Landscape and Urban Planning 218 (2022) 104286

Contents lists available at ScienceDirect

Landscape and Urban Planning

journal homepage: www.elsevier.com/locate/landurbplan

Mapping abundance distributions of allergenic tree species in urbanized landscapes: A nation-wide study for Belgium using forest inventory and citizen science data

Sebastien Dujardin [a,b,c,*], Michiel Stas [c,d], Camille Van Eupen [c,e], Raf Aerts [c,f,g,h,i], Marijke Hendrickx [i], Andy W. Delcloo [j,k], François Duchêne [j,k], Rafiq Hamdi [j,k], Tim S. Nawrot [h,l], An Van Nieuwenhuyse [l,m], Jean-Marie Aerts [d], Jos Van Orshoven [c], Ben Somers [c], Catherine Linard [a,b], Nicolas Dendoncker [a,b]

[a] Department of Geography, University of Namur, Rue de Bruxelles 61, BE-5000 Namur, Belgium
[b] Institute for Life, Earth and Environment (ILEE), University of Namur, Rue de Bruxelles 61, BE-5000 Namur, Belgium
[c] Division Forest, Nature and Landscape, Department Earth and Environmental Sciences, KU Leuven, Celestijnenlaan 200E-2411, BE-3001 Leuven, Belgium
[d] Measure, Model & Manage Bioresponses (M3-BIORES), Division Animal and Human Health Engineering, Department of Biosystems (BIOSYST), KU Leuven, Kasteelpark Arenberg 30, BE-3001 Leuven, Belgium
[e] Ghent University, Department of Data Analysis and Mathematical Modelling, Coupure Links 653, BE-9000 Gent, Belgium
[f] Risk and Health Impact Assessment, Sciensano (Belgian Institute of Health), J. Wytsmanstraat 14, BE-1050 Brussel, Belgium
[g] Division Ecology, Evolution and Biodiversity Conservation, KU Leuven, Kasteelpark Arenberg 31-3245, BE-3001 Leuven, Belgium
[h] Center for Environmental Sciences, Hasselt University, Campus Diepenbeek, Agoralaan Gebouw D, BE-3590 Hasselt, Belgium
[i] Mycology and Aerobiology, Sciensano (Belgian Institute of Health), J. Wytsmanstraat 14, BE-1050 Brussels, Belgium
[j] Royal Meteorological Institute of Belgium, Ringlaan 3 Avenue Circulaire, BE-1180 Brussel, Belgium
[k] Department of Physics and Astronomy, Ghent University, Proeftuinstraat 86, BE-9000 Gent, Belgium
[l] Centre Environment and Health, Department of Public Health and Primary Care, KU Leuven, Kapucijnenvoer 35 blok d box 7001, BE-3000 Leuven, Belgium
[m] Department of Health Protection, Laboratoire National de santé (LNS) , 1, Rue Louis Rech, LU-3555 Dudelange, Luxembourg

HIGHLIGHTS

- We computed a 1-ha resolution map of tree abundance at the national scale.
- The selected allergenic tree species included alder, birch and hazel.
- The combination of forest inventory with citizen science data was successful.
- Maps show fine-scale variations across the urban-rural landscape of Belgium.
- It offers many applications in the field of urban health and forest management.

ARTICLE INFO

Keywords:
Urban vegetation
Forest mapping
Allergenic trees
Respiratory health
Species distribution modelling
Citizen science

ABSTRACT

Mapping the distribution of allergenic plants in urbanized landscapes is of high importance to evaluate its impact on human health. However, data is not always available for the allergy-relevant species such as alder, birch, hazel, especially within cities where systematic inventories are often missing or not readily available. This research presents an approach to produce high-resolution abundance maps of allergenic tree species using existing forest inventories and opportunistic open-access citizen science data. Following a two-step approach, we first built species distribution models (SDMs) to predict species habitat suitability, using environmental characteristics as predictors. Second, we used statistical regressions to model the relationships between abundance, the habitat suitability predicted by the SDMs, and additional vegetation cover covariates. The combination of

* Corresponding author at: Department of Geography, University of Namur, Rue de Bruxelles 61, BE-5000 Namur, Belgium.
E-mail addresses: sebastien.dujardin@unamur.be (S. Dujardin), michiel.stas@kuleuven.be (M. Stas), camille.vaneupen@kuleuven.be (C. Van Eupen), raf.aerts@kuleuven.be (R. Aerts), andy.delcloo@meteo.be (A.W. Delcloo), duchenef@meteo.be (F. Duchêne), rafiq.hamdi@meteo.be (R. Hamdi), tim.nawrot@uhasselt.be (T.S. Nawrot), an.vannieuwenhuyse@kuleuven.be (A. Van Nieuwenhuyse), jean-marie.aerts@kuleuven.be (J.-M. Aerts), jos.vanorshoven@kuleuven.be (J. Van Orshoven), ben.somers@kuleuven.be (B. Somers), catherine.linard@unamur.be (C. Linard), nicolas.dendoncker@unamur.be (N. Dendoncker).

https://doi.org/10.1016/j.landurbplan.2021.104286
Received 11 June 2021; Received in revised form 30 September 2021; Accepted 10 October 2021
Available online 31 October 2021

forest inventory data with citizen science data improves the accuracy of abundance distribution models of allergenic tree species. This produces a continuous, 1-hectare resolution map of alder, birch, and hazel showing spatial variations of abundance distributions both within the urban fabric and along the urban–rural gradient. Species abundance modelling can offer a better understanding of the existing and potential future allergy risk posed by green spaces and pave the way for a wide variety of applications at fine-scale, which is indispensable for evidence-based urban green space policy and planning in support of public health.

1. Introduction

Mapping tree distributions of allergenic plants in urbanized landscapes is of high importance to evaluate their impact on micro-climate, air-quality and human health (Battisti, Pille, Wachtel, Larcher, & Säumel, 2019; Ottosen, Petch, Hanson, & Skjøth, 2020). By providing shade and reducing radiant temperatures, trees can mitigate the urban heat-island effect (Hamdi et al., 2020; Oke, 1982) and improve the thermal comfort of humans (Lee, Mayer, & Chen, 2016; Teshnehdel, Akbari, Di Giuseppe, & Brown, 2020). In addition, urban trees and peri-urban forests can improve air quality through trapping of particulate matter on the leaf surface and by absorbing gaseous pollutants (Baumgardner, Varela, Escobedo, Chacalo, & Ochoa, 2012; Grote et al., 2016). The pollution removal by trees is valued as an important respiratory health benefit for city-dwellers (Nowak, Crane, & Stevens, 2006; Nowak, Hirabayashi, Bodine, & Greenfield, 2014; Nowak, Hirabayashi, Doyle, McGovern, & Pasher, 2018). However, trees can also have an allergenic potential and produce volatile organic compounds (VOCs). These can entail respiratory health risks for populations living in both urban (Battisti et al., 2019; Eisenman, Jariwala, & Lovasi, 2019) and rural areas (Grewling et al., 2019; McInnes et al., 2017).

Human populations are not exposed to the same pollen patterns (Majkowska-Wojciechowska et al., 2007; Priftis et al., 2007), which can be strongly explained by the urbanization gradient (Bosch-Cano et al., 2011; McDonnell & Pickett, 1990). In a study of asthma prevalence in Northern Europe, Timm et al. (2016) showed that people growing up on livestock farms had less asthma than subjects growing up in inner cities, and observed a significant urban–rural gradient across six urbanization levels. For Belgium, the last national health interview survey data (Drieskens et al., 2018) show that allergy is most prevalent in suburban areas and lowest in rural areas, with major cities and urbanized municipalities in an intermediate position. Besides, urban forests exhibit unique ecosystem structure and their functioning often depends on the relations with the nearby suburban and rural forest stands (McDonnell et al., 2008). Therefore, mapping allergenic vegetation and source areas of pollen across the urban and rural fabric is crucial for a comprehensive assessment of respiratory health risks.

The distribution and abundance of allergenic trees at the regional scale has been found to impact mental (Stas, Aerts, Hendrickx, Dendoncker, et al., 2021) and physical health (Stas, Aerts, Hendrickx, Delcloo, et al., 2021) of adults with a birch pollen allergy in Belgian studies. Cariñanos et al. (2019) proposed an index of urban green zones allergenicity by considering, among other factors, the allergenicity of present trees, their number, and their corresponding area. This index has further been found to be strongly affected by allergenic tree abundance for assessing the allergenic risk of street trees and urban green spaces (Aerts et al., 2021; Zong, Yao, Tang, & Chen, 2020). Accurate distribution maps of allergenic vegetation are valuable for urban green space management and planning (Česnakaitė et al., 2019), especially in regions where knowledge on allergenic pollen is still limited such as the Middle East (Mansouritorghabeh, Jabbari-Azad, Sankian, Varasteh, & Farid-Hosseini, 2019).

In addition, assessment of the spatial distribution of pollen sources is an important prerequisite for the application of pollen dispersion models (Maya-Manzano, Skjøth, et al., 2021; Maya-Manzano, Smith, et al., 2021; Shin et al., 2020). Pollen transport models require source maps and, depending on the type of vegetation, these maps require seasonal updates (Mimić, Podraščanin, Lugonja, & Šikoparija, 2021; Verstraeten et al., 2021). Source maps for birch pollen distribution modelling can often be obtained from forest inventories (Maya-Manzano, Skjøth, et al., 2021; Maya-Manzano, Smith, et al., 2021; Verstraeten et al., 2019), although urban green spaces, which are also important sources of birch pollen, are in general not included in national forest inventories. There is clearly a need for abundance maps of allergenic vegetation that include vegetated areas outside forests and that are potentially time-dependant (i.e. covering several time periods across the year). However, such maps are rarely available for the allergy-relevant species such as alder, birch, or hazel. This is particularly critical for small- and medium-sized cities where systematic inventories are often missing or not readily available.

Making maps of abundance distributions at fine-scale is challenging. It requires robust information on the distribution and abundance of allergenic tree species. Several recent studies have suggested promising approaches using species distributions models (SDMs) that combine structured surveys from forest inventories with less structured data from citizen science initiatives (Feldman et al., 2021; Miller, Pacifici, Sanderlin, & Reich, 2019; Robinson, Ruiz-Gutierrez, Reynolds, Golet, Strimas-Mackey, & Fink, 2019). On the one hand, forest inventories are structured, with systematic surveys collected via a network of geo-located, ground-based sampling plots in which forest structure and tree species composition are measured to calculate estimates of forest attributes. The advantage of forest inventory data is that it is gathered via repeated visits at fixed sites. Records from different time periods can be more easily compared with one another. Forest inventory data also provide more consistent information because they contain a complete set of records that constitutes all the observed species in forest areas.

On the other hand, biological records from open citizen science initiatives are mainly opportunistic, mostly incidental and often less structured observations. Citizen science data is thus suspected to be information-poor compared to information-rich databases from the above-mentioned forest inventories (Bird et al., 2014; Dobson et al., 2020; Isaac & Pocock, 2015). Yet, it contains a greater number of records that have the potential to cover a large number of species at a fine-scale, over potentially long periods, and over large geographical areas, including places outside forest areas (Kosmala, Wiggins, Swanson, & Simmons, 2016). Since the quantity and extent of this data can never be reached by systematic surveys, opportunistic data can make a valuable contribution to science if processed correctly (Giraud, Calenge, Coron, & Julliard, 2016; Soroye, Ahmed, & Kerr, 2018; Steen, Elphick, & Tingley, 2019; Van Eupen et al., 2021). However, the literature lacks studies that focus on plant taxa and combine both types of datasets to explicitly explore tree distributions in relation to the pattern of urbanization.

As a response, the main goal of this research is to produce high-resolution maps of allergenic tree species using forest inventories and opportunistic citizen science data that better capture the spatial variations of species abundances across heterogeneous urban–rural landscapes such as Belgium. The country's contemporary landscape includes 54 sub-types of landscapes ranging from urban to forest landscapes, along with suburban, industrial, and pastural landscapes (Van Eetvelde & Antrop, 2009), which makes the nation an ideal candidate for measuring the spatial variations of allergenic tree species across the urban–rural gradient. Besides, it has well-structured forest inventory surveys covering the entire country and a growing number of citizen science initiatives providing datasets of which the potential for ecological modelling has not been fully exploited yet.

2. Data

We modelled allergenic tree species across Belgium using forest inventory data and citizen science data gathered across both the Flemish and the Walloon regions. We focused on three indigenous common taxa of allergenic tree species, namely hazel (*Corylus avellana*), alder (*Alnus* spp., including *A. glutinosa* and *A. incana*) and/or birch (*Betula* spp.). These are found to be the most important sources of allergenic tree pollen in urban green spaces of NW Europe (Damialis, Traidl-Hoffmann, & Treudler, 2019).

2.1. Forest inventory data

The most detailed and up-to-date information about tree abundance currently available in Belgium is contained within forest inventories. In Belgium, forest covers about 21% of the territory and the distribution of the forest over Flanders (north) and Wallonia (south) is 77,0% and 22,8% respectively (with 0,2% for the Brussels region). In both administrative regions, a region-wide survey characterizing the type of tree species observed within forested areas has been undertaken over the period of 2009–2017 in Flanders and 2008–2017 in Wallonia, respectively, (see Bos, 2001; Alderweireld, Burnay, Pitchugin, & Lecomte, 2015; see Rondeux & Lecomte, 2010). While Wallonia and Flanders conducted separate forest inventories, the sampling design and variables are similar. About 696.300 ha of forested areas are regularly sampled across both regions on the basis of a regular grid of 500 m × 1000 m. The sampling is limited to forests with a surface area equal to or greater than 0.1 ha (Alderweireld et al., 2015). For each sampling plot, a GPS coordinate is recoded and detailed inventory of tree characteristics is undertaken within a 36-meter radius around the centre of each sampling plot (age, type of tree stand, species, trunk circumference at 1.5 m, and total height). The datasets compiled contained 2.664 (17%) sampling plots situated in Flanders and 13.228 (83%) in Wallonia and included information on hazel and alder trees at the genus level. Yet, there is no distinction between the different types of birch species, although survey authors assume that *B. pendula* and *B. pubescens* are the most common species encountered in the country.

The purpose of our analysis was to: (i) extract presence-absence information about allergenic trees from regional forest inventories and (ii) provide an indicator of their relative abundance within a given tree stand. Abundance was measured by calculating a basal area, i.e. the area of a given section of land that is occupied by the cross-section of tree trunks and stems at the base. In practical terms, we calculated the specific basal area of each allergenic tree species studied and divided this value by the total basal area of all trees observed within the same sampling plot (both expressed in m^2/ha) (see Dujardin, Linard, and Dendoncker, 2017 for more details). We chose using the basal area as a metric of abundance instead of the number of individuals because it is best suited to quantify the size of trees and thus pollen productivity. The larger the crown volume of the tree, the more pollen can be produced. The percentage of basal area per sampling plot thus best allow comparing the relative abundance of allergenic tree species within a tree stand. As Stas, Aerts, Hendrickx, Delcloo, et al. (2021) showed, the domination (or high relative basal area) of allergenic trees is problematic for the health of allergy sufferers. Yet, when allergenic trees are present with a low relative basal area, green space can still contribute to positive health outcomes. Presence-absence information was extracted by classifying null or non-null abundance as an absence or presence data-point respectively.

2.2. Citizen science data

Data records of allergenic tree species were also retrieved from the *observations.org* website. This platform gathers observations of plants made by volunteers and scientists. Its aim is to facilitate observers to make their experiences with the environment even more valuable for conservation, research, education or policy making. Citizen science (CS) data is used here as a valuable open access data source providing opportunistic information about the presence of allergenic tree species in areas not visited by surveyors from forest inventories. Data selection criteria included observations made between 2008 and 2018, spatial accuracy of 100 m (or better), and verified data only (i.e. data for which the quality was controlled and approved on the basis of evidence, expert judgement or knowledge rules). Field attributes describing trees' stage (e.g. vegetative, flowering, fruit-bearing) and number were not used for filtering observation records.

In order to effectively combine the selected CS dataset with the FI dataset, we first checked for overlapping observations and removed 46 entries from the CS dataset because these fell within the 36-meter radius of forest inventories' sampling plots. Then, we generated pseudo-absences out of opportunistic observation records from the full CS dataset. These pseudo-absences were drawn upon the notion of sampling effort, i.e. the degree to which a set of records is an accurate reflection of the organisms that were actually present (Isaac & Pocock, 2015). Starting from the dataset containing all tree species from *observations.org*, we considered the centroid of a 100 × 100 m grid cell as a pseudo-absence data-point when 10 or more other tree species were observed within the cell, but no targeted allergenic tree was found. This search effort analysis allowed for creating a unified dataset compatible with presence-absence data from forest inventories. Lastly, for each allergenic tree species, we kept only one observation per 100 × 100 m grid cell from the CS database. Applying this spatial filtering (or thinning) technique (Boria, Olson, Goodman, & Anderson, 2014) reduces sampling bias and discards potential duplicate observations made by different observers at the same location.

The combined database contains 14.926 presence data-points and 38.710 absence data-points (see details in Table S1. The total amount of data-points is fairly similar by tree species with 17.834, 17.774, and 18.028 entries for *Alnus*, *Betula*, and *Corylus* respectively. The number of data-points in the FI dataset is much larger (40.452 entries) than in the CS dataset (13.184 entries). The proportion of absence data-points is larger in the FI dataset (92.5%) because it comes from a systematic survey. For CS data, the number of pseudo-absence data-points is proportionally lower (9.8%) as pseudo-absences rely upon the density of observation records reported by volunteers. Respectively, 527, 435, and 334 pseudo-absences for *Alnus*, *Betula*, and *Corylus* were created from the sampling effort analysis.

Fig. 1 shows the spatial pattern of geo-located tree records of either *Alnus*, *Betula*, and *Corylus* from both forest inventory and citizen science data across Belgium. The number of data points is significantly higher in the southern part of the country (35.488 records in the Walloon region) than in the northern part (18.148 records in the Flemish region) as the southern region shows a higher proportion of forested areas sampled intensively in forest inventories. The spatial footprint of citizen science data is larger than forest inventory data, although the number of data points per square kilometre is lower compared with the forest inventory data. The CS dataset covers non-forested areas, including major urban areas where a higher number of people potentially sensitive to pollen are located.

3. Methods

We followed a two-step ecological modelling approach (see Fig. 2. First, we used forest inventory and citizen science data to build a species distribution model (SDM) (Elith & Leathwick, 2009; Guisan & Thuiller, 2005; Guisan et al., 2007; Thuiller, 2003) and predicted species habitat suitability, using environmental characteristics as predictors (see Table 1. Second, we used abundance records from forest inventories to build statistical regressions and model the relationships between abundance, the habitat suitability predicted by the SDMs, and additional vegetation cover covariates (see Table 2 assumed to significantly explain tree abundance (Hill et al., 2017). Such a type of approach proved

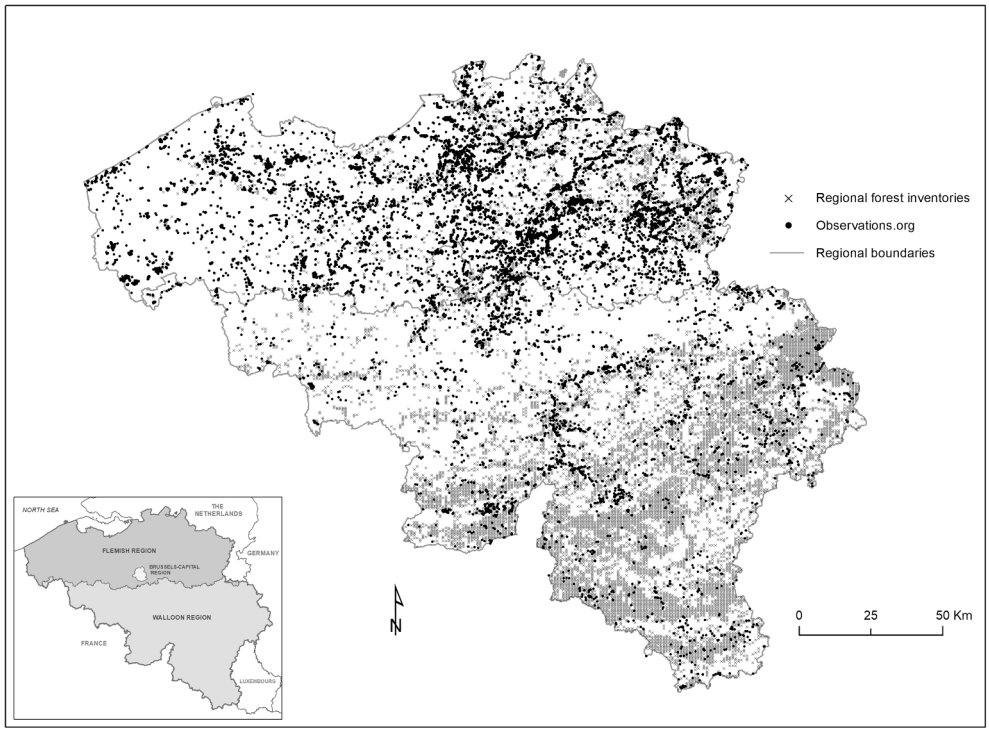

Fig. 1. Spatial distribution of data-point records per type of data source for three selected allergenic tree species (Alnus, Betula, Corylus) across Belgium.

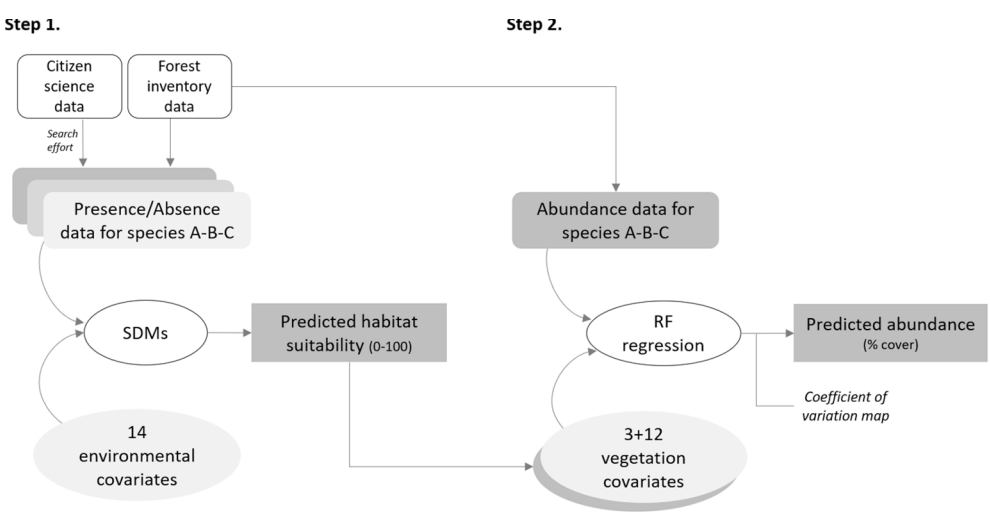

Fig. 2. Outline of the two-step method followed for predicting abundance distributions of three selected allergenic tree species. The first step uses species distribution modelling (SDM) to produce maps of predicted habitat suitability for Alnus, Betula, and Corylus (A-B-C). The second step integrate these maps as inputs and uses random forest regressions to produce maps of predicted abundance. Distribution data inputs are shown in rounded square boxes and model covariates in round boxes, and model outputs are shaded in solid grey and modelling processing in white.

successful for modelling abundance over large spatial extents, especially when large amount of presence or presence absence data is available (e.g. Barras, Braunisch, & Arlettaz, 2021). Yet, we focused here on highlighting differences in model outcomes depending on the type of dataset used for modelling allergenic tree species distribution. Besides, for each step, we chose the random forest algorithm because it is less sensitive to data distribution (Hastie, Tibshirani, & Friedman, 2009; Prasad, Iverson, & Liaw, 2006), which vary among tree species and datasets. Besides, it can take a large number of potentially collinear covariates and reduces the risk of model overfitting (Breiman, 2001). This novel approach was expected to yield a fairly accurate allergenic tree distribution map encompassing both the urban and rural landscapes of Belgium.

3.1. Step 1: predicting habitat suitability

In step 1 we used the R package *biomod2* (see Thuiller, 2003) to compute a SDM with random forest showing the habitat suitability of three targeted allergenic tree species in Belgium at 1-hectare resolution (100 m x100m). We started the analysis with the collection of forest inventory data from the Flemish and Walloon regions in combination with citizen data from *observations.org* available in open access. This allowed building a representative sample of allergenic tree observation records covering both forested and non-forested areas of Belgium. Then, we created a set of independent covariates to predict habitat suitability. Drawing upon the literature and previous studies, we chose 14 environmental covariates that are important for plant growth (see Guisan et al., 2007; McInnes et al., 2017; Prentice et al., 1992).

We paid specific attention to include covariates capturing natural and human factors that affect plant growth and cover both urban and rural environments. The altitude, slope, and soil type covariates are expected to better suit the description of rural environments, while land use types and distance to roads are expected to best describe the

Table 1
List of environmental covariates for modelling species distributions.

Layer name	Description	Map unit	Data source and processing
r_altitude	Altitude	m × 10	NGI-IGN Digital Terrain Model (DTM)
r_aspect	Aspect	Degrees	NGI-IGN. Derived from Altitude using ArcGIS (Slope)
r_slope	Slope	%	NGI-IGN. Derived from Altitude using ArcGIS (Slope)
r_sra	Direct radiation: incoming direct solar radiation	Watt hr m^{-2}	NGI-IGN. Derived from Altitude using ArcGIS (Solar Radiation Analysis)
r_bio2	Mean diurnal temperature range: mean of monthly (max temp – min temp)	°C × 10	KMI-IRM. Computed from hourly temperature records over the 2008–2018 period using R (dismo package)
r_bio4	Temperature seasonality: standard deviation × 100	°C × 1000	KMI-IRM. Computed from hourly temperature records over the 2008–2018 period using R (dismo package)
r_bio15	Precipitation Seasonality: coefficient of variation	Fraction	KMI-IRM. Computed from hourly precipitation records over the 2008–2018 period using R (dismo package)
r_soil_drainage	Soil drainage class	Nominal	Regional soil maps. Compiled using R (plyr and sf package)
r_soil_texture	Soil surface textural class	Nominal	Regional soil maps. Compiled using R (plyr and sf package)
r_soil_profile	Soil profile development class	Nominal	Regional soil maps. Compiled using R (plyr and sf package)
r_LU_y2018	Land use/land cover (LULC) type	Nominal	VITO. Integrated LULC map of Belgium used as such
r_AncientF	Ancient woodland	Binary	Regional maps derived from Ferraris maps 1770 and 1778
r_distRORA	Distance to roads and railways	m	NGI-IGN Top10vGIS map. Computed using ArcGIS.
r_distHYD	Distance to water bodies (e.g. rivers, sources, lakes)	m	NGI-IGN Top10vGIS map. Computed using ArcGIS

conditions of urban environments. The temperature and precipitation covariates are derived from a downscaled climate model simulation that integrate the urban heat-island effect and thus provide adjusted values depending on the built-up density across the country.

We downloaded data describing environmental characteristics from multiple sources. We pre-processed layers using ArcGIS v. 10.0 (ESRI, 2014) and R Studio (v1.4.1106) software and produced a set of 14 environmental covariates at 100 m resolution with identical coordinate system, spatial resolution (100 m) and extent (Table 1. Only four vector datasets had to be rasterized to 100 m resolution (soil and ancient forest covariates), limiting distortion error and the modifiable areal unit problem (Dark & Bram, 2007; Hanberry, 2013; Dendoncker, Schmit, & Rounsevell, 2008). Each pair of covariates had a pairwise Pearson's correlation coefficient lower than 0.7 (see correlation matrix in Table S4.

Next, we built SDMs to predict the habitat suitability across the wider urban–rural landscape of Belgium. We used random forest to fit a model that best describes the statistical relationship between observation records of allergenic tree species' and our set of environmental covariates. As Hill et al. (2017, p. 1047) showed in their study about abundance

Table 2
List of vegetation cover covariates for modelling abundance

Set	Layer name	Description	Scale	Value	Data source and processing
	r_1_S	Coniferous woodland	100 m	Average m^2/ha within a 1 km buffer	NGI_IGN Top10vGIS. Computed using R (raster package) from polygon features.
	r_2_Sf	Predominantly coniferous mixed woodland	100 m	Average m^2/ha within a 1 km buffer	NGI_IGN Top10vGIS. Computed using R (raster package) from polygon features.
	r_3_FS	Mixed woodland	100 m	Average m^2/ha within a 1 km buffer	NGI_IGN Top10vGIS. Computed using R (raster package) from polygon features.
C_F	r_4_Fs	Predominantly broad-leaved mixed woodland	100 m	Average m^2/ha within a 1 km buffer	NGI_IGN Top10vGIS. Computed using R (raster package) from polygon features.
	r_5_F	Broad-leaved woodland	100 m	Average m^2/ha within a 1 km buffer	NGI_IGN Top10vGIS. Computed using R (raster package) from polygon features.
	r_15_Kb	Unspecified herbaceous vegetation with brushwood	100 m	Average m^2/ha within a 1 km buffer	NGI_IGN Top10vGIS. Computed using R (raster package) from polygon features.
	r_19_J	Gardens	100 m	Average m^2/ha within a 1 km buffer	NGI_IGN Top10vGIS. Computed using R (raster package) from polygon features.
L_F	r_LinVeg	Proportion of edge rows	100 m	Average m/ha within a 1 km buffer	NGI_IGN Top10vGIS. Linear vegetation features
I_F	r_IsoVe	Density of isolated trees	100 m	Average N/ha within a 1 km buffer	NGI_IGN Top10vGIS. Point vegetation features.
	r_NDVI_min	Normalized difference vegetation index (2008–2016)	1 km	Min NDVI	LIFEWATCH project database
N_F	r_NDVI_max	Normalized difference vegetation index (2008–2016)	1 km	Max NDVI	LIFEWATCH project database
	r_NDVI_mean	Normalized difference vegetation	1 km	Long Term Average NDVI	LIFEWATCH project database

(continued on next page)

Table 2 (*continued*)

Set	Layer name	Description	Scale	Value	Data source and processing
		index (2008–2016)			
	r_HS_Alnus	Habitat suitability of Alnus (computed in step 1)	100 m	Probability on a 0–100 scale	Regional forest inventories and Observation.org
HS	r_HS_Betula	Habitat suitability of Betula (computed in step 1)	100 m	Probability on a 0–100 scale	Forest inventories and Observation.org
	r_HS_Corylus	Habitat suitability of Corylus (computed in step 1)	100 m	Probability on a 0–100 scale	Forest inventories and Observation.org

distributions for tree species in Great Britain, the random forest algorithm allows for building ensemble distribution models (Araújo & New, 2007) with high accuracy, including for tree species such as *Alnus*, *Betula*, and *Corylus*.

We ran the random forest algorithm 15 times for each species using the 14 environmental covariates and compared the results obtained when using forest inventory data only (FI), citizen science data only (CS), and from the combination of both (FI + CS). We produced 135 modes in total (3 species × 15 repeats × 3 datasets). We ran each model with 70% of the presence–absence data (Heikkinen, Marmion, & Luoto, 2012; Thuiller, 2003) and used the remaining 30% for cross-validation and assessing model performance. The model assessment criteria included the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC – Swets (1988)) and the sensitivity (SENS) as a measure of the correctly predicted presences. We kept the best-performing model to build an ensemble distribution model (Araújo & New, 2007), i.e. we calculated for each allergenic tree species a mean of the raw model results weighted by the model AUC scores. Finally, model results were projected on a continuous grid map showing the habitat suitability for the three targeted allergenic tree species across Belgium.

3.2. Step 2: modelling abundance

In step 2 we built random forest regressions using the caret and randomForest packages in R-software (Kuhn, Wind, Weston, Williams, Keefer, Engelhardt, and Benesty (2016) and Liaw and Wiener (2002), respectively) to link the results of the SDMs to abundance data from forest inventories and predict abundance across Belgium at the same resolution. We extracted abundance data (the dependent variable) from regional forest inventories and implemented a separate random forest regression for Alnus, Betula and Corylus. We used 15 vegetation cover covariates in the random forest regression (see Table 2. These included the habitat suitability maps of each allergenic tree species in order to account for biotic effects and potential interactions between species. The remaining covariates described environmental characteristics that are expected to influence abundance but not occupancy (Hill et al., 2017).

Drawing up the vectorised topographic map (Top10vGIS) provided by the National Geographic Institute on a cartographic scale of 1:10:000, we computed the following vegetation cover layers from the polygon features dataset: coniferous woodland, predominantly coniferous mixed woodland, mixed woodland, predominantly broad-leaved mixed woodland, broad-leaved woodland, unspecified herbaceous vegetation with brushwood, and gardens. In addition, we derived the number of isolated trees and the length of hedgerows per hectare from the point and linear features dataset respectively. We then computed the average value of each metric per buffer of 1 km as to capture the broader environmental conditions surrounding each one-hectare raster cell. We also added normalized difference vegetation indices (min, max, and Long-Term Average) for the 2008–2016 period as to consider plant phenology and live green vegetation.

Random forest regressions can be expressed as follows: $\text{Abundance}_{\text{sp.A}} \sim HS_{\text{sp.A}} + HS_{\text{sp.B}} + HS_{\text{sp.C}} + C_F + I_F + L_F + N_F$; where HS is the habitat suitability computed in stage 1, C_F are the 7 selected land cover covariates, I_F the proportion of hedge rows, L_F the density of isolated trees, and N_F are the three selected vegetation indices (see Table 2. F indicates that we used the same covariate to model each of the three targeted allergenic tree species. For each allergenic tree species, we computed 3 models in order to compare results obtained when using habitat suitability maps computed from the FI, CS and FI + SC datasets (3 species × 3 datasets = 9 models). We evaluated model performance by implementing a 10-fold cross-validation that leaves out 10% of our dataset at each run (Gareth, Daniela, Trevor, & Robert, 2013). The evaluation metrics retained for evaluating model's average error included the root-mean-square error (RMSE) and mean absolute error (MAE).

This allowed for producing a map that represents an estimation of abundance for each allergenic tree species across Belgium at 1-hectare resolution. However, these estimates are based on forest inventory data as dependent variable. Hence, each value estimated relies upon the assumption that each raster cell across the country is entirely covered by forest. Therefore, we rescaled the abundance values obtained from random forest regressions (in percentage of basal area) as to provide a final metric that is representative of the actual vegetation cover present within each raster grid cell (in number of hectares covered by a species per square kilometre). This operation took the form of: (abundance value in percent of basal area) × (percent cover of vegetation). The information regarding the percent cover of vegetation was extracted from the vectorised topographic map (Top10vGIS) and included all polygon features describing any type of vegetation such as woodland, heathland, grassland, and gardens. In this way, the final map products are more realistic and better represent the actual tree abundance found either in areas where vegetation covers an important part of land (e.g. rural areas) or in areas where there is a high proportion of artificial land (e.g. city centres).

We used R-software version 3.2.3 for all modelling and data processing (R Core Team, 2020).

4. Results

4.1. Species distribution modelling

Species distribution models from step 1 yielded good results providing high-resolution predictions of the spatial distribution of habitat suitability for the three selected allergenic tree species (*Alnus*, *Betula*, and *Corylus* – See Fig. S1. The prediction capability of all 15 random forest models was fairly high (see Table 3). AUC scores ranged from 0.68 to 0.94 and sensitivity scores from 61.4% to 89.5%. Models using only FI data allowed for good predictions of these three allergenic tree species, especially for *Alnus* (average AUC = 0.90 and sensitivity = 80.5%). Models using only CS data suffered from the lower number of data-point records. These had lower predictive performance, especially for *Corylus* that showed the lowest model performance score (AUC = 0.68) and percentage of predicted presences (sensitivity = 76.9%). The species distribution models produced with both FI and CS data consistently showed stronger links between the dataset and selected environmental covariates. For each species, all 15 models had an AUC score greater than 0.8. The combination of both datasets improved model performances for all three species. AUC scores were significantly higher for *Alnus* and *Corylus*. For *Betula*, both AUC and sensitivity scores were significantly higher. For each species, all 15 random forest models were used for building the final weighted mean ensemble model as their AUC scores were all above 0.70.

Table 3
Summary statistics of the 15 models built for producing habitat suitability maps. Distinction is made between models computed with forest inventory data only (FI), citizen science data only (CS), and the combination of both dataset (FI + CS). AUC = average Area Under the Curve (AUC ± SD). SENS = percentage of correctly predicted presences or sensitivity. n80 = Number of models with AUC greater than 0.80. N = number of modelled data-points. Asterisks placed at the highest value indicate outcomes of one-way ANOVA tests and show significant differences for the average AUC and sensitivity between the FI, CS and FI + CS models (***$p < 0.001$ or **$p < 0.01$).

	FI				CS				FI + CS			
	AUC	SENS (%)	n80	N	AUC	SENS (%)	n80	N	AUC	SENS (%)	n80	N
Alnus	0.90 ± 0.01	80.5 ± 4.1	15	13,484	0.84 ± 0.02	83.0 ± 5.6	15	4350	0.94 ± 0.00 ***	89.5 ± 1.3	15	17,834
Betula	0.78 ± 0.01	61.4 ± 3.9	0	13,484	0.80 ± 0.02	74.5 ± 10.8	9	4290	0.86 ± 0.00 ***	78.5 ± 1.9 ***	15	17,774
Corylus	0.76 ± 0.01	82.8 ± 5.8 **	0	13,484	0.68 ± 0.03	76.9 ± 5.4	0	4544	0.87 ± 0.01 ***	78.7 ± 1.4	15	18,028

Fig. 3 shows variations of model predictions of habitat suitability for each source dataset. The habitat suitability derived from FI data shows high probabilities of finding *Corylus* trees outside cities where vegetation and main forest patches are located. The habitat suitability map derived from CS data in turn shows a high probability of occupancy of *Corylus* trees both in highly urbanized areas and across the less urbanized areas located outside the city. Habitat suitability maps combining both FI data and CS data show high probabilities of finding *Corylus* at the bottom of river valleys, within urbanized areas, as well as along road networks.

4.2. Abundance modelling

Random forest regressions performed in step 2 led to good predictions of species' abundance of the three selected allergenic tree species across the country. Table 4 below shows the statistics obtained from

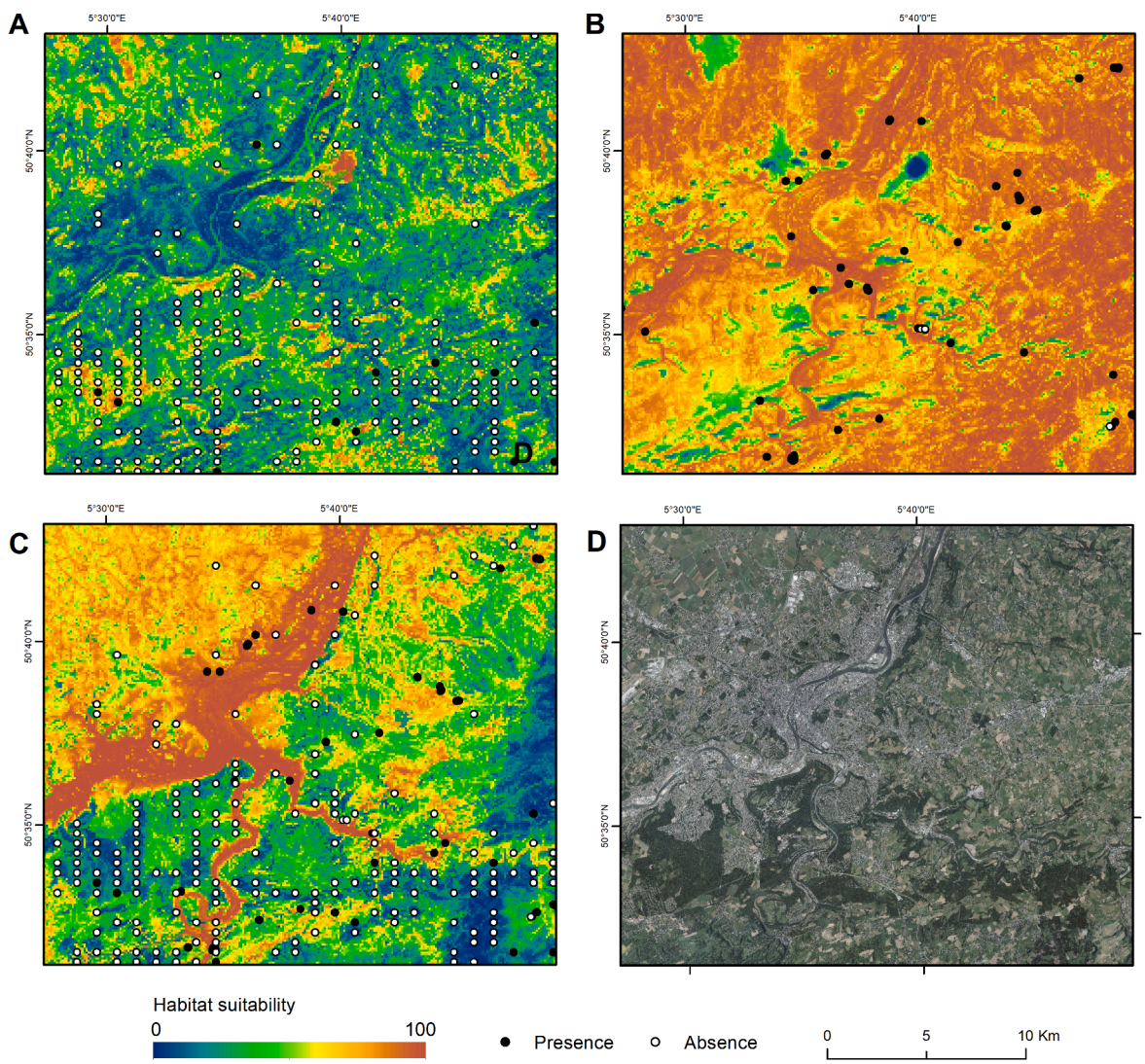

Fig. 3. Maps showing differences in model predictions of habitat suitability for Corylus in the area of Liège city in the South East part of Belgium. Map A is obtained using presence-absence data from forest inventory and shows high probability of presence in forest areas. Map B is computed using presence and pseudo-absence data from citizen science data and shows the highest probability of presence in urban and countryside areas. Map C shows predictions obtained using both forest inventory and citizen science data and presents a more balanced distribution of Corylus both within the urban fabric and along the urban–rural gradient. As a reference, a satellite image (Map D) highlights the pattern of the heterogeneous urban–rural landscape in the area.

Table 4
Descriptive statistics of the root mean squared error (RMSE), mean absolute error (MAE) and coefficient of determination (R^2)

	FI			CS			FI + CS		
	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2
Alnus	2.87	0.56	0.93	4.43	1.24	0.93	2.77	0.53	0.94
Betula	4.03	1.19	0.93	6.22	2.77	0.94	3.99	1.17	0.93
Corylus	1.51	0.27	0.91	2.30	0.96	0.75	1.53	0.27	0.91

model outputs and provide indicators to assess model performance. The root-mean-square error (RMSE) and mean absolute (MAE) scores are provided at the same scale as the dependent covariate, i.e. a percentage of basal area. All models have a RMSE score lower than 7 and most of them have a score of 4 or below. MAE scores are all below 3, suggesting that the average error of prediction is under 3%. RMSE and MAE scores are different, but to a reasonable scale. This suggests the variance in error scores is limited. Differences between very good predictions and very bad predictions in the model is somehow moderate. Such difference is lower for the model using FI data and stronger with the model using CS data. The model using a combination of the two datasets shows intermediate differences. R^2 scores are above 0.90 for all three allergenic tree species, suggesting that the model predictions are very close to the actual abundance measured in forest inventories.

Fig. 4 shows the final maps of predicted abundance for *Alnus*, *Betula* and *Corylus* in percent cover using the combined dataset (FI + CS). While high species abundance may be expected in the southern part of the country where forest landscapes prevail, abundance is higher in urban and agricultural landscapes of the northern part of the country because topographic elevation is generally lower there and forests in the south are dominated by other tree species such as conifers. High abundance values are also shown at the bottom of river valleys. The pattern of *Alnus* abundance distributions mainly follows the river network across the country, especially in lowlands. A larger area of *Alnus* densities is also located along the seashore, especially in West-Flanders. *Betula* abundance is also high in the northern part of the country. Hotspots are located in the north-eastern part (Campine) and north-western area (East-Flanders) of Brussels. Highly urbanized areas, including cities located in the main river valleys of Wallonia, show higher abundances values. For *Corylus*, abundance values tend to be higher in cities, especially where building density is lower, allowing for the development of shrubs and small trees. The species is abundant in peripheral areas and places not suitable for the growth of large trees (steep slopes areas). *Corylus* is especially less abundant in the Ardennes uplands, as well as in the north-eastern part of the country. The final maps products are viewable and downloadable at full resolution from https://s-dujardin.sh

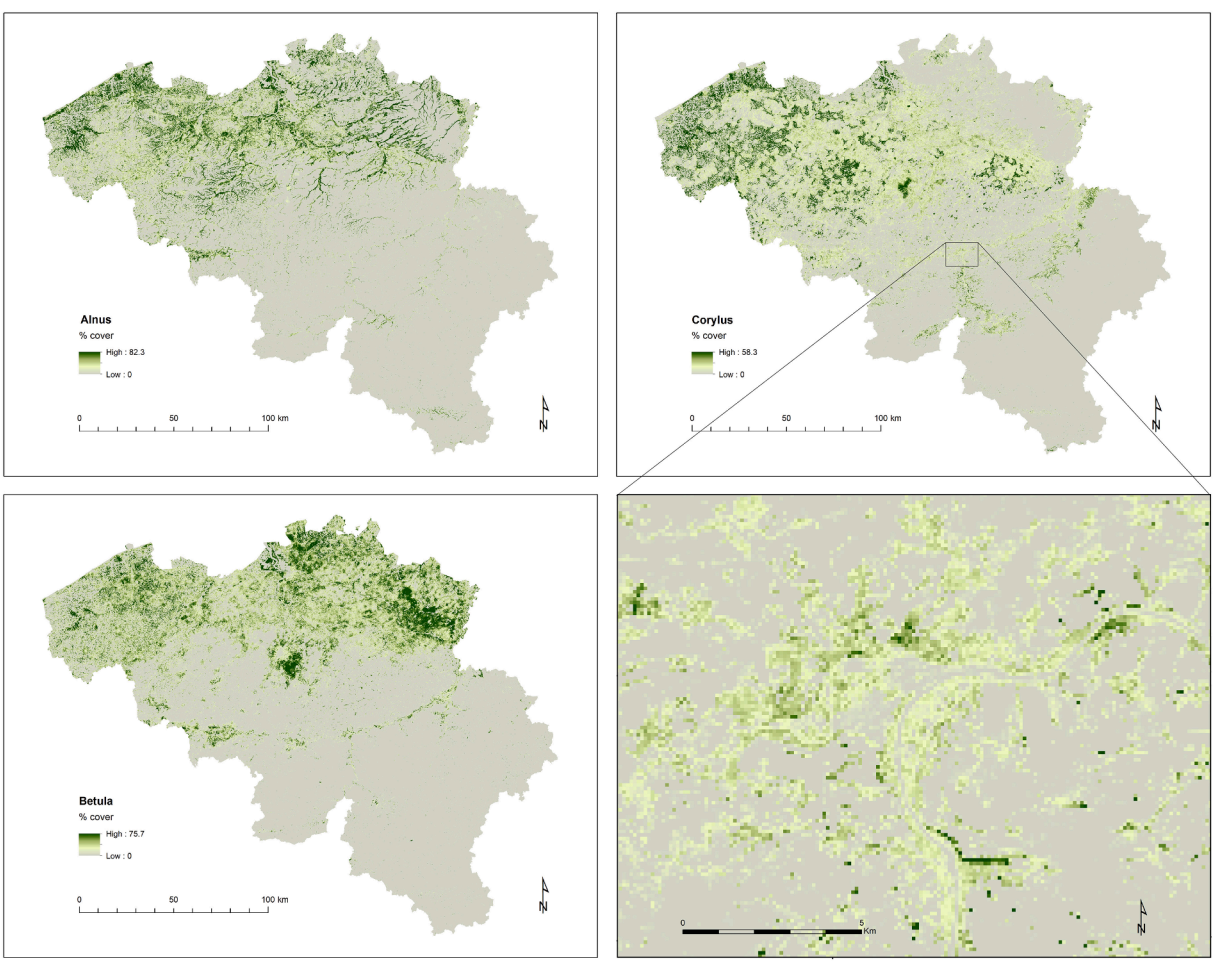

Fig. 4. Percent cover map for Alnus, Betula and Corylus. These maps are produced from a combination of species distribution models and random forest regressions and show the spatial distribution of three targeted allergenic tree species across Belgium at a 100 m × 100 m resolution. A detailed view centred on the small-size city of Namur is provided, showing spatial variations of Corylus density within the city centre and its peri-urban, less urbanized areas.

inyapps.io/shinyapprespirit.

Such a species abundance distribution pattern is in line with model responses obtained from covariate importance scores calculated from the Out Of the Bag (OOB) data (see Table S2 and Table S3 both in step 1 and 2. For all tree species, altitude is a major factor limiting plant growth and the distribution of allergenic tree species where topographic elevation is low (importance scores of 18.4, 20.3, and 28.5 for *Alnus*, *Betula*, and *Corylus* respectively). For *Alnus*, distance to water is of high importance (importance = 8.6), suggesting a higher probability of occupancy close to rivers, streams, and lakes. In the case of *Betula* trees, precipitation and temperature seasonality (importance = 7.4 and 6.7, respectively) and the type of land use land cover (importance = 4.6) are also important for explaining their distribution. The habitat suitability of *Corylus*, in turn, is driven by temperature seasonality (importance = 11.9) and land use land cover (importance = 7.6). Covariate importance from random forest regressions shows that the habitat suitability of the modelled species is of utmost importance (importance scores of 100 for *Alnus*, *Betula* and *Corylus*) as well as the habitat suitability of the two other tree species, especially for *Alnus* and *Corylus*. The minimum NDVI contributes well to the modelling of abundance for *Betula* (importance = 18.4). The spatial pattern of abundance distributions can be further explored in Fig. S2 where a detailed view around three small and medium sized cities of Belgium is provided.

4.3. Spatial uncertainty

We computed a coefficient of variation map with the best performing model (FI + CS) to further explore the spatial uncertainty of predictions. The coefficient of variation (CV) is computed on the prediction of 25 models runs in step 2 (modelling abundance) and is presented in Fig. 5. CV values were obtained by dividing the standard deviation of the predicted values by the mean of the predicted values. For all three species, coefficients of variation values are higher in areas located in the south-eastern part of the country, suggesting high spatial prediction uncertainty in areas where large tree stands exist. Low CV values, on the opposite, tend to be clustered in areas were vegetation is less present, including major cities and highly urbanized areas of Belgium. The range of CV values (0 to 5) is the same for all three allergenic tree species, showing consistency between species in the spatial uncertainty of predictions.

5. Discussion

This research aimed at producing a detailed cartography of allergenic tree species abundance using both forest inventory data and citizen science data. While an increasing number of studies are combining structured data from planned surveys with opportunistic data from open citizen science initiatives (Miller et al., 2019; Pacifici et al., 2017; Robinson et al., 2019), studies on vascular plants remain rare compared to animal taxa (Feldman et al., 2021). To our knowledge, no other previous studies combined these two datasets to predict abundance distributions of allergenic tree species at a fine spatial resolution.

The ecological modelling approach presented here has a number of advantages for capturing allergenic tree species distribution in urbanized landscapes.

First and foremost, combining forest inventory data and citizen science data offered a better coverage of data-points across the country, including within urbanized areas where a higher number of people potentially sensitive to pollen are located. This improved model predictions and cartographic results by showing a more balanced pattern of habitat suitability both within the urban fabric and along the urban–rural gradient. Such results suggest that citizen science data allow for better modelling allergenic trees in highly human-modified landscapes where tree distributions have been strongly influenced by human land use and preferences (Hopkins & Kirby, 2007; Rackham, 2008). This was particularly clear with the modelling of *Corylus*' habitat suitability for which land use types are of high importance to explain its distribution.

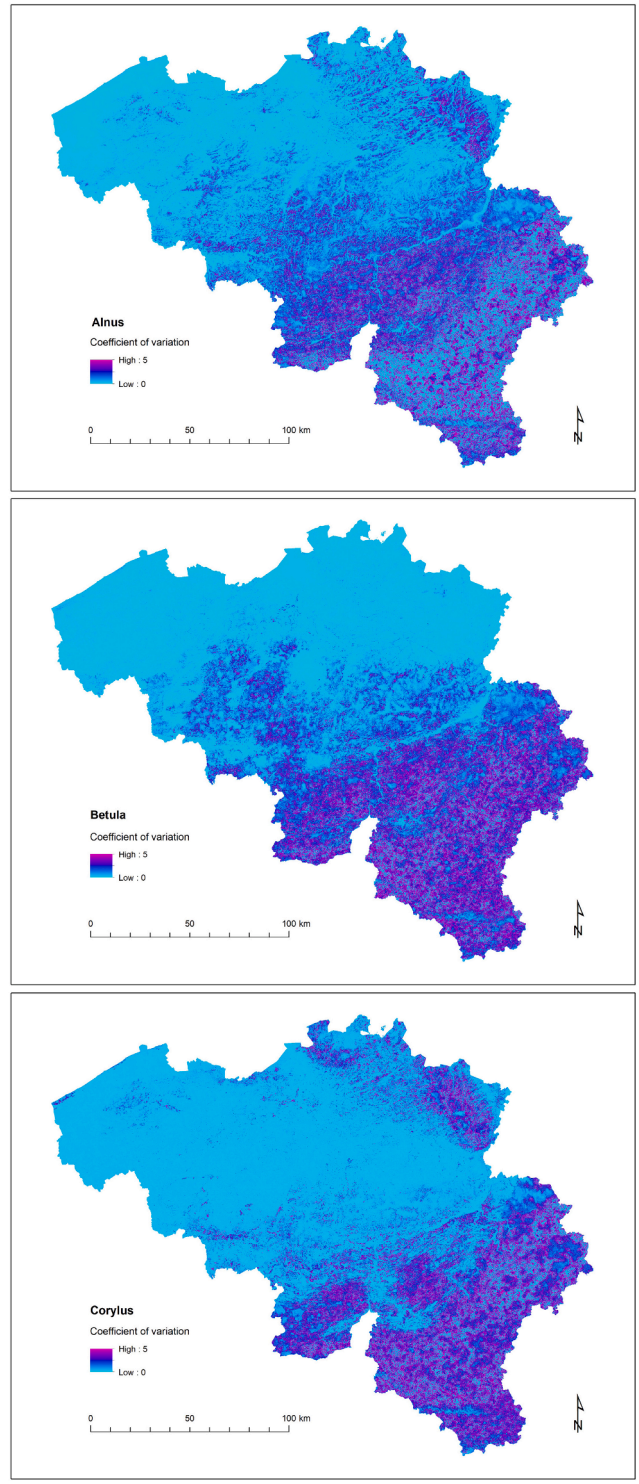

Fig. 5. Coefficient of variation map of abundance predictions for Alnus, Betula, and Corylus. The coefficient of variation is computed on the predictions from 25 model runs for each allergenic tree species.

Another advantage is the greater availability of citizen science data compared to the structured survey data. Citizen science data makes the access to large, up-to-date tree observation records easier, with less constraints on the time period studied. The high temporal resolution of data-point records used in this study helped reducing sampling bias through both the creation of absence data-points from the search effort analysis and the aggregation of individual observations at 1-hectare

resolution. These findings add up to previous studies such as Robinson et al. (2019) where the integration of filtered citizen science data with survey data increased the extent and accuracy of distribution models on shorebirds.

The ecological modelling approach can be reproduced in any region of the world where structured data and citizen science data are available. In some regions, however, volunteers may prefer to report on special or exotic species rather than common species (e.g. hazel trees). To compensate this gap, scientists can launch specific calls for citizens to report observations made in their gardens (see for instance the garden monitor project: https://mijntuinlab.be/). Beyond data availability issues, the method also presents a set of limitations that should be considered as opportunities for further research.

First, the information about tree abundance used for the random forest regression analysis in step 2 is only representative of the conditions observed in forests. This is because most data-points from forest inventories (83,5%) fall within the land use categories of mixed forest, deciduous forest and coniferous forest. In order to collect more information about abundance outside forests, the variable 'number of individuals' in the CS dataset should further exploited for modelling abundance. However, this information could not be retrieved form the CS dataset because 92.8% of data point entries contained only one individual, which leaves too much uncertainty for differentiating an actual count from unfilled or missing data.

Second, observations from citizen science data often contain an important number of records along pathways and roads where opportunistic observations often occur. This was very useful to capture presences of allergenic tree species in understudied areas where no systematic forest inventory will ever be made. However, one species observed on a woodland edge is not necessarily representative of the core of this woodland. The type of geographic distribution (core, satellite, see Collins, Glenn, and Roberts (1993)) may affect model performance and should be further explored for each type of allergenic tree species.

Third, the coefficient of variation maps shows that variability in predictions was greater in areas where large forest landscapes dominate. This indicates that some refinement in the environmental covariates compiled is needed and the subsequent modelling process revised. One way to explore this further is to add additional covariates that account for anthropogenic factors such as distance to city centres or the median age of buildings. Another way is to broaden the scope of species considered in the modelling for better capturing biotic interactions between allergenic tree species and non-allergenic tree species (König et al., 2021; Wisz et al., 2013). In the UK, Hill et al. (2017) for instance highlighted that the most important covariate in the random forest regressions of abundance for *Alnus glutinosa* was the habitat suitability of *Crataegus monogyna*. In Belgium, *Fraxinus* could also be an important covariate that further captures the *Alder-Fraxinus* forests of moist valleys. *Betula* and *Corylus* are fast growing species with weak competitive power. Yet, *Betula* can grow along with *Quercus* trees or in forests of sandy soils where *Pinus sylvestris* has often been planted. *Corylus* is an understory plant of broadleaved forests growing on rich soils alongside *Quercus* and *Fagus* trees. Further studies should include these species as additional biotic covariates in order to improve the accuracy of abundance modelling.

By covering the entire urban–rural landscape of Belgium at a fine spatial resolution, the abundance distribution maps produced here opens the way for a wide range of applications for urban forestry and respiratory health research.

Allergenic tree abundance is essential for the simulation of pollen levels in the air and pollen dispersion models, which requires the quantification of the spatio-temporal distributions of emission sources of pollen in the model domain (Verstraeten et al., 2019). Previously, studies looking at pollen concentrations in Belgium had to build pollen samplers located at a few fixed sites (4 in total in Belgium) and pollen concentration models were available at coarse resolution (1–4 km). The maps presented here already improved previous pollen concentration models by providing pollen emissions sources for birch trees at higher spatial resolution (Delcloo et al., 2018; Verstraeten et al., 2019).

For respiratory health studies, it can provide a meaningful indicator to assess how much allergenic tree species can emit pollen and thus trigger allergies and asthma (Stas, Aerts, Hendrickx, Delcloo, et al., 2021; Stas, Aerts, Hendrickx, Dendoncker, et al., 2021). In particular, these maps already opened up innovative ways of studying interactions between whereabouts of patients sensitive to tree pollen. The density maps of *Alnus*, *Corylus* and *Betula* were used in the exposure study of a Belgian cohort of 144 adults suffering from a tree pollen allergy. By combining the smartphone GPS tracks of the participants with the allergenic tree density maps the true exposure to source vegetation of allergenic pollen could be determined at high spatio-temporal detail revealing risk effects and protective effects (Stas, Aerts, Hendrickx, Delcloo, et al., 2021).

Finally, tree abundance distribution maps can also support the study of possible future health effects onto the growing urban population of Belgium and Europe. The modelling approach can be reproduced to support the development of scenarios of change in allergenic plant distributions. When coupled to climate and land use change scenarios (Dujardin, Linard, & Dendoncker, 2020), it can provide insights into the respiratory health implications of current and anticipated allergenic tree changes, meanwhile highlighting the potential effect of urban planning and green space management decisions in urbanized landscapes.

In practice, results can contribute to the well-being of urban and suburban communities in many ways. For the prevention of respiratory health risks, distribution maps of allergenic tree species could be used in a mobile application that would suggest 'safe areas' for physical activity during the pollen season, and would give 'warning signals' when a user is entering an area with high densities of allergenic tree species. For urban forest management, knowing the abundance distribution of allergenic plants both within cities and along the urban–rural gradient can guide urban planning and green space management decisions, for instance, by supporting a careful tree species selection in the creation of new green spaces, the development of green infrastructures, or the management of existing parks and wooded spaces. As suggested from the results of Stas, Aerts, Hendrickx, Delcloo, et al. (2021), large green spaces with a low density of allergenic trees should be promoted to guarantee health benefits that can also be enjoyed by tree-pollen allergy patients.

6. Conclusions

This research aimed at producing a high-resolution map of allergenic tree species from existing forest inventories and opportunistic open-access citizen science data. We focused on three main types of allergenic trees, namely hazel (*Corylus avellana*), alder (*Alnus* spp., including *A. glutinosa* and *A. incana*) and birch (*Betula* spp.). We explicitly explored the results obtained when using forest inventory data or citizen science data alone, and when combining both datasets. The latter setting brought about the best results and allowed producing a continuous, 1-hectare resolution map of alder, birch, and hazel showing spatial variations of abundance distributions both within the urban fabric and along the urban–rural gradient.

Such an ecological modelling approach is highly relevant for landscape and urban planning because it offers a better understanding of the existing and potential future allergy risk posed by green spaces. This opens the way for a wide range of applications that require fine-scale vegetation maps, including estimates of pollen concentrations or impact studies of allergenic trees on respiratory health. Knowing the local abundance distribution of allergenic plants can guide urban planning and green space management decisions, for instance by supporting a careful tree species selection and the development of large green spaces with low density of allergenic trees. This is indispensable for evidence-based urban green space policy and planning in support of

public health.

7. Data accessibility

- Elevation data: NGI-IGN Digital Terrain Model (DTM) available from http://www.ngi.be/FR/FR1-5–5.shtm (accessed 12/11/2018).
- Soil maps: Regional dataset from the Flemish and Walloon administrations available from http://www.geopunt.be/catalogus/datasetfolder/5c129f2d-4498-4bc3-8860-01cb2d513f8f (Accessed 17/07/2017) and https://geoportail.wallonie.be/catalogue/38c2a87e-d38a-4359–9899-9d4a6b9f0c2a.html (accessed 16/09/2016) respectively.
- Precipitation and temperature data: Meteorological data can be obtained upon request to the user-interface department of the KMI-IRM (freely available for research purpose).
- Land use and land cover map 2018: Integrated LULC map of Belgium produced by the Vlaamse Instelling voor Technologisch Onderzoekas (VITO) as part of the research project entitled "GDrought-related vulnerability and risk assessment of groundwater resources in Belgium (GroWaDRISK)" (Verbeiren, Huysmans, Tychon, Jacquemin, Canters, Vanderhaegen, and Tsakiris (2013). Available from https://growadrisk.marvin.vito.be/growadrisk_results.html (accessed 12/11/2018).
- LIFEWATCH data: Open e-Data for biodiversity and ecosystems continuously generated. Available from http://maps.elie.ucl.ac.be/lifewatch/geoviewer.html (accessed 02/04/2021).
- Ancient woodland maps: Regional dataset available from the Flemish administration (Vlaamse overheif, Instituut voor Natuur- en Bosonderzoek) and the Walloon administration (Service public de Wallonie, Direction de la Nature et de l'Eau (DNE)); available from http://www.geopunt.be/catalogus/datasetfolder/656a43e0-3254-47c4-976d-a71ee2e616cb (Accessed 18/09/2019) and https://geoportail.wallonie.be/catalogue/6ff283ae-8d33-48c6-9af9-b620939095b3.html (Accessed 20/09/2019) respectively.
- Abundance distribution maps. Final map products of *Alnus*, *Betula* and *Corylus* abundance distribution are available for download at https://s-dujardin.shinyapps.io/shinyapprespirit/

8. Data statement

The data from the topographic map used to compute model covariates (Top10Vector, identifier BE.NGI-IGN/5F4130E6-DF5C-41E6-A956-BB9F04088D11) are copyrighted (©Institut Géographique National) and were used under federal use license 2016_F014 granted by the Nationaal Geografisch Instituut (NGI-IGN) to the Belgian Science Policy Office (BELSPO). Observations of tree species from the data portal observations.org was granted by the Natuurpunt vzw and the Natagora asbl under the agreement contract 2019-131 ("Records tree and bush species").

CRediT authorship contribution statement

Sebastien Dujardin: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Visualization. **Michiel Stas:** Methodology, Investigation, Writing –review & editing. **Camille Van Eupen:** Methodology, Investigation, Writing –review & editing. **Raf Aerts:** Conceptualization, Writing – review & editing. **Marijke Hendrickx:** Funding acquisition. **Andy Delcloo:** Writing – review & editing, Funding acquisition. **François Duchêne:** Data Curation. **Rafiq Hamdi:** Writing – review & editing, Funding acquisition. **Tim Nawrot:** Funding acquisition. **An Van Nieuwenhuyse:** Funding acquisition. **Jean-Marie Aerts:** Writing – review & editing, Funding acquisition. **Jos Van Orshoven:** Writing – review & editing, Funding acquisition. **Ben Somers:** Writing – review & editing, Project administration, Funding acquisition. **Catherine Linard:** Conceptualization, Writing – review & editing, Supervision, Funding acquisition. **Nicolas Dendoncker:** Conceptualization, Writing – review& editing, Visualization, Supervision, Funding acquisition.

Acknowledgements

Results presented in this study were carried out in the framework of the RespirIT project supported by a national grant from the Belgian Science Policy Office BELSPO (grant nr. BR/154/A1/RespirIT). It was also supported by the Flemish Research Foundation FWO–SB [grant number 1S92118N]. Authors would like to warmly thank S. Bauwens (BIOSE department - Biosytem Engineering, Forest Resources Management, Gembloux Agro-Bio Tech, Université de Liège) as well as L. Govaere (Cel Beheerplanning en Monitoring, Agentschap voor Natuur en Bos, Vlaams Gewest) for their many advices and feedback for selecting and processing forest inventory data. We are also very thankful to C. Visée (Geography Department, UNamur) for contributing to the development of the WebGis application and making abundance distribution maps readily available for download.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.landurbplan.2021.104286.

References

Aerts, R., Bruffaerts, N., Somers, B., Demoury, C., Plusquin, M., Nawrot, T. S., & Hendrickx, M. (2021). Tree pollen allergy risks and changes across scenarios in urban green spaces in Brussels, Belgium. *Landscape and Urban Planning, 207*, Article 104001. https://doi.org/10.1016/j.landurbplan.2020.104001

Afdeling Bos & Groen. (2001). *De bosinventarisatie van het Vlaamse Gewest. Resultaten van de eerste inventarisatie 1997-1999*. Retrieved from.

Alderweireld, M., Burnay, F., Pitchugin, M., & Lecomte, H. (2015). *Inventaire Forestier Wallon-Résultats 1994-2012*.

Araújo, M. B., & New, M. (2007). Ensemble forecasting of species distributions. *Trends in Ecology & Evolution, 22*(1), 42–47. https://doi.org/10.1016/j.tree.2006.09.010

Barras, A. G., Braunisch, V., & Arlettaz, R. (2021). Predictive models of distribution and abundance of a threatened mountain species show that impacts of climate change overrule those of land use change. *Biodiversity Research, 27*(6), 989–1004. https://doi.org/10.1111/ddi.13247

Battisti, L., Pille, L., Wachtel, T., Larcher, F., & Säumel, I. (2019). Residential greenery: state of the art and health-related ecosystem services and disservices in the City of Berlin. *Sustainability, 11*(6), 1815.

Baumgardner, D., Varela, S., Escobedo, F. J., Chacalo, A., & Ochoa, C. (2012). The role of a peri-urban forest on air quality improvement in the Mexico City megalopolis. *Environmental Pollution, 163*, 174–183. https://doi.org/10.1016/j.envpol.2011.12.016

Bird, T. J., Bates, A. E., Lefcheck, J. S., Hill, N. A., Thomson, R. J., Edgar, G. J., … Frusher, S. (2014). Statistical solutions for error and bias in global citizen science datasets. *Biological Conservation, 173*, 144–154. https://doi.org/10.1016/j.biocon.2013.07.037

Boria, R. A., Olson, L. E., Goodman, S. M., & Anderson, R. P. (2014). Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling, 275*, 73–77. https://doi.org/10.1016/j.ecolmodel.2013.12.012

Bosch-Cano, F., Bernard, N., Sudre, B., Gillet, F., Thibaudon, M., Richard, H., … Ruffaldi, P. (2011). Human exposure to allergenic pollens: A comparison between urban and rural areas. *Environmental Research, 111*(5), 619–625. https://doi.org/10.1016/j.envres.2011.04.001

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Cariñanos, P., Grilo, F., Pinho, P., Casares-Porcel, M., Branquinho, C., Acil, N., … Vilhar, U. (2019). Estimation of the allergenic potential of urban trees and urban parks: Towards the healthy design of urban green spaces of the future. *International Journal of Environmental Research and Public Health, 16*(8), 1357.

Česnakaitė, I., Dėdelė, A., & Miškinytė, A. (2019). *Distribution of allergenic tree species in Kaunas.* Paper presented at the Smart Bio: ICSB 3rd international conference, 02-04 May 2019, Kaunas, Lithuania: abstract book/Vytautas Magnus university. Panevėžys: UAB" Reklamos forma", 2019, Nr. 3.

Collins, S. L., Glenn, S. M., & Roberts, D. W. (1993). The hierarchical continuum concept. *Journal of Vegetation Science, 4*(2), 149–156. https://doi.org/10.2307/3236099

Damialis, A., Traidl-Hoffmann, C., & Treudler, R. (2019). Climate change and pollen allergies. In R. M. Marselle, J. Stadler, H. Korn, N. K. Irvine, & A. Bonn (Eds.), *Biodiversity and health in the face of climate change* (pp. 47–66). Cham, Switzerland: Springer.

Dark, S. J., & Bram, D. (2007). The modifiable areal unit problem (MAUP) in physical geography. *Progress in Physical Geography: Earth and Environment, 31*(5), 471–479. https://doi.org/10.1177/0309133307083294

Delcloo, A., Verstraeten, W. W., Dujardin, S., Bruffaerts, N., Hendrickx, M., Hamdi, R., & Sofiev, M. (2018). Spatio-Temporal Monitoring and Modelling of Birch Pollen in Belgium. *Paper presented at the International Technical Meeting on Air Pollution Modelling and its Application, Cham*.

Dendoncker, N., Schmit, C., & Rounsevell, M. (2008). Exploring spatial data uncertainties in land-use change scenarios. *International Journal of Geographical Information Science, 22*(9), 1013–1030. https://doi.org/10.1080/13658810701812836

Dobson, A. D. M., Milner-Gulland, E. J., Aebischer, N. J., Beale, C. M., Brozovic, R., Coals, P., … Keane, A. (2020). Making messy data work for conservation. *One Earth, 2*(5), 455–465. https://doi.org/10.1016/j.oneear.2020.04.012

Drieskens S., Braekman E., Charafeddine R., Demarest S., Berete F., Gisle L., & J., V. d. H. (2018). *Health Interview Survey, Belgium, 1997 - 2001 - 2004 - 2008 - 2013 - 2018: Health Interview Survey Interactive Analysis*. from Sciensano https://hisia.wiv-isp.be/.

Dujardin, S., Linard, C., & Dendoncker, N. (2017). *Research note on mapping allergenic tree species distributions in Belgium*. Available from: https://researchportal.unamur.be/en/publications/research-note-on-mapping-future-abundance-distributions-of-allerg.

Dujardin, S., Linard, C., & Dendoncker, N. (2020). *Research note on mapping future abundance distributions of allergenic tree species for risk allergy prediction*. Available from: https://researchportal.unamur.be/en/publications/research-note-on-mapping-allergenic-tree-species-distributions-in.

Eisenman, T. S., Jariwala, S. P., & Lovasi, G. S. (2019). Urban trees and asthma: A call for epidemiological research. *The Lancet Respiratory Medicine, 7*(7), e19–e20. https://doi.org/10.1016/S2213-2600(19)30193-6

Elith, J., & Leathwick, J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics, 40*, 677–697.

ESRI. (2014). *ArcGIS Desktop (Version Version 10.0)*. Redlands, CA: ESRI.

Feldman, M. J., Imbeau, L., Marchand, P., Mazerolle, M. J., Darveau, M., & Fenton, N. J. (2021). Trends and gaps in the use of citizen science derived data as input for species distribution models: A quantitative review. *PLoS ONE, 16*(3), Article e0234587. https://doi.org/10.1371/journal.pone.0234587

Gareth, J., Daniela, W., Trevor, H., & Robert, T. (2013). *An introduction to statistical learning: with applications in R*. Spinger.

Giraud, C., Calenge, C., Coron, C., & Julliard, R. (2016). Capitalizing on opportunistic data for monitoring relative abundances of species. *Biometrics, 72*(2), 649–658. https://doi.org/10.1111/biom.12431

Grewling, Ł., Bogawski, P., Kryza, M., Magyar, D., Šikoparija, B., Skjøth, C. A., … Smith, M. (2019). Concomitant occurrence of anthropogenic air pollutants, mineral dust and fungal spores during long-distance transport of ragweed pollen. *Environmental Pollution, 254*, Article 112948. https://doi.org/10.1016/j.envpol.2019.07.116

Grote, R., Samson, R., Alonso, R., Amorim, J. H., Cariñanos, P., Churkina, G., … Calfapietra, C. (2016). Functional traits of urban trees: air pollution mitigation potential. *Frontiers in Ecology and the Environment, 14*(10), 543–550. https://doi.org/10.1002/fee.1426

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters, 8*(9), 993–1009. https://doi.org/10.1111/j.1461-0248.2005.00792.x

Guisan, A., Zimmermann, N. E., Elith, J., Graham, C. H., Phillips, S., & Peterson, A. T. (2007). What matters for predicting the occurences of trees: Techniques, data, or species' characteristics? *Ecological Monographs, 77*(4), 615–630.

Hamdi, R., Kusaka, H., Doan, Q.-V., Cai, P., He, H., Luo, G., … Termonia, P. (2020). The State-of-the-Art of Urban Climate Change Modeling and Observations. *Earth Systems and Environment, 4*(4), 631–646. https://doi.org/10.1007/s41748-020-00193-3

Hanberry, B. B. (2013). Finer grain size increases effects of error and changes influence of environmental predictors on species distribution models. *Ecological Informatics, 15*, 8–13. https://doi.org/10.1016/j.ecoinf.2013.02.003

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. New York, USA: Springer Science & Business Media.

Heikkinen, R. K., Marmion, M., & Luoto, M. (2012). Does the interpolation accuracy of species distribution models come at the expense of transferability? *Ecography, 35*(3), 276–288.

Hill, L., Hector, A., Hemery, G., Smart, S., Tanadini, M., & Brown, N. (2017). Abundance distributions for tree species in Great Britain: A two-stage approach to modeling abundance using species distribution modeling and random forest. *Ecology and Evolution, 7*(4), 1043–1056.

Hopkins, J. J., & Kirby, K. J. (2007). Ecological change in British broadleaved woodland since 1947. *Ibis, 149*, 29–40.

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

König, C., Wüest, R. O., Graham, C. H., Karger, D. N., Sattler, T., Zimmermann, N. E., & Zurell, D. (2021). Scale dependency of joint species distribution models challenges interpretation of biotic interactions. *Journal of Biogeography*. https://doi.org/10.1111/jbi.14106

Kosmala, M., Wiggins, A., Swanson, A., & Simmons, B. (2016). Assessing data quality in citizen science. *Frontiers in Ecology and the Environment, 14*(10), 551–560. https://doi.org/10.1002/fee.1436

Kuhn, M., Wind, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., … Benesty, M. (2016). *Caret: Classification and regression training*. package ver. 6.0–68. In.

Lee, H., Mayer, H., & Chen, L. (2016). Contribution of trees and grasslands to the mitigation of human heat stress in a residential district of Freiburg, Southwest Germany. *Landscape and Urban Planning, 148*, 37–50. https://doi.org/10.1016/j.landurbplan.2015.12.004

Liaw, A., & Wiener, M. (2002). Classification and regression by random. *Forest., 2*(3), 18–22.

Majkowska-Wojciechowska, B., Pełka, J., Korzon, L., Kozłowska, A., Kaczała, M., Jarzębska, M., … Kowalski, M. L. (2007). Prevalence of allergy, patterns of allergic sensitization and allergy risk factors in rural and urban children. *Allergy, 62*(9), 1044–1050. https://doi.org/10.1111/j.1398-9995.2007.01457.x

Mansouritorghabeh, H., Jabbari-Azad, F., Sankian, M., Varasteh, A., & Farid-Hosseini, R. (2019). The most common allergenic tree pollen grains in the middle east: A narrative review. *Iranian journal of medical sciences, 44*(2), 87–98.

Maya-Manzano, J. M., Skjøth, C. A., Smith, M., Dowding, P., Sarda-Estève, R., Baisnée, D., … O'Connor, D. J. (2021). Spatial and temporal variations in the distribution of birch trees and airborne Betula pollen in Ireland. *Agricultural and Forest Meteorology, 298–299*, Article 108298. https://doi.org/10.1016/j.agrformet.2020.108298

Maya-Manzano, J. M., Smith, M., Markey, E., Hourihane Clancy, J., Sodeau, J., & ÓConnor, D. J. (2021). Recent developments in monitoring and modelling airborne pollen, a review. *Grana, 60*(1), 1–19. https://doi.org/10.1080/00173134.2020.1769176

McDonnell, M. J., & Pickett, S. T. A. (1990). Ecosystem structure and function along urban-rural gradients: An unexploited opportunity for ecology. *Ecology, 71*(4), 1232–1237. https://doi.org/10.2307/1938259

McDonnell, M. J., Pickett, S. T. A., Groffman, P., Bohlen, P., Pouyat, R. V., Zipperer, W. C., … Medley, K. (2008). Ecosystem processes along an urban-to-rural gradient. In J. M. Marzluff, E. Shulenberger, M. Alberti, G. Bradley, C. Ryan, U. Simon, … C. ZumBrunnen (Eds.), *Urban Ecology: An International Perspective on the Interaction Between Humans and Nature* (pp. 299–313). Boston, MA: Springer, US.

McInnes, R. N., Hemming, D., Burgess, P., Lyndsay, D., Osborne, N. J., Skjøth, C. A., … Vardoulakis, S. (2017). Mapping allergenic pollen vegetation in UK to study environmental exposure and human health. *Science of the Total Environment, 599*, 483–499.

Miller, D. A. W., Pacifici, K., Sanderlin, J. S., & Reich, B. J. (2019). The recent past and promising future for data integration methods to estimate species' distributions. *Advances in Modelling Demographic Processes, 10*(1), 22–37. https://doi.org/10.1111/2041-210X.13110

Mimić, G., Podraščanin, Z., Lugonja, P., & Šikoparija, B. (2021). The influence of source maps on SILAM performance in modeling ragweed pollen concentrations in the area of a major European source. *International Journal of Biometeorology*. https://doi.org/10.1007/s00484-021-02075-3

Nowak, D. J., Crane, D. E., & Stevens, J. C. (2006). Air pollution removal by urban trees and shrubs in the United States. *Urban Forestry & Urban Greening, 4*(3), 115–123. https://doi.org/10.1016/j.ufug.2006.01.007

Nowak, D. J., Hirabayashi, S., Bodine, A., & Greenfield, E. (2014). Tree and forest effects on air quality and human health in the United States. *Environmental Pollution, 193*, 119–129. https://doi.org/10.1016/j.envpol.2014.05.028

Nowak, D. J., Hirabayashi, S., Doyle, M., McGovern, M., & Pasher, J. (2018). Air pollution removal by urban forests in Canada and its effect on air quality and human health. *Urban Forestry & Urban Greening, 29*, 40–48. https://doi.org/10.1016/j.ufug.2017.10.019

Oke, T. R. (1982). The energetic basis of the urban heat island. *Quarterly Journal of the Royal Meteorological Society, 108*(455), 1–24.

Ottosen, T.-B., Petch, G., Hanson, M., & Skjøth, C. A. (2020). Tree cover mapping based on Sentinel-2 images demonstrate high thematic accuracy in Europe. *International Journal of Applied Earth Observation and Geoinformation, 84*, Article 101947. https://doi.org/10.1016/j.jag.2019.101947

Pacifici, K., Reich, B. J., Miller, D. A. W., Gardner, B., Stauffer, G., Singh, S., … Collazo, J. A. (2017). Integrating multiple data sources in species distribution modeling: a framework for data fusion*. *Ecology, 98*(3), 840–850. https://doi.org/10.1002/ecy.1710

Prasad, A. M., Iverson, L. R., & Liaw, A. (2006). Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems, 9*(2), 181–199. https://doi.org/10.1007/s10021-005-0054-1

Prentice, I. C., Cramer, W., Harrison, S. P., Leemans, R., Monserud, R. A., & Solomon, A. M. (1992). Special paper: A global biome model based on plant physiology and dominance, soil properties and climate. *Journal of biogeography, 19*(2), 117–134. https://doi.org/10.2307/2845499

Priftis, K. N., Anthracopoulos, M. B., Nikolaou-Papanagiotou, A., Mantziou, V., Paliatsos, A. G., Tzavelas, G., … Mantzouranis, E. (2007). Increased sensitization in urban vs. rural environment – Rural protection or an urban living effect? *Pediatric Allergy and Immunology, 18*(3), 209–216. https://doi.org/10.1111/j.1399-3038.2006.00514.x

R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Core Team.

Rackham, O. (2008). Ancient woodlands: Modern threats. *New Phytologist, 180*(3), 571–586.

Robinson, O. J., Ruiz-Gutierrez, V., Reynolds, M. D., Golet, G. H., Strimas-Mackey, M., & Fink, D. (2019). *Integrating citizen science data with expert surveys increases accuracy and spatial extent of species distribution models*. 806547. doi:10.1101/806547.

Rondeux, J., & Lecomte, H. (2010). *Inventaire Permanent des Ressources Forestières de Wallonie-Guide méthodologique.*.

Shin, J.-Y., Han, M. J., Cho, C., Kim, K. R., Ha, J.-C., & Oh, J.-W. (2020). Allergenic Pollen calendar in Korea based on probability distribution models and up-to-date observations. *Allergy, Asthma & Immunology Research, 12*(2), 259–273.

Soroye, P., Ahmed, N., & Kerr, J. T. (2018). Opportunistic citizen science data transform understanding of species distributions, phenology, and diversity gradients for global change research. *Global Change Biology, 24*(11), 5281–5291. https://doi.org/10.1111/gcb.14358

Stas, M., Aerts, R., Hendrickx, M., Delcloo, A., Dendoncker, N., Dujardin, S., … Somers, B. (2021). Exposure to green space and pollen allergy symptom severity: A

case-crossover study in Belgium. *Science of the Total Environment, 781*, Article 146682. https://doi.org/10.1016/j.scitotenv.2021.146682

Stas, M., Aerts, R., Hendrickx, M., Dendoncker, N., Dujardin, S., Linard, C., & Somers, B. (2021). Residential green space types, allergy symptoms and mental health in a cohort of tree pollen allergy patients. *Landscape and Urban Planning, 210*, Article 104070. https://doi.org/10.1016/j.landurbplan.2021.104070

Steen, V. A., Elphick, C. S., & Tingley, M. W. (2019). An evaluation of stringent filtering to improve species distribution models from citizen science data. *Biodiversity Research, 25*(12), 1857–1869. https://doi.org/10.1111/ddi.12985

Swets, J. (1988). Measuring the accuracy of diagnostic systems. *Science, 240*(4857), 1285–1293. https://doi.org/10.1126/science.3287615 %J Science

Teshnehdel, S., Akbari, H., Di Giuseppe, E., & Brown, R. D. (2020). Effect of tree cover and tree species on microclimate and pedestrian comfort in a residential district in Iran. *Building and Environment, 178*, Article 106899. https://doi.org/10.1016/j.buildenv.2020.106899

Thuiller, W. (2003). BIOMOD–optimizing predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology, 9*(10), 1353–1362.

Timm, S., Frydenberg, M., Janson, C., Campbell, B., Forsberg, B., Gislason, T., … Schlünssen, V. (2016). The Urban-Rural Gradient In Asthma: A Population-Based Study in Northern Europe. *International Journal of Environmental Research and Public Health, 13*(1), 93.

Van Eetvelde, V., & Antrop, M. (2009). A stepwise multi-scaled landscape typology and characterisation for trans-regional integration, applied on the federal state of Belgium. *Landscape and Urban Planning, 91*(3), 160–170. https://doi.org/10.1016/j.landurbplan.2008.12.008

Van Eupen, C., Maes, D., Herremans, M., Swinnen, K. R. R., Somers, B., & Luca, S. (2021). The impact of data quality filtering of opportunistic citizen science data on species distribution model performance. *Ecological Modelling, 444*, Article 109453. https://doi.org/10.1016/j.ecolmodel.2021.109453

Verbeiren, B., Huysmans, M., Tychon, B., Jacquemin, I., Canters, F., Vanderhaegen, S., … Tsakiris, G. (2013). *Drought-related vulnerability and risk assessment of groundwater resources under temperate conditions*.

Verstraeten, W. W., Dujardin, S., Hoebeke, L., Bruffaerts, N., Kouznetsov, R., Dendoncker, N., & Delcloo, A. W. (2019). Spatio-temporal monitoring and modelling of birch pollen levels in Belgium. *Aerobiologia, 35*(4), 703–717. https://doi.org/10.1007/s10453-019-09607-w

Verstraeten, W. W., Kouznetsov, R., Hoebeke, L., Bruffaerts, N., Sofiev, M., & Delcloo, A. W. (2021). Modelling grass pollen levels in Belgium. *Science of the Total Environment, 753*, Article 141903. https://doi.org/10.1016/j.scitotenv.2020.141903

Wisz, M. S., Pottier, J., Kissling, W. D., Pellissier, L., Lenoir, J., Damgaard, C. F., … Guisan, A. (2013). The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biological Reviews, 88*(1), 15–30.

Zong, H., Yao, M., Tang, Y., & Chen, H. (2020). Assessing the composition, diversity, and allergenic risk of street trees in Qingyang District of Chengdu City. *Urban Forestry & Urban Greening, 54*, Article 126747. https://doi.org/10.1016/j.ufug.2020.126747