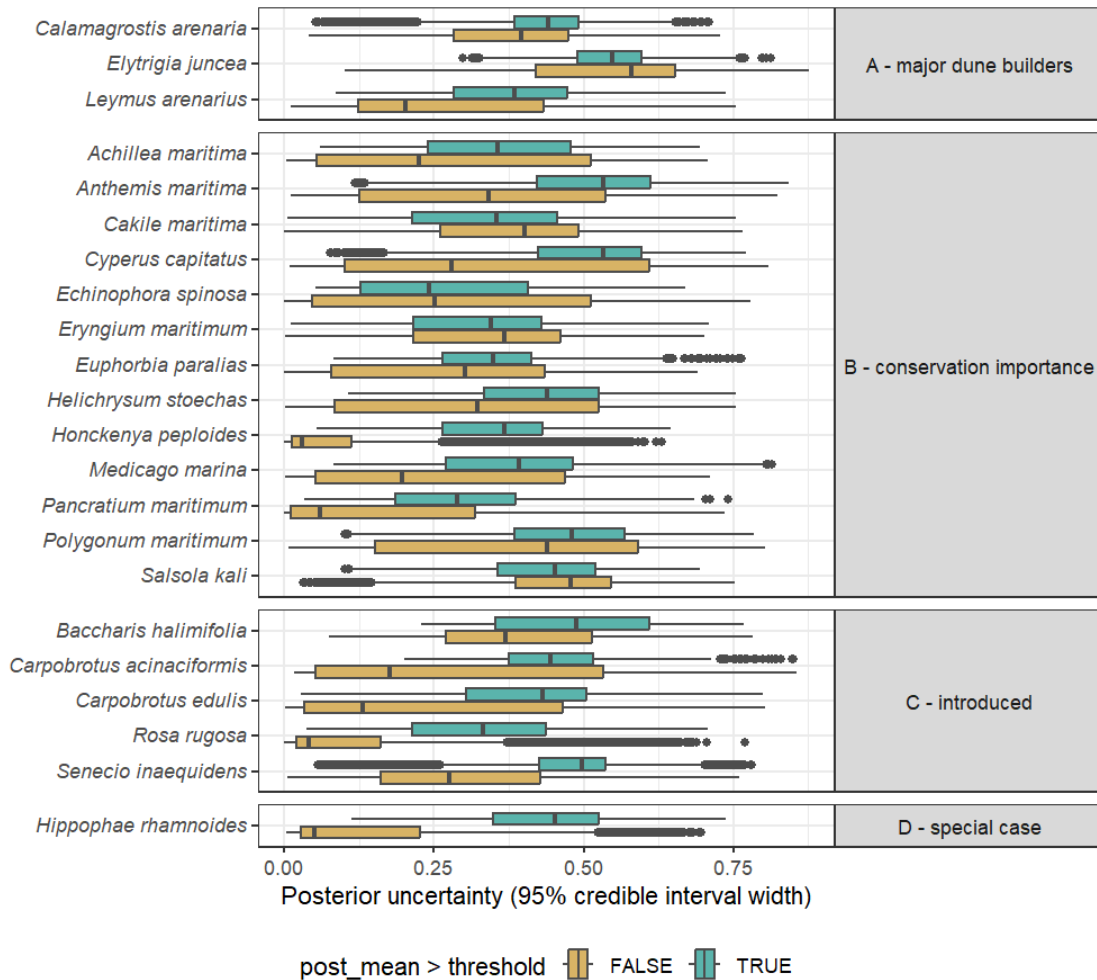


**Figure 5.** Maps of predicted posterior mean suitability for four selected species. Note that overlapping points may obscure local patterns in highly variable regions. These are probabilities of suitability not occupancy, and as such may or may not be directly comparable between species. For maps that are directly comparable between species, see the discretized maps **Fig. 6**.



**Figure 6.** Maps of sites with posterior mean suitability  $>$  the species-specific discretization threshold for four selected species.

However, using BART models allows us to see that these broadly accurate predictions come with substantial amounts of uncertainty, and that this uncertainty is not randomly distributed: in many though not all species, models tend to be much more certain of their predictions in areas of low suitability than in areas of predicted high suitability (**Fig. 7**). This suggests in our opinion that the models accurately recover the general bioclimatic patterns that would be expressed at broad European scale, while struggling to predict with certainty the local distribution within that generally suitable area. This is despite including relatively fine scale climate data as well as local geophysical characteristics in our predictors. This should be taken into account when considering practical applications of these maps, or when identifying potential pathways to ameliorate them.

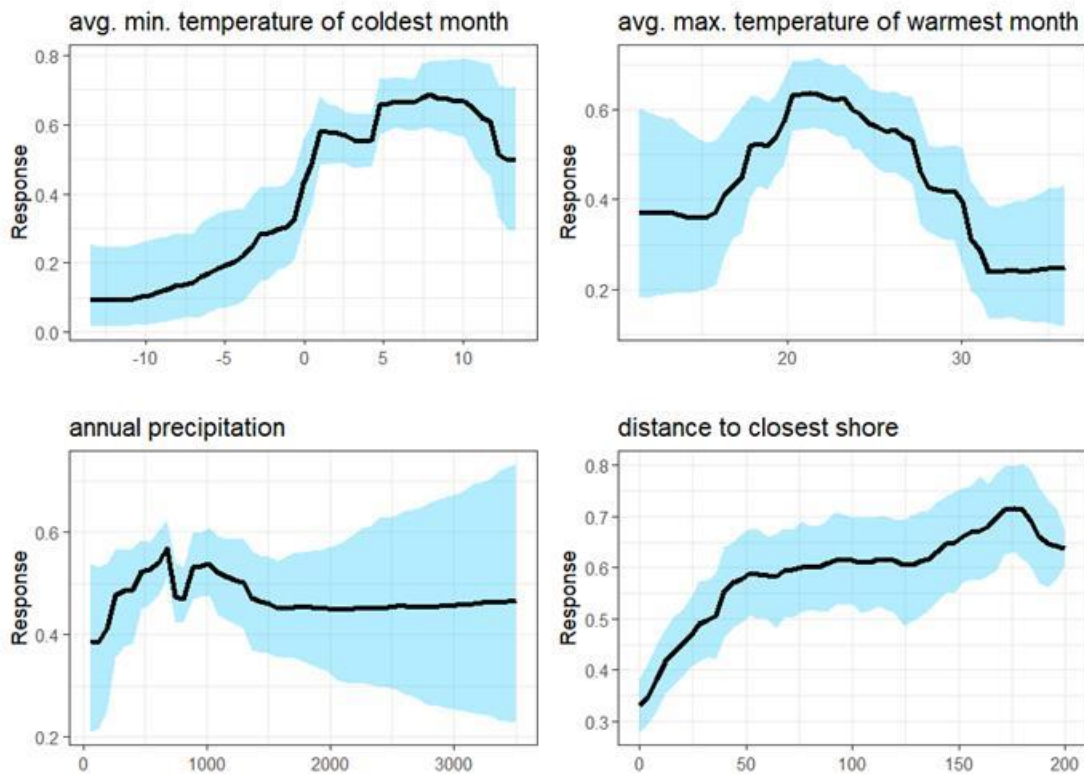


**Figure 7.** Differences in predicted posterior uncertainty between sites above (TRUE) and below (FALSE) the discretization threshold, by species.

This uncertainty in geographic space is paralleled in environmental space: partial dependence plots showing the effect of each predictor also show quite wide posteriors in many cases (Fig. 8). The models performing well at broad scale might indicate that macroclimate is the primary driver of distributions, and indeed, at least one climate variable is retained for all species (Fig. 9). This does not mean that non-climatic variables are always dropped from the final models during model selection, on the contrary (Fig. 9). Notably, orthogonal distance to the closest coast, which indicates whether the local coastline is open to the sea (high distances) or embayed (low distances), is a key predictor retained in most species (Fig. 9). Further, its effect consistently goes in the same direction, with open sandy coasts being more suitable to dune species than embayed sandy coasts.

Interestingly, distance to the nearest city is dropped as a predictor in many species, suggesting that it does not inform the model much. Importantly, this does not mean there is no sampling

bias; sampling bias (leading to an oversampling in cities) and the negative effects of urbanization on true presence may well cancel each other out.



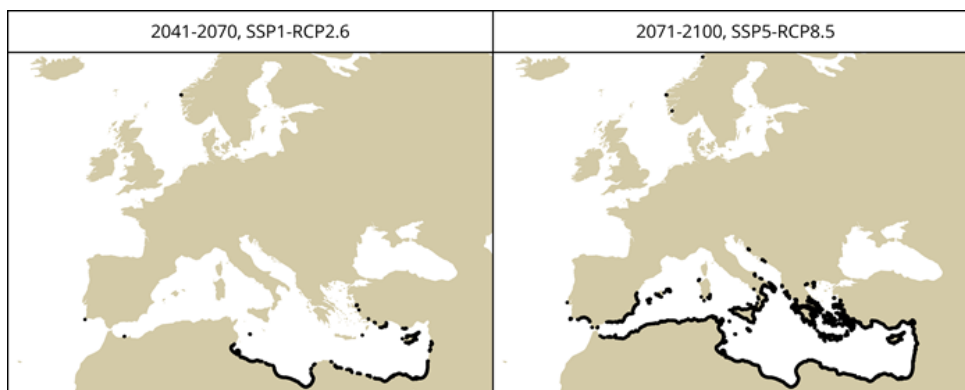
**Figure 8.** Four representative examples of partial dependence plots, all for *Cakile maritima*. Blue bands represent the 95% credible intervals. Given the very large number of species x predictor combinations, a qualitative summary is provided for all cases **Fig. 9**.

		coastal orientation	shoreline trend	distance to closest shore	100-year return storm surge	wave height	wave period	highest astronomical tide	tidal range	avg. daily temperature	avg. min. temperature of coldest month	avg. max. temperature of warmest month	annual precipitation	avg. wind speed	avg. wind speed of windiest month	distance to nearest city
A - major dune builders	<i>Calamagrostis arenaria</i>			↗	∩	↘	∩		—	∩	↗	↘			↗	
	<i>Elytrigia juncea</i>						∩			—	↗		—	—		—
	<i>Leymus arenarius</i>	—		↗	↗	—	—	—	—	↘	—	↘	—	↘		
B - conservation importance	<i>Achillea maritima</i>			↗		↘	↗			∩	↗	∩				
	<i>Anthemis maritima</i>			↗			∩	—		∩		∩				
	<i>Cakile maritima</i>	—		↗	↗	↘	∩	↘	↘	∩	↗	∩	—	∩	∩	
	<i>Cyperus capitatus</i>	—		↗		—	—	—	—	↗	↗	∩		—	—	—
	<i>Echinophora spinosa</i>			↗		↘	↘	—		∩	∩	∩				
	<i>Eryngium maritimum</i>			↗	∩	↗	↘			∩	↗	∩				↗
	<i>Euphorbia paralias</i>				↗	↘	↗			∩		∩	↗		↗	
	<i>Helichrysum stoechas</i>			↗	—	—	↗	—	—	↗	↗	∩	↘			
	<i>Honckenya peploides</i>			↗	∩	↗	↘			∩	↗	∩		—	—	
	<i>Medicago marina</i>			↗		↘	∩			∩		—	↗			↘
	<i>Pancratium maritimum</i>			↗		↘	↗			∩	↗	—	↗		↘	
	<i>Polygonum maritimum</i>	—		↗	↘	—	↗			∩	↗	—	—			
	<i>Salsola kali</i>	—		↗	↗	∩	—	—	—	—	↗	∩	—			
C - introduced	<i>Baccharis halimifolia</i>					—			↗	∩	↘	↗				
	<i>Carpobrotus acinaciformis</i>						∩			∩	↗		—	↘		↘
	<i>Carpobrotus edulis</i>	—		↗	∩	↘	↗		↗	∩	↘	↘		↘		
	<i>Rosa rugosa</i>			↗	↗	↗	↘		—	∩	—	∩	↘	↘	—	
	<i>Senecio inaequidens</i>				↗			—		∩					↘	
D - special case	<i>Hippophae rhamnoides</i>	—	↗	↗	∩		—			∩	∩	∩	↘	—	—	↘

**Figure 9.** Qualitative summary of the partial dependence plots for all species × predictor combinations. Predictors are ordered as in **Table 1**, species as in **Table 2**. Symbols indicate whether suitability is overall increasing when the predictor increases, overall decreasing or whether it shows a bell-shaped response. Combinations for which the variable is not retained show no symbol; combinations for which the variable is retained, but high posterior uncertainty leads to no clear pattern are shown with flat lines. Given coastal orientation is an angular variable, an interpretation in terms of overall increases/decreases is not meaningful; we therefore only flag whether or not this variable is retained or not. Note that because of the sometimes complex shape of partial dependencies, these are to some extent subjective interpretations.

#### 4.1.3 Expected future shifts in suitability

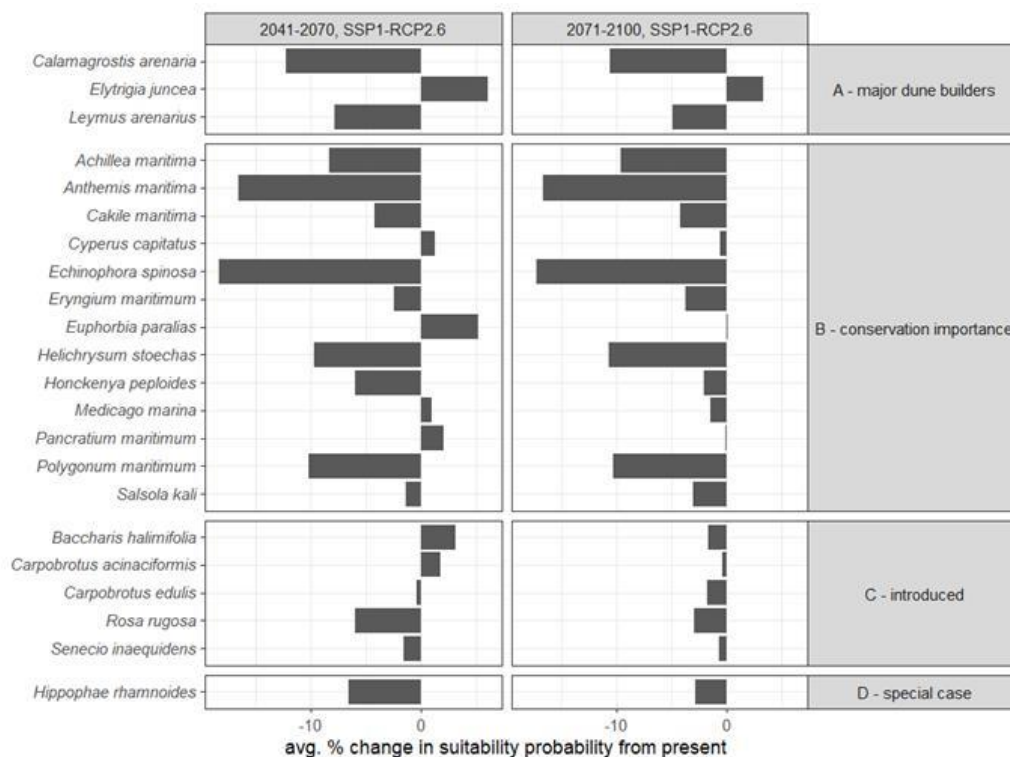
As a preamble, we note that these maps should be interpreted even more carefully than the present-day ones. First, as many have remarked before, changes in distribution in response to environmental change depend not only on how these changes make a site more or less suitable for a focal species, but also on whether the species can disperse away from newly unfavourable and to newly favourable sites (Essl et al., 2024). Second, using SDMs to forecast future suitabilities assumes that the present-day climate-suitability relationships seen **Figs 8, 9** are both causal and transferrable from space to time (Lovell et al., 2023); these assumptions are rarely tested (but see Lovell et al., 2025), and *cannot* be tested with our own data. Third, in any case only a small set of our environmental predictors have been projected into the future; our future maps still assume for instance that storm surge or coastal erosion/accretion rates, or even the shape of the coastline itself, will stay at the same level than in the 20th century, which is unlikely (van de Wal et al., 2024). Finally, with climate change many sites are predicted to experience climate conditions that have no analogue in the present-day study area on which the models were trained. Predictions at these sites should be regarded with the utmost caution, or even disregarded. Indeed, even if all the other assumptions were valid, the correctness of the model extrapolation beyond its training range cannot be guaranteed. We provide additional map layers indicating which sites show no-analogue conditions under the different climate change scenarios (predicted temperatures and/or precipitation outside the present-day range in the study area; **Fig. 10**). We find that even under the 2071-2100 SSP5-RCP8.5 scenario, no-analogue conditions remain for the most part limited to the extra-European, south Mediterranean area (though Malta, Cyprus, as well as Southern Spain, Italy and Greece start showing no-analog conditions; **Fig. 10**). Nonetheless, these expected future suitability maps may provide some guidance about changes in potential suitability in the coming decades, especially for the species in our set that most strongly respond to temperature and precipitation.



**Figure 10.** Sites with no-analogue conditions (predicted temperatures and/or precipitation outside the present-day range in the study area) under the climate scenario × period combinations least (left) and most (right) divergent from present climate.

We send the interested reader to the archived map files (**Data availability**) for access to all the future scenario maps. We highlight here some salient elements, using the lower emissions scenario (SSP1-RCP2.6) as an example and excluding no-analogue sites from calculations (but patterns are qualitatively similar if we include them):

- In the near-term (2041-2070), the overall suitability of the study area generally declines for most species, though some species see increases (**Fig. 11**; note that these estimates ignore posterior variation for simplicity and are likely to be quite uncertain);
- Decreases in suitability are overall of higher magnitude than increases;
- There is no clear pattern where e.g. the study area becomes systematically more suitable for non-native species and/or less suitable for species of conservation concern. Note that this reflects in part our species set, which was chosen to balance northern and southern European species. There is no guarantee at all whether this would hold considering the entire flora of European coastal dunes;
- In the longer term (2071-2100), almost no focal species should expect increases in suitability (**Fig. 11**).



**Figure 11.** Expected proportional shifts in the number of occupied sites (based on discretized probabilities) between present and future climates, by species.

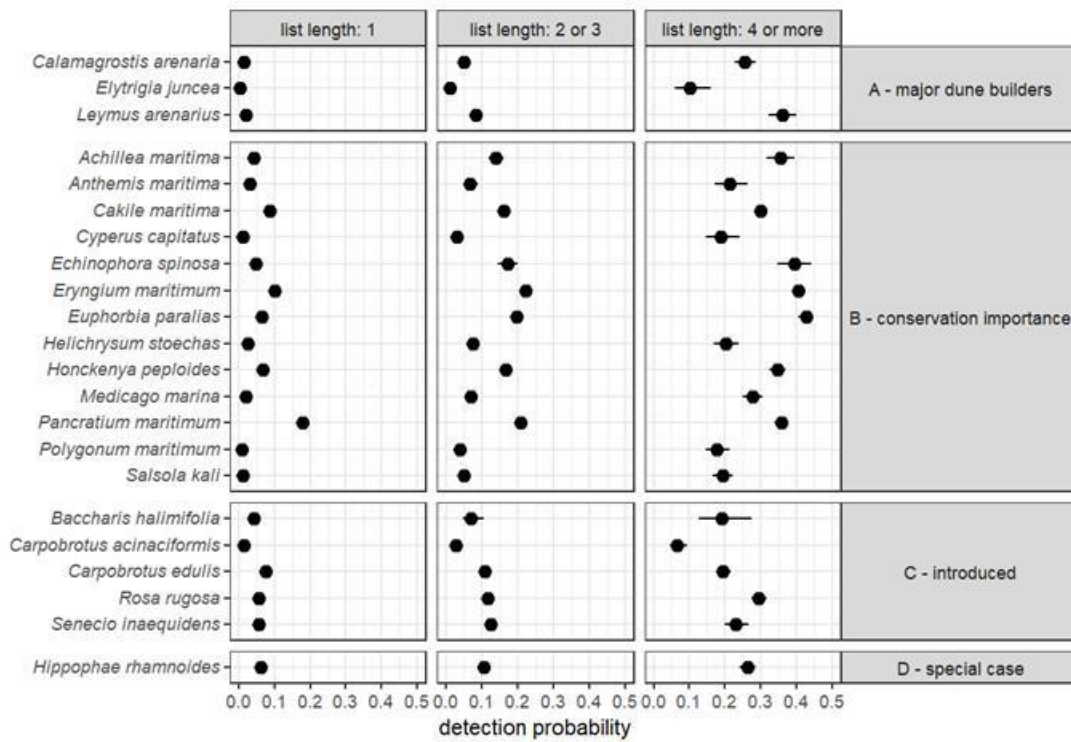
## 4.2 Occupancy models

### 4.2.1 Differences in detection between species

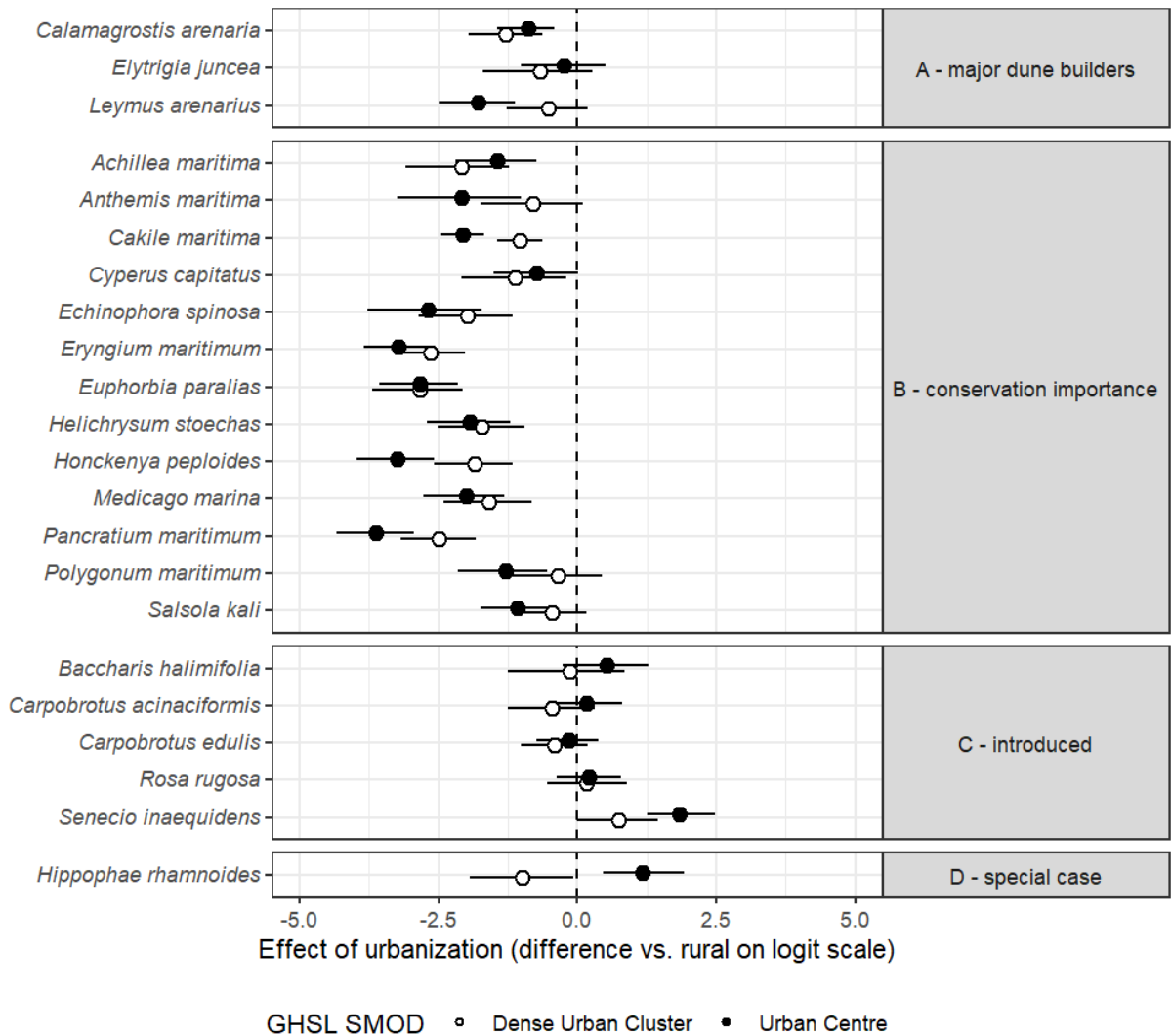
Detection probabilities, and how they change with list length, is shown **Fig. 12**. We find that as expected, detectability generally increases with the number of species recorded per visit (the list length). However, average detectability and the effect of list length do vary between species. Interestingly, we find indication that the main dune-building grasses (*Calamagrostis arenaria*, *Leymus arenarius*, and the *Elytrigia juncea* species complex) have lower than average detectability at small list lengths. This is also reflected on the number of sites with records of these species (**Table 2**). This is in line with some expectations about the expected behaviour of opportunistic recorders, which might preferentially record more visually striking species (in animals, see Callaghan et al., 2021; Goldstein et al., 2024). This may lead to a paradox where we have less spatial information (at least from individual, on the ground biodiversity records) on what are actually among the most abundant and structurally important species of the dune ecosystems. To solve this, the presence of high-coverage dune-building species might be detected by alternate methods such as remote sensing, with spectral or RGB signature, aggregation patterns and location on the shore potentially differentiating species. Work in that direction is ongoing in other DuneFront work packages (WP5, 6), and combining these two methodologies may yield more complete insights on dune communities at European scale.

### 4.2.2 Effect of urbanization

The effect of urbanization was highly variable between species, and depending on whether the city was a smaller Dense Urban Cluster, or a larger Urban Centre according to the GHSL (**Fig. 13**). Overall, however, urbanization had a generally negative impact on dune-building species and species of conservation concern, with effects stronger on the latter. Notably though, introduced species were an apparent exception to this, all showing either neutral or positive responses to urbanization. These results are fully in line with the expected role of cities as major hubs shaping introduction and invasion dynamics (Potgieter et al., 2024).



**Figure 12.** Relationship between list length and detectability, by species. Points are posterior means and bars 95% credible intervals.



**Figure 13.** Effect of urbanization on predicted occupancy (posterior differences vs. rural, on the logit scale), by species. Points are posterior means and bars 95% credible intervals.

#### 4.2.3 Important note

At this point, one might have noticed that outputs from both the BART and occupancy models are expressed as probabilities, and wondered whether it might be possible to somehow quantitatively combine these probabilities to inject the effect of urbanization back into the BART outputs. We explicitly advise against doing so, and to focus instead on qualitatively discussing their interplay. Indeed, while the BART outputs are expressed as probabilities, SDMs fitted from presence-only models might not directly reflect true probabilities of presence, only relative suitability, as the pseudo-absence points are not true absences (the species may be present but not observed). While our use of an environmentally-informed method to choose pseudo-

absences may mitigate that issue and make them more likely to reflect true absences (Da Re et al., 2023), it may not be enough.

## 5. Conclusion and perspectives

This report shows that meaningful and usable ecological information about key dune species can be extracted and summarized at the European level from opportunistic biodiversity data. Our work confirms the usefulness of BART as a Species Distribution Modelling method, and reveals substantial variation in ecological responses between species sharing the same space. BART models were especially valuable in highlighting the uncertainty of our suitability estimates and how it is distributed, which points to the need for more detailed local data. Furthermore, our application of occupancy models to crowdsourced data shows that impacts of urbanization are both strong and species-dependent. The differing responses to urbanization between native and non-native species point to potential conflicts in goals for Dune-Dike hybrids in urban settings: while urban DD-hybrids might be able to achieve their coastal protection goals, they might also favour the installation and connectivity of non-native species compared to non-urban dunes. Future work, including in other DuneFront work packages, might help shed light on these issues.

From an ecologist's perspective, future studies and updates might expand on the work presented here in the following ways:

- We relied here on a relatively coarse detection of sandy coasts, with a simple sandy/not-sandy binary classification based on a few transects per site. This means we likely missed from our analysis sites with very small sandy beaches (i.e. smaller than the typical  $\approx$  300-500 m between the reference transects), but which could have harboured our species of interest. Given the continued debates about the benefits of small habitat patches for conservation (in urbanized contexts, see e.g. Vega & Küffer, 2021), a re-analysis that accounts for the *quantity* of sandy habitat available per site may prove useful. There are however major technical and data hurdles here, and there should be careful evaluations of the costs and benefits of attempting to do this at the European scale, rather than at more local scales where accessing that quantitative information is more achievable.
- We were quite stringent in our selection criteria for biodiversity dataset inclusion; there is reason to believe that slightly loosening the filtering here might enable us to increase our dataset size without compromising quality. For instance, we excluded datasets that were limited to one country, to avoid injecting even more spatial bias in records, as not all countries have open recording schemes in GBIF. Extensions of BART with country-specific random effects might here be helpful (Carlson et al., 2022).
- A larger range of environmental variables may improve the predictive success of our BART models. For instance, the physical and chemical properties of the substrate have an impact

in shaping plant distributions (Mod et al., 2016). However, while there has been significant effort to develop open maps of these data for both marine (e.g. <https://www.coastal.hub.copernicus.eu/>) and terrestrial (e.g. <https://soilgrids.org/>) environments, there are no European-wide equivalent, as far as we know, specifically tailored to coastal systems (the physico-chemical layers available in the Copernicus Coastal Hub for instance, actually mostly relate to the marine realm). It is not necessarily clear how to relate marine and terrestrial patterns in any useable way to physico-chemical variation experienced on the coastal sandy interface in-between. Even if not as a proxy of coastal characteristics, these variables may still be useful as indicators of the influence of inland land use on the coast (e.g. runoffs...). With an increase in the number of predictors, issues around variable selection and overfitting will become more important, especially for species with fewer records. Modifications of BART with priors that allow for principled in-model variable selection might here be useful (for an example, see e.g. Becker et al., 2022).

- While we compared species responses informally and *qualitatively* in this report, there is significant added value in doing so systematically and *quantitatively*. Comparison of species responses, for instance through meta-analyses, can yield major insights on the higher-level mechanisms behind biodiversity responses, including the role of species traits and evolutionary history/ phylogeny (e.g. Capilla-Lasheras et al., 2022; Liang et al., 2023; Merckx et al., 2018; Murray et al., 2019). BART is a black-box non-parametric method, which means it does not provide meaningful effect sizes that would be analysable in that way across models/species. However, various summary metrics could be generated along with their uncertainty and compared in this framework. By contrast, parameters from the occupancy models (detectabilities, baseline occupancies, effects of urbanization...) can be readily and meaningfully compared between species. Rather than meta-analysing summary statistics from series of single-species models, occupancy models can even be expanded to multi-species models where traits and phylogeny are included directly (Bürkner, 2018; Socolar & Mills, 2023). New methodological developments, reducing the computational needs for fitting such multi-species models to crowdsourced data, are here very promising (Fajgenblat et al., 2025).

## 6. Data availability statement

The BART suitability maps for present and future scenarios are archived on Zenodo as geopackages, along with a copy of this report (<https://doi.org/10.5281/zenodo.15044553>). The processed environmental and species data tables, as well as the R code to reproduce the BART and occupancy models are available on GitHub ([https://github.com/mdahirel/DuneFront\\_WP4-2\\_analysis](https://github.com/mdahirel/DuneFront_WP4-2_analysis)) and archived separately in Zenodo (<https://doi.org/10.5281/zenodo.15044685>). The “raw” environmental data layers (i.e. before any processing described here) are openly available from a variety of sources all cited in this report. The reference coastline data are © OpenStreetMap contributors and were made available under the Open Database License (ODbL). The raw GBIF downloads are archived under the following DOIs: <https://doi.org/10.15468/dl.z8sdfd>, <https://doi.org/10.15468/dl.xc3pkg>, and <https://doi.org/10.15468/dl.m635ds> for iNaturalist, Observation.org and Pl@ntNet, respectively.

## 7. References

- Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, 43(6), Article 6. <https://doi.org/10.1111/j.1365-2664.2006.01214.x>
- Aronson, C. H. N., Myla F. J. (Ed.). (2023). *Routledge Handbook of Urban Biodiversity*. Routledge. <https://doi.org/10.4324/9781003016120>
- Athanasίου, P., van Dongeren, A., Pronk, M., Giardino, A., Vousdoukas, M., & Ranasinghe, R. (2024). Global Coastal Characteristics (GCC): a global dataset of geophysical, hydrodynamic, and socioeconomic coastal indicators. *Earth System Science Data*, 16(7), 3433–3452. <https://doi.org/10.5194/essd-16-3433-2024>
- Banfi, E. (2018). A survey of the *Elymus* L. s. l. species complex (Triticeae, Poaceae) in Italy: taxa and nothotaxa, new combinations and identification key. *Natural History Sciences*, 5(2), 57–64. <https://doi.org/10.4081/nhs.2018.392>
- Becker, D. J., Albery, G. F., Sjodin, A. R., Poisot, T., Bergner, L. M., Chen, B., Cohen, L. E., Dallas, T. A., Eskew, E. A., Fagre, A. C., Farrell, M. J., Guth, S., Han, B. A., Simmons, N. B., Stock, M., Teeling, E. C., & Carlson, C. J. (2022). Optimising predictive models to prioritise viral discovery in zoonotic reservoirs. *The Lancet Microbe*, 3(8), e625–e637. [https://doi.org/10.1016/S2666-5247\(21\)00245-7](https://doi.org/10.1016/S2666-5247(21)00245-7)
- Bensettiti, F., Bioret, F., Roland, J., & Lacoste, J.-P. (Eds.). (2004). « Cahiers d’habitats » *Natura 2000. Connaissance et gestion des habitats et des espèces d’intérêt communautaire. Tome 2 - Habitats côtiers*. Documentation française.

- Bond, T. E. T. (1952). *Elymus arenarius* L. *Journal of Ecology*, 40(1), 217–227. <https://doi.org/10.2307/2258031>
- Boyd, R. J., Bowler, D. E., Isaac, N. J. B., & Pescott, O. L. (2024). On the trade-off between accuracy and spatial resolution when estimating species occupancy from geographically biased samples. *Ecological Modelling*, 493, 110739. <https://doi.org/10.1016/j.ecolmodel.2024.110739>
- Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, 10(1), Article 1. <https://doi.org/10.32614/RJ-2018-017>
- Callaghan, C. T., Poore, A. G. B., Hofmann, M., Roberts, C. J., & Pereira, H. M. (2021). Large-bodied birds are over-represented in unstructured citizen science data. *Scientific Reports*, 11(1), Article 1. <https://doi.org/10.1038/s41598-021-98584-7>
- Campoy, J. G., Acosta, A. T. R., Affre, L., Barreiro, R., Brundu, G., Buisson, E., González, L., Lema, M., Novoa, A., Retuerto, R., Roiloa, S. R., & Fagúndez, J. (2018). Monographs of invasive plants in Europe: *Carpobrotus*. *Botany Letters*, 165, 440–475. <https://doi.org/10.1080/23818107.2018.1487884>
- Capilla-Lasheras, P., Thompson, M. J., Sánchez-Tójar, A., Haddou, Y., Branston, C. J., Réale, D., Charmantier, A., & Dominoni, D. M. (2022). A global meta-analysis reveals higher variation in breeding phenology in urban birds than in their non-urban neighbours. *Ecology Letters*, 25(11), 2552–2570. <https://doi.org/10.1111/ele.14099>
- Carlson, E., & Dominoni, D. M. (2024). The role of urbanization in facilitating the introduction and establishment of non-native animal species: a comprehensive review. *Journal of Urban Ecology*, 10(1), juae015. <https://doi.org/10.1093/jue/juae015>
- Carlson, C. J. (2020). embarcadero: species distribution modelling with Bayesian additive regression trees in R. *Methods in Ecology and Evolution*, 11(7), 850–858. <https://doi.org/10.1111/2041-210X.13389>
- Carlson, C. J., Bevins, S. N., & Schmid, B. V. (2022). Plague risk in the western United States over seven decades of environmental change. *Global Change Biology*, 28(3), 753–769. <https://doi.org/10.1111/gcb.15966>
- Castelle, B., & Dahirel, M. (2024). *DuneFront deliverable D4.1 - Physical boundary conditions over European coasts* (Version 1.0) [Dataset]. Zenodo. <https://doi.org/10.5281/zenodo.14525766>
- Castelle, B., Kras, E., Masselink, G., Scott, T., Konstantinou, A., & Luijendijk, A. (2024). Satellite-derived sandy shoreline trends and interannual variability along the Atlantic coast of Europe. *Scientific Reports*, 14(1), 13002. <https://doi.org/10.1038/s41598-024-63849-4>

Chipman, H. A., George, E. I., & McCulloch, R. E. (2010). BART: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4(1), 266–298. <https://doi.org/10.1214/09-AOAS285>

Commission Implementing Regulation (EU) 2022/1203 of 12 July 2022 Amending Implementing Regulation (EU) 2016/1141 to Update the List of Invasive Alien Species of Union Concern, 186 OJ L (2022). [http://data.europa.eu/eli/reg\\_impl/2022/1203/oj/eng](http://data.europa.eu/eli/reg_impl/2022/1203/oj/eng)

Da Re, D., Tordoni, E., Lenoir, J., Lembrechts, J. J., Vanwambeke, S. O., Rocchini, D., & Bazzichetto, M. (2023). USE it: Uniformly sampling pseudo-absences within the environmental space for applications in habitat suitability models. *Methods in Ecology and Evolution*, 14(11), 2873–2887. <https://doi.org/10.1111/2041-210X.14209>

Davy, A. J., Scott, R., & Cordazzo, C. V. (2006). Biological flora of the British Isles: *Cakile maritima* Scop. *Journal of Ecology*, 94(3), 695–711. <https://doi.org/10.1111/j.1365-2745.2006.01131.x>

Demichele, D., Belcore, E., Piras, M., & Camporeale, C. (2025). Species-by-species pattern analysis of coastal dune vegetation. *Journal of Geophysical Research: Biogeosciences*, 130(2), e2024JG008419. <https://doi.org/10.1029/2024JG008419>

Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N. D., Wikramanayake, E., Hahn, N., Palminteri, S., Hedao, P., Noss, R., Hansen, M., Locke, H., Ellis, E. C., Jones, B., Barber, C. V., Hayes, R., Kormos, C., Martin, V., Crist, E., ... Saleem, M. (2017). An ecoregion-based approach to protecting half the terrestrial realm. *BioScience*, 67(6), 534–545. <https://doi.org/10.1093/biosci/bix014>

Directorate-General for Environment (European Commission). (2013). *Interpretation Manual of European Union Habitats, version EU-28*.

Doody, J. P. (Ed.). (1991). *Sand dune inventory of Europe*. Joint Nature Conservation Committee.

Dorie, V., Chipman, H., & McCulloch, R. (2024). *dbarts: Discrete Bayesian Additive Regression Trees Sampler* (Version 0.9-30) [Computer software]. <https://doi.org/10.32614/CRAN.package.dbarts>

Essl, F., García-Rodríguez, A., Lenzner, B., Alexander, J. M., Capinha, C., Gaüzère, P., Guisan, A., Kühn, I., Lenoir, J., Richardson, D. M., Rumpf, S. B., Svenning, J.-C., Thuiller, W., Zurell, D., & Dullinger, S. (2024). Potential sources of time lags in calibrating species distribution models. *Journal of Biogeography*, 51(1), 89–102. <https://doi.org/10.1111/jbi.14726>

Eurostat (European Commission). (2021). *Applying the degree of urbanisation: a methodological manual to define cities, towns and rural areas for international comparisons : 2021 edition*. Publications Office of the European Union. <https://doi.org/10.2785/706535>

Fajgenblat, M., Wijns, R., De Knijf, G., Stoks, R., Lemmens, P., Herremans, M., Vanormelingen, P., Neyens, T., & De Meester, L. (2025). Leveraging massive opportunistically collected datasets to study species communities in space and time. *Ecology Letters*, 28(3), e70094. <https://doi.org/10.1111/ele.70094>

Gabry, J., Češnovar, R., & Johnson, A. (2024). *cmdstanr: R interface to “CmdStan”* [Computer software]. <https://mc-stan.org/cmdstanr/>

Geurts, E. M., Reynolds, J. D., & Starzomski, B. M. (2023). Turning observations into biodiversity data: Broadscale spatial biases in community science. *Ecosphere*, 14(6), e4582. <https://doi.org/10.1002/ecs2.4582>

Giulio, S., Acosta, A. T. R., Carboni, M., Campos, J. A., Chytrý, M., Loidi, J., Pergl, J., Pyšek, P., Isermann, M., Janssen, J. A. M., Rodwell, J. S., Schaminée, J. H. J., & Marcenò, C. (2020). Alien flora across European coastal dunes. *Applied Vegetation Science*, 23(3), 317–327. <https://doi.org/10.1111/avsc.12490>

Goldstein, B. R., Stoudt, S., Lewthwaite, J. M., Shirey, V., Mendoza, E., & Guzman, L. M. (2024). Logistical and preference bias in participatory science butterfly data. *Frontiers in Ecology and the Environment*, 22(8), e2783. <https://doi.org/10.1002/fee.2783>

Hennekens, S. (2019). *Distribution and habitat suitability maps of revised EUNIS coastal and wetland habitats* (p. 102). European Topic Centre on Biological Diversity (ETC/BD), European Environment Agency (EEA).

Hernández-Cordero, A. I., Gracia Prieto, F. J., Hernández-Calvento, L., Pérez-Chacón Espino, E., & Alonso, I. (2015). Proposal for new EU habitats associated with coastal dune fields of the Macaronesian region. A case study in the Canary Islands (Spain). *Journal of Coastal Conservation*, 19(2), 213–225. <https://doi.org/10.1007/s11852-015-0382-z>

Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C., & Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, 199(2), 142–152. <https://doi.org/10.1016/j.ecolmodel.2006.05.017>

Huiskes, A. H. L. (1979). *Ammophila arenaria* (L.) Link (*Psamma arenaria* (L.) Roem. et Schult.; *Calamagrostis arenaria* (L.) Roth). *Journal of Ecology*, 67(1), 363–382. <https://doi.org/10.2307/2259356>

Hulskamp, R., Luijendijk, A., van Maren, B., Moreno-Rodenas, A., Calkoen, F., Kras, E., Lhermitte, S., & Aarninkhof, S. (2023). Global distribution and dynamics of muddy coasts. *Nature Communications*, 14(1), 8259. <https://doi.org/10.1038/s41467-023-43819-6>

- IPCC. (2023). *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]*. (First). Intergovernmental Panel on Climate Change (IPCC). <https://doi.org/10.59327/IPCC/AR6-9789291691647>
- Isaac, N. J. B., & Pocock, M. J. O. (2015). Bias and information in biological records. *Biological Journal of the Linnean Society*, 115(3), 522–531. <https://doi.org/10.1111/bij.12532>
- Isaac, N. J. B., van Strien, A. J., August, T. A., de Zeeuw, M. P., & Roy, D. B. (2014). Statistics for citizen science: extracting signals of change from noisy ecological data. *Methods in Ecology and Evolution*, 5(10), 1052–1060. <https://doi.org/10.1111/2041-210X.12254>
- Isermann, M., Diekmann, M., & Heemann, S. (2007). Effects of the expansion by *Hippophaë rhamnoides* on plant species richness in coastal dunes. *Applied Vegetation Science*, 10(1), 33–42. <https://doi.org/10.1111/j.1654-109X.2007.tb00501.x>
- Joint Research Centre (European Commission). (2023). *GHSL data package 2023*. Publications Office of the European Union. <https://doi.org/10.2760/098587>
- Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., & Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, 4(1), 170122. <https://doi.org/10.1038/sdata.2017.122>
- Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., & Kessler, M. (2021). *Climatologies at high resolution for the earth's land surface areas* [Dataset]. EnviDat. <https://doi.org/http://dx.doi.org/10.16904/envidat.228>
- Karger, D. N., Schmatz, D. R., Dettling, G., & Zimmermann, N. E. (2020). High-resolution monthly precipitation and temperature time series from 2006 to 2100. *Scientific Data*, 7(1), 248. <https://doi.org/10.1038/s41597-020-00587-y>
- Karger, D. N., Schmatz, D., Dettling, G., & Zimmermann, N. E. (2019). *High resolution monthly precipitation and temperature timeseries for the period 2006-2100* [Dataset]. EnviDat. <https://doi.org/http://dx.doi.org/10.16904/envidat.124>
- Kéry, M., & Royle, J. A. (2015). *Applied Hierarchical Modeling in Ecology: Analysis of Distribution, Abundance and Species Richness in R and Bugs: Volume 1: Prelude and Static Models* (Illustrated édition). Academic Press. <https://doi.org/10.1016/C2015-0-04070-9>
- Kéry, M., Royle, J. A., Schmid, H., Schaub, M., Volet, B., Häfliger, G., & Zbinden, N. (2010). Site-occupancy distribution modeling to correct population-trend estimates derived from opportunistic observations. *Conservation Biology*, 24(5), 1388–1397. <https://doi.org/10.1111/j.1523-1739.2010.01479.x>

Knapp, S., Aronson, M. F. J., Carpenter, E., Herrera-Montes, A., Jung, K., Kotze, D. J., La Sorte, F. A., Lepczyk, C. A., MacGregor-Fors, I., MacIvor, J. S., Moretti, M., Nilon, C. H., Piana, M. R., Rega-Brodsky, C. C., Salisbury, A., Threlfall, C. G., Trisos, C., Williams, N. S. G., & Hahs, A. K. (2021). A research agenda for urban biodiversity in the global extinction crisis. *BioScience*, *71*(3), 268–279. <https://doi.org/10.1093/biosci/biaa141>

Kuhn, M., & Wickham, H. (2020). *Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles*. [Computer software]. <https://www.tidymodels.org>

Leonardi, M., Colucci, M., Pozzi, A. V., Scerri, E. M. L., & Manica, A. (2024). tidysdm: Leveraging the flexibility of tidymodels for species distribution modelling in R. *Methods in Ecology and Evolution*, *15*(10), 1789–1795. <https://doi.org/10.1111/2041-210X.14406>

Liang, H., He, Y.-D., Theodorou, P., & Yang, C.-F. (2023). The effects of urbanization on pollinators and pollination: A meta-analysis. *Ecology Letters*, *26*(9), 1629–1642. <https://doi.org/10.1111/ele.14277>

Lojek, O., Schweiger, C., & Santos, P. R. (2024). *DuneFront deliverable D4.3 - Compilation of freely existing socio- economic/administrative boundaries*. Technische Universität Braunschweig.

Lovell, R. S. L., Collins, S., Martin, S. H., Pigot, A. L., & Phillimore, A. B. (2023). Space-for-time substitutions in climate change ecology and evolution. *Biological Reviews*, *98*(6), 2243–2270. <https://doi.org/10.1111/brv.13004>

Lovell, R. S. L., Powney, G. D., Botham, M. S., & Phillimore, A. B. (2025). Testing space-for-time transferability of climate effects on occupancy and abundance. *Functional Ecology*, *n/a*(*n/a*). <https://doi.org/10.1111/1365-2435.70005>

Luijendijk, A., Hagenaars, G., Ranasinghe, R., Baart, F., Donchyts, G., & Aarninkhof, S. (2018). The state of the world's beaches. *Scientific Reports*, *8*(1), Article 1. <https://doi.org/10.1038/s41598-018-24630-6>

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Andrew Royle, J., & Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, *83*(8), 2248–2255. [https://doi.org/10.1890/0012-9658\(2002\)083\[2248:ESORWD\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2002)083[2248:ESORWD]2.0.CO;2)

Marcenò, C., Guarino, R., Loidi, J., Herrera, M., Isermann, M., Knollová, I., Tichý, L., Tzonev, R. T., Acosta, A. T. R., FitzPatrick, Ú., Iakushenko, D., Janssen, J. A. M., Jiménez-Alfaro, B., Kaçki, Z., Keizer-Sedláková, I., Kolomiychuk, V., Rodwell, J. S., Schaminée, J. H. J., Šilc, U., & Chytrý, M. (2018). Classification of European and Mediterranean coastal dune vegetation. *Applied Vegetation Science*, *21*(3), 533–559. <https://doi.org/10.1111/avsc.12379>

Maun, M. A. (2009). *The biology of coastal sand dunes*. Oxford University Press.

McCallum, I., Kyba, C. C. M., Bayas, J. C. L., Moltchanova, E., Cooper, M., Cuaresma, J. C., Pachauri, S., See, L., Danylo, O., Moorthy, I., Lesiv, M., Baugh, K., Elvidge, C. D., Hofer, M., & Fritz, S. (2022). Estimating global economic well-being with unlit settlements. *Nature Communications*, 13(1), 2459. <https://doi.org/10.1038/s41467-022-30099-9>

McElreath, R. (2020). *Statistical rethinking: a Bayesian course with examples in R and Stan* (2nd edition). Chapman and Hall/CRC.

McKinney, M. L. (2008). Effects of urbanization on species richness: a review of plants and animals. *Urban Ecosystems*, 11(2), Article 2. <https://doi.org/10.1007/s11252-007-0045-4>

Merckx, T., Souffreau, C., Kaiser, A., Baardsen, L. F., Backeljau, T., Bonte, D., Brans, K. I., Cours, M., Dahirel, M., Debortoli, N., Wolf, K., Engelen, J. M. T., Fontaneto, D., Gianuca, A. T., Govaert, L., Hendrickx, F., Higuti, J., Lens, L., Martens, K., ... Van Dyck, H. (2018). Body-size shifts in aquatic and terrestrial urban communities. *Nature*, 558(7708), Article 7708. <https://doi.org/10.1038/s41586-018-0140-0>

Mod, H. K., Scherrer, D., Luoto, M., & Guisan, A. (2016). What we use is not what we know: environmental predictors in plant distribution models. *Journal of Vegetation Science*, 27(6), 1308–1322. <https://doi.org/10.1111/jvs.12444>

Murray, M. H., Sánchez, C. A., Becker, D. J., Byers, K. A., Worsley-Tonks, K. E., & Craft, M. E. (2019). City sicker? A meta-analysis of wildlife health and urbanization. *Frontiers in Ecology and the Environment*, 17(10), Article 10. <https://doi.org/10.1002/fee.2126>

Nawarat, K., Reyns, J., Vousdoukas, M. I., Duong, T. M., Kras, E., & Ranasinghe, R. (2024). Coastal hardening and what it means for the world's sandy beaches. *Nature Communications*, 15(1), 10626. <https://doi.org/10.1038/s41467-024-54952-1>

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'Amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., & Kassem, K. R. (2001). Terrestrial ecoregions of the world: a new map of life on Earth. *BioScience*, 51(11), 933. [https://doi.org/10.1641/0006-3568\(2001\)051\[0933:TEOTWA\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2)

OpenStreetMap contributors. (2025). *OpenStreetMap data - coastlines* [Dataset].

Outhwaite, C. L., Powney, G. D., August, T. A., Chandler, R. E., Rorke, S., Pescott, O. L., Harvey, M., Roy, H. E., Fox, R., Roy, D. B., Alexander, K., Ball, S., Bantock, T., Barber, T., Beckmann, B. C., Cook, T., Flanagan, J., Fowles, A., Hammond, P., ... Isaac, N. J. B. (2019). Annual estimates of occupancy for bryophytes, lichens and invertebrates in the UK, 1970–2015. *Scientific Data*, 6(1), 259. <https://doi.org/10.1038/s41597-019-0269-1>

- Pacifici, M., Rondinini, C., Rhodes, J. R., Burbidge, A. A., Cristiano, A., Watson, J. E. M., Woinarski, J. C. Z., & Di Marco, M. (2020). Global correlates of range contractions and expansions in terrestrial mammals. *Nature Communications*, *11*(1), 2840. <https://doi.org/10.1038/s41467-020-16684-w>
- Perret, J., Besnard, A., Charpentier, A., & Papuga, G. (2023). Plants stand still but hide: Imperfect and heterogeneous detection is the rule when counting plants. *Journal of Ecology*, *111*(7), 1483–1496. <https://doi.org/10.1111/1365-2745.14110>
- Potgieter, L. J., Li, D., Baiser, B., Kühn, I., Aronson, M. F. J., Carboni, M., Celesti-Gradow, L., Matos, A. C. L. de, Lososová, Z., Montaña-Centellas, F. A., Pyšek, P., Richardson, D. M., Tsang, T. P. N., Zenni, R. D., & Cadotte, M. W. (2024). Cities shape the diversity and spread of nonnative species. *Annual Review of Ecology, Evolution, and Systematics*, *55*, 157–180. <https://doi.org/10.1146/annurev-ecolsys-102722-012749>
- R Core Team. (2024). *R: a language and environment for statistical computing* (Version 4.4.0) [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rezaei, N., & Millard-Ball, A. (2023). Urban form and its impacts on air pollution and access to green space: A global analysis of 462 cities. *PLOS ONE*, *18*(1), e0278265. <https://doi.org/10.1371/journal.pone.0278265>
- Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., Lutz, W., Popp, A., Cuaresma, J. C., Kc, S., Leimbach, M., Jiang, L., Kram, T., Rao, S., Emmerling, J., ... Tavoni, M. (2017). The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*, *42*, 153–168. <https://doi.org/10.1016/j.gloenvcha.2016.05.009>
- Richards, E. G., & Burningham, H. (2011). *Hippophae rhamnoides* on a coastal dune system: a thorny issue? *Journal of Coastal Conservation*, *15*(1), 73–85. <https://doi.org/10.1007/s11852-010-0122-3>
- Richardson, J., Lees, A. C., Miller, E. T., & Marsden, S. J. (2023). Avian diversity and function across the world's most populous cities. *Ecology Letters*, *26*(8), 1301–1313. <https://doi.org/10.1111/ele.14238>
- Santini, L., Benítez-López, A., Maiorano, L., Čengić, M., & Huijbregts, M. A. J. (2021). Assessing the reliability of species distribution projections in climate change research. *Diversity and Distributions*, *27*(6), 1035–1050. <https://doi.org/10.1111/ddi.13252>
- Sanyal, A., & Decocq, G. (2015). Biological Flora of the British Isles: *Crambe maritima*. *Journal of Ecology*, *103*(3), 769–788. <https://doi.org/10.1111/1365-2745.12389>

Schiavina, M., Melchiorri, M., & Pesaresi, M. (2023). *GHS-SMOD R2023A - GHS settlement layers, application of the Degree of Urbanisation methodology (stage I) to GHS-POP R2023A and GHS-BUILT-S R2023A, multitemporal (1975-2030)* [Dataset]. European Commission, Joint Research Centre (JRC). <https://doi.org/10.2905/A0DF7A6F-49DE-46EA-9BDE-563437A6E2BA>

Smith, J. R., & Levine, J. M. (2025). Linking relative suitability to probability of occurrence in presence-only species distribution models: Implications for global change projections. *Methods in Ecology and Evolution*, *n/a*(*n/a*). <https://doi.org/10.1111/2041-210X.70003>

Socular, J. B., & Mills, S. C. (2023). *Introducing flocker: an R package for flexible occupancy modeling via brms and Stan* (p. 2023.10.26.564080). bioRxiv. <https://doi.org/10.1101/2023.10.26.564080>

Stan Development Team. (2024). *Stan Modeling Language Users Guide and Reference Manual* (Version 2.33.1) [Computer software]. <https://mc-stan.org>

Szulkin, M., Munshi-South, J., & Charmantier, A. (Eds.). (2020). *Urban Evolutionary Biology* (1st ed.). Oxford University Press. <https://doi.org/10.1093/oso/9780198836841.001.0001>

Tison, J.-M., & de Foucault, B. (2014). *Flora Gallica: Flore de France*. Biotope Editions.

van de Wal, R., Melet, A., Bellafiore, D., Camus, P., Ferrarin, C., Oude Essink, G., Haigh, I. D., Lionello, P., Luijendijk, A., Toimil, A., Staneva, J., & Vousdoukas, M. (2024). Sea Level Rise in Europe: Impacts and consequences. *State of the Planet, 3-slre1*, 1–33. <https://doi.org/10.5194/sp-3-slre1-5-2024>

Van Der Maarel, E., & Van Der Maarel-Versluys, M. (1996). Distribution and conservation status of littoral vascular plant species along the European coasts. *Journal of Coastal Conservation*, *2*(1), 73–92. <https://doi.org/10.1007/BF02743039>

van Strien, A. J., Termaat, T., Groenendijk, D., Mensing, V., & Kéry, M. (2010). Site-occupancy models may offer new opportunities for dragonfly monitoring based on daily species lists. *Basic and Applied Ecology*, *11*(6), 495–503. <https://doi.org/10.1016/j.baae.2010.05.003>

van Strien, A. J., van Swaay, C. A. M., & Termaat, T. (2013). Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. *Journal of Applied Ecology*, *50*(6), 1450–1458. <https://doi.org/10.1111/1365-2664.12158>

Vega, K. A., & Küffer, C. (2021). Promoting wildflower biodiversity in dense and green cities: The important role of small vegetation patches. *Urban Forestry & Urban Greening*, *62*, 127165. <https://doi.org/10.1016/j.ufug.2021.127165>

Venter, Z. S., Chakraborty, T., & Lee, X. (2021). Crowdsourced air temperatures contrast satellite measures of the urban heat island and its mechanisms. *Science Advances*, 7(22), eabb9569. <https://doi.org/10.1126/sciadv.abb9569>

von Hirschheydt, G., Stofer, S., & Kéry, M. (2023). “Mixed” occupancy designs: When do additional single-visit data improve the inferences from standard multi-visit models? *Basic and Applied Ecology*, 67, 61–69. <https://doi.org/10.1016/j.baae.2023.01.003>

Whittingham, M. J., Stephens, P. A., Bradbury, R. B., & Freckleton, R. P. (2006). Why do we still use stepwise modelling in ecology and behaviour? *Journal of Animal Ecology*, 75(5), 1182–1189. <https://doi.org/10.1111/j.1365-2656.2006.01141.x>