# Incorporation of pollen data in source maps is vital for pollen dispersion models

Alexander Kurganskiy[1,2,3], Carsten Ambelas Skjøth[3], Alexander Baklanov[4], Mikhail Sofiev[5], Annika Saarto[6], Elena Severova[7], Sergej Smyshlyaev[2], and Eigil Kaas[1]

[1]Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark
[2]Russian State Hydrometeorological University, St. Petersburg, Russia
[3]School of Science and the Environment, University of Worcester, Worcester, United Kingdom
[4]World Meteorological Organization, Geneva, Switzerland
[5]Finnish Meteorological Institute, Helsinki, Finland
[6]Biodiversity Unit of the University of Turku, Turku, Finland
[7]Lomonosov Moscow State University, Moscow, Russia

**Correspondence:** Alexander Kurganskiy (a.kurganskiy@worc.ac.uk)

**Abstract.** Information about distribution of pollen sources, i.e. their presence and abundance in a specific region, is important especially when atmospheric transport models are applied to forecast pollen concentrations. The goal of this study is to evaluate three pollen source maps using an atmospheric transport model and study the effect on the model results by combining these source maps with pollen data. Here we evaluate three maps for the birch taxon: (1) a map derived by combining land cover data and forest inventory; (2) a map obtained from land cover data and calibrated using model simulations and pollen observations; (3) a statistical map resulting from analysis of forest inventory and forest plot data. The maps were introduced to the Enviro-HIRLAM (Environment – High Resolution Limited Area Model) as input data to simulate birch pollen concentrations over Europe for the birch pollen season 2006. 18 model runs were performed using each of the selected maps in turn with and without calibration with observed pollen data from 2006. The model results were compared with the pollen observation data at 12 measurement sites located in Finland, Denmark and Russia. We show that calibration of the maps using pollen observations significantly improved the model performance for all three maps. The findings also indicate the large sensitivity of the model results to the source maps and agree well with other studies on birch showing that pollen or hybrid-based source maps provide the best model performance. This study highlights the importance of including pollen data in the production of source maps for pollen dispersion modelling and for exposure studies.

*Copyright statement.*

## 1 Introduction

Aeroallergens are a specific type of atmospheric aerosols causing allergic reactions among people suffering from allergic rhinitis and it is often connected with asthma (Bachert et al., 2004). The amount of allergic patients sensitive to pollen is assessed

1

to be 20 % of the European population (WHO, 2003). There are substantial variations in sensitization levels towards specific aeroallergens in Europe (Heinzerling et al., 2009). In Northern Europe, pollen from Poaceae (grasses) and Betulaceae (e.g. hazel, alder and birch) families show the highest sensitization among patients (Heinzerling et al., 2009). Birch pollen is generally the most abundant pollen type of those four pollen types (Skjøth et al., 2013b), often with large interannual fluctuations in the overall pollen integral (e.g., Piotrowska and Kaszewski, 2011). In particular, for Northern Europe the pollen season of 2006 was characterized by abundant birch pollen concentrations that resulted in an increase of patient calls for medical assistance during the spring period (Sofiev et al., 2011). Such episodes can potentially be described and forecast using atmospheric dispersion models (Sofiev et al., 2006) helping sensitized individuals to minimize their symptoms. Several European countries, such as Denmark, have seen an increase in sensitized individuals over the past few decades (e.g., Rasmussen, 2002; Linneberg et al., 2007). Generally, skin prick testing (SPT) is performed to identify sensitization of patients to common pollen allergens (e.g., Heinzerling et al., 2009). Haahtela et al. (2015) showed that the percentage of the population with positive SPT was about 35 % and 20 % in Finish and Russian Karelia, respectively. These facts demonstrate the importance of pollen studies using atmospheric dispersion models applied for research and operational purposes (Klein et al., 2012).

A variety of pollen types are considered as biological components by several atmospheric dispersion models such as SILAM, COSMO-ART, WRF/CMAQ, Enviro-HIRLAM and others (see e.g., Zink et al., 2012, 2013; Zhang et al., 2014; Sofiev et al., 2015a; Baklanov et al., 2017; Sofiev, 2017; Zink et al., 2017). Information about distribution of pollen sources, i.e. their presence and abundance in a region of study, plays a crucial role in aerobiological modelling and forecasting using atmospheric dispersion models (e.g., Zink et al., 2017). In other words, it is one of the key input data used to simulate or forecast pollen emission, its atmospheric transport and depositions in these models. For this reason, recent studies (e.g., Pauling et al., 2012; Bonini et al., 2018) are devoted to developing pollen source inventory/maps. Two main approaches aiming to build/construct the pollen source inventories are suggested in the literature: bottom-up and top-down (Skjøth et al., 2013b). The bottom-up approach is based on statistical data of pollen source distribution combined with land cover data obtained from remote sensing (Skjøth et al., 2008a). The limitation of this approach is that statistical data are either unavailable or do cover too large territory (e.g. the whole country) in some regions. The top-down approach uses pollen observations as starting point to estimate the abundance of different pollen taxa and it can be combined with land use data and/or models for more detailed assessment of pollen sources on a regional scale. Poor geographical and temporal coverage of the pollen observation data can limit this approach in some regions.

Skjøth et al. (2008a) used the bottom-up approach to create the Tree Species Inventory (TSI) for 39 species (including birch, oak, alder etc.) redistributed to a 50 km grid to be used in atmospheric dispersion models. The top-down approach has been used to produce ragweed pollen inventories for the Pannonian Basin (Skjøth et al., 2010), France (Thibaudon et al., 2014), Austria (Karrer et al., 2015) and Italy (Bonini et al., 2018). The inventories are based on combination of pollen observation data, knowledge on ragweed ecology and detailed land cover information. The latter one (i.e., Bonini et al., 2018) also includes influence of the Ophraella communa beetle on the plants. Prank et al. (2013) as well as Hamaoui-Laguel et al. (2015) have both used combined occupancy and climatic habitat quality maps for ragweed from an ecological model. The map covering Europe has been introduced to the atmospheric models as input data, run for several years and calibrated using pollen obser-

vation data. A birch pollen source map has been derived by Pauling et al. (2012). The methodology includes using a forest inventory combined with local and global land use datasets. The obtained map has been calibrated by observational pollen data and it covers Europe. An urban-scale grass pollen source inventory has been obtained by Skjøth et al. (2013a) using GIS and remote sensing data. A high resolution (1 km) pollen source inventory based on 12 taxa of trees, grass and weeds has been

5    produced by McInnes et al. (2017) for the UK to be mainly used in exposure studies as well as other possible applications. Many of these different pollen source maps show substantial variation, exemplified by the birch maps over the UK produced by McInnes et al. (2017) and Pauling et al. (2012) or the European scale by comparing the maps by Pauling et al. (2012) with Siljamo et al. (2013). This suggests substantial disagreement or uncertainty in the production of the pollen source maps – the key input data for pollen dispersion models. An additional problem is the large variability of the pollen production per plant in

10   each specific year. Several approaches have been proposed to address this issue (e.g., Masaka and Maguchi, 2001; Ranta et al., 2005; Ritenberga et al., 2018).

Zink et al. (2017) applied an atmospheric dispersion model to compare different ragweed pollen source maps using the COSMO-ART model. The results show that the pollen source map resulting from combining land cover and pollen observation data provides the best model performance. The advantage of including a dispersion model is that, in principle, such

15   models take into account atmospheric transport and its effect on pollen concentrations during a specific year (e.g., Prank et al., 2013). A recent study showed high potential of extended 4-dimensional variational data assimilation as a rigorous procedure for obtaining the source maps calibrations (Sofiev, 2019).

The literature review presented in this section ~~show~~ shows there are multiple pollen source maps produced for different pollen types, regions and using different approaches and data. Despite this, they are rarely applied to, and intercompared using at-

20   mospheric dispersion models. This study aims to evaluate three pollen source maps for the birch taxon using an atmospheric dispersion model and study the effect on the model results by combining these source maps with pollen data. The maps and model are described in the methods section below.


## 2    Methods

### 2.1    Birch pollen source maps

25   Three birch pollen source maps have been chosen in this study: (1) a map derived by combining land cover data and forest inventory; (2) a map obtained from land cover data and calibrated using model simulations and pollen observations; (3) a statistical map resulting from analysis of forest inventory and forest plot data ~~.~~ (see Fig. A1 in the appendix section). The maps have been pre-processed for two modelling domains: P15 and T15. P15 covers part of Spain in the south to part of Scandinavia in the North, with Denmark near the center of the model domain (Fig. 1a). T15 covers Finland, parts of Sweden, North-West

30   Russia, Baltic States, Belarus (Fig. 1b).

Map 1 (see Fig. 2a,b) has been obtained by combining three datasets: Global Land Cover Characteristics (GLCC, edc2.usgs.gov/glcc) version 2 (Belward et al., 1999), European Forest Institute tree cover (EFI, Päivinen et al. (2001)) and Tree Species Inventory (TSI, Skjøth et al. (2008a)). GLCC has global coverage and includes geographical distribution of land cover, soil type and

other various properties with 1 km horizontal resolution. EFI covers Europe, has 1 km horizontal resolution and it contains spatial distribution of forest with sub-classes, i.e. broadleaved forest and coniferous forest. TSI is presented by 39 types of tree species (birch, oak, alder, etc.) and covers Europe with 50 km horizontal resolution and it is based on national forest inventories and national statistics. The following layers have been extracted from these databases and combined to derive the birch pollen
5   source: land-mask, forest, vegetation, urban territories (from GLCC), broadleaved forest (from EFI) and birch forest fraction in broadleaved forest (from TSI). For details of the combining procedure see (Kurganskiy et al., 2015; Kurganskiy, 2017).

Map 2 (see Fig. 2c,d) is a product of two datasets and a calibration using model simulations. The two datasets are EFI tree cover and ECOCLIMAP (Masson et al., 2003). ECOCLIMAP is a global dataset with 1 km resolution and it includes 215 ecosystems. The model simulations have been performed with the System for Integrated modeLing of Atmospheric coMposi-
10  tion (SILAM, http://silam.fmi.fi, Sofiev et al. (2015b)) and pollen observation data for several years to ensure average unbiased representation of the Seasonal Pollen Integral (SPIn) with the SILAM model, where SPIn is the integral of pollen concentrations over time (Galán et al., 2017). The calibrated map is calculated for a 1 km grid covering the whole Europe with nearby regions. Details of the map calibration procedure is described by Prank et al. (2013) using ragweed and its application in the SILAM model on birch can be found in Sofiev (2017).

15  Map 3 (see Fig. 2e,f) is the result of combining ICP-Level-I forest plot data (www.icp-forests.org) and National Forest Inventory (NFI) plot data as well as the NFI statistics (Brus et al., 2012). Two techniques were used to derive the map: compositional kriging in areas with NFI plot data and regression modelling outside of the areas. The National Forest Inventory statistics ~~was~~ were used to scale the regression model results. The map has 1 km horizontal resolution and covers the most part of Europe. However, information about birch pollen sources in Russia is missing in Map 3. Therefore, the Russian sources were extracted
20  from Map 1 and combined with Map 3 to fill the gap.

The maps were introduced to the Enviro-HIRLAM (Environment – High Resolution Limited Area Model) as input data to simulate birch pollen concentrations for the modelling domains.

## 2.2   Model description

Enviro-HIRLAM is an online-coupled meteorology-chemistry model allowing to take into account spatial-temporal evolution
25  of atmospheric chemical (Korsholm, 2009; Baklanov et al., 2017) and biological (Kurganskiy, 2017) aerosols simulated inside the meteorological HIRLAM model (Undén et al., 2002). The current version of the Enviro-HIRLAM model is based on HIRLAM v7.2 (https://hirlam.org). The detailed model description of all components can be found in (Korsholm, 2009; Kurganskiy, 2017; Baklanov et al., 2017) while this article is focused on description of the Enviro-HIRLAM features relevant to pollen. In order to simulate birch pollen concentrations Enviro-HIRLAM requires external information/data on the
30  spatial pollen source distribution, phenology, i.e. start of birch flowering, pollen production, characteristics of pollen release (emissions) within the given study region, where the spatial pollen source distribution is represented by the maps described above in section 2.1. Atmospheric transport and dispersion as well as dry and wet depositions are calculated within Enviro-HIRLAM at each time step. In numerical models, the start of birch pollen flowering is often determined by applying simple crop growth models (e.g. McMaster and Wilhelm (1997)), here the Growing Degree-Day (GDD) method is implemented in

**4**

Enviro-HIRLAM. GDD is based on accumulation of 2-meter daily mean air temperatures above a given threshold (cut-off temperature) as this is the WMO definition of GDD. It is assumed here that birch flowering starts as soon as the accumulated temperature (GDD) reaches a certain value depending on geographical location. These temperature sum threshold maps have been produced by (Sofiev et al., 2013) Sofiev et al. (2013) and have been implemented and used as input data for the GDD pa-

5  rameterization in Enviro-HIRLAM. The data were obtained by the fitting procedure applied using the observed leaf unfolding dates (Siljamo et al., 2008) and modelled ones by the GDD parameterization and ECMWF's ERA-40 reanalysis air temperature data (Uppala et al., 2005) utilized as input for GDD. It was shown in Kurganskiy (2017) that the temperature sum threshold maps are applicable for use to simulate of the birch flowering start in Enviro-HIRLAM, even though another model (ERA-40) has been originally used to obtain the maps. GDD is an integral part of Enviro-HIRLAM and GDD calculations are performed

10  for the same dates as the model is run over. Birch pollen emission is meteorology-dependent and uses parameterizations proposed by (Sofiev et al., 2013). It is based on dimensionless functions correcting the seasonal pollen productivity, here chosen to be $3.7 \times 10^8$ pollen $m^{-2}$ season$^{-1}$ ) as a default value obtained for SILAM (Sofiev et al., 2013). The main meteorological parameters affecting the emission are air temperature, relative humidity, wind speed and accumulated precipitation. The emitted birch pollen particles are subject to atmospheric transport which is parameterized in the same way as for aerosols using the Lo-

15  cally Mass Conserving Semi- Lagrangian (LMCSL) scheme (Kaas, 2008; Sørensen et al., 2013). The scheme provides proper mass conservation, shape preservation, and multi-tracer efficiency required from numerical point of view. Dry deposition of birch pollen particles is determined by gravitational settling (Seinfeld and Pandis, 2006). The gravitational settling parameterization is based on calculation of settling velocity according to Stokes law and taking into account density of birch pollen particles (800 $kg\ m^{-3}$) and an estimated size of 22 µm (e.g., Mäkelä, 1996) with near spherical shape (Sofiev et al., 2006).

20  Wet deposition distinguishes between in-cloud (Stier et al., 2005) and below-cloud (Baklanov and Sørensen, 2001) scavenging, i.e. removal of pollen particles from the atmosphere by precipitation. Both scavenging types are parameterized through the scavenging coefficients used for aerosol particles with r ≥ 10 µm.

## 2.3  Model configuration

The Enviro-HIRLAM model is used to calculate pollen concentrations in two different modelling domains shown in Fig. 1.

25  Both domains have about 15 km horizontal resolution and 40 hybrid vertical levels. Input meteorological initial and boundary conditions were taken from the ECMWF-IFS model (Persson, 2011) with 15 km resolution and 6-hour interval. The birch pollen emission module has the same settings as described in (Sofiev et al., 2013; Siljamo, 2013). The Enviro-HIRLAM model has been run for the birch pollen season 2006. from 1st March until 15 June, 2006 with the ECMWF input data covering the same period. The two model domains have been simulated with three types of calculations: 1) A standard calculation, 2) A

30  calculation that includes a correction for 2-meter air temperature ($T_{2m}$) bias between assimilated and simulated $T_{2m}$ (termed COR), 3) A calculation that includes both the correction for temperature and a grid based scaling factor for pollen emissions (termed SCF). These calculations were performed using each of the selected pollen source maps, hence 9 model calculations for each of the two domains. The model results from model simulations 2 and 3 are presented in the results section. The scaling factor correction for $T_{2m}$ bias is done in two steps. Firstly, the model output from the simulations of type 1 is used to

5

calculate the biases (differences between simulated and assimilated $T_{2m}$) for each assimilation window (i.e. 6 hours). Secondly, the calculated biases are introduced in the model simulations of type 2 (COR) where the simulated $T_{2m}$ is corrected at each model time step using the average bias between 2 nearest/closest assimilation windows. Further details of this procedure can be found in Kurganskiy (2017). The scaling factor for pollen emissions is based on the ratio between simulated and observed

5 Seasonal Pollen Integral (SPIn) (see next section) from the simulations of type ~~2.~~ 2 (COR). This provides 15 point values that have been interpolated to the entire model domain, thereby providing a grid based scaling factor for each of the tree source maps. The ~~grid~~ interpolation procedure takes into account the weighted distance between each observation point and grid cell with a constant radius of influence (1 km in this study). The procedure also ensures the scaling factors are equal to around 1 in the areas located far away from the observation points. This is especially visible in P15 domain, e.g. in France (Fig. 3

10 a,c,e). Further details of the interpolation procedure can be found in Kurganskiy (2017). The grid based scaling factor has then been extracted for each model domain, thus providing 6 grid based scaling factors (Fig. 3) applied for the SCF simulations for all maps. The principal difference between the runs is that the COR ~~ones~~ runs include a constant value for birch pollen productivity ($3.7 \times 10^8$ pollen m$^{-2}$ season$^{-1}$) ~~)~~ in each grid cell whereas the productivity is calibrated using the grid based scaling factor (obtained from the COR runs and shown in Fig. 3) in the SCF runs. The model results were compared with the

15 pollen observation data at 12 measurement sites located in Finland, Denmark and Russia (see Fig. 1) for all model runs.

## 2.4 Used observations and statistical evaluation of model results

Daily concentrations of birch pollen were used from 12 sites (Fig. 1a,b) and furthermore annual SPIn were obtained from 3 additional sites in Lithuania, published by Veriankaite (2010) in order to strengthen the calculation of the grid based scaling factor (section 2.3). The observational data were obtained using 7-day volumetric samplers (Hirst, 1952) and analysed ac-

20 cording to standard methods in aerobiology (Galán et al., 2014). Station names and coordinates are presented in the appendix section along with statistical summaries for each ~~stations~~ station (see Tables A1–A4), while overall results are presented in the results section. The statistical evaluation includes standard and threshold-based statistical approaches, e.g. previously used by (Siljamo et al., 2013; Sofiev et al., 2015a) and calculated in this paper for each map (1,2,3) and run (COR, SCF) types. The start and end of the birch pollen season are calculated using the retrospective method according to Nilsson and Persson (1981), here

25 defined as the dates when the total sums of birch pollen concentrations (SPIn) reach 5 % and 95 %, correspondingly. The criteria are chosen to provide filtration of LRT (Long Range Transport) episodes which have significant impact in Northern Europe, especially in the beginning of the birch pollen season (Ranta et al., 2006; Skjøth et al., 2007; Mahura et al., 2007; Veriankaite, 2010). The standard statistical approach includes calculations of correlation coefficient (R), coefficient of determination ($R^2$), mean bias (MB) and root mean squared error (RMSE) ~~.~~ and the evaluation primarily concerns correlation and RMSE. Two

30 additional metrics Normalized Mean Bias Factor (NMBF) and Normalized Mean Absolute Error Factor (NMAEF) are also calculated according to Yu et al. (2006) since RMSE is highly sensitive to outliers. The $R^2$ values are shown on the time series plots for each station, map and run type as well as on the global scatter plots. The standard statistical metrics have been calculated for the "main" pollen season (30 ~~April - 15~~ April–15 June, 2006) due to non-ergodicity (i.e. non-similarity in space and time) and non-stationarity of the birch pollen time series (Sofiev et al., 2015a). The threshold-based statistical approach

(Siljamo et al., 2013) comprises calculations of the model accuracy (MA), probability of detection (POD), false alarm ratio (FAR), probability of false detection (POFD) and odds ratio (OR) using the threshold value for birch pollen concentrations $C_{th}$ = 50 pollen m$^{-3}$. The threshold value $C_{th}$ is chosen since most of the pollen allergy sensitive population might start suffering from allergic reactions when daily mean birch pollen concentration, $C \geq C_{th}$ in the air (Jantunen et al., 2012). Additionally, the modelled and observed birch pollen concentrations have been sorted into classes in order to estimate the number of cases for low (1–10 pollen m$^{-3}$), moderate (10–100 pollen m$^{-3}$), high (100–1000 pollen m$^{-3}$), and very high (> 1000 pollen m$^{-3}$) birch pollen concentrations. The sorted data plotted with corresponding standard errors are shown in the result section. Hit rates (HR) are calculated for the classes and their results are shown in tables in the results section. The results of statistical significance test (chi-squared test) performed for the HR values are also discussed. For summary of the thresholds, classes and metrics used to calculate the threshold-based statistics the reader is referred to Tables A5–A6 in the appendix section.

## 3 Results

The simulated time series of birch pollen concentrations have been compared against the selected pollen observation sites by determining the start and end of the birch pollen season and applying the standard statistics as well as threshold-based statistics described in the methods section. The modelled and observed time series are also shown in Fig. 4–5 for COR and SCF runs, respectively. The start of the birch pollen season is simulated quite well for both runs and all maps. Station-wise the bias does not exceed 2–3 days at all stations and indicates 2 days too early flowering on average for all maps. Simulation of the last flowering day turns to be the most challenging for Enviro-HIRLAM. The bias shows too late flowering at almost all stations. On average, it varies from 15 days (Map 3) to 9 days (Map 2) according to the COR run. Rescaling of the pollen productivity (SCF run) reduced the bias by 2-3 2–3 days for each map. The largest biases for the last flowering day are found at five Finnish sites (Helsinki, Joutseno, Kangasala, Vaasa and Kuopio) and two Russian sites (Moscow and Smolensk) and they are related to overestimation of the observed birch pollen concentrations at the end of the season for both COR and SCF runs (see Fig. 4 and 5). It can be addressed to large variations in the year to year pollen production in that region which, in turn, introduce a large bias in the model simulations.

Generally, the time series of the modelled birch pollen concentrations ( Fig. 4 and 5) show good agreement with observations for all maps at most of the selected stations. COR run indicates that the model behaves similar with respect to describing variance in observations (see $R^2$ values in Fig. 4), except some stations (e.g. Moscow, Helsinki) where Map 2 provides higher $R^2$ values in comparison with Map 1 and Map 3. The SCF run did not affect the $R^2$ values significantly, but it reduced the mean bias (MB, NMBF) at the stations (see Tables A1–A4). The highest $R^2$ values are found at two Danish sites (Copenhagen and Viborg) where the model explains more than 70 % of variance found in observations for all considered maps and model runs. It should here be noted that the $R^2$ values decrease slightly at the Danish sites according to the SCF runs. The reason for this is that the simulated concentrations in the SFC runs are increased on almost every single day, which has the effect of reducing the underestimation found in the central season (e.g. Fig. 5a) and increasing LRT, most likely originating from sources found in Germany and Poland in the beginning of the season and from Scandinavia in the end of the season. For more detailed statistical

summaries at the stations the reader is referred to tables in the appendix section.

The global standard statistical metrics for all stations are presented in Table 1. According to the COR run the model performs slightly better with Map 2. It is supported by higher correlation coefficient (R = 0.59) and lower mean bias (MB = 69.08 pollen m$^{-3}$, NMBF = 0.09) in comparison with two other maps. One can assume that this result should be expected since pollen observations are already included to the methodology used to obtain Map 2. Rescaling of the pollen productivity (SCF run) provides improvement of the Enviro-HIRLAM model performance by increasing the R values (R = 0.72 for Map 1,2 and R = 0.68 for Map 3) and decreasing the MB, NMBF and NMAEF values. However, RMSE is large (i.e. > 1000 pollen m$^{-3}$) for all maps and both model runs. It can be explained by sensitivity of RMSE to large differences between modelled and observed values. The large bias for the last flowering day simulation contributes significantly to the RMSE values.

The global scatter plots (Fig. 6) also indicate improvements in the modelled-observed concentration agreement expressed in increase of the R$^2$ for the SCF model setup. As seen the coefficient of determination do not exceed 0.34 (Map 2) for all maps in COR run. SCF run provides R$^2$ values around 0.5 for all maps. This means that the model can explain about 50 % of variance found in observations after the rescaling using any of the maps.

The threshold-based statistical metrics calculated for each map and model run are shown in Table 2. The results indicate similar behavior of the model when comparing performance of the maps. The values of the statistical metrics are slightly different between the maps, but the difference does not exceed 7 %. The model accuracy (MA) – the fraction of correct simulations – is higher than 70 % for all model simulations. The POD values show that the Enviro-HIRLAM model simulates the high concentrations (> 50 pollen m$^{-3}$) correctly in more than 70 % of cases. The fraction of incorrect simulations for the high concentrations (FAR values) are less than 30 % for the model setups. The POFD values show that the fraction of the cases when concentrations are simulated as high, but observed as low (i.e. < 50 pollen m$^{-3}$) is more than 30 % for the COR runs and it decreases below 30 % after rescaling of the pollen productivity (SCF run). As expected the odds ratio (OD) values increases in SCF run for all maps. The modelled and observed birch pollen concentrations have been grouped in order to estimate the number of cases for low, moderate, high and very high birch pollen concentrations (see Fig. 7). As seen the Enviro-HIRLAM model simulates the low and very high pollen concentrations quite well with the difference less than ± 5 % between the modelled and observed number of cases for both COR (Fig. 7a) and SCF (Fig. 7b) runs and all maps. The largest differences between the modelled and observed number of cases are found for the moderate and high pollen concentrations, but the differences are within the error bars (see Fig. 7). Map 2 underestimates the number of cases for the moderate concentrations by 10 % whereas Map 3 overestimates the high ones by more than 10 % (COR run). SCF run slightly improves the results for the moderate and high number of cases.

The hit rate (HR ) HR calculations (see Tables 3–4) show that simulations of the moderate and high birch pollen concentrations is the most challenging for all model setups. SCF run (Table 4) indicates improvements of the model score for the moderate and high birch pollen concentrations. However, the hit rates still do not exceed 50 % for all maps. This feature is also visible on the scatter plots in Fig. 6. The HR values for the low pollen concentrations are quite high and exceed 90 % for COR and 80 % for SCF runs. The very high pollen concentrations are simulated slightly better by using Map 3 with HR = 70 %. However,

according to the chi square test the differences between the HR values for different maps are not statistically significant for both COR and SCF runs.

## 4   Discussion

Analysis of the start of birch pollen season showed good model performance for all maps with the bias not exceeding 2 days on average. Such error is acceptable by state-of-the-art atmospheric dispersion models (e.g., Sofiev et al., 2015a). The analysis did not reveal significant dependency of the start of the birch pollen season on the underlying pollen source map and/or pollen productivity. This could be explained by the fact that the start of the birch pollen season is calculated using meteorological parameters (i.e. temperate sum thresholds) employed in GDD parameterization in the model. However, long distance transport of birch pollen has often been observed in that region (e.g., Hjelmroos, 1992; Mahura et al., 2007) which can systematically affect the start of the pollen season (Skjøth et al., 2007). It could therefore be assumed that changing the source term in remote regions such as Poland could affect the start of the calculated pollen season in Denmark and Sweden. This phenomenon was not observed, which may be explained by the fact that the scaling factor for Poland was near to unity for all three maps, while the scaling factor for Copenhagen and the near surroundings was substantially higher. This observation therefore complements the suggestion from other studies that long distance transport of pollen is an episodic phenomenon (e.g., Smith et al., 2008; Fernández-Rodríguez et al., 2014) and that the overall contribution to locally observed birch pollen concentration is due to sources found within the region (Sofiev, 2017). Add to this, that urban areas have previously been identified as areas that may enhance pollen production of flowering plants (Ziska et al., 2003). This may be particular relevant for Copenhagen as a substantial fraction of the birch pollen observed is expected to originate from a source found in the city itself (Skjøth et al., 2008b). The last flowering day simulation was the most challenging for Enviro-HIRLAM. Map 2 showed the lowest bias estimated as 6 days (too late flowering end delay of end of flowering) on average for the stations according to the SCF runs. However, such bias is a known feature for state-of-the-art atmospheric models (e.g., Sofiev et al., 2015a). Introducing more pollen observation stations and performing more iterations for rescaling of pollen emissions (i.e. repeating SCF run several times) would strengthen the current approach and it could potentially improve the modelling of the last flowering day. Station-wise the Finish and Russian sites had the bias $\geq$ 2 weeks in all model runs, which . One of the possible reasons for this is that the large pollen emission area found in Russia, North of Moscow and Smolensk is overestimated for this particular year. The source is found in all three source maps and the overestimation is further supported by the fact that the model simulates large amounts of birch pollen concentration very late in the season at Moscow and Smolensk (Fig. 4g and Fig. 4j), most likely originating from more Northerly regions. This area will occasionally also affect the Finnish stations as it is well known that this region receives LRT from Russia (e.g., Sofiev et al., 2006). The study therefore complements a multi-model ensemble, that demonstrated generally good results for Europe but which also highlighted that large variations in the year to year pollen production in that region can introduce a large bias in the model simulations (Sofiev et al., 2015a). Large annual variations in the birch pollen integral is well known for a number of European countries (Latałowa et al., 2002; Dahl et al., 2013; Piotrowska and Kaszewski, 2011), as these annual fluctuations are generally not accounted for in the atmospheric

models (e.g., Ritenberga et al., 2018).

The standard statistical metrics indicated a slightly better performance of the Enviro-HIRLAM with Map 2 on average for the stations according to the COR run. However, even for this map, the model explained less than 35 % of variance in observations. It is therefore plausible to attribute the remaining part of the disagreement to small-scale patterns of the birch habitation, which are neither captured by existing calibration methods nor satisfactorily reproduced by the land-cover datasets. In this study, a scaling of the pollen productivity improved the model results significantly for all source terms and provided near to identical results for Map 1 and Map 2. ~~This suggest,~~ The small-scale pattern in birch habitation will be difficult to capture by the land cover data sets, which will have the effect that a small but diffuse emission source found in many locations is not present, causing the model to underestimate medium concentration levels. Contrary the large emission source found in parts of Russia will cause the model to overestimate periods, when long distance transport is present. This effect of over and underestimation will vary from year to year, depending on pollen productivity which is known to vary from year to year and in between regions (Ranta et al., 2005). This suggests, that accurate exposure calculations that use dispersion models preferably should use data fusion that combine a detailed inventory based source term (e.g., Skjøth et al., 2008a; Pekkarinen et al., 2009; Sofiev, 2017) along with observed pollen ~~concentration~~ concentrations covering the period of interest. Other possible sources of error (e.g., timing of pollen release, quality of meteorological parameters, simulation of atmospheric transport and removal processes) could also be taken into account. However, their effect is less important in comparison with the quality of emission (source) maps – the main uncertainty in air pollution and pollen dispersion modelling.

The threshold-based statistics calculated ~~relatively~~ relative to one threshold value ($C_{th}$ = 50 pollen m$^{-3}$) showed a good model agreement with observations independently of the map used. The threshold-based analysis was also done for low (1–10 pollen m$^{-3}$), moderate (10–100 pollen m$^{-3}$), high (100–1000 pollen m$^{-3}$) and very high (> 1000 pollen m$^{-3}$) birch pollen concentrations. According to the analysis simulations of the moderate and high birch pollen concentrations appeared to be the most challenging for Enviro-HIRLAM for all maps. The hit rates were less than 50 % for all model setups. It reveals a need in improvement of the rescaling of birch pollen productivity by introducing more observational points, in particular in areas sparse with data or regions with high emissions such as parts of Russia and/or performing more iterations. The differences of the HR values appeared to be not significant during intercomparsion of the maps. This study concerning birch pollen complements other studies on ragweed (Prank et al., 2013; Hamaoui-Laguel et al., 2015), that also demonstrated the need for recalibration of the source term. However, it has since been shown by Zink et al. (2017) that source terms combining pollen data from several years with detailed land cover data can outperform other approaches minimizing the need for local calibrations. This approach is also limited by the fact that only few regions have a sufficiently high number of pollen observations (Zink et al., 2017). Add to this that abrupt annual changes in the underlying pollen integral can be caused by changes in the pollen emission (Bonini et al., 2015, 2018).

The insufficient density of the observational network also became the methodological limitation of the current study: we have used the same data from the same stations and for the same years for both calibration and evaluation. However, we separated them frequency-wise: calibration uses annual or multi-annual average whereas the evaluation primarily concerned correlation and RMSE. It should be noted, that an independent evaluation that do not use pollen data is shown by analysis of the COR runs

in the study. Another approach could be to carry out a cross-validation procedure using the so-called leave one out procedure as this tests the sensitivity of individual data points. However, such sensitivity study requires 15 additional model simulations for each map and each modelling domain, hence 90 simulations. The increased computational cost makes this exercise unfeasible suggesting that other approaches needs to be developed.

5

## 5 Conclusions

The aim of the study was to evaluate three birch pollen source maps using the atmospheric dispersion model Enviro-HIRLAM. The evaluation has been performed for 12 pollen observation sites located in Denmark, Finland and Russia. The modelled and observed time series of birch pollen concentrations have been analyzed for the start and end of flowering as well as by
10 calculating the standard and threshold-based statistical metrics for two sets of the Enviro-HIRLAM model runs: without and with calibration using pollen observations.

The analysis did not reveal significant dependency of the start/end of the birch pollen season on the underlying pollen source map. The statistical analysis showed a good model agreement with the observed birch pollen concentrations in the region studied. The model can explain up to 50 % of variance found in observations after the calibration with all maps. It was shown
15 that the remaining part of the disagreement should be addressed to small-scale patterns of the birch habitation, which are neither captured by existing calibration methods nor satisfactorily reproduced by the land-cover datasets. The analysis also revealed that the insufficient density of the observational network was the methodological limitation of the study.

Generally, it was shown that calibration of the maps using pollen observations covering the investigation year significantly improved the model performance for all three maps. The findings also indicate the large sensitivity of the model results to
20 the source maps and agree well with other studies on birch showing that pollen or hybrid-based source maps provide the best model performance. This study highlights the importance of including pollen data in the production of source maps for pollen dispersion modelling and for exposure studies.

11

# References

Bachert, C., Vignola, A. M., Gevaert, P., Leynaert, B., Van Cauwenberge, P., and Bousquet, J.: Allergic rhinitis, rhinosinusitis, and asthma: one airway disease, Immunology and Allergy Clinics, 24, 19–43, 2004.

Baklanov, A. and Sørensen, J.: Parameterisation of radionuclide deposition in atmospheric long-range transport modelling, Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere, 26, 787 – 799, https://doi.org/http://dx.doi.org/10.1016/S1464-1909(01)00087-9, 2001.

Baklanov, A., Smith Korsholm, U., Nuterman, R., Mahura, A., Nielsen, K. P., Sass, B. H., Rasmussen, A., Zakey, A., Kaas, E., Kurganskiy, A., Sørensen, B., and González-Aparicio, I.: Enviro-HIRLAM online integrated meteorology–chemistry modelling system: strategy, methodology, developments and applications (v7.2), Geoscientific Model Development, 10, 2971–2999, https://doi.org/10.5194/gmd-10-2971-2017, 2017.

Belward, A., Estes, J., and Kline, K.: The IGBP-DIS 1-Km Land-Cover Data Set DISCover: A Project Overview, Photogrammetric Engineering and Remote Sensing, 65, 1013–1020, 1999.

Bonini, M., Šikoparija, B., Prentović, M., Cislaghi, G., Colombo, P., Testoni, C., Grewling, L., Lommen, S. T. E., Müller-Schärer, H., and Smith, M.: Is the recent decrease in airborne Ambrosia pollen in the Milan area due to the accidental introduction of the ragweed leaf beetle Ophraella communa?, Aerobiologia, 31, 499–513, 2015.

Bonini, M., Šikoparija, B., Skjøth, C. A., Cislaghi, G., Colombo, P., Testoni, C., A.I.A.-R.I.M.A.®, POLLnet, and Smith, M.: Ambrosia pollen source inventory for Italy: a multi-purpose tool to assess the impact of the ragweed leaf beetle (Ophraella communa LeSage) on populations of its host plant, International Journal of Biometeorology, 62, 597–608, https://doi.org/10.1007/s00484-017-1469-z, 2018.

Brus, D. J., Hengeveld, G. M., Walvoort, D. J. J., Goedhart, P. W., Heidema, A. H., Nabuurs, G. J., and Gunia, K.: Statistical mapping of tree species over Europe, European Journal of Forest Research, 131, 145–157, https://doi.org/10.1007/s10342-011-0513-5, 2012.

Dahl, Å., Galán, C., Hajkova, L., Pauling, A., Sikoparija, B., Smith, M., and Vokou, D.: The Onset, Course and Intensity of the Pollen Season, pp. 29–70, Springer Netherlands, Dordrecht, 2013.

Fernández-Rodríguez, S., Skjøth, C. A., Tormo-Molina, R., Brandao, R., Caeiro, E., Silva-Palacios, I., Gonzalo-Garijo, Á., and Smith, M.: Identification of potential sources of airborne Olea pollen in the Southwest Iberian Peninsula, International Journal of Biometeorology, 58, 337–348, 2014.

Galán, C., Smith, M., Thibaudon, M., Frenguelli, G., Oteros, J., Gehrig, R., Berger, U., Clot, B., Brandao, R., and EAS QC Working Group: Pollen monitoring: minimum requirements and reproducibility of analysis, Aerobiologia, 30, 385–395, https://doi.org/10.1007/s10453-014-9335-5, 2014.

Galán, C., Ariatti, A., Bonini, M., Clot, B., Crouzy, B., Dahl, A., Fernandez-González, D., Frenguelli, G., Gehrig, R., Isard, S., Levetin, E., Li, D. W., Mandrioli, P., Rogers, C. A., Thibaudon, M., Sauliene, I., Skjoth, C., Smith, M., and Sofiev, M.: Recommended terminology for aerobiological studies, Aerobiologia, 33, 293–295, https://doi.org/10.1007/s10453-017-9496-0, 2017.

Haahtela, T., Laatikainen, T., Alenius, H., Auvinen, P., Fyhrquist, N., Hanski, I., Hertzen, L., Jousilahti, P., Kosunen, T. U., Markelova, O., Mäkelä, M. J., Pantelejev, V., Uhanov, M., Zilber, E., and Vartiainen, E.: Hunt for the origin of allergy – comparing the Finnish and Russian Karelia, 45, 891–901, https://doi.org/10.1111/cea.12527, 2015.

Hamaoui-Laguel, L., Vautard, R., Liu, L., Solmon, F., Viovy, N., Khvorostyanov, D., Essl, F., Chuine, I., Colette, A., Semenov, M. A., et al.: Effects of climate change and seed dispersal on airborne ragweed pollen loads in Europe, Nature Climate Change, 5, 766, 2015.

Heinzerling, L., Burbach, G., Edenharter, G., Bachert, C., Bindslev-Jensen, C., Bonini, S., Bousquet, J., Bousquet-Rouanet, L., Bousquet, P., Bresciani, M., et al.: GA2LEN skin test study I: GA$^2$LEN harmonization of skin prick testing: novel sensitization patterns for inhalant allergens in Europe, Allergy, 64, 1498–1506, 2009.

Hirst, J. M.: An automatic volumetric spore trap, Annals of Applied Biology, 39, 257–265, https://doi.org/10.1111/j.1744-7348.1952.tb00904.x, 1952.

Hjelmroos, M.: Long-distance transport ofBetula pollen grains and allergic symptoms, Aerobiologia, 8, 231–236, 1992.

Jantunen, J., Saarinen, K., and Rantio-Lehtimäki, A.: Allergy symptoms in relation to alder and birch pollen concentrations in Finland, Aerobiologia, 28, 169–176, https://doi.org/10.1007/s10453-011-9221-3, 2012.

Kaas, E.: A simple and efficient locally mass conserving semi-Lagrangian transport scheme., Tellus, 60, 305–320, 2008.

Karrer, G., Skjøth, C., Šikoparija, B., Smith, M., Berger, U., and Essl, F.: Ragweed (Ambrosia) pollen source inventory for Austria, Science of The Total Environment, 523, 120 – 128, https://doi.org/https://doi.org/10.1016/j.scitotenv.2015.03.108, 2015.

Klein, T., Kukkonen, J., Dahl, Å., Bossioli, E., Baklanov, A., Vik, A. F., Agnew, P., Karatzas, K. D., and Sofiev, M.: Interactions of physical, chemical, and biological weather calling for an integrated approach to assessment, forecasting, and communication of air quality, Ambio, 41, 851–864, https://doi.org/10.1007/s13280-012-0288-z, 2012.

Korsholm, U. S.: Integrated modeling of aerosol indirect effects, DMI scientific report 1, DMI, http://www.dmi.dk/dmi/print/sr09-01.pdf, 2009.

Kurganskiy, A.: Integrated modelling of physical, chemical and biological weather, Ph.D. thesis, Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, 2100, Copenhagen, Denmark, 2017.

Kurganskiy, A., Mahura, A., Nuterman, R., Saarto, A., Rasmussen, A., Baklanov, A., Smyshlyaev, S., and Kaas, E.: Enviro-HIRLAM birch pollen modeling for Northern Europe, in: Report series in Aerosol Science, edited by Kulmala, M., Zilitinkevich, S., Lappalainen, H., Kyrö, E.-M., and Kontkanen, J., vol. 163, pp. 229–234, 2015.

Latałowa, M., Miętus, M., and Uruska, A.: Seasonal variations in the atmospheric Betula pollen count in Gdańsk (southern Baltic coast) in relation to meteorological parameters, Aerobiologia, 18, 33–43, 2002.

Linneberg, A., Gislum, M., Johansen, N., Husemoen, L., and Jørgensen, T.: Temporal trends of aeroallergen sensitization over twenty-five years, Clinical & Experimental Allergy, 37, 1137–1142, 2007.

Mahura, A. G., Korsholm, U. S., Baklanov, A. A., and Rasmussen, A.: Elevated birch pollen episodes in Denmark: contributions from remote sources, Aerobiologia, 23, 171–179, https://doi.org/10.1007/s10453-007-9061-3, 2007.

Mäkelä, E. M.: Size distinctions between Betula pollen types — A review, Grana, 35, 248–256, https://doi.org/10.1080/00173139609430011, 1996.

Masaka, K. and Maguchi, S.: Modelling the Masting Behaviour of Betula platyphylla var. japonica using the Resource Budget Model , Annals of Botany, 88, 1049–1055, https://doi.org/10.1006/anbo.2001.1547, 2001.

Masson, V., Champeaux, J.-L., Chauvin, F., Meriguet, C., and Lacaze, R.: A Global Database of Land Surface Parameters at 1-km Resolution in Meteorological and Climate Models, Journal of Climate, 16, 1261–1282, https://doi.org/10.1175/1520-0442(2003)16<1261:AGDOLS>2.0.CO;2, 2003.

McInnes, R. N., Hemming, D., Burgess, P., Lyndsay, D., Osborne, N. J., Skjøth, C. A., Thomas, S., and Vardoulakis, S.: Mapping allergenic pollen vegetation in UK to study environmental exposure and human health, Science of The Total Environment, 599-600, 483 – 499, https://doi.org/https://doi.org/10.1016/j.scitotenv.2017.04.136, 2017.

McMaster, G. S. and Wilhelm, W.: Growing degree-days: one equation, two interpretations, Agricultural and Forest Meteorology, 87, 291 – 300, https://doi.org/https://doi.org/10.1016/S0168-1923(97)00027-0, 1997.

Nilsson, S. and Persson, S.: Tree pollen spectra in the stockholm region (sweden), 1973–1980, Grana, 20, 179–182, https://doi.org/10.1080/00173138109427661, 1981.

5  Päivinen, R., Lehikoinen, M., Schuck, A., Häme, T., Väätäinen, S., Kennedy, P., and Folving, S.: Combining Earth Observation Data and Forest Statistics, Tech. Rep. 14, EFI, Joensuu and Joint Research Centre / European Commission, 2001.

Pauling, A., Rotach, M. W., Gehrig, R., Clot, B., and Contributors to the European Aeroallergen Network (EAN): A method to derive vegetation distribution maps for pollen dispersion models using birch as an example, International Journal of Biometeorology, 56, 949– 958, https://doi.org/10.1007/s00484-011-0505-7, 2012.

10  Pekkarinen, A., Reithmaier, L., and Strobl, P.: Pan-European forest/non-forest mapping with Landsat ETM+ and CORINE Land Cover 2000 data, ISPRS Journal of Photogrammetry and Remote Sensing, 64, 171 – 183, https://doi.org/https://doi.org/10.1016/j.isprsjprs.2008.09.004, 2009.

Persson, A.: User Guide to ECMWF forecast products, Tech. rep., ECMWF, 2011.

Piotrowska, K. and Kaszewski, B. M.: Variations in birch (Betula spp.) pollen seasons in Lublin and correlations with meteorological factors
15     in the period 2001-2010. A preliminary study, Acta agrobotanica, 64, 2011.

Prank, M., Chapman, D. S., Bullock, J. M., Belmonte, J., Berger, U., Dahl, A., Jäger, S., Kovtunenko, I., Magyar, D., Niemelä, S., Rantio-Lehtimäki, A., Rodinkova, V., Sauliene, I., Severova, E., Sikoparija, B., and Sofiev, M.: An operational model for forecasting ragweed pollen release and dispersion in Europe, Agricultural and Forest Meteorology, 182-183, 43 – 53, https://doi.org/https://doi.org/10.1016/j.agrformet.2013.08.003, 2013.

20  Ranta, H., Oksanen, A., Hokkanen, T., Bondestam, K., and Heino, S.: Masting by Betula-species; applying the resource budget model to north European data sets, International Journal of Biometeorology, 49, 146–151, https://doi.org/10.1007/s00484-004-0228-0, 2005.

Ranta, H., Kubin, E., Siljamo, P., Sofiev, M., Linkosalo, T., Oksanen, A., and Bondestam, K.: Long distance pollen transport cause problems for determining the timing of birch pollen season in Fennoscandia by using phenological observations, Grana, 45, 297–304, https://doi.org/10.1080/00173130600984740, 2006.

25  Rasmussen, A.: The effects of climate change on the birch pollen season in Denmark, Aerobiologia, 18, 253–265, https://doi.org/10.1023/A:1021321615254, 2002.

Ritenberga, O., Sofiev, M., Siljamo, P., Saarto, A., Dahl, A., Ekebom, A., Sauliene, I., Shalaboda, V., Severova, E., Hoebeke, L., and Ramfjord, H.: A statistical model for predicting the inter-annual variability of birch pollen abundance in Northern and North-Eastern Europe, Science of The Total Environment, 615, 228 – 239, https://doi.org/https://doi.org/10.1016/j.scitotenv.2017.09.061, 2018.

30  Seinfeld, J. and Pandis, S.: Atmospheric Chemistry and Physics: From Air Pollution to Climate Change, A Wiley-Interscience publication, Wiley, 2006.

Siljamo, P.: Numerical modelling of birch pollen emissions and dispersion on regional and continental scales, Ph.D. thesis, Finnish Meteorological Institute, Erik Palménin Aukio 1 (P.O. Box 503) FI-00101 Helsinki, 2013.

Siljamo, P., Sofiev, M., Ranta, H., Linkosalo, T., Kubin, E., Ahas, R., Genikhovich, E., Jatczak, K., Jato, V., Nekovár, J., Minin, A., Severova,
35     E., and Shalaboda, V.: Representativeness of point-wise phenological Betula data collected in different parts of Europe, Global Ecology and Biogeography, 17, 489–502, https://doi.org/10.1111/j.1466-8238.2008.00383.x, 2008.

Siljamo, P., Sofiev, M., Filatova, E., Grewling, Ł., Jäger, S., Khoreva, E., Linkosalo, T., Ortega Jimenez, S., Ranta, H., Rantio-Lehtimäki, A., Svetlov, A., Veriankaite, L., Yakovleva, E., and Kukkonen, J.: A numerical model of birch pollen emission and dispersion in the atmo-

sphere. Model evaluation and sensitivity analysis, International Journal of Biometeorology, 57, 125–136, https://doi.org/10.1007/s00484-012-0539-5, 2013.

Skjøth, C. A., Sommer, J., Stach, A., Smith, M., and Brandt, J.: The long-range transport of birch (Betula) pollen from Poland and Germany causes significant pre-season concentrations in Denmark, 37, 1204–1212, https://doi.org/10.1111/j.1365-2222.2007.02771.x, 2007.

5  Skjøth, C. A., Geels, C., Hvidberg, M., Hertel, O., Brandt, J., Frohn, L. M., Hansen, K. M., Hedegaard, G. B., Christensen, J. H., and Moseholm, L.: An inventory of tree species in Europe-An essential data input for air pollution modelling, Ecological Modelling, 217, 292–304, https://doi.org/10.1016/j.ecolmodel.2008.06.023, 2008a.

Skjøth, C. A., Sommer, J., Brandt, J., Hvidberg, M., Geels, C., Hansen, K. M., Hertel, O., Frohn, L. M., and Christensen, J. H.: Copenhagen – a significant source of birch (Betula) pollen?, International Journal of Biometeorology, 52, 453–462, https://doi.org/10.1007/s00484-007-0139-y, 2008b.

10  Skjøth, C. A., Ørby, P. V., Becker, T., Geels, C., Schlünssen, V., Sigsgaard, T., Bønløkke, J. H., Sommer, J., Søgaard, P., and Hertel, O.: Identifying urban sources as cause of elevated grass pollen concentrations using GIS and remote sensing, Biogeosciences, 10, 541–554, https://doi.org/10.5194/bg-10-541-2013, 2013a.

Skjøth, C. A., Šikoparija, B., Jäger, S., et al.: Pollen sources, in: Allergenic pollen, pp. 9–27, Springer, 2013b.

15  Skjøth, C. A., Smith, M., Šikoparija, B., Stach, A., Myszkowska, D., Kasprzyk, I., Radišić, P., Stjepanović, B., Hrga, I., Apatini, D., Magyar, D., Páldy, A., and Ianovici, N.: A method for producing airborne pollen source inventories: An example of Ambrosia (ragweed) on the Pannonian Plain, Agricultural and Forest Meteorology, 150, 1203 – 1210, https://doi.org/https://doi.org/10.1016/j.agrformet.2010.05.002, 2010.

Smith, M., Skjøth, C., Myszkowska, D., Uruska, A., Puc, M., Stach, A., Balwierz, Z., Chlopek, K., Piotrowska, K., Kasprzyk, I., 20  and Brandt, J.: Long-range transport of Ambrosia pollen to Poland, Agricultural and Forest Meteorology, 148, 1402 – 1411, https://doi.org/https://doi.org/10.1016/j.agrformet.2008.04.005, 2008.

Sofiev, M.: On impact of transport conditions on variability of the seasonal pollen index, Aerobiologia, pp. 1–13, https://doi.org/10.1007/s10453-016-9459-x, 2017.

Sofiev, M.: On possibilities of assimilation of near-real-time pollen data by atmospheric composition models, Aerobiologia, 25  https://doi.org/10.1007/s10453-019-09583-1, 2019.

Sofiev, M., Siljamo, P., Ranta, H., and Rantio-Lehtimäki, A.: Towards numerical forecasting of long-range air transport of birch pollen: theoretical considerations and a feasibility study, International Journal of Biometeorology, 50, 392–402, https://doi.org/10.1007/s00484-006-0027-x, 2006.

Sofiev, M., Siljamo, P., Ranta, H., Linkosalo, T., Jaeger, S., Jaeger, C., Rassmussen, A., Severova, E., Oksanen, A., Karppinen, A., and 30  Kukkonen, J.: From Russia to Iceland: an evaluation of a large-scale pollen and chemical air pollution episode during April and May, 2006, in: Aerobiological Monographs, Towards a comprehensive vision, edited by Clot B, Comtois P, E.-G. B., vol. 1, pp. 95–114, MeteoSwiss and University of Montreal, 2011.

Sofiev, M., Siljamo, P., Ranta, H., Linkosalo, T., Jaeger, S., Rasmussen, A., Rantio-Lehtimaki, A., Severova, E., and Kukkonen, J.: A numerical model of birch pollen emission and dispersion in the atmosphere. Description of the emission module, International Journal of 35  Biometeorology, 57, 45–58, https://doi.org/10.1007/s00484-012-0532-z, 2013.

Sofiev, M., Berger, U., Prank, M., Vira, J., Arteta, J., Belmonte, J., Bergmann, K.-C., Chéroux, F., Elbern, H., Friese, E., Galan, C., Gehrig, R., Khvorostyanov, D., Kranenburg, R., Kumar, U., Marécal, V., Meleux, F., Menut, L., Pessi, A.-M., Robertson, L., Ritenberga, O., Rodinkova, V., Saarto, A., Segers, A., Severova, E., Sauliene, I., Siljamo, P., Steensen, B. M., Teinemaa, E., Thibaudon, M., and Peuch,

V.-H.: MACC regional multi-model ensemble simulations of birch pollen dispersion in Europe, Atmospheric Chemistry and Physics, 15, 8115–8130, https://doi.org/10.5194/acp-15-8115-2015, 2015a.

Sofiev, M., Vira, J., Kouznetsov, R., Prank, M., Soares, J., and Genikhovich, E.: Construction of the SILAM Eulerian atmospheric dispersion model based on the advection algorithm of Michael Galperin, Geoscientific Model Development, 8, 3497–3522, https://doi.org/10.5194/gmd-8-3497-2015, 2015b.

Sørensen, B., Kaas, E., and Korsholm, U. S.: A mass-conserving and multi-tracer efficient transport scheme in the online integrated Enviro-HIRLAM model, Geoscientific Model Development, 6, 1029–1042, https://doi.org/10.5194/gmd-6-1029-2013, 2013.

Stier, P., Feichter, J., Kinne, S., Kloster, S., Vignati, E., Wilson, J., Ganzeveld, L., Tegen, I., Werner, M., Balkanski, Y., Schulz, M., Boucher, O., Minikin, A., and Petzold, A.: The aerosol-climate model ECHAM5-HAM, Atmospheric Chemistry and Physics, 5, 1125–1156, https://doi.org/10.5194/acp-5-1125-2005, 2005.

Thibaudon, M., Šikoparija, B., Oliver, G., Smith, M., and Skjøth, C. A.: Ragweed pollen source inventory for France – The second largest centre of Ambrosia in Europe, Atmospheric Environment, 83, 62 – 71, https://doi.org/https://doi.org/10.1016/j.atmosenv.2013.10.057, 2014.

Undén, P., Rontu, L., Järvinen, H., Lynch, P., Calvo, J., Cats, G., Cuxart, J., Eerola, K., Fortelius, C., Garcia-Moya, J. A., Jones, C., Lenderlink, G., McDonald, A., Mc-Grath, R., Navascues, B., Nielsen, N. W., Øidegaard, V., Rodriguez, E., Rummukainen, M., Rõõm, R., Sattler, K., Sass, B. H., Savijärvi, H., Schreur, B. W., Sigg, R., The, H., and Tijm, A.: HIRLAM-5 Scientific Documentation. The HIRLAM project, 2002.

Uppala, S. M., KÅllberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Berg, L. V. D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., Mcnally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, Quarterly Journal of the Royal Meteorological Society, 131, 2961–3012, https://doi.org/10.1256/qj.04.176, 2005.

Veriankaite, L.: Pollen concentration in the air: circulating and phenological aspects, Ph.D. thesis, Vilnius University, 3 Universiteto St, LT-01513 Vilnius, Lithuania, 2010.

WHO: Phenology and human health: allergic disorders, Tech. rep., 2003.

Yu, S., Eder, B., Dennis, R., Chu, S.-H., and Schwartz, S. E.: New unbiased symmetric metrics for evaluation of air quality models, Atmospheric Science Letters, 7, 26–34, https://doi.org/10.1002/asl.125, http://dx.doi.org/10.1002/asl.125, 2006.

Zhang, R., Duhl, T., Salam, M. T., House, J. M., Flagan, R. C., Avol, E. L., Gilliland, F. D., Guenther, A., Chung, S. H., Lamb, B. K., and VanReken, T. M.: Development of a regional-scale pollen emission and transport modeling framework for investigating the impact of climate change on allergic airway disease, Biogeosciences, 11, 1461–1478, https://doi.org/10.5194/bg-11-1461-2014, 2014.

Zink, K., Vogel, H., Vogel, B., Magyar, D., and Kottmeier, C.: Modeling the dispersion of Ambrosia artemisiifolia L. pollen with the model system COSMO-ART, International journal of biometeorology, 56, 669–680, 2012.

Zink, K., Pauling, A., Rotach, M. W., Vogel, H., Kaufmann, P., and Clot, B.: EMPOL 1.0: a new parameterization of pollen emission in numerical weather prediction models, Geoscientific Model Development, 6, 1961–1975, https://doi.org/10.5194/gmd-6-1961-2013, 2013.

Zink, K., Kaufmann, P., Petitpierre, B., Broennimann, O., Guisan, A., Gentilini, E., and Rotach, M. W.: Numerical ragweed pollen forecasts using different source maps: a comparison for France, International journal of biometeorology, 61, 23–33, 2017.

**17**

Ziska, L. H., Gebhard, D. E., Frenz, D. A., Faulkner, S., Singer, B. D., and Straka, J. G.: Cities as harbingers of climate change: Common ragweed, urbanization, and public health, Journal of Allergy and Clinical Immunology, 111, 290 – 295, https://doi.org/https://doi.org/10.1067/mai.2003.53, 2003.
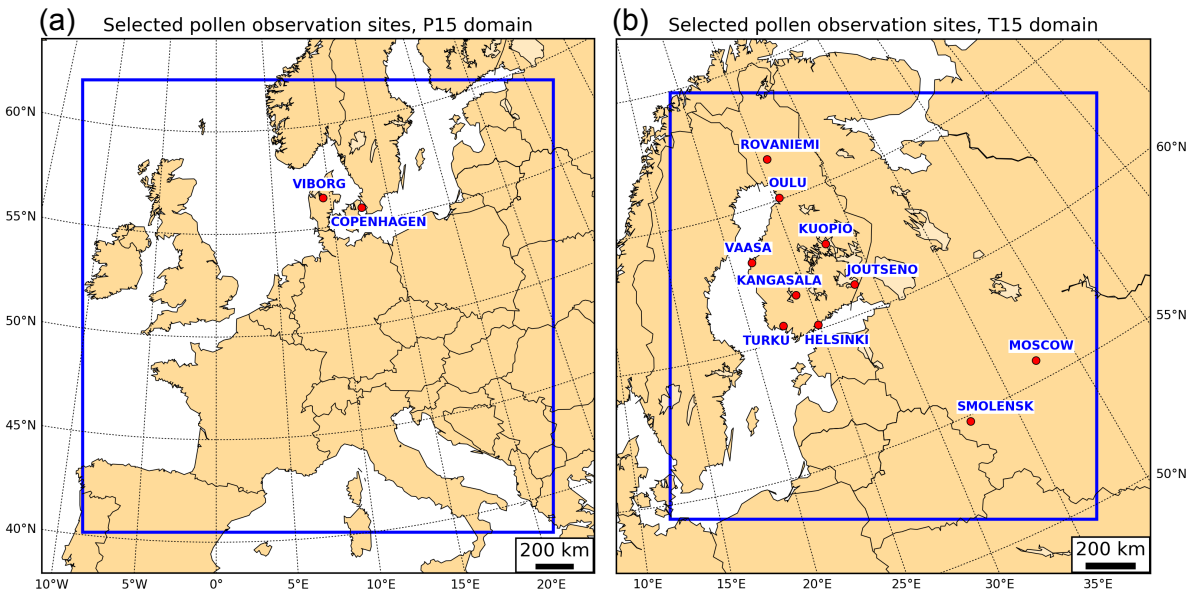
**Figure 1.** Location of the selected birch pollen observation sites with boundaries of the modelling domains: P15 (a) and T15 (b).
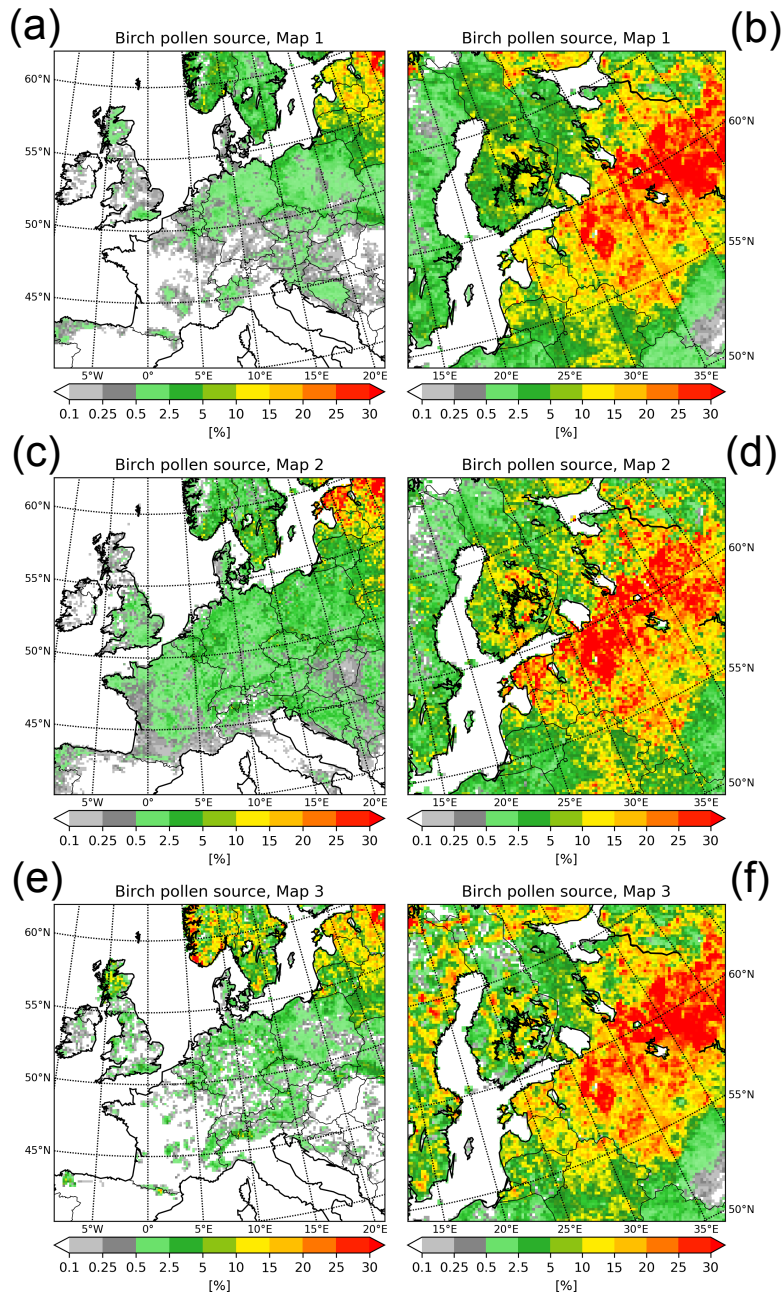
**Figure 2.** Birch pollen source maps processed for two different modelling domains P15 (left) and T15 (right) with 15 km grid resolution. The upper panel corresponds to Map 1 (a,b), the middle panel ~to Map 2 (c,d) and the lower panel ~to Map 3 (e,f). The maps are shown in [%] ~ percentage cover of birch in each 15 km x 15 km grid square.
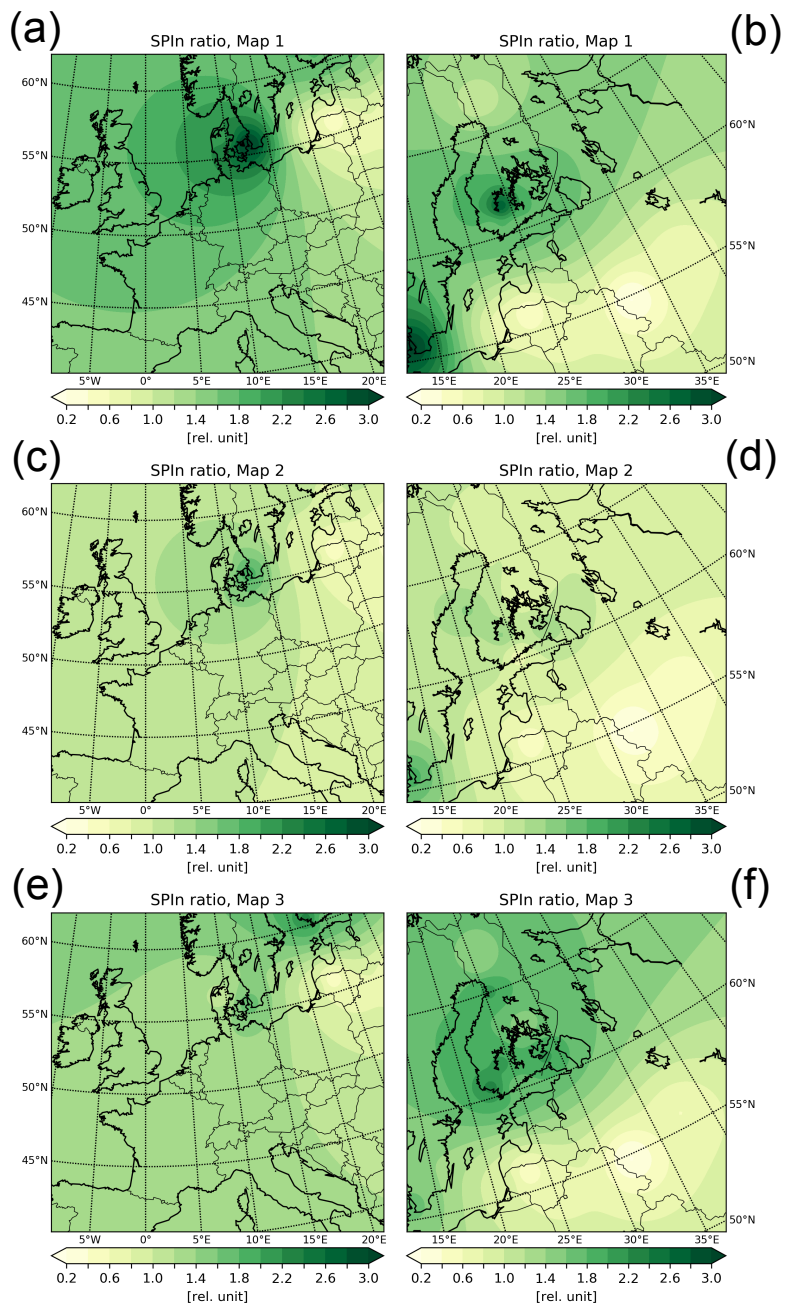
**Figure 3.** SPIn ratio used to calibrate pollen productivity in the SCF runs for two different domains P15 (left) and T15 (right). The upper panel corresponds to Map 1 (a,b), the middle panel — to Map 2 (c,d) and the lower panel — to Map 3 (e,f).
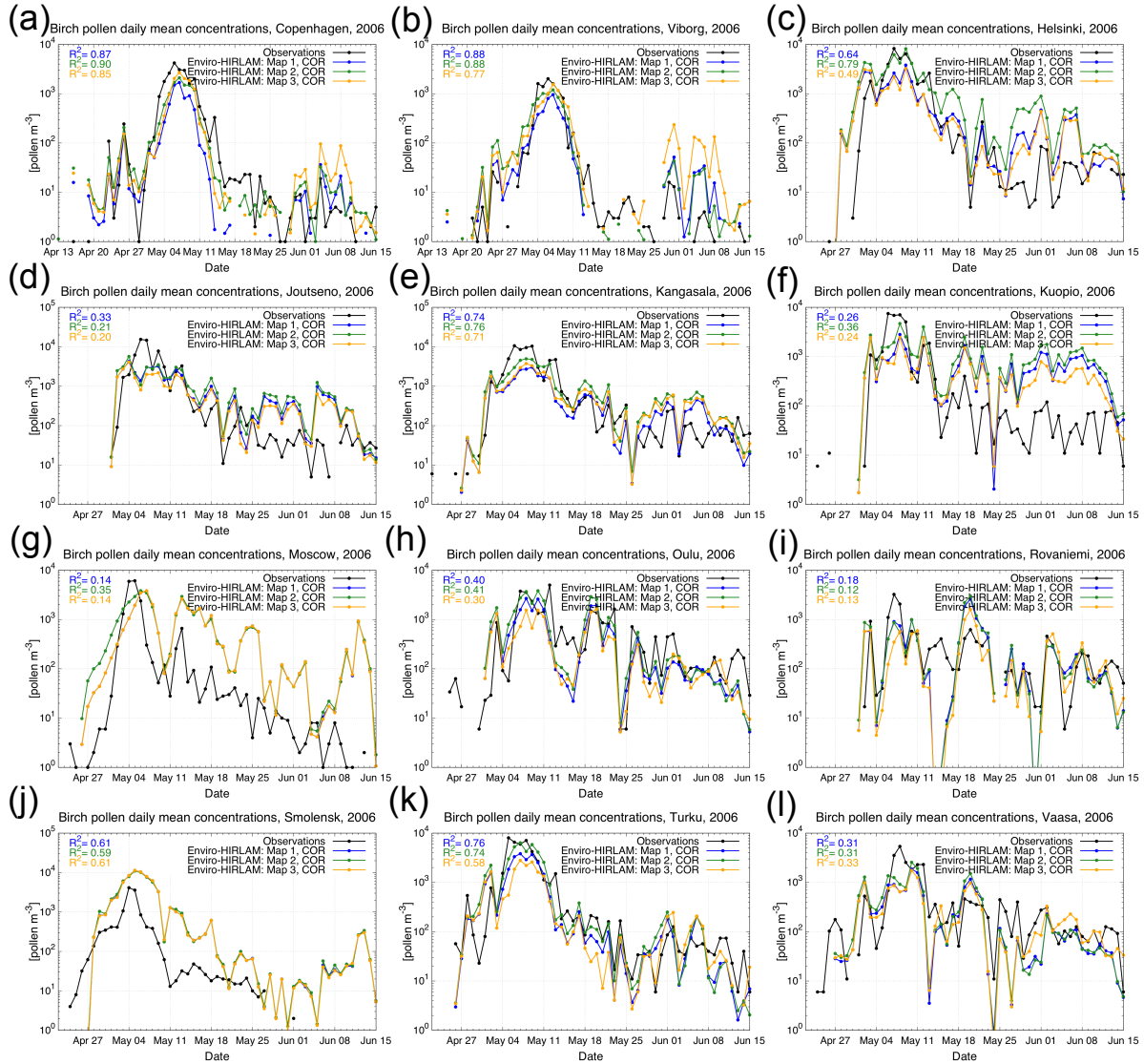
**Figure 4.** Time series of the modelled and observed birch pollen daily mean concentrations for the selected pollen observation sites. The modelled concentrations correspond to each simulated map according to the COR run. A base-10 log scale is used for the Y axis.
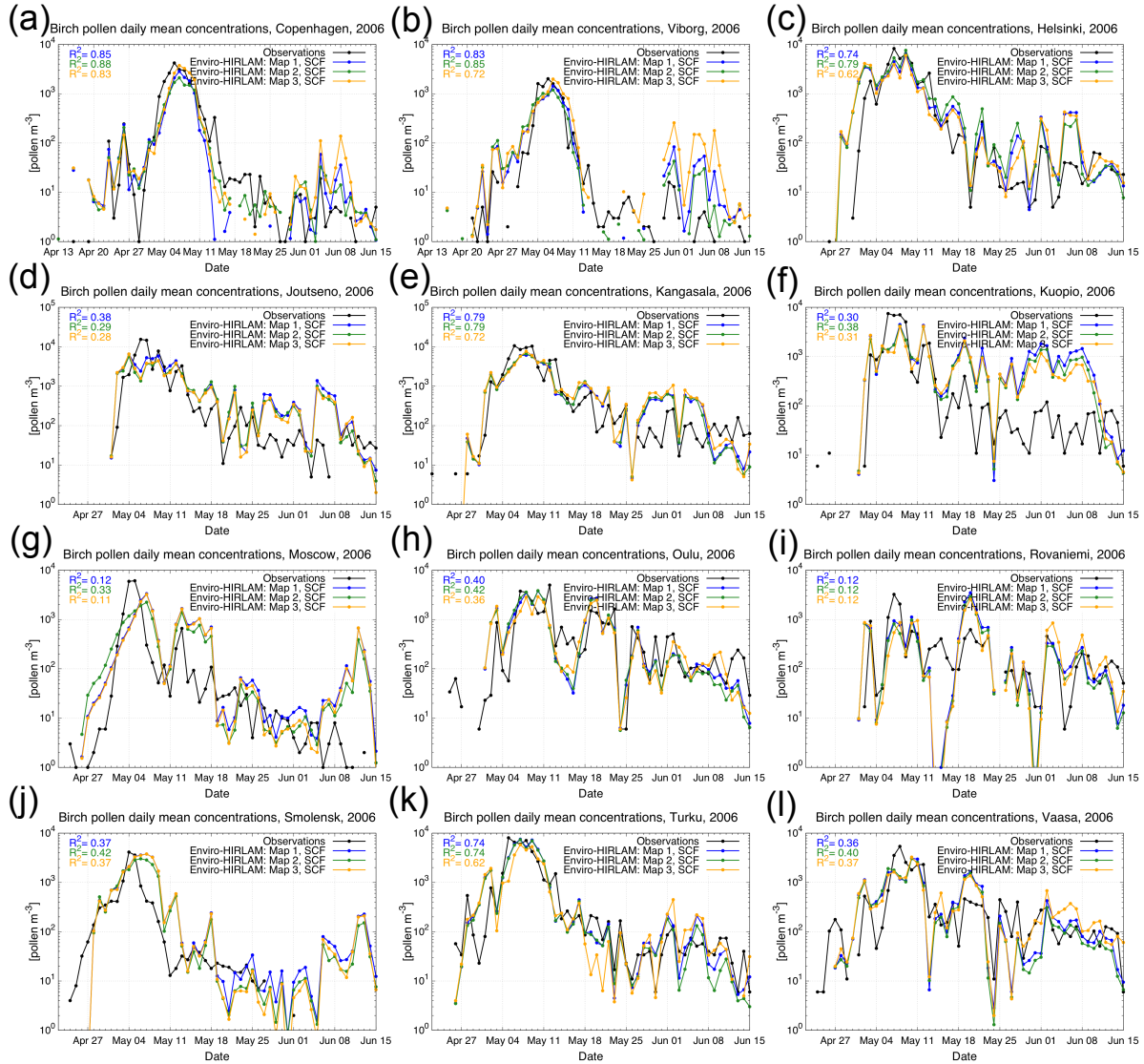
**Figure 5.** Time series of the modelled and observed birch pollen daily mean concentrations for the selected pollen observation sites. The modelled concentrations correspond to each simulated map according to the SCF run. A base-10 log scale is used for the Y axis.

23
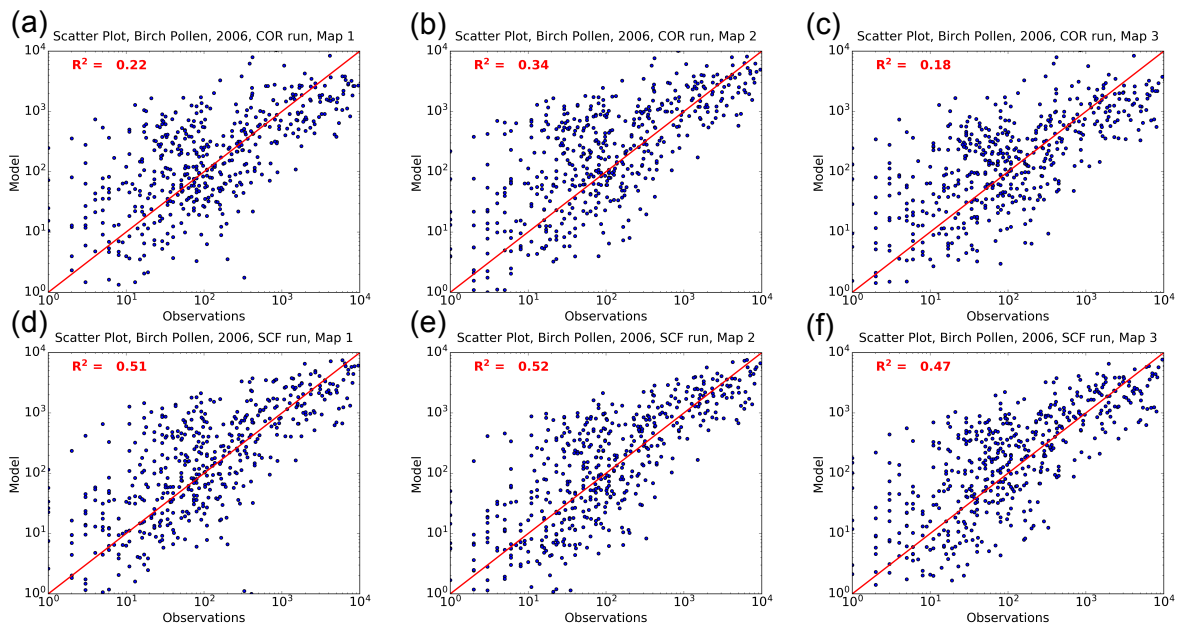
**Figure 6.** Scatter plots for birch pollen concentrations according to comparing COR (the upper panel) and SCF (the lower panel) model runs with observations for Map 1 (a,d), Map 2 (b,e) and Map 3 (c,f).
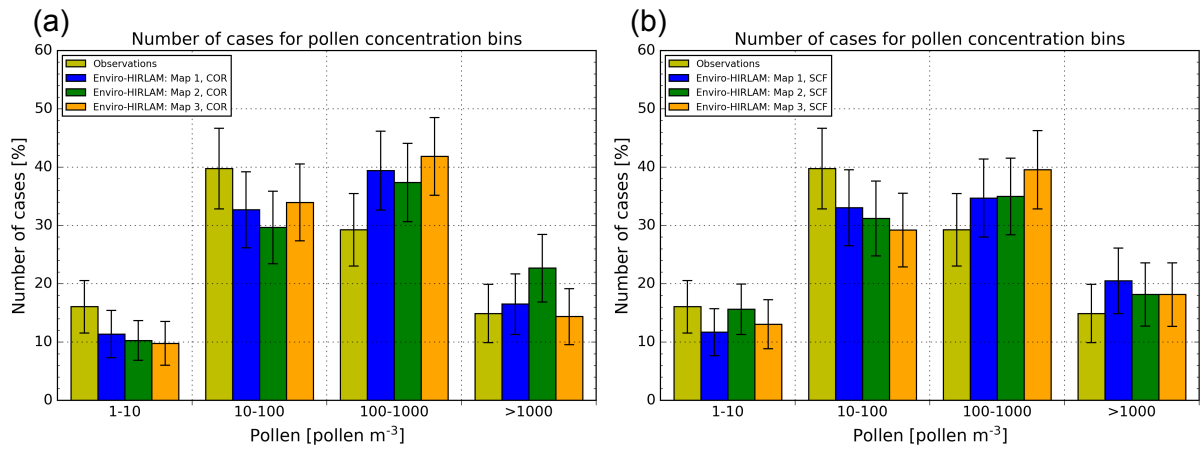
**Figure 7.** Histogram of the observed and modelled birch pollen concentrations sorted into low (1–10 pollen $m^{-3}$), moderate (10–100) pollen $m^{-3}$, high (100–1000 pollen $m^{-3}$) and very high (> 1000 pollen $m^{-3}$) classes. The left panel (a) corresponds to COR and the ~~rght~~ right panel (b) - to the SCF model simulations.

**Table 1.** Global "standard" statistical metrics obtained by comparing the modelled ($\overline{C_m}$) and observed ($\overline{C_o}$) daily mean birch pollen concentrations. The metrics were calculated using all available station data for both simulated domains, over 30 April – 15 June, 2006 period. $\overline{C_m}$, $\overline{C_o}$, MB and RMSE are given in [pollen m$^{-3}$], p-value < 0.01

| Run | R | $\overline{C_m}$ | $\overline{C_o}$ | MB | RMSE | NMBF | NMAEF |
|-----|-----|--------|--------|---------|---------|-------|-------|
| COR, Map 1 | 0.47 | 545.89 | 683.63 | -137.74 | 1587.81 | -0.25 | 1.12 |
| COR, Map 2 | 0.59 | 752.71 | 683.63 | 69.08 | 1465.79 | 0.09 | 0.81 |
| COR, Map 3 | 0.42 | 509.49 | 683.63 | -174.14 | 1639.47 | -0.34 | 1.21 |
| SCF, Map 1 | 0.72 | 661.64 | 683.63 | -21.99 | 1229.16 | -0.03 | 0.77 |
| SCF, Map 2 | 0.72 | 630.02 | 683.63 | -53.61 | 1223.95 | -0.09 | 0.76 |
| SCF, Map 3 | 0.68 | 614.41 | 683.63 | -69.22 | 1290.04 | -0.10 | 0.75 |

**Table 2.** The threshold-based statistical metrics obtained by comparing model simulations with observed daily mean concentrations. MA, POD, FAR and POFD are given in [%].

| Run, map | MA | POD | FAR | POFD | OR |
|---|---|---|---|---|---|
| COR, Map 1 | 76.3 | 74.6 | 25.4 | 31.9 | 2.3 |
| COR, Map 2 | 74.7 | 71.1 | 28.9 | 38.3 | 1.9 |
| COR, Map 3 | 75.4 | 72.7 | 27.3 | 35.0 | 2.1 |
| SCF, Map 1 | 79.5 | 77.7 | 22.3 | 27.6 | 2.8 |
| SCF, Map 2 | 79.8 | 79.1 | 20.9 | 24.8 | 3.2 |
| SCF, Map 3 | 79.7 | 77.1 | 22.9 | 29.1 | 2.7 |

**Table 3.** Hit rates of the Enviro-HIRLAM birch pollen simulations (COR run) for classes:
zero or low, moderate, high and very high.

The classes are based on daily mean pollen concentrations and presented in units: [pollen $\mathrm{m}^{-3}$].

| Observation Model | 0–10 | 10–100 | 100–1000 | > 1000 |
|---|---|---|---|---|
| COR run, Map 1 | | | | |
| 0–10 | 94.5 % | 4.6 % | 0.9 % | 0.0 % |
| 10–100 | 29.1 % | 46.0 % | 24.9 % | 0.0 % |
| 100–1000 | 6.1 % | 41.3 % | 38.5 % | 14.1 % |
| > 1000 | 0.0 % | 11.5 % | 25.3 % | 63.2 % |
| COR run, Map 2 | | | | |
| 0–10 | 95.2 % | 3.9 % | 0.9 % | 0.0 % |
| 10–100 | 29.0 % | 47.9 % | 23.1 % | 0.0 % |
| 100–1000 | 9.2 % | 44.4 % | 37.7 % | 8.7 % |
| > 1000 | 0.0 % | 15.1 % | 28.6 % | 56.3 % |
| COR run, Map 3 | | | | |
| 0–10 | 95.0 % | 4.2 % | 0.8 % | 0.0 % |
| 10–100 | 34.0 % | 44.2 % | 21.8 % | 0.0 % |
| 100–1000 | 6.9 % | 40.5 % | 40.1 % | 12.5 % |
| > 1000 | 0.0 % | 10.0 % | 20.0 % | 70.0 % |

**Table 4.** Hit rates of the Enviro-HIRLAM birch pollen simulations (SCF run) for classes:
zero or low, moderate, high and very high.
The classes are based on daily mean pollen concentrations and presented in units: [pollen m$^{-3}$].

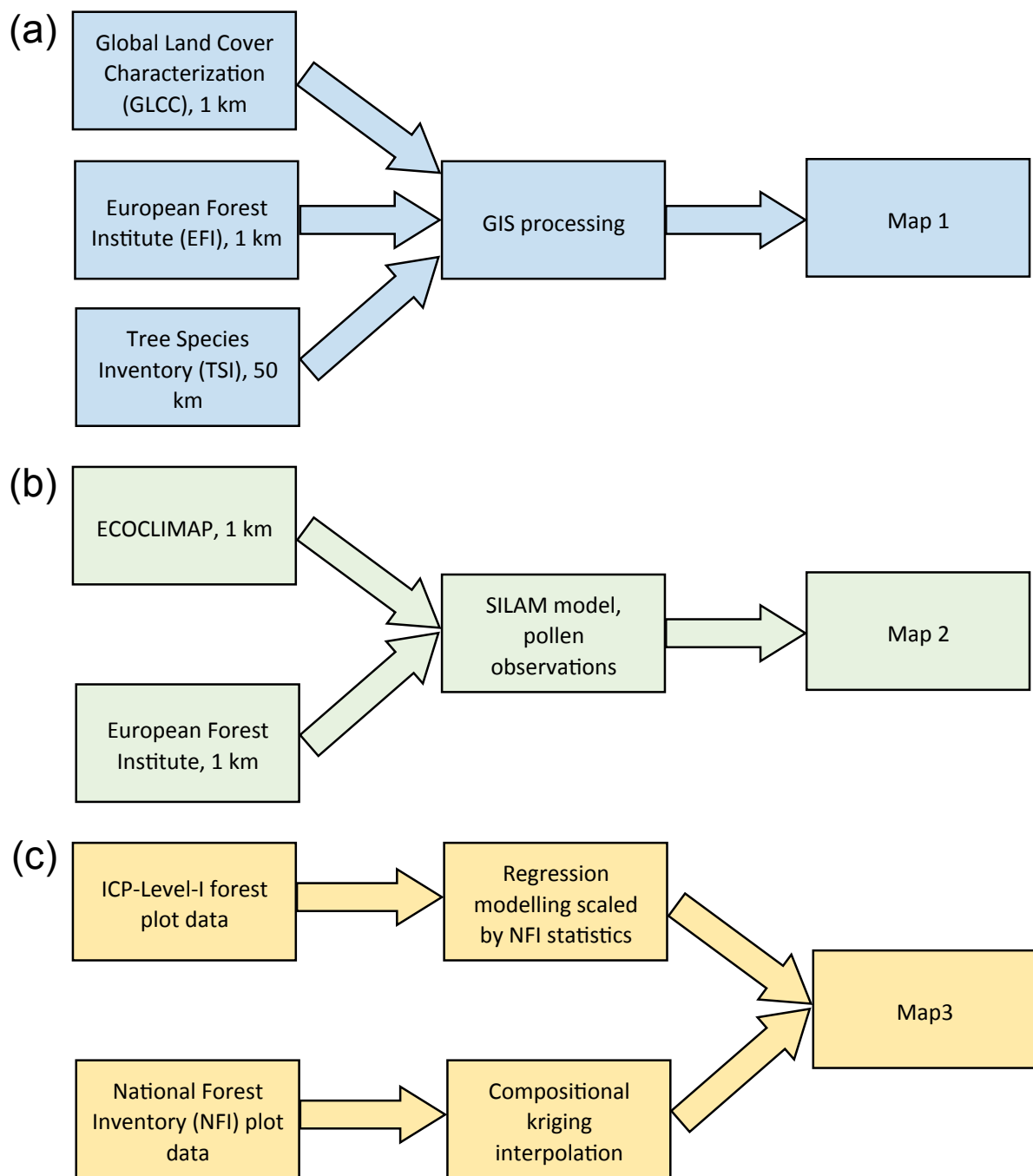| Observation Model | 0–10 | 10–100 | 100–1000 | > 1000 |
|---|---|---|---|---|
| SCF run, Map 1 | | | | |
| 0–10 | 85.5 % | 12.1 % | 2.4 % | 0.0 % |
| 10–100 | 32.0 % | 49.5 % | 18.5 % | 0.0 % |
| 100–1000 | 5.3 % | 40.0 % | 46.8 % | 7.9 % |
| > 1000 | 0.0 % | 12.7 % | 23.6 % | 63.7 % |
| SCF run, Map 2 | | | | |
| 0–10 | 84.8 % | 13.2 % | 2.0 % | 0.0 % |
| 10–100 | 29.6 % | 48.5 % | 21.9 % | 0.0 % |
| 100–1000 | 5.7 % | 41.8 % | 46.3 % | 6.2 % |
| > 1000 | 0.0 % | 11.4 % | 23.7 % | 64.9 % |
| SCF run, Map 3 | | | | |
| 0–10 | 84.0 % | 14.3 % | 1.7 % | 0.0 % |
| 10–100 | 34.9 % | 46.3 % | 18.8 % | 0.0 % |
| 100–1000 | 6.4 % | 42.7 % | 44.0 % | 6.9 % |
| > 1000 | 0.0 % | 5.1 % | 24.2 % | 70.7 % |

**Figure A1.** Flowcharts outlining the data and methods used to obtain Map 1 (a), Map 2 (b) and Map 3 (c).

**Table A1.** Birch pollen statistical metrics for the selected observation sites: COR run.
$\overline{C_m}$, $\overline{C_o}$, MB, RMSE are in pollen $m^{-3}$; p-value $< 0.05$.

| St. name | Map | Lat | Lon | R | $\overline{C_m}$ | $\overline{C_o}$ | MB | RMSE | NMBF | NMAEF |
|---|---|---|---|---|---|---|---|---|---|---|
| Helsinki | | 60.17° N | 24.90° E | | | | | | | |
| | Map 1 | | | 0.80 | 574.69 | 879.28 | -299.59 | 1269.56 | -0.53 | 1.05 |
| | Map 2 | | | 0.89 | 1233.00 | 879.28 | 353.72 | 895.53 | 0.40 | 0.64 |
| | Map 3 | | | 0.70 | 515.97 | 879.28 | -363.31 | 1412.44 | -0.70 | 1.28 |
| Joutseno | | 61.10° N | 28.50° E | | | | | | | |
| | Map 1 | | | 0.57 | 838.61 | 1345.81 | -507.2 | 2865.20 | -0.60 | 1.30 |
| | Map 2 | | | 0.46 | 1002.84 | 1345.81 | -342.97 | 2945.21 | -0.34 | 1.21 |
| | Map 3 | | | 0.45 | 659.18 | 1345.81 | -686.63 | 3070.78 | -1.04 | 1.71 |
| Kangasala | | 61.47° N | 24.08° E | | | | | | | |
| | Map 1 | | | 0.86 | 569.86 | 1442.55 | -872.69 | 2369.17 | -1.53 | 1.80 |
| | Map 2 | | | 0.87 | 1018.23 | 1442.55 | -424.32 | 1841.15 | -0.42 | 0.87 |
| | Map 3 | | | 0.84 | 701.75 | 1442.55 | -740.8 | 2252.54 | -1.06 | 1.46 |
| Kuopio | | 62.88° N | 27.63° E | | | | | | | |
| | Map 1 | | | 0.51 | 672.75 | 777.23 | -104.48 | 1627.23 | -0.16 | 1.28 |
| | Map 2 | | | 0.60 | 1032.18 | 777.23 | 254.95 | 1508.64 | 0.33 | 1.25 |
| | Map 3 | | | 0.49 | 530.67 | 777.23 | -246.56 | 1665.51 | -0.46 | 1.45 |
| Oulu | | 65.07° N | 25.52° E | | | | | | | |
| | Map 1 | | | 0.63 | 469.77 | 682.43 | -212.66 | 845.99 | -0.45 | 0.85 |
| | Map 2 | | | 0.64 | 686.56 | 682.43 | 4.13 | 883.44 | 0.01 | 0.67 |
| | Map 3 | | | 0.55 | 333.85 | 682.43 | -348.58 | 950.66 | -1.04 | 1.33 |
| Rovaniemi | | 66.55° N | 25.73° E | | | | | | | |
| | Map 1 | | | 0.42 | 302.92 | 309.64 | -6.72 | 574.93 | -0.02 | 0.92 |
| | Map 2 | | | 0.35 | 324.04 | 309.64 | 14.40 | 646.57 | 0.05 | 0.98 |
| | Map 3 | | | 0.36 | 211.50 | 309.64 | -98.14 | 544.39 | -0.46 | 1.14 |

**Table A2.** Birch pollen statistical metrics for the selected observation sites: COR run. $\overline{C_m}, \overline{C_o}$, MB, RMSE are in pollen m$^{-3}$; p-value < 0.05.

| St. name | Map | Lat | Lon | R | $\overline{C_m}$ | $\overline{C_o}$ | MB | RMSE | NMBF | NMAEF |
|---|---|---|---|---|---|---|---|---|---|---|
| Turku | | 60.53° N | 22.47° E | | | | | | | |
| | Map 1 | | | 0.87 | 548.80 | 925.02 | -376.22 | 1269.02 | -0.69 | 0.89 |
| | Map 2 | | | 0.86 | 862.73 | 925.02 | -62.29 | 1011.47 | -0.07 | 0.50 |
| | Map 3 | | | 0.76 | 404.71 | 925.02 | -520.31 | 1605.97 | -1.29 | 1.58 |
| Vaasa | | 63.10° N | 21.62° E | | | | | | | |
| | Map 1 | | | 0.56 | 307.10 | 548.43 | -241.33 | 902.99 | -0.79 | 1.24 |
| | Map 2 | | | 0.56 | 389.07 | 548.43 | -159.36 | 871.11 | -0.41 | 1.06 |
| | Map 3 | | | 0.57 | 294.52 | 548.43 | -253.91 | 913.76 | -0.86 | 1.29 |
| Moscow | | 55.74° N | 37.55° E | | | | | | | |
| | Map 1 | | | 0.37 | 647.42 | 388.28 | 259.14 | 1269.87 | 0.67 | 1.74 |
| | Map 2 | | | 0.59 | 733.42 | 388.28 | 345.14 | 1100.18 | 0.89 | 1.65 |
| | Map 3 | | | 0.37 | 647.40 | 388.28 | 259.12 | 1270.04 | 0.67 | 1.74 |
| Smolensk | | 54.78° N | 32.05° E | | | | | | | |
| | Map 1 | | | 0.78 | 1382.37 | 259.32 | 1123.05 | 2542.82 | 4.33 | 4.33 |
| | Map 2 | | | 0.77 | 1368.29 | 259.32 | 1108.97 | 2466.38 | 4.28 | 4.28 |
| | Map 3 | | | 0.78 | 1377.17 | 259.32 | 1117.85 | 2537.15 | 4.31 | 4.31 |
| Copenhagen | | 55.72° N | 12.57° E | | | | | | | |
| | Map 1 | | | 0.93 | 144.63 | 445.83 | -301.2 | 717.53 | -2.08 | 2.10 |
| | Map 2 | | | 0.95 | 231.68 | 445.83 | -214.15 | 560.20 | -0.92 | 0.95 |
| | Map 3 | | | 0.92 | 263.55 | 445.83 | -182.28 | 503.85 | -0.69 | 0.79 |
| Viborg | | 56.45° N | 9.40° E | | | | | | | |
| | Map 1 | | | 0.94 | 91.72 | 199.70 | -107.98 | 325.90 | -1.18 | 1.34 |
| | Map 2 | | | 0.94 | 150.53 | 199.70 | -49.17 | 229.32 | -0.33 | 0.64 |
| | Map 3 | | | 0.88 | 173.58 | 199.70 | -26.12 | 251.93 | -0.15 | 0.58 |

**Table A3.** Birch pollen statistical metrics for the selected observation sites: SCF run.
$\overline{C_m}$, $\overline{C_o}$, MB, RMSE are in pollen m$^{-3}$; p-value $< 0.05$.

| St. name | Map | Lat | Lon | R | $\overline{C_m}$ | $\overline{C_o}$ | MB | RMSE | NMBF | NMAEF |
|---|---|---|---|---|---|---|---|---|---|---|
| Helsinki | | 60.17° N | 24.90° E | | | | | | | |
| | Map 1 | | | 0.86 | 841.46 | 879.28 | -37.82 | 942.66 | -0.04 | 0.55 |
| | Map 2 | | | 0.89 | 981.76 | 879.28 | 102.48 | 833.56 | 0.12 | 0.51 |
| | Map 3 | | | 0.79 | 785.22 | 879.28 | -94.06 | 1122.77 | -0.12 | 0.68 |
| Joutseno | | 61.10° N | 28.50° E | | | | | | | |
| | Map 1 | | | 0.62 | 1224.33 | 1345.81 | -121.48 | 2601.02 | -0.10 | 0.92 |
| | Map 2 | | | 0.54 | 989.56 | 1345.81 | -356.25 | 2823.36 | -0.36 | 1.15 |
| | Map 3 | | | 0.53 | 1015.47 | 1345.81 | -330.34 | 2812.42 | -0.33 | 1.12 |
| Kangasala | | 61.47° N | 24.08° E | | | | | | | |
| | Map 1 | | | 0.89 | 1100.24 | 1442.55 | -342.31 | 1595.21 | -0.31 | 0.73 |
| | Map 2 | | | 0.89 | 1135.93 | 1442.55 | -306.62 | 1575.66 | -0.27 | 0.69 |
| | Map 3 | | | 0.85 | 1200.14 | 1442.55 | -242.41 | 1633.02 | -0.20 | 0.72 |
| Kuopio | | 62.88° N | 27.63° E | | | | | | | |
| | Map 1 | | | 0.55 | 953.83 | 777.23 | 176.6 | 1559.38 | 0.23 | 1.23 |
| | Map 2 | | | 0.62 | 817.90 | 777.23 | 40.67 | 1474.81 | 0.05 | 1.07 |
| | Map 3 | | | 0.56 | 755.99 | 777.23 | -21.24 | 1543.05 | -0.03 | 1.09 |
| Oulu | | 65.07° N | 25.52° E | | | | | | | |
| | Map 1 | | | 0.63 | 669.43 | 682.43 | -13.00 | 876.15 | -0.02 | 0.68 |
| | Map 2 | | | 0.65 | 676.29 | 682.43 | -6.14 | 861.12 | -0.01 | 0.66 |
| | Map 3 | | | 0.60 | 583.34 | 682.43 | -99.09 | 857.18 | -0.17 | 0.78 |
| Rovaniemi | | 66.55° N | 25.73° E | | | | | | | |
| | Map 1 | | | 0.35 | 368.22 | 309.64 | 58.58 | 691.80 | 0.19 | 1.06 |
| | Map 2 | | | 0.34 | 309.19 | 309.64 | -0.45 | 632.14 | 0.00 | 0.97 |
| | Map 3 | | | 0.34 | 332.60 | 309.64 | 22.96 | 611.50 | 0.07 | 0.98 |

**Table A4.** Birch pollen statistical metrics for the selected observation sites: SCF run.

$\overline{C_m}$, $\overline{C_o}$, MB, RMSE are in pollen m$^{-3}$; p-value $< 0.05$.

| St. name | Map | Lat | Lon | R | $\overline{C_m}$ | $\overline{C_o}$ | MB | RMSE | NMBF | NMAEF |
|---|---|---|---|---|---|---|---|---|---|---|
| Turku | | 60.53° N | 22.47° E | | | | | | | |
| | Map 1 | | | 0.86 | 958.24 | 925.02 | 33.22 | 1032.33 | 0.04 | 0.49 |
| | Map 2 | | | 0.86 | 924.51 | 925.02 | -0.51 | 1033.30 | 0.00 | 0.49 |
| | Map 3 | | | 0.79 | 732.00 | 925.02 | -193.02 | 1241.52 | -0.26 | 0.66 |
| Vaasa | | 63.10° N | 21.62° E | | | | | | | |
| | Map 1 | | | 0.60 | 507.24 | 548.43 | -41.19 | 831.66 | -0.08 | 0.84 |
| | Map 2 | | | 0.63 | 475.72 | 548.43 | -72.71 | 811.14 | -0.15 | 0.86 |
| | Map 3 | | | 0.61 | 503.97 | 548.43 | -44.46 | 817.61 | -0.09 | 0.82 |
| Moscow | | 55.74° N | 37.55° E | | | | | | | |
| | Map 1 | | | 0.34 | 392.03 | 388.28 | 0.01 | 1206.26 | 0.01 | 1.25 |
| | Map 2 | | | 0.57 | 334.68 | 388.28 | -0.16 | 1047.91 | -0.16 | 1.18 |
| | Map 3 | | | 0.33 | 375.92 | 388.28 | -0.03 | 1208.18 | -0.03 | 1.28 |
| Smolensk | | 54.78° N | 32.05° E | | | | | | | |
| | Map 1 | | | 0.61 | 505.21 | 259.32 | 245.89 | 859.77 | 0.95 | 1.34 |
| | Map 2 | | | 0.65 | 415.25 | 259.32 | 155.93 | 693.79 | 0.60 | 1.14 |
| | Map 3 | | | 0.61 | 494.46 | 259.32 | 235.14 | 854.60 | 0.91 | 1.32 |
| Copenhagen | | 55.72° N | 12.57° E | | | | | | | |
| | Map 1 | | | 0.92 | 261.39 | 445.83 | -184.44 | 494.49 | -0.71 | 0.77 |
| | Map 2 | | | 0.94 | 312.42 | 445.83 | -133.41 | 407.55 | -0.43 | 0.50 |
| | Map 3 | | | 0.91 | 370.93 | 445.83 | -74.90 | 418.39 | -0.20 | 0.47 |
| Viborg | | 56.45° N | 9.40° E | | | | | | | |
| | Map 1 | | | 0.91 | 158.00 | 199.70 | -41.7 | 239.74 | -0.26 | 0.57 |
| | Map 2 | | | 0.92 | 187.03 | 199.70 | -12.67 | 201.17 | -0.07 | 0.43 |
| | Map 3 | | | 0.85 | 222.83 | 199.70 | 23.13 | 261.80 | 0.12 | 0.64 |

**Table A5.** Thresholds and classes used to calculate the threshold-based statistical metrics for 12 selected pollen observational sites.

| Threshold/Class | Value/Value range |
|---|---|
| $C_{th}$ - the threshold for daily mean pollen concentration | 50 pollen m$^{-3}$ |
| Low pollen concentration | 1–10 pollen m$^{-3}$ |
| Moderate pollen concentration | 10–100 pollen m$^{-3}$ |
| High pollen concentration | 100–1000 pollen m$^{-3}$ |
| Very high pollen concentration | $> 1000$ pollen m$^{-3}$ |

**Table A6.** Definitions of the threshold-based statistical metrics used to compare the modelled and observed daily mean birch pollen concentrations. See Siljamo et al. (2013) for more details.

| Metric | Abbreviation | Definition |
|---|---|---|
| Model Accuracy | MA | The fraction of correct model predictions. |
| Probability of Detection | POD | The fraction of correct model predictions for concentrations $> C_{th}$. |
| False Alarm Ratio | FAR | The fraction of incorrect model predictions for concentrations $> C_{th}$. |
| Probability of False Detection | POFD | The fraction of the observed pollen concentrations $< C_{th}$ predicted as $> C_{th}$. |
| Odds Ratio | OR | OR indicates how much higher are the chances to get pollen concentration $> C_{th}$ than $< C_{th}$ if the model prediction is $> C_{th}$. |
| Hit Rates | HR | The percentage of correct model predictions for the classes of pollen concentrations. |