

Journal Pre-proof

Numerical modelling of birch pollen dispersion in Canada

Alain Robichaud, Paul Comtois



PII: S0013-9351(20)31451-1

DOI: <https://doi.org/10.1016/j.envres.2020.110554>

Reference: YENRS 110554

To appear in: *Environmental Research*

Received Date: 21 May 2020

Revised Date: 23 November 2020

Accepted Date: 25 November 2020

Please cite this article as: Robichaud, A., Comtois, P., Numerical modelling of birch pollen dispersion in Canada, *Environmental Research*, <https://doi.org/10.1016/j.envres.2020.110554>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Crown Copyright © 2020 Published by Elsevier Inc. All rights reserved.

Alain Robichaud wrote the article, proposed the design and did all the experiments and figures. Paul Comtois took care of the pollen measurement at UdeM, participated into the design, provided advices and revised the research article.

Numerical modelling of birch pollen dispersion in Canada

Alain Robichaud¹
Paul Comtois²

¹Environnement and Climate Change Canada, 2121 Trans-Canada, Dorval, Canada, H9P 1J3
alain.robichaud@canada.ca

²Department of Geography, Université de Montréal, 1375 Avenue Thérèse-Lavoie-Roux, Montréal, Canada, H2V 0B3, paul.comtois@canada.ca

ABSTRACT

Simulating allergenic tree pollen is important to protect sensitive population and to support bioaerosols monitoring effort. Using the regional air quality model GEM-MACH, a simulation was conducted adopting two new main hypotheses: 1) the use of vertical correlation concept to force the vertical dispersion (a method normally used in tracer data assimilation) and, 2) the use of a puff instead of a continuous pollen release. The simulation was compared with pollen observations in Montreal and with the corresponding statistical forecasts (issued daily by the Weather Network) at several locations in the province of Quebec and elsewhere. The comparison with the simulation was found satisfactory (outperform forecasts based on persistence or pollen calendar and is also superior to numerical simulation of tree pollen done elsewhere in North America). Simulation shows that, for the 2012 pollen season, the majority (88%) of the *Betula* pollen measured in Montreal originated from the Laurentides region. Another result of scientific importance obtained here is that *Betula* pollen episodes (observed or simulated birch pollen) in Montreal occur only when the average daily temperature is in the range of 10° to 18°C. This research is considered as a first step in forecasting bioaerosols in Canada within an air quality model.

1. Introduction

Pollinosis is among one of the most common respiratory disease worldwide and it is likely to increase in the future due to growing urbanization and climate change (D'Amato

et al., 2001; 2007; 2015; EPA, 2008). According to WHO (2003), in Europe, approximately 15 % of the population is sensitized to birch (*Betula*) pollen. In Canada, Dales et al. (2004; 2008) reported that birch tree pollen is responsible for a significant number of hospitalization in urban environments. Therefore, improving pollen monitoring and modelling is needed in Canada.

The goal of this study is to describe numerical simulations of emission, transport and dispersion of birch pollen with a focus in eastern Canada and more precisely in the region of Montreal (province of Quebec) during a whole pollen season (11th April - 31st May 2012). As a prototype study here, birch¹ (*Betula*) was selected because 1) the *Betula* pollen measured in the Montreal region is, at times, present in high concentration and likely brought by the transport/dispersion at medium-long distance (TDMLD) from major source regions (Figure 1), 2) the pollen emitted quantity is much larger than most other tree species (Ranta and Satri, 2007), and 3) birch is an important allergy plant (Guérin, 1993; WHO, 2003; D'Amato et al., 2007; Sierra-Heredia et al., 2018) and therefore has a significant public health issue especially over polluted urban environment such as Montreal (second highest urban population of Canada). Although this paper focuses to a small region the prototype could be exported to any regions of the world provided pollen measurements are available.

2. Modelling system

Recent advancements in modelling (including coupled meteorological-chemical models and vegetation databases) have allowed some progress in simulation and prediction of pollen and air pollutants affected by TDMLD (see a review in Efstathiou et al., 2011; Sofiev and Bergmann, 2013; Zhang et al., 2014). The meteorological driver GEM, a numerical weather prediction (NWP) model is currently used for weather forecast in Canada (Côté et al., 1998a,b) and is based on the fundamental conservation principles of fluid dynamics, namely, those of mass, momentum and energy. The chemical module MACH provides the air quality forecast (Moran et al., 2011; 2012). GEM-MACH (Global Environmental Multiscale – Model for Air quality and CHemistry) model solves

¹ Birch is a broadleaved deciduous hardwood tree of the family of *Betulaceae*.

the standard advection-diffusion equation and uses the semi-Lagrangian transport algorithm to disperse the pollutants according to the following equation:

$$\partial C / \partial t = -\nabla(\mathbf{v}C) + \nabla(\mathbf{K} \cdot \nabla C) + Q + S \quad (1)$$

where C , $\mathbf{v}$, $\mathbf{K}$, ∇ , Q and S represent, respectively, the pollen concentration, the advection velocity 3-D vector field (wind field), the three-dimensional sub-grid eddy diffusivity vector, the spatial gradient mathematical operator, source (emission) and sink (removal) terms. Eq. 1 means that the local change in pollen concentration at a given point depends on the transport, diffusion (horizontal and vertical) and source and sinks respectively. This equation is applied to every point of a domain of 744 X 644 covering most of North America. Details of the GEM-MACH platform are presented in Supplementary material S1.

To be successful, numerical simulation of pollen dispersal must reproduce four complex phenomena: pollen emission flux, transport by the mean wind, diffusion by atmospheric turbulence, and dry/wet deposition (Helbig et al., 2004; Chamecki et al., 2009). Since GEM-MACH model already includes the treatment of $PM_{2.5-10}$ (coarse particles having diameter between 2.5 to 10 microns), this capability is used to include a new tracer (birch/*Betula* pollen) and consider it as a coarse aerosol ($PM_{2.5-10}$) to take advantage of physical and chemical processes such as wet deposition and coagulation. As other authors (Helbig et al., 2004; Sofiev et al., 2006a; Efstathiou et al., 2011), we have assumed that deposition of pollen grains is negligible compared to their sedimentation and imposed a fixed sedimentation velocity of 1.3 cm/s (see Supplementary material S2 for details). Finally, tree pollen has been found to be transported by the wind and turbulence over longer distances than expected (Mandrioli et al., 1984; Campbell et al. 1999; Skj  th et al. 2007) so the use of a meteorological model framework is appropriate (Sofiev et al. 2006a,b).

2.1. Study area

In here, the model simulations cover the whole Canada and US territory (see Figure 1A) but the focus is on an area depicted in Figure 1B (centred over Montreal, latitude 45.51 N, longitude 73.55 W) where the phenology model was developed (see Section 4.2 for more details). The focus is on Montreal since pollen data was available and of good quality but also for the following additional reasons:

- 1) in Montreal, most of the pollen is not locally emitted (low density of birch in the city, i.e. less than 5%, Figure 1B) and TDMLD impact could be studied (which requires the use of NWP model),
- 2) one of the most important source region of birch pollen is located upwind of Montreal to the north and north-west (Laurentide mountains),
- 3) Montreal has the largest urban population of Quebec and the second highest in Canada (after Toronto) so the combined exposure of both pollen and urban pollution on humans is very significant.

The climate of the study area (i.e. Greater Montreal) is semi-continental, with a warm, humid summer and a very cold winter (Dfb in the Köppen climate classification). It has a sunny climate in spring during the birch pollen season (mid-April/May) which is ideal for pollen dispersion with intermittent low pressure and frontal systems.

2.2. Vegetation distribution

A high-resolution (1 km X 1 km) distribution of birch (*Betula*) trees inventory covering North America has been obtained from EPA (Environmental Protection Agency). This inventory is based on the **Biogenic Emissions Landuse Database**, version 3 (BELD3 of US/EPA). Kinnee et al. (1997) provided data for distributions of 230 vegetation classes over most of North America and included in BELD3. More details concerning the vegetation file and the source of data are given in Supplementary material S3. Note that the distribution area of birch extends over most of Canada and some parts of northeast US (Figure 1A) but the three major source regions likely to affect Montreal (Figure 1B) are all located within few hundred kilometres: the Quebec Laurentides (to the NW of Montreal), Eastern Townships (to the east) and New England forests in the USA (to the SE).

Figure 1A,B here.

2.3. Emission modelling

In this study, we extended the parametrization of Helbig et al. (2004) developed for alder pollen emission to the case of birch. According to numerous authors (Laaidi et al. 1997; Helbig et al. 2004; Dahl et al. 2013; Zink et al. 2013), meteorological factors have the most impact in controlling anther dehiscence and aerosolization. Relatively high temperature, low humidity and moderate wind speed favour a passive dehydration that leads to the bursting of the anthers and dehiscence (Helbig et al., 2004; Dahl et al., 2013). We have followed this trend but adopted a specific hypothesis that slightly diverts from the general tendency (i.e. Helbig et al., 2004; Sofiev et al., 2013a; Zhang et al. 2014) and follows the idea that the pollen emission does not follow a smooth diurnal profile. Hence, only a pulse of emission in the morning (pulse of pollen at 12Z²) was allowed. This should reflect the nature of puff release rather than continuous emission over a 24-hour window (as pointed by Jackson and Lyford, 1999) and is also consistent with Zink et al. (2013) who suggest that there is a limited amount of pollen available in the anthers and after pollen emission takes place (usually in the morning), the pollen emission should be turned off for the rest of the day. This is also supported by observations that suggest that pollen emission takes place mostly in the early morning (Bianchi et al., 1959 for ragweed; Dahl et al., 2013 for tree pollen) since at that period decreasing humidity and increasing temperature most likely drive dehiscence and increasing wind speed after sunrise produces pollen detachment.

2.3.1. Phenology model

The flowering phenological season of birch is short and highly peaked (Dahl et al., 2013). Moreover, phenology models are likely to apply over a limited region. In this study, we have taken the input phenology parameters based on the following approach. For the length of the season, a combination of persistence and calendar's method was shown to give the best results whereas for the start of the season a multiple regression model was

² Z indicates Zulu time and is identical to UTC (Universal Coordinated Time). This convention (Z) is used in meteorological models such as GEM and other models and among NATO countries. To convert Z time to local time in Montreal, one has to subtract 4 hours, i.e. 12Z corresponds to 8AM.

used to predict input phenology parameters (see Robichaud and Comtois, 2017 for more details).

2.3.2 Flux of pollen release

Over a model grid cell, the total pollen emission flux obeys the following equation (Helbig et al., 2004; Siljamo, 2013):

$$\text{Emission} = F_e(t, i, j) \times SA(i, j) \times D(i, j) \quad (2)$$

where $F_e(t, i, j)$, $SA(i, j)$ and $D(i, j)$ are, respectively, the emission flux per unit area per amount of time, the surface area of a model grid cell (about 100 km² for GEM-MACH) and the birch coverage density (a fraction between 0 and 1; see Figure 1A,B). In that formulation, F_e is evaluated at the top of vegetation and is proportional to the product of a characteristic concentration (c^*) and a characteristic velocity (u^* , i.e. friction velocity):

$$F_e(t, i, j) = C_e(t) \times c^* \times K_e(u^*, T, RH) \times u^* \quad (3)$$

where u^* is given by the GEM model averaged over a tile (10 km X 10 km). $C_e(t)$ and K_e are, respectively, the probability of emission of etamines and a meteorological switch defined below. In Helbig et al. (2004), the value of C_e is given as a function of time:

$$\begin{aligned} C_e(t) &= 4 \text{ E-}4 \times (d/S - d^2/S^2) \quad \text{during the pollen season} \\ C_e &= C_o \quad \text{before and after the season} \end{aligned} \quad (4)$$

where d , S are, respectively, the number of days since the beginning of the pollen season and the length of the season. C_o is a very small value related to the pollen remote background. Note that the pollen season is defined here by 2.5 - 97.5% of pollen sum over the whole period (as in Robichaud and Comtois, 2017). The parameter c^* is defined as:

$$c^* = Q_p / (LAI \times h_c) \quad (5)$$

where LAI, Q_p and h_c are, respectively, the mean leaf area index, the total yearly pollen quantity (pollen grains /m²) and the canopy height. The product $C_e(t) \times c^*$ in Equation 3 is assumed constant here over the domain and is a symmetrical function which only depends on the day of the season (see Supplementary material S4, Figure S1A) or could be an asymmetrical function (see Figure S1B) in accordance to the so-called *open-pocket* principle (Sofiev et al., 2013a). The latter profile is adopted to avoid overestimation of the pollen emission at the end of the season as reported by Kawashima and Takahashi (1999) and Efstathiou et al. (2011); i.e. the bloom probability is forced here to decrease twice as fast in the second half of the season (after the anthesis point, i.e. maximum emission). Finally, the parameter K_e is a non-dimensional switch that is defined in Equation 6 (see below). As in Helbig et al. (2004), the switch K_e depends on a threshold u^*_{th} , temperature, relative humidity and surface wind velocity:

$$\begin{aligned} K_e &= (1 - u^*_{th}/u^*) \text{ if } u^* > u^*_{th} \\ K_e &= 0 \text{ otherwise} \end{aligned} \quad (6)$$

with the threshold value $u^*_{th} = u^*_t \alpha$. The value of u^*_t is taken as 0.323 (following Shao and Lu, 2000) and the expression for α is given below:

$$\alpha = 3 / (\alpha_v + \alpha_T + \alpha_D) \quad (7)$$

with the threshold wind, temperature and dew point temperature, respectively are:

$$\alpha_v = V/V_{th}, \alpha_T = T/T_{th}, \alpha_D = D/D_{th}.$$

Table S2 (Supplementary material S4) summarizes the calibrated parameters adopted for the simulation and the threshold values for meteorological parameters as defined in Equations 1-7. Note that no further modification to the core of the formulation of Helbig et al. (2004) was attempted mostly because of great uncertainties of individual values and parameters. Moreover, sensitivity tests have confirmed that most of the parameters taken from Helbig et al. (2004) for alder are appropriate for birch except for the threshold temperature which is taken higher for birch (11 °C was taken instead of 8 °C, the former value reducing the bias and RMSE when compared to observations, see details in Supplementary material S5 for the validation metrics). The pollen pool (Q_p) was taken as

2.0 billion per square meter which is a considered a typical value for birch (Schueler and Schlünzen, 2006; Helbig et al., 2004). However, the latter has high uncertainty as shown by Zhang et al. (2014) and pointed out by Siljamo (2013). Resuspension is ignored here since it is at least one order of magnitude less than the emission flux (Helbig et al., 2004). It is also considered as a rare phenomenon (occurs whenever the wind speed > 15 m/s) and inclusion into the simulation code is not expected to give any gain given large uncertainties of other factors (Laaidi et al., 1997; Sofiev et al., 2013a,b). At night, the high humidity, low temperature and low wind speed are the factors reducing the emissions (Sofiev et al., 2013a). Recent lidar observations confirm that tree pollen concentrations in the air at night are low (Noh et al., 2013). Therefore, it is reasonable to assume no pollen emission at night and a peak emission as a puff release in the morning as done here. Following Sofiev et al. (2013a), when the precipitation > 0.5 mm/h or the relative humidity > 80 % (corresponding to $T - T_d < 5^\circ\text{C}$, where T_d is the dew-point temperature) no pollen emission takes place (prohibitive thresholds). Pollen emission is then allowed as puff emission at 12UTC (8AM local time) as the model run is restarted for a new cycle of 24 hours forecast.

Note that solar radiation might have an impact on pollen emission, but we believe that this variable is already taken into account indirectly through meteorological variables sensitive to sunshine such as temperature, u^* , relative humidity, etc. Therefore, to avoid redundancy or over-modelling it was decided that this variable will not explicitly appear in Eqs 3-7 (due to collinearity with several meteorological variables). This is line with numerous authors who do not include solar radiation into the pollen emission modelling (Helbig et al., 2004; Damialis et al. 2017).

2.4 Source apportionment

One of the goal of this study is to identify the major source regions of birch pollen affecting Montreal and the rest of the domain. To achieve this, pollen emissions in each region are shut-off except for a given sub-region from which the contribution is evaluated (see Figure 2). The average pollen load is then obtained in Montreal during the simulation period (April 11-May 31st 2012) performed for each sub-regions. More details about the methodology are given in Supplementary material S6.

Figure 2 here.

2.5 Observations and methods of validation

Validation of model pollen concentration will be made using observations of birch pollen collected at Université de Montréal (UdeM). The collecting instrument is a volumetric pollen sampler based on Hirst design (Hirst, 1952). The instrument is located about 20 m above ground level on the roof of Université de Montréal (Geography Department). Pollen measurements are also made by Aerobiological Research but only the forecast (based on the measurements) could be accessed to the public. The forecasts (in units of pollen counts per cubic meter) have been downloaded daily from their public website³ during spring 2012 for Montreal, Sherbrooke, Quebec City, Lac St-Jean (in the province of Québec), and Ottawa, Kingston, Barrie, Toronto and London (in the province of Ontario, Canada). ROTOROD[®] measurements and forecast meteorological parameters provide the basic inputs for the statistical forecast delivered by Weather Network (INSPQ, 2013). More details are given in Supplementary material S7. Figure S2 shows the location of the sites where WXN forecasts based on observations are available and were used in this study (i.e. all sites except MSM and MST).

2.6 Other considerations

Hypothesis about how the surface pollen emission is converted into near-surface concentration and vertically transported in the boundary layer is important and discussed in Supplementary material S8 and S9. The concept of vertical correlation length used in the domain of data assimilation has been utilized here (see Robichaud, 2017) to circumvent numerous problems linked with boundary layer schemes of chemical transport models (Lin and McElroy, 2010; Bosveld et al., 2014).

³ <https://www.theweathernetwork.com/outdoors/pollen/canada/>

3. Numerical simulation of airborne birch pollen

3.1 Emission flux

Using Equations 1-7 and the gridded vegetation density for birch (Figure 1), an average gridded pollen emission flux F_e (gr/m²/s) for May 2012 is shown at 12Z in Figure 3. It shows the major source regions for *Betula* pollen emission: 1) to the north of Montreal, that is the Laurentide Mountains (L), 2) to the east and southeast, Eastern Townships (E) and New-England (NE), respectively (Figure 3B).

Figure 3A,B here.

3.2 Time series of simulated pollen

Time series of simulated pollen concentration obtained near the surface for the period April 11 through May 31st of 2012 are presented in Figure 4. The model simulation time series is compared with observation (UdeM site) and with a statistical public forecast produced by WXN for Montreal (hereafter WXN-YUL). A moving average was applied to observations, WXN-YUL forecasts and model simulations using three different time frames, A: average for the current day of simulation (lag0), B: average of today's simulation and previous day simulation (lag0 and lag1), and C: average of today and the two previous days (lag0, lag1, lag2). The choice of 3 days moving average is based on the fact that temporal resolution below 3 days is difficult to achieve given uncertainties about phenology and meteorology (Robichaud and Comtois, 2017).

The three averaging scenarios are shown respectively in Figure 4A, B and C. In Figure 4A, the agreement is satisfactory (peaks coincide within three days; the amplitude is within a factor of two). Mismatch in the synchronicity of pollen peaks between model and observation could be due to temporal misalignment between modelled pollen phenology and observed concentrations. For example, the presence of asynchronous flowering of the different trees or species at the local scale within the study area or elsewhere is not reproduced by the model simulation which considers synchronous flowering (everywhere at all locations). Another reason for mismatch is that the

simulated pollen model plume can slightly miss the receptor point corresponding to the observation location or be temporally delayed or being advected too fast by the model. For the case of two days moving average window (Figure 4B), some noise and uncertainties relative to the synchronicity are removed and the simulated pollen time series versus observation and WXN-YUL statistical forecast are more similar. With three days moving average, the independent sources of information (model simulation, observation and WXN statistical forecast) almost coincide (Figure 4C) showing that the model captures well the two observed pollen episodes within ± 1 day. Spearman's coefficient of correlation is $R=0.68$ between average model and observation and $R=0.73$ between average model and WXN-YUL. When the daily temperature is superimposed in Figure 4 (dashed curve with squares), a relation seems to be present with pollen major peaks: whenever the observed pollen episodes occurred in Montreal (i.e. mean daily > 80 grains/ m^3), the mean daily temperature was in a range 10° to $18^\circ C$ (the range is indicated by horizontal lines in Figure 4). A similar observation (i.e. narrow range of temperature) was also noted in the study of Kawashima and Takahasi (1995) for Japanese cedar and Ant  paria et al. (1995) for grass pollen in European Atlantic Seaboard. The latter authors obtained a range of $15.7^\circ - 21.7^\circ C$ (i.e. $18.7 \pm 3^\circ C$) in Spain for the presence of grass pollen episodes (i.e. > 50 grains / m^3). This information is very useful for atopic patients which can better protect themselves by avoiding exposure whenever ambient temperatures are within these ranges. In presence of precipitation (indicated by a dot at the top of Figure 4), the model's value is attenuated to different degree as expected (washout of pollen by precipitation).

Figure 4A,B here.

3.3 Vertical cross section of simulated pollen

Vertical cross section outputs help to validate the three-dimensional model simulation performance (Figure 5). The cross-section X-Y-Z for May 13th at 18 UTC shows that the pollen cloud vertically extends up to cover the whole boundary layer in both cases (i.e. up to $HY = 0.8$ hybrid vertical coordinate or $\sim 2-3$ km altitude). The point Y corresponds to

the location of Montreal in Figure 5A and to Québec City in Figure 5B. The highest pollen concentrations lie within the lowest part of the atmospheric boundary layer (HY=0.90-1.0). At Quebec City (Figure 5B), near surface hourly values exceeded 1000 grains per cubic meter (one-hour average) on May 13th at 18UTC when the “pollen cloud” passes over the city. Overall, these results agree reasonably well with the literature which reports that pollen grains that are released during warm sunny days are often well mixed throughout the whole atmospheric boundary layer (hereafter ABL) which reaches about 1-2 km depth during a typical sunny afternoon (Raynor et al., 1970; Gregory, 1973; 1978; Gruber et al., 1998; Mandrioli et al., 1984; Helbig et al., 2004). However, Kawashima and Takahashi (1995) reported that more than 50% of the pollen grains of Japanese cedar were observed at a height of 300 meters or lower (height 0.95-1.0 HY, Figure 5). This suggests that most of the cedar pollen is transported in the lower part of the convective boundary layer whereas the more buoyant (less dense) birch pollen could extend to higher altitude. Comtois et al. (2000) found observational evidence for a layer of pollen transport at about 500 m above the ground with pollen concentrations higher by 30 % compared to ground level. Mandrioli et al. (1984) show vertical profile of pollen (their Figure 8) which corresponds to the decrease shown in our Figure 5 (that is, pollen concentration is maximum near the surface and decreases to few grains/m³ about two km of altitude above ground level). More recently, Noh et al. (2013) observed by lidar (light detection and ranging) that daily tree pollen is lifted to 1-2 km of altitude consistent with Figure 5. Vertical dispersion of up to 1-2 km as shown in our simulations are also consistent with TDMLD of up to 1000 km as reported in Europe for *Betula* (Sofiev et al. 2006a,b). For example, a pollen particle at altitude of 1.5 km under a sedimentation velocity of 1.3 cm/s would be airborne for about 32 hours. Under a moderate wind speed of 30 km/h, the pollen would then be transported to about 1000 km in agreement with the literature. Nevertheless, more measurements of pollen vertical profile such as done in Europe (e.g. the MicroPoem initiative, <http://edoc.unibas.ch/31204> or Damialis et al. 2017) are needed in the future to understand the flow of biological aerosols but also to better study the behaviour of the complex motion within the boundary layer.

Figure 5 here.

3.4 Simulated pollen diurnal profile

Knowledge of diurnal variations can help atopic patients to avoid exposure to allergenic pollen during critical hours of the day. Simulated averaged diurnal profiles are shown in Figure 6 for the study period (April 11th - May 31st May 2012) for different locations in the study area (see location of the sites indicated by a triangle in Figure S2) and for Montreal. To make the analysis easier, we must distinguish first here between two types of sites: a source site (located in a zone of high density of *Betula* vegetation, i.e. a major emission area source) and a receptor site (located away from a zone of high density of *Betula*, i.e. low emission area source). In Figure 6, as expected, a strong peak in the morning (around 8AM, local time) reveals the signature of source sites (i.e. Mont-St-Michel, Mont-Sutton, and Sherbrooke). For sites not located within the source areas, (i.e. receptor sites: Montreal, Ottawa and Quebec), there is a time lag necessary to reach maximum pollen concentration later in the day (i.e. afternoon). This lag is due to the transit time from source areas to receptor areas associated with TDMLD so that concentrations build up in the morning and reach a plateau in the afternoon at the time of maximum mixing and height of the ABL for the latter sites. Notice that nocturnal concentrations are minimum due to absence of pollen release during that period.

Afternoon pollen maximum concentrations are common in urban centres (which usually are not emission regions). Peaks between 14-16 UTC (2 to 4 PM local time) in North London were found by Skjøth et al. (2009). On the other hand, the minimum value found in Neuchâtel (Switzerland) was between 4 and 8 AM (local time) according to a study by Clot (2001). Similarly, Mahura et al. (2007; 2009) show that the diurnal profile of *Betula* has a minimum value in early morning hours and start to increase after that time at two different receptors sites in Denmark. Finally, diurnal variation of total airborne pollen in Taipei City (1993 and 1994) also has a very similar profile to what is shown in Figure 6. The presented simulation also seems to reproduce a similar pattern with a minimum shortly before 7 AM (local time) and increasing values after the following hour or so for receptor sites (MTL, YQB, YOW) and a peak value in the morning for source sites (MSM, MTS and YSC). Therefore, the hypothesis of a puff release at 12Z (8AM local time in Montreal) made here is supported by observations in the literature.

Moreover, boundary layer processes and their representation in models have a central role in shaping the vertical and spatial distribution of tracers in the atmosphere (Lin et al., 2008; Bosveld et al., 2014).

Figure 6 here.

4 Numerical validation

In section 3, model simulation outputs were qualitatively compared with observations, WXN statistical forecasts or results from the literature. In this section, numerical metrics are used to examine in greater details the performance of the simulation. The observations from UdeM and statistical forecasts (WXN) for birch pollen concentration will be used in this section.

4.1 Numerical comparison of pollen counts with observations

We present a summary of the results in Table 1 (see Suppl. Material S5 for definition of the metrics). The model simulation has the lowest mean bias (OmP= -9.9) and the highest Pearson's coefficient correlation (model avg lag0-2; $R=0.63$). However, overall, the statistical forecast provided by WXN-YUL (avg lag0-2) is the most accurate for several metrics: highest Spearman's correlation coefficient ($R=0.60$), the lowest RMSE (49.2), lowest MAE (29.3%), the lowest false alarm ratio (FAR80=0.02) and highest percent correct (PC80=0.94). By comparison, a pollen forecast based on persistence, LTAA (Long term average of aerobiological measurements) or a combination of both (LTAA/persistence) provide less overall accuracy than that of the model simulation or compared to a WNX-YUL forecast. Although WXN-YUL provides more accuracy it is defined at only one point whereas the model is defined everywhere in the domain which represents a huge advantage.

Note that the lag 0-2 metric (i.e. mean of day 0, day -1 and day-2) adopted here filters the noise due to natural variations linked with phenology and meteorology (i.e. equivalent to 3 days moving average). The choice of a 3-day moving average is based on the following: 1) it has been shown that phenology models for birch pollen are precise within

a timescale of 2-3 days (Siljamo 2013), and, 2) 3 days is the approximate time for meteorological turnover of weather systems in Canada.

Table 1 here.

4.2 Representativeness of the phenology model

In this section, we evaluate the spatial representativeness of the phenology model (based on Robichaud and Comtois, 2017) by comparing model simulations with several statistical forecasts provided by Weather Network. Figure 7 presents time series of both model simulation and WXN statistical forecast at various points. The comparisons are made at eight sites (see location of the sites in Figure S2, Supplementary material S7). These sites were selected because a daily forecast was available from WXN. Table S3 (Supplementary material S7) presents the Spearman's coefficient of correlation between model simulation and WXN forecast for each site. Results of the table are plotted in Figure 8 as a form of variogram (R versus distance) shows a linear decrease of the correlation between model simulation versus statistical forecast with distance from Montreal. Figure 8 suggests that beyond a certain distance (about 400-500 km) the model simulation and Weather Network forecasts diverge and, consequently, there is no longer coincidence between the two sources of information ($R < 0.5$ between model and WXN time series). This representativeness distance found is surprisingly large giving the various climatic diversity and the climatic gradients within the region. Note that the lack of representativeness ($R < 0.5$) for larger scales (> 500 km) does not affect the quality of the simulation presented here since the major area sources (Laurentides, Eastern Townships and New England) are mostly located within 400-500 km of Montreal. However, it should be noted that this limits the extension of the model simulation to other regions of Canada and North America (located beyond 500 km of Montreal). For example, sites in Ontario (Toronto, Barrie and London) tend to have their pollen season starting earlier than that in the model simulation (see Figures 7F, G, H) and consequently the model simulation based on phenology developed for Montreal will significantly lag in time as compared with the starting date for those sites in Ontario. To extend the

simulations to other regions (e.g. the whole North America), pollen data is needed to calibrate the phenology models that would apply to those regions.

The result obtained for the area of representativeness found here (if we consider R^2 as a measure of spatial coherence) is quite compatible with results of Ranta et al. (2011) who showed that the coefficient of correlation of the inter-annual variability between sites for the *Betula* pollen in Finland is over 0.5 for most sites lying within 400-500 km of distance. Note that the WXN forecasts are slightly delayed compared to the model simulation for Sherbrooke (YSC), Ottawa (YOW), Quebec (YQB) and Kingston (YKG) in Figure 7. The same observation holds for Montreal as well (Figure 4). Finally, WXN forecasts also seem to systematically under-forecast when compared to model or observations.

Figure 7 here.

Figure 8 here.

5. Contribution of major source regions (source apportionment)

It is important to know for pollen sufferers, the spatial origin of the allergenic pollen. Using the methodology described in Section 2.4, it is possible to know from each major region (as depicted in Figure 2) what is the percentage of pollen from each sub-region contributing to the pollen load in Montreal. Table 2 shows the results of different numerical experiments (allowing only one sub-region at the time to emit pollen while the other are shut-off). According to these source apportionment experiments, the regions which contribute the most to the pollen load in Montreal are: the Southern Laurentides (region 4 in Figure 2) with 55% of the total load and the Northern Laurentides, with 33% (region 3). Note that the local pollen (Montreal Island, region 6) contributes to less than 1% (0.6%) of the load indicating that the pollen is associated with TDMLD rather than having a local origin. Adding 2.8% load originating from the Appalachian mountain (region 2), means that more than 90% of the pollen load in Montreal comes from sources within 400-500 km of distance under TDMLD.

Table 2 here.

6. Discussion and conclusion

Overall, the model simulation presented validates satisfactorily with data. Figure 4 suggests that the model reproduces the two major observed pollen peaks (mean daily concentration > 80 grains/m³) reasonably well (using a 3-day average, i.e. lag0-2). Compared to other authors who did similar studies in North America (Pasken and Pietrowicz, 2005; Efsthathiou et al., 2011; Zhang et al., 2014) our tree pollen simulation (for sites within 500 km of Montreal) seems to be much closer to observations than that in the above studies. For example, the overall model performance obtained here is significantly superior to that obtained for *Betula pendula* in California in a similar numerical modelling study (Zhang et al. 2014). Our results (Figure 4) show less misalignment of the time series (see Zhang et al., their Figure 8a) and smaller difference with observations with respect to the maximum pollen value (their Figure 9a). Similarly, in a numerical simulation of birch over Northeast US (Efsthathiou et al., 2011), the CMAQ model simulation shows a pollen peak about 4.5 times higher than observations for birch (their Figure 12) compared to an overestimation of less than 2 in our study (Figure 4A). We believe that using puff release (morning window of pulse of emission at 12UTC) instead of continuous release (all day emission window) as well as using the concept of vertical correlation length to mimic full vertical mixing hypothesis (as done in Robichaud, 2017) attenuate the deficiency of vertical transport of tracers within GEM-MACH and related boundary layer problems leading to a successful simulation in our case. These two verified hypotheses here are considered an added value to the scientific knowledge. By comparison of the results, birch pollen simulations made by Zink et al. (2013) show an average coefficient of determination (r^2) of 0.435 with observations (using a new emission module called EMPOL). This value for r^2 is similar to the coefficient of determination obtained here by comparison of statistical forecast of Weather Network (if we consider the forecast as a proxy for observation, which is a

reasonable hypothesis here in absence of other information). In fact, results presented in Table S3 (for sites within 500 km from Montreal), shows an average r^2 of 0.44.

Note that the presence of model simulated secondary peaks near Julian day 120-121 and near day 138-139 are not present in the observation (UdeM) and not synchronous either with the statistical forecast produced by Weather Network for Montreal. As found in several studies, numerical or statistical models have weak skills to predict the beginning and end of the season (Sofiev et al., 2013a,b) and further research is needed.

6.1 Evaluation of the uncertainty

Despite reasonable results shown in this research, large uncertainties remain associated with pollen emission, phenology, boundary conditions, transport inaccuracies (wind speed and direction), atmospheric instability, solar radiation, characteristics of the boundary layer, precipitation and dry and wet deposition. Other possible sources of errors are the following: vegetation misrepresentation, small pollination window, phenology heterogeneity among different birch variety or even different trees within the same species, north-south and altitude gradients of phenology not considered here, inadequate spatial information in current land cover databases in North America.

Uncertainties associated with the pollen emission module are also important (Duhl et al., 2013; Sofiev et al., 2013a,b) and include pollen emission probability (likelihood to bloom, i.e. parameter C_e in Equation 3) and the pollen pool Q_p parameter (Zhang et al., 2014). It was found in preliminary results that a symmetrical curve with respect to the anthesis point (Figure S1A) produces over-forecast in the second half of the season whereas an *open-pocket* scenario (Figure S1B) would give better results in the second half of the season. However, uncertainty remains concerning this issue and more research is needed.

Other sources of errors include vegetation distribution maps (Sofiev et al., 2006a,b; Efstathiou et al., 2011; Sofiev et al., 2013a) and forest canopy heterogeneity (Lefsky, 2010). For example, it is important to note that pollen production may vary by at least an order of magnitude between species of a given genus (Duhl et al., 2013). Note

that according to Siljamo (2013), Q_p is very uncertain, varying with latitude and altitude and quite variable from one year to another (Ranta et al., 2008) which further complicates the issue.

The pollen size distribution (we have considered only one size, i.e. 22 microns diameter for all pollen grains), pollen re-suspension processes, rebound as well as re-entrainment, have not been treated in this study. Finally, uncertainties of input meteorological factors (winds, temperature and precipitation) embedded in the NWP and chemical transport models are also significant. Sofiev et al. (2006a,b) found that the numerical meteorological input data set was the most important factor for *Betula* pollen numerical model performance. For example, the surface (2 m above the ground) temperature, humidity or 10 m wind speed are all used in the emission model whereas the canopy is located between 6-30 meters above the ground. Moreover, small errors in the model wind direction can make a large difference over a long transport distance. Microclimate exists especially in mountainous areas where most birch pollen sources lie (e.g. in Laurentian or Appalachian Mountains) but the meteorological model cannot forecast correctly at the local and micro-scale (less than few km) leading to high uncertainty on the sub-grid scale especially over irregular topography (due to insufficient lack of model resolution). Again, we believe that the use of the vertical correlation concept (see Suppl. Material S8 and S9) is likely to be the best approach in the context here. More observation studies about the horizontal and vertical pollen distribution in different locations around the world should be undergone (such as Damialis et al. 2017 for the Mediterranean region or Comtois et al. 2000 for Spain) to better validate numerical models.

The concept of pollen emission as a “puff release” adopted here considers that the entire mass of pollen is emitted during a small morning window. However, the atmospheric pollen dispersion takes place at any moment and does not “rule-out” or shut-off to zero in any circumstance pollen concentration in our simulation platform. Pollen is allowed to be dispersed at any time of the day and long-range transport is effective at any time, day or night, as well. The morning puff release is a reasonable assumption since it is known that pollen is not emitted at night although at times significant concentrations of pollen can be observed during that period. Prior experience done with this model simulation platform

allowing emission at any time of the day turned out to give poor results (results not shown). In that regard, botanists or plant experts recommend allowing pollen emission at one time of the day (more likely in the morning) and thereafter shut-off emission for the rest of the day (Zink et al. 2013).

Only one season was examined here as a preliminary study with the goal of developing a prototype. Future work will apply this prototype presented here to other years, regions and will integrate it to more recent version of the GEM-MACH. In particular, the pollen pool interannual variation and spatio-temporal variation should be examined more closely to achieve that goal. However, as a first step towards the feasibility of numerical modelling of tree pollen in Canada, we consider that the initial goal of demonstrating a proof of the concept was successfully achieved in here.

The main conclusions of this research are:

- 1) given many sources of uncertainties, the numerical simulations shown reproduce relatively well pollen observations in Montreal and agree with the statistical forecasts produced by Weather Network (Figure 4C; Tables 1 and Table S3),
- 2) vertical profiles of simulated pollen concentrations are realistic and agree with literature observations (Figure 5),
- 3) simulated diurnal profiles are realistic reproducing morning peaks in major source regions and afternoon maxima in receptor regions (Figure 6),
- 4) the spatial representativeness of the phenology corresponds to a distance of about 500 km (Figures 7 and 8; Table S3),
- 5) a method to evaluate the contribution of different major source areas of pollen suggests that the Laurentide mountain regions provided 88% of the pollen load in Montreal in 2012 and that the local sources provide less than 1% (Table 2),
- 6) a video of the simulation shows the complexity of motions of the “pollen clouds” and reveals the sources of *Betula* pollen in Montreal (video link in Supplementary material S10) and the effect of medium and long range transport of this bioaerosol
- 7) This prototype could be included in the NWP suite to forecast birch pollen concentration for Canada

- 8) The methodology and techniques described in this paper could be extended anywhere and for other tree pollen species as long as the phenology parameters and pollen data are available for the chosen location. Developing simulation for larger regions such as North America is left as future work.

Acknowledgments. The main author wishes to thank internal reviewers at Environment and Climate Change Canada (Janusz Pudykiewicz, Kirill Semeniuk and Rodrigo Muñoz-Alpizar) and François Courchesne from Université de Montréal for their in depth revision of the article. Thanks are also given to the two anonymous reviewers of this journal.

7. References

- Bosveld et al., 2014 F.C. Bosveld, P. Baas, G.J., Steeneveld, A. Holtslag, W.M. Angevine, E. Bazile, I.F. Evert, , A. de Bruijn, D. Deacu, J.M. Edwards, M. Ek, V.E. Larson, J.E. Pleim, M. Raschendorfer, G. Svensson **The Third GABLS Intercomparison Case for Evaluation Studies of Boundary-Layer Models. Part A: Case Selection and Set-Up** Boundary-Layer Meteorol, doi 10.1007/s10546-014-9917-3
- Campbell et al., 1999 I.D. Campbell, K. McDonald, M.D. Flannigan, J. Kringayark **Long-distance transport of pollen into the Arctic.** Nature 399 29-30
- Chamecki et al. 2009 M. Chamecki, C. Meneveau, M.B. Parlange **Large eddy simulation of pollen transport in the atmospheric boundary layer** Aerosol Science 40:241-255, doi:10.1016/j.jaerosci.2008.11.004
- Clot, 2001 B. Clot **Airborne birch pollen in Neuchatel (Switzerland): onset, peak and daily patterns.** Aerobiologia 17:25-29
- Comtois et al., 2000 P. Comtois, D. Fernández-Gonzales, R.Valencia-Barrera, J. Sánchez, R. Fraile,S. Rodier **Pollen content study of the lower atmosphere in León (Spain) by use of a tethered balloon** Aerobiologia 16, 187–191
- Dahl et al., 2013 A. Dahl, C. Gálan, L. Hajkova, A. Pauling,B. Sikoparija, M. Smith, D. Vokou . **The onset, course and intensity of the pollen season.** In: Allergenic Pollen. A Review of the Production, Release, Distribution and Health Impacts. Chap. 3. Sofiev, Mikhail, Bergmann, Karl-Christian (Eds.), doi 10.1007/978-94-007-4881-16
- Dales et al., 2004 R.E. Dales, S. Cakmak, S. Judek, T. Dann, F. Coates, J.R. Brook, R.T. Burnett **Influence of outdoor aeroallergens on hospitalization for asthma in**

- Canada.** Environmental and occupational disorders, doi:10.1016/j.jaci.2003.11.016
- Dales et al. 2008, R.E. Dales, S. Cakmak, S. Judek, F. Coates **Tree Pollen and Hospitalization for Asthma in urban Canada** Int. Arch. Allergy Immunol. 146:241–247, doi:10.1159/000116360
- D’Amato et al., 2001 G. D’Amato, M. Liccardi, M. Cazzola **The role of outdoor air pollution and climatic changes on the rising trends in respiratory allergy** Respiratory Medicine 95:606-611
- D’Amato et al. 2007 G. D’Amato, L. Cecchi, S. Bonini, C. Nunes, I. Annesi-Maesano, H. Behrendt, G. Liccardi, T. Popov, P. van Cauwenber **Review article: Allergenic pollen and pollen allergy in Europe** Allergy 62:976-990, doi:10.1111/j.1398-9995.2007.01393.x
- D’Amato et al. 2015 G. D’Amato, S.T. Holgate, R. Pawankar, D.K. Ledford, L. Cecchi, M. Al-Ahmad, F. Al-Enezi, S. Al-Muhsen, I. Ansotegui, C.E. Baena-Cagnani, D.J. Baker, H. Bayram, K.C. Bergmann, L.P. Boulet, J.T. Buters, M. D’Amato, S. Dorsano, J. Douwes, S.E. Finlay, D. Garrasi, M. Gómez, T. Haahtela, R. Halwani, Y. Hassani, B. Mahboub, G. Marks, P. Michelozzi, M. Montagni, C. Nunes, J.J. Oh, T.A. Popov, J. Portnoy, E. Ridolo, N. Rosário, M. Rottem, M. Sánchez-Borges, E. Sibanda, J.J. Sienra-Monge, C. Vitale, I. Annesi-Maesano **Meteorological conditions, climate change, new emerging factors, and asthma and related allergic disorders. A statement of the World Allergy Organization** World Allergy Organization Journal 8:25, doi:10.1186/s40413-015-0073-0
- Damialis et al. 2017, A. Damialis, E. Kaimakamis, M. Konoglo **Estimating the abundance of airborne pollen and fungal spores at variable elevations using an aircraft: how high can they fly?** Sci. Rep. 7, 44535. doi:10.1038/srep44535
- Duhl et al. 2013 T.R. Duhl, R. Zhang, A. Guenther, S.H. Chun, M.T. Salam, J.M. House, R.C. Flagan, E.L. Avol, F.D. Gilliland, B.K. Lamb, T.M. Van Reken, Y. Zhang, E. Salathé **The Simulation of the Timing and Magnitude of Pollen Season (STaMPS) model: a pollen production model for regional emission and transport modeling** Geosci. Model Dev. Discuss. 6:2325-2368, doi:10.5194/gmdd-6-2325-2013
- Efstathiou et al., 2011 C. Efstathiou, S. Isukapalli, P. Georgopoulos **A mechanistic modeling system for estimating large scale emissions and transport of pollen and co-allergens** Atmos. Environ. 45(13):2260-2276, doi:10.1016/j.atmosenv.2010.12.008
- EPA (US Environmental Protection Agency) 2008 **Review of the impacts of climate variability and change on aeroallergens and their associated effects.** Report EPA/600/R-06/164F. Available at www.epa.gov. Last accessed January 31st 2017
- Gregory, 1973 H.P. Gregory **The microbiology of the atmosphere** Second edition. John Wiley & Sons, New-York, Toronto
- Gregory, 1978 H.P. Gregory. **Distribution of airborne pollen and spores and their long-distance transport** Pure and Applied Geophysics 116, 309-315

- Gruber et al., 1998 S. Gruber, S. Matthias-Maser, J. Brinkmann, R. Jaenicke **Vertical distribution of biological aerosol particles above the North Sea** J. Aerosol Science, 29, S771-772
- Guérin, 1993 B. Guérin **Pollen et allergies** Édition Allerbio, Varennes-en-Argonne, 279 pages
- Helbig et al. 2004, N. Helbig, B. Vogel, H. Vogel, F. Fiedler **Numerical modelling of pollen dispersion on the regional scale** Aerobiologia 3:3-19
- Hirst, 1952 J.M. Hirst **An automatic volumetric spore trap** Annals of Applied Biology 39:257-265
- Institut national de santé publique du Québec (INSPQ) 2013 **États des connaissances sur le pollen et les allergies. Les assises pour une gestion efficace** ISBN 978-2-550-68403-9 (available at <http://www.inspq.qc.ca>)
- Jackson and Lyford, 1999 S.T., Jackson, M.E. Lyford **Pollen dispersal models in Quaternary plant ecology: assumptions, parameters, and prescriptions** The Botanical Review 65:39-75
- Kawashima and Takahashi, Y. 1995 S. Kawashima, Y. Takahashi **Modelling and simulation of mesoscale dispersion processes for airborne cedar pollen** Grana 34:142-150
- Kawashima and Takahashi 1999- S. Kawashima, Y. Takahashi **An improved simulation of mesoscale dispersion of airborne cedar pollen using a flowering-time map** Grana 38:316-324
- Laaidi et al., 1997 K. Laaidi, M. Laaidi, J.P. Besancenot **Pollens, pollinoses et météorologie** La météorologie 20(8):41-55
- Lin and McElroy, 2010 J.T. Lin, M.B. McElroy **Impacts of boundary layer mixing on pollutant vertical profiles in the lower troposphere: implications to satellite remote sensing** Atmospheric Environment 44:1726-1739, doi: 10.1016/j.atmosenv.2010.02.009
- Lefsky, 2010 M.A. Lefsky **A global forest canopy height map from the Moderate Resolution Imaging Spectroradiometer and the Geoscience Laser Altimeter System** Geophysical Research Letters 37: L15401, doi:10.1029/2010GL043622, 2010
- Mahura et al. 2009 A.G. Mahura, A. Baklanov, U. Korsholm **Parameterization of the birch pollen diurnal cycle** Aerobiologia 25:203-208, doi 10.1007/s10453-009-9125-7
- Mahura et al., 2007 A.G. Mahura, U. Korsholm, A. Baklanov, A. Rasmussen, **Elevated birch pollen episodes in Denmark: contributions from remote sources** Aerobiologia 23:171-179, doi 10.1007/s10453-007-9061-3

- Mandrioli et al. 1984 P. Mandrioli, M.G. Negrini G. Cesari, G. Morgan **Evidence for long range transport of biological and anthropogenic aerosol particles in the atmosphere** Grana 23:43-53
- Moran et al. 2011, M.D. Moran, J. Chen, S. Ménard, R. Pavlovic, H. Landry, P.A. Beaulieu, S. Gilbert, P.A. Makar, W. Gong, C. Stroud, A. Kallaur, A. Robichaud, S. Gong, D. Anselmo **Two years of operational AQ forecasting with GEM-MACH15: A look back and look ahead** Presented at the 10th annual CMAS Conference, Chapel Hill, NC, October 24-26
- Moran et al. 2012, M.D. Moran, S. Ménard, R. Pavlovic, D. Anselmo, S. Antonopoulos, A. Robichaud S. Gravel, P.A. Makar W. Gong, C. Stroud, J. Zhang Q. Zheng, H. Landry, P.A. Beaulieu, S. Gilbert, J. Chen, A. Kallaur **Recent Advances in Canada's National Operational Air Quality Forecasting System** Proceedings of the 32nd NATO-SPS ITM Conference, 7-11 May 2012 Utrecht, NL
- Noh et al. 2013 Y.M. Noh, D. Müller, H. Lee, T.J. Choi **Influence of biogenic pollen on optical properties of atmospheric aerosols observed by lidar over Gwangju, South Korea** Atmospheric Environment 69:139-147
- Pasken and Pietrowicz, 2005 R. Pasken, J. Pietrowicz **Using dispersion and mesoscale meteorological models to forecast pollen concentrations** Atmos. Environ. 39:7689-7701, doi:10.1016/j.atmosenv.2005.04.043
- Ranta and Satri, 2007 H. Santri, P. Satri **Synchronised inter-annual fluctuations of flowering intensity affect the exposure to allergenic tree pollen in North Europe** Grana 46(4):274-284
- Ranta et al., 2008, H. Ranta, T. Hokkanen, T. Linkosalo, L. Laukkanen, K. Bondestam, A. Oksanen **Male flowering of birch: spatial synchronization, year-to-year variation and relation of catkin numbers and airborne pollen counts** Forest Ecology and Management 255:643-650
- Ranta et al., 2011 H. Ranta, P. Siljamo, A. Oksanen, M. Sofiev, T. Linkosalo K.C. Nergmann, E. Bucher, A. Ekeboom, J. Emberlin, R. Gehrig, M. Hallsdottir, V. Jato, S. Jäger, D. Myszkowska, A. Paldy, H. Ramfjord, E. Severova, M. Thibaudon, M. **Aerial and annual variation of birch pollen loads and a modelling system for simulating and forecasting pollen emissions and transport at European scale** In: Aerobiological Monographs. Towards a comprehensive vision. Edited by B. Clot, P. Comtois, and B. Escamilla-Garcia, 1:115-131. Université de Montréal, ISBN 978-2-8399-0466-7
- Raynor et al., 1970 G.S. Raynor, E.C. Ogden, J.V. Hayes **Dispersion and deposition of ragweed pollen from experimental sources.** *J. Applied Meteorol. Climatol.* 9:885-895, doi: 10.1175/1520-0450(1970)009
- Robichaud, 2017 A. Robichaud **Surface data assimilation of chemical compounds over North America and its impact on air quality and Air Quality Health Index**

- (AQHI) forecasts** Air Qual Atmos Health 10: 955. <https://doi.org/10.1007/s11869-017-0485-9>
- Robichaud and Comtois, 2017 A. Robichaud, P. Comtois **Statistical modeling, forecasting and time series analysis of birch phenology in Montreal, Canada** Aerobiologia 33, 529–554, doi:10.1007/s10453-017-9488-0
- Schueler and Schlüzen, 2006 S. Schueler, K.H. Schlüzen. **Modelling oak pollen dispersal on the landscape level with a mesoscale atmospheric model** Environ. Model Assess. 11: 179-194, doi: 10.1007/s10666-006-9044-8
- Shao and Lu 2000 Y. Shao, H. Lu **A simple expression for wind erosion threshold friction velocity** J. Geophys. Res. 105(D17), 22437-22443
- Sierra-Heredia et al., 2018 C. Sierra-Heredia, M. North, J. Brook, C. Daly, A. K. Ellis, D. Henderson, S. B. Henderson, É. Lavigne, T. K. Takaro **Aeroallergens in Canada: Distribution, Public Health Impacts, and Opportunities for Prevention** Int. J. Res. Public Health 15(8):1577, doi: 10.3390/ijerph15081577
- Siljamo, 2013 P.Siljamo **Numerical modelling of birch pollen emissions and dispersion on regional and continental scales** Finnish Meteorological Institute Contributions no. 99, FMI-CONT-99. Helsinki, Finland. ISBN 978-951-697-795-2
- Skjøth et al., 2007 C.A. Skjøth, J. Sommer, A. Stach, M. Smith, J. Brandt **The long-range transport of birch (*Betula*) pollen from Poland and Germany causes significant pre-season concentrations in Denmark** Clin. Exp. Allergy 37, 1204-1212
- Sofiev and Bergmann, 2013 M. Sofiev, K.C. Bergmann **Allergenic pollen. A review of the production, release, distribution and health impacts** Springer Science+Business Media, Dordrecht, ISBN978-94-007-4880-4e-ISBN 978-94-94-007-4881-1
- Sofiev et al., 2006a M. Sofiev, P. Siljamo, H. Ranta, A. Rantio-Lehtimäki **Towards numerical forecasting of long-range air transport of birch pollen: theoretical considerations and a feasibility study** Int. J. Biometeorol. 50:392-402, doi:10.1007/s00484-006-0027-X
- Sofiev et al., 2006b M. Sofiev, P. Siljamo I. Valkama M. Ilvonen, J. Kukkonen **A dispersion modelling system SILAM and its evaluation against ETEX data** Atmos. Env. 40:674-685
- Sofiev et al., 2013a M. Sofiev, J. Belmonte, R. Gehrig, R. Izquierdo, M. Smith, A. Dahl, P. Siljamo, **Airborne pollen transport** Chap 5. In: Allergenic pollen, A review of the production, release, distribution and health impacts. Editors: M. Sofiev and K.C. Bergmann. Springer Science+Business Media, Dordrecht, 256 pp. ISBN 978-94-007-4880-4e-ISBN 978-94-94-007-4881-1

- Sofiev et al., 2013b M. Sofiev, P. Siljamo, H. Ranta, T. Linkosalo, S. Jaeger, A. Rasmussen, A. Rantio-Lehtimäki, E. Severova, J. Kukkonen **A numerical model of birch pollen emission and dispersion in the atmosphere. Description of the emission module** Int. J. Biometeorol. 57:45-58, doi:10.1007/s00484-012-0532-z
- WHO (World Health Organization) 2003. **Phenology and human health: Allergic disorders** Rome; WHO regional office for Europe, 55p. Available at <http://apps.who.int/iris/bitstream/10665/107479/1/e79129.pdf> (last access 22 August 2014).
- Zhang et al. 2014, R. Zhang, T. Duhl, M.T. Salam, J.M. House, R.C. Flagan, E.L. Avol, F.D. Gilliland, A. Guenther, S.H. Chung, B. Lamb, T.M. VanReken **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, doi:10.5194/bg-11-1461-2014
- Zink et al. 2013 K. Zink, A. Pauling, M.W. Rotach, H. Vogel, P. Kaufmann, B. Clot **EMPOL 1.0: A new parameterization of pollen emission in numerical weather prediction models** Geosci. Model Dev. 6: 1961-1975

List of figures.

Figure 1 Distribution of vegetation fraction for birch in two domains A) larger domain: North America, B) sub-domain of study centred over Montreal (white star), which covers the southern parts of province of Quebec, parts of southern Ontario as well as New-England (NE) states of USA. The birch density colour bar is from cyan (fraction > 0.05) to deep red (max. 1.0) averaged over a given model tile (10 km X 10 km). Note that wherever there is no colour, the vegetation fraction is less than 5% (birch could be present but likely as an ornamental tree or in isolated spots). Note: L stands for Laurentide mountain region, E: Eastern Townships and NE, New-England (USA).

Figure 2 Sub-regions defined to evaluate the contribution of each major sources. 1: Labrador/Newfoundland, 2: Appalachian region, 3: Northern Laurentides, 4: Southern Laurentides, 5: Prairies/Great-Lakes, 6: Local Montreal region.

Figure 3 Averaged flux emission (grains/m²/s) for *Betula* pollen at 12Z for the month of May 2012 obtained for North America (top) and a zoom for the study area (bottom). The location of Montreal is indicated by a white star. Main source regions are: Laurentide Mountains (L), Eastern Townships (E) and New-England (NE). Note. Values below 5 grain/m²/s are considered as noise and therefore not shown.

Figure 4 A) Time series of numerical model for pollen concentration in Montreal as compared to observations taken at UdeM and Weather Network forecast using A) current day average (mean of lag0). Note. Grey dots at the top of each panel means presence of significant precipitation (more than 0.5 mm). Reference lines correspond to 10°C and 18°C, respectively (see text for details), **B)** as an A) but with two days moving average, **C)** As in A) but with three days moving average.

Figure 5 A) Model outputs (#/m³) (left) and vertical cross section (X-Y-Z) (right) for May 13th at 18UTC through Montreal. **B)** Model outputs (gr/m³) (left) and vertical cross section (X-Y-Z) (right) for May 13th at 18UTC through Quebec City. **Note.** The vertical coordinate is expressed with HY (hybrid vertical coordinate, HY=1.0: surface, HY=0.8: ~2 km altitude).

Figure 6 Average diurnal profile of *Betula* counts (April 11th – May 31st 2012) for simulated pollen for receptor sites: Montreal (YUL), Quebec (YQB), Ottawa (YOW) and source sites: Mont-St-Michel (MSM), Mont-Sutton (MST) and Sherbrooke (YSC). Note: see location of sites in Figure 3.

Figure 7 Time series of moving daily averaged pollen count (*Betula*) model output simulation (solid black) compared with moving average statistical weather forecast produced WXN (dotted grey) versus Julian day for A) Sherbrooke (YSC), B) Ottawa (YOW), C) Kingston (YGK), D) Quebec (YQB), E) Lac St-jean (YSJ), F) Toronto (YYZ), G) Barrie (WCU) and H) London (YXU).

Figure 8 Variogram showing correlation between model pollen output concentration and WXN statistical forecast time series at a given location (see Figure 10) versus its distance from Montreal.

Table 1. Performance of the modelling system for pollen concentration (lag 0 and lag 0-2) compared to: i) statistical forecast (obtained from Weather Network, i.e. WXN-YUL), ii) LTAA, iii) persistence (last year's value for the same day) and iv) from a combination of persistence and LTAA evaluated for Montreal. The metrics used are R (Pearson's and Spearman's coefficient of correlation), average OmP, standard deviation of OmP (RMSE), average absolute value of OmP, false alarm rate for low threshold (FAR30) and high threshold (FAR80) and percent correct for low threshold (PC30) and high threshold (PC80). Units for OmP (mean and RMSE are grains/m³).

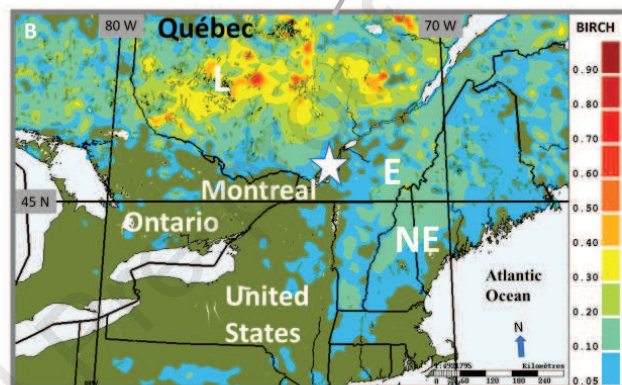
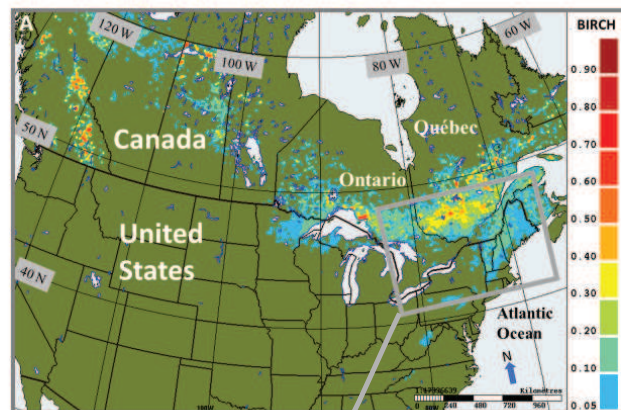
	R (Pearson)	R (Spearman)	Mean OmP	RMSE OmP	MAE (%) OmP	FAR30	FAR80	PC30	PC80
Model (lag0)	0.35	0.52	-10.4	119.2	61.2	0.21	0.12	0.82	0.86
Model	0.63	0.54	-9.9	57.4	37.0	0.21	0.12	0.82	0.89
Avg (lag0-2)									
WXNYUL	0.39	0.60	12.9	105.3	48.1	0.18	0.05	0.84	0.91
WXNYUL	0.38	0.60	11.9	49.2	29.3	0.18	0.02	0.85	0.94
Avg (lag0-2)									
LTAA (lag0-2)	0.25*	0.41	-30.2	66.8	53.0	0.14	0.17	0.92	0.87
Persistence	0.20*	0.34	-50.6	159.1	86.2	0.22	0.08	0.82	0.88
Persistence									
(lag0-2)									
LTAA/persistence	0.24*	0.38	-40.4	101.2	62.7	0.18	0.13	0.86	0.90
(lag0-2)									

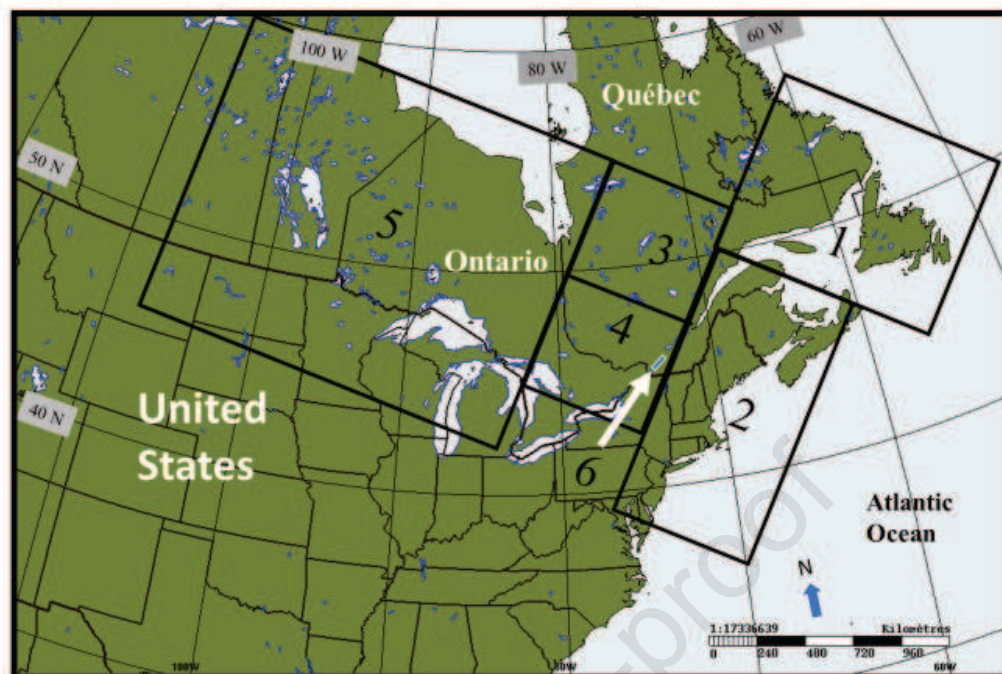
Note: lag0 average and the 3-days moving average (lag 0-2) are evaluated for both model and statistical forecast (WXN-YUL), * indicates non-significant correlation with observations. Bold numbers indicate the best scores among all forecast methods.

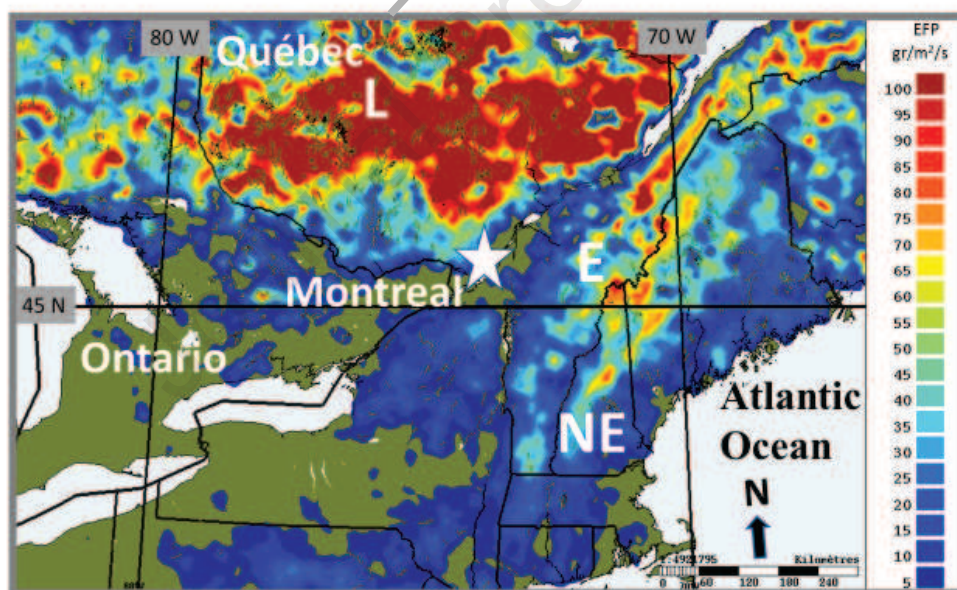
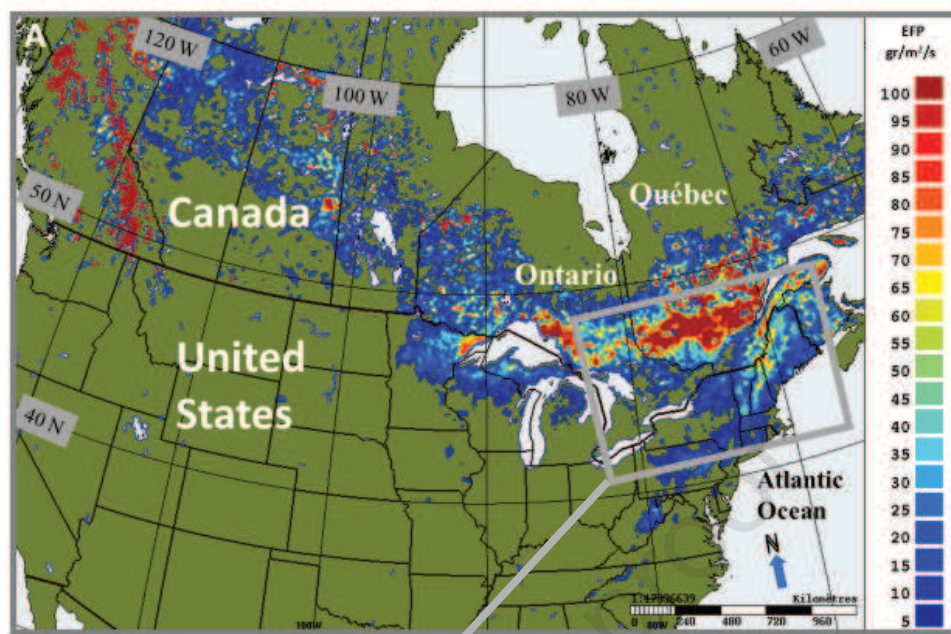
Table 2. Contribution (in % of total) from different area sources to the total *Betula* pollen count SPI (seasonal pollen index) arriving in Montreal. The corresponding different regional source areas are shown in Figure 2.

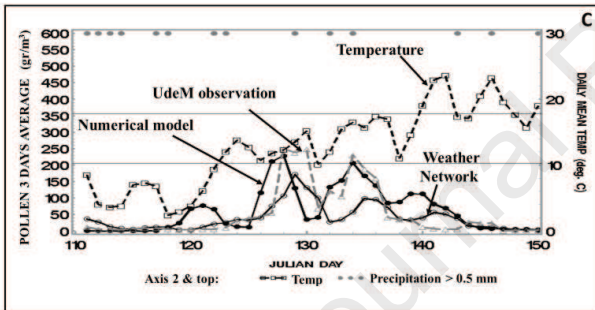
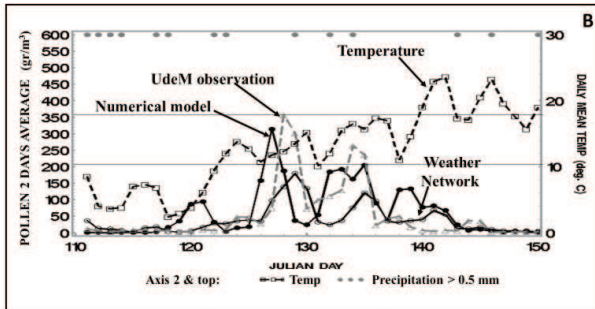
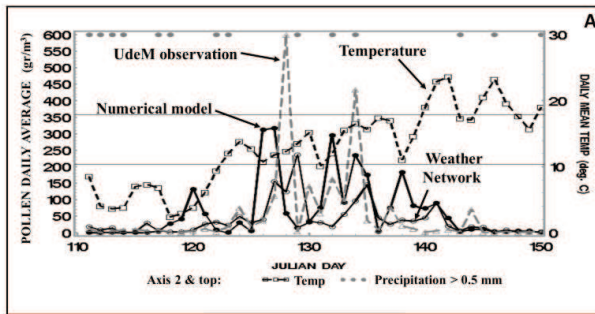
	Source	Transport Distance	SPI (model)	% of total
Region 1	Newfoundland & Labrador	LDD	~ 0.7	~0
Region 2	Appalachian mountain	MDD-LDD	37.7	2.8
Region 3	Northern Laurentides	LDD	431.2	33.3
Region 4	Southern Laurentides	MDD-LDD	711.8	55.0
Region 5	W. Ontario & Prairies	LDD	68.5	5.1
Region 6	Montreal Island	Local (< 30 km)	8.3	0.6
Elsewhere	Other	LDD	45.3	3.2

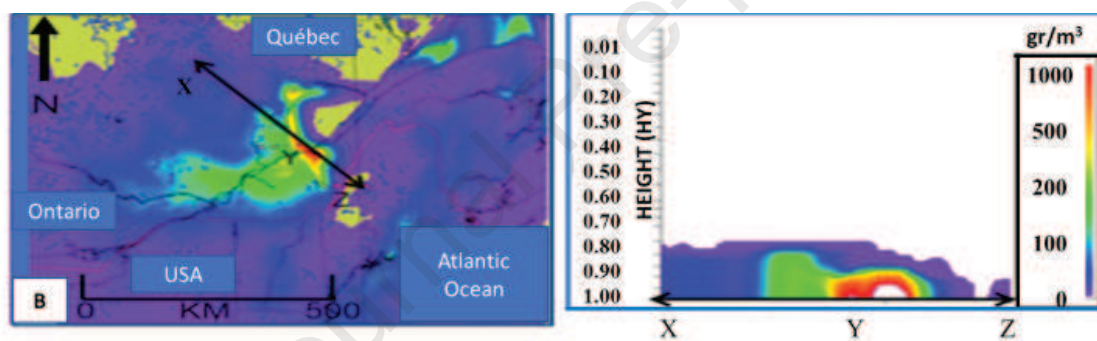
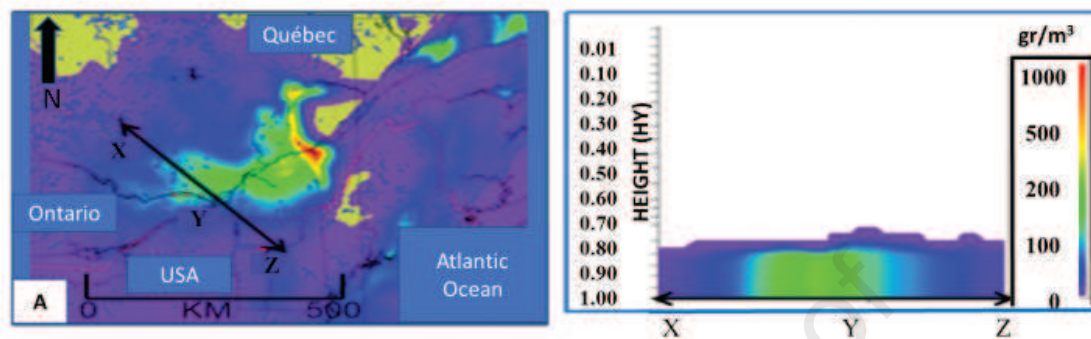
Journal Pre-proof

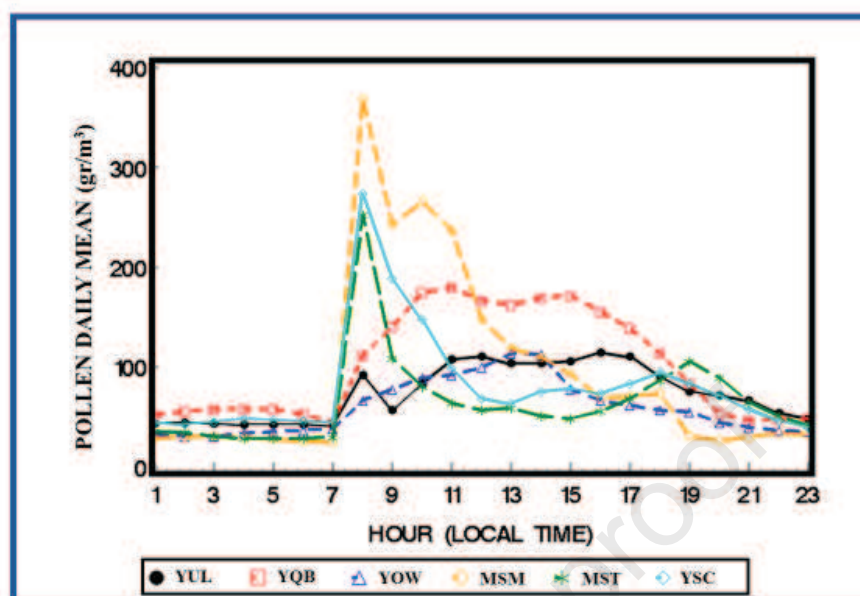


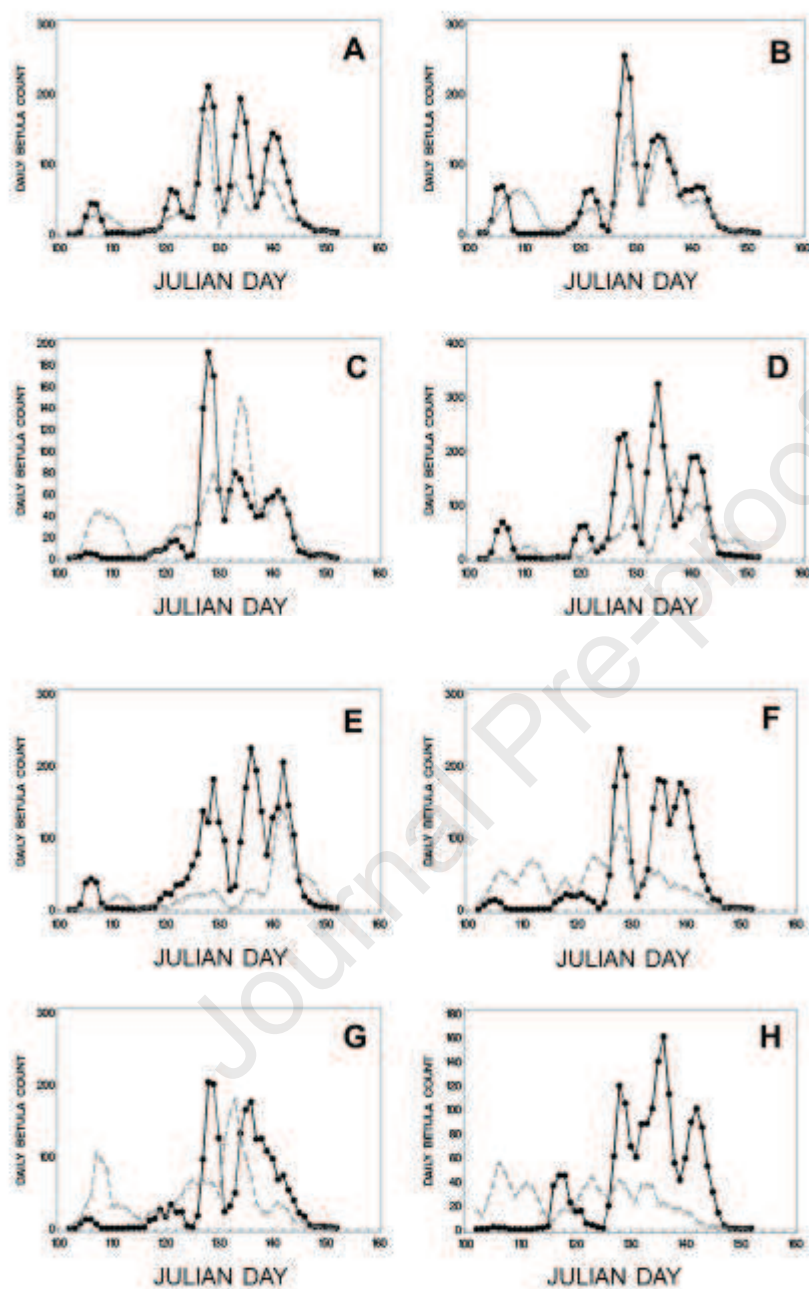


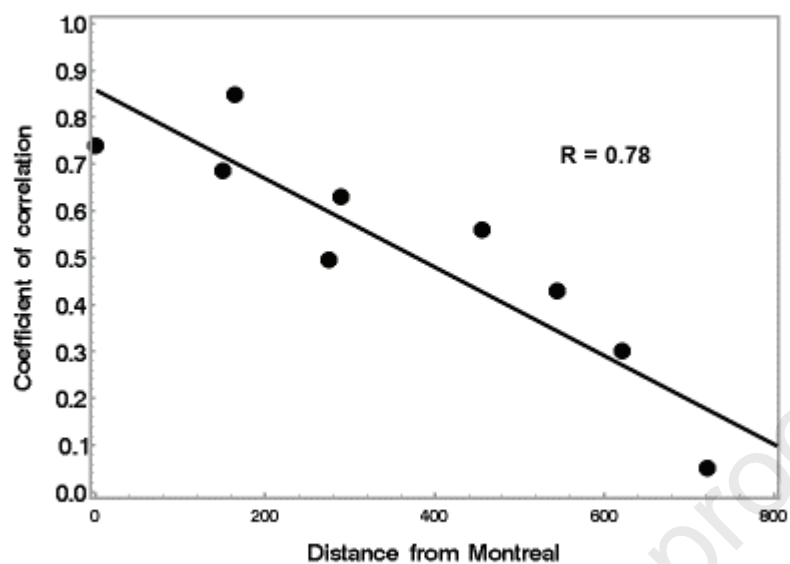












Environmental Research Journal

Highlights

- The birch airborne pollen simulations presented shows reasonable agreement with observations ($R > 0.60$)
- About 88 % of the birch pollen measured in Montreal originated from the Laurentide mountains
- Birch pollen episodes occur within a narrow temperature range (10°C - 18°C) in Montreal
- The concepts of puff emission and vertical correlation length have been used and found successful
- For the first time, simulation of birch pollen dispersion is published in Canada with possible extension to any other region of the world given pollen data availability

The authors declare no conflict of interest.

Journal Pre-proof

Supplementary material

S1. Summary of characteristics of the simulation model GEM-MACH model

GEM-MACH (version 1.3.8) uses a simple 2-bin sectional representation of the particulate size representation (bin 1 is for 0-2.5 μm aerodynamic diameter and bin 2 is for 2.5-10 μm particles, that is fine and coarse particles respectively). The main characteristics of version 1.3.8 are described here:

Regional assimilation system

- Spin-up cycle (12 hours) with trial field every 6-h from GEM global version
- 28 vertical levels
- 4DVAR assimilation of meteorological observations

Dynamical

- Primitive hydrostatic equations
- Temporal integration semi-implicit and semi-lagrangian
- 80 hybrid levels with a model top at 0.1 hPa
- 48 hours forecast with a resolution of 10 km (timestep of 300 sec)
- Horizontal diffusion of momentum with a sponge at the model top

Physics

- Atmospheric boundary layer based on TKE (Total Kinetic Energy) subgrid parametrization
- Moist TKE in presence of clouds
- Implicit vertical diffusion
- ISBA model for terrestrial schemes
- Kuo convective scheme for shallow convection
- Fritsch-Chappel scheme for deep convection
- Sundqvist for condensation scheme
- GWD (gravity wave drag) and topography block option

Main characteristics of GEM-MACH model used here

Characteristics/model	GEM-MACH 10 (v.1.3.8)
In operational use at CMC	2012-2014
Meteorological driver	GEM (on-line)
Resolution	10 km
Advection scheme	Semi-Lagrangian method [<i>Côté et al</i> , 1998a,b]
Number of vertical levels	80
PM composition and representation	9 species
Anthropogenic emissions	Major and minor point sources, area and mobile sources for 17 gas and 2 bins for particles (PM _{2.5} , PM ₁₀). Inventories: Canada, 2006; US, 2005 projected to 2012; Mexico 1999.
Biogenic emissions	Four emission factors: NO, isoprene, monoterpenes & other volatile organic compounds. Uses BEIS system with BELD3 vegetation database (231 categories) + Canadian National Forest Inventory
Gas-phase chemistry	ADOM-II (47 species advected; 114 chemical reactions) [<i>Lurmann et al.</i> , 1986]
Gas-phase chemical solver	Vectorized version of <i>Young and Boris</i> [1977]; <i>Makar</i> [1995].
Aqueous-phase chemistry	Based on ADOM (20 reactions, 7 gases and 13 aqueous species)
Aqueous-phase solver	<i>Makar</i> [1995]
Heterogeneous chemistry	Based on ISORROPIA
Secondary organic aerosol	IAY scheme, <i>Jiang</i> [2003]
Dry deposition	modified scheme of <i>Wesely</i> [1989] for gas; <i>Zhang et al.</i> , [2001] for particles
Wet deposition	<i>Gong et al.</i> , [2003]
Chemical boundary condition	Lateral and upper climatological profiles
Table S1	

S2. Dry and wet deposition

Birch pollen was taken as 22 microns of diameter although its aerodynamic diameter is not really defined in the literature and size distribution is not available. Note that particles with diameter bigger than 10 microns are deposited independent of the roughness of the surface as the sedimentation velocity dominates [*Seinfeld and Pandis*, 2006]. In this study, we suppose that birch pollen has nearly a spherical shape with approximately known densities. Therefore, settling of the pollen depends only on the pollen size and Stokes law applies:

$$V_s = \frac{2}{9}(\rho_{pollen} - \rho_{air})gr^2/\mu \quad (S1)$$

where $\rho_{pollen} - \rho_{air}$ is the difference of pollen density and air density, g the gravitational constant, r , the radius of the aerosol and μ , the air viscosity. Taking typical values [see *Gregory*, 1973]: $\rho_{pollen} = 0.8 \text{ g/cm}^3$ and $r = 0.11 \text{ }\mu\text{m}$ for the *Betula* pollen, the dynamic air viscosity $\mu = 1.8\text{E-}4 \text{ g/cm/sec}$, $g = 981 \text{ cm/s}^2$ and $\rho_{air} = 1.25 \text{ E-}3 \text{ g/cm}^3$ gives $V_s \approx 1.3 \text{ cm/sec}$ for the settling velocity. Note that effects of temperature and humidity in Equation S1 are neglected here given all uncertainties. The washout of pollen by precipitation and the washout by clouds are treated the same way as for coarse particles (PM_{2.5-10}) by the GEM-MACH model in our simulations. Pollen are considered not soluble in water but easily scavenged via impaction and precipitation.

S3. Vegetation distribution

The raw input data used has been obtained from <https://www.epa.gov/air-emissions-modeling/biogenic-emissions-landuse-database-version-3-beld3>. Data from the United States Forest Service (USFS) on individual tree species for each county or parish in US was utilized. Data from the Canadian National Forest Inventory was also aggregated. The final vegetation file is provided in terms of the area covered by individual tree species as a fraction (from 0 to 1) of total area covered by all tree species at a specific gridpoint. The file obtained however needed to be further processed to be accepted as input in the GEM-MACH model. The format file (tiff) obtained from EPA needed to be converted into a CMC (Canadian Meteorological Centre) standard file to be processed as an input for GEM-MACH. The original 1 km by 1 km resolution vegetation data was re-distributed and transformed into a 10 km by 10 km resolution standard file for input in GEM-MACH. The gridded version of the birch distribution is given as a fraction from 0 to 1 in Figure 1A,B.

S4. Shape of the bloom probability curve versus Julian date

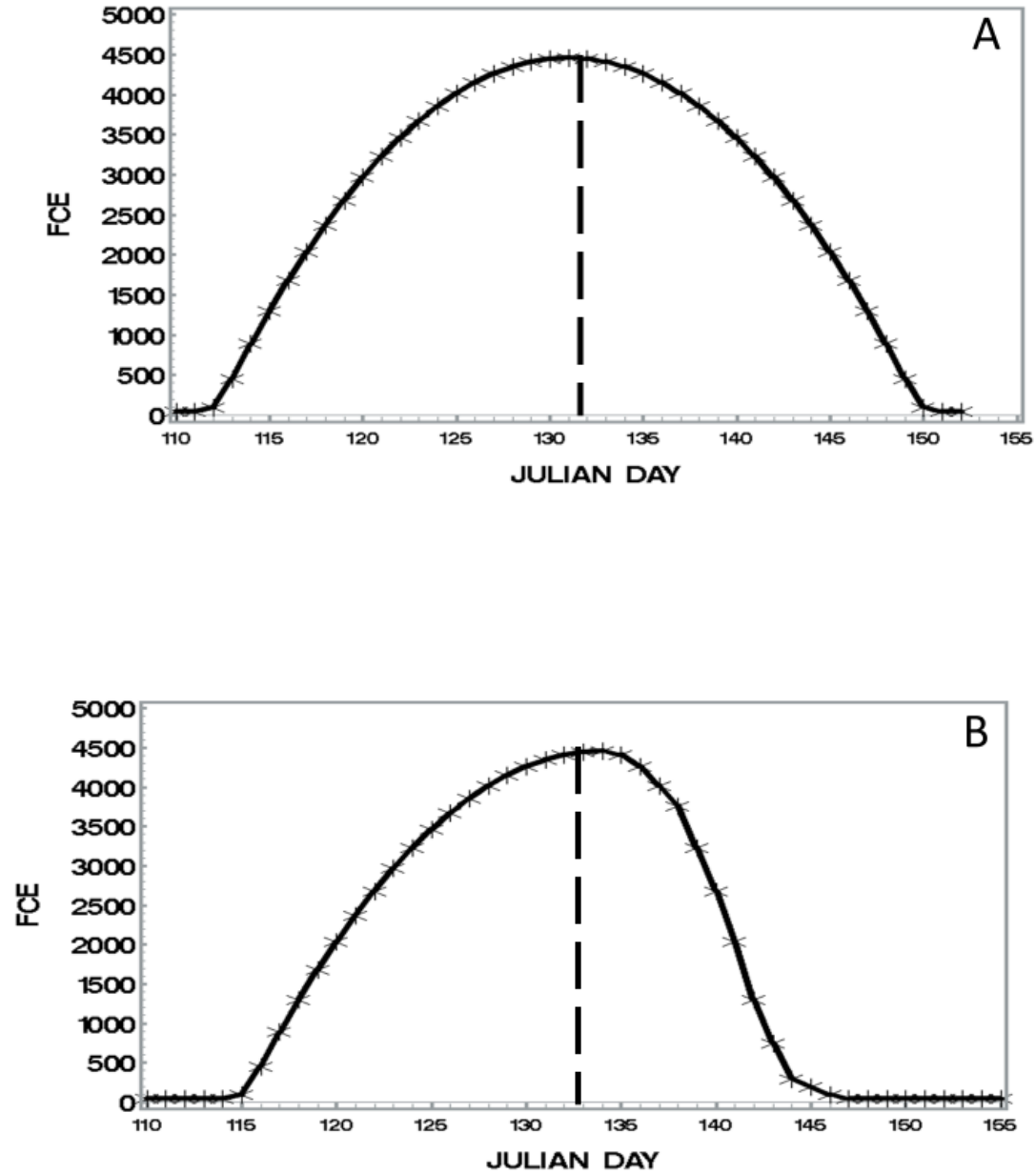


Figure S1. Proposed curve of FCE ($=C_e(t) \times c^*$; see Eqs. 4 and 5); **A)** similar to *Helbig et al.* [2004], **B)** modified to consider the *open-pocket* concept in this study. The vertical line indicates the mid-season (anthesis point).

According to the *open-pocket* concept, after the anthesis point (maximum emission), the wind speed (which is linked to u^* in Equation 3) becomes much less effective in spreading pollen in the air because the amount of pollen left in the anther and catkins is limited and becomes much less abundant (i.e. the pocket is becoming empty). Therefore, to mimic this effect, the curve, FCE, is set to decrease twice faster in the second part of the season.

Table S2. Adopted parameters used in the pollen modelling simulation.

Parameter	Symbol	Value	Units	Reference
Total emission per unit surface	Q_p	2.0E9	grain/m ²	<i>Siljamo</i> [2013]
Length duration	S	42	days	<i>Climatology</i>
Average height	H_c	22.5	meters	From Table 1
Mean leaf area index	LAI	2	-	<i>Helbig et al.</i> [2004]
Pollen density	ρ_{pollen}	800	kg/m ³	<i>Erdtman</i> (1943)
Pollen aerodynamic Diameter	D_a	22	microns	Taken same as alder [<i>Helbig et al.</i> 2004]
Sedimentation velocity	V_d	1.3	cm/s	See Suppl. Material S2
Temperature Threshold*	T_{th}	11	°C	From calibration tests (for alder, <i>Helbig et al.</i> [2004] proposed 8 °C)
Dew pt. depression threshold*	D_{th}	5	°C	Adapted from <i>Helbig et al.</i> [2004] (Ute=80% R.H. corresponds to T-Td of 5°C).
Wind threshold*	V_{th}	2.9	m/s	From sensitivity tests (same result given for alder, <i>Helbig et al.</i> , 2004]
Threshold friction velocity*	u^*_t	0.323	m/s	Taken as <i>Shao and Lu</i> [2000]
Precipitation threshold**	P_{th}	0.5	mm/h	<i>Sofiev et al.</i> [2013]
Vertical correlation length	L_z	10	Hybrid levels	From calibration tests (corresponds to average height of boundary layer)

*No emission below this threshold, **No emission above this threshold

S5. Metrics used for evaluation of model performance

Many metrics exist to validate model simulations. *Chang and Hanna* [2004] provides guidance for selecting appropriate metrics. The following set of metrics were chosen in order to avoid metric duplication but also to be able to compare with results presented in other studies : 1) Pearson's coefficient of correlation and rank coefficient of correlation (Spearman), the latter giving different information and is more appropriate for non-parametric distribution [SAS[®], 1989; *Maritz*, 1981], 2) average bias (average of observation minus model prediction, hereafter OmP), 3) average absolute bias, 4) standard deviation of OmP, 5) false alarm rate for two thresholds FAR30 and FAR80 which correspond to low threshold (30 µg/m³) and high threshold (80 µg/m³) respectively, and 6) PC30 and PC80, percent correct for low and high threshold respectively. The low threshold PC30 corresponds to a value where most of the sensitive patients start showing symptoms and the high threshold PC80 where 90% of the patients show symptoms [*Emberlin*, 1997; *de Weger et al.*, 2013]. The mathematical definition of the metrics are given below:

Definitions

Given $\mathbf{O}=\{O_i\}$ and $\mathbf{X}=\{A_i\}$ or $\{P_i\}$, respectively the observation, the corresponding analysis and the short-term forecast interpolated to an ensemble of measurement sites N , $i=1,2,...,N$.

Continuous indices

Bias (B) : Mean of OmP (Observation minus Prediction)

$$B = (1/N) \sum (O_i - X_i) \quad (\text{S.1})$$

The definition of the absolute bias is the same as Eq. S.1 except that the absolute value of $O - X_i$ is taken (before the summation).

The root mean squared error ($RMSE$) (equivalent to standard deviation of $O-P$) :

$$RMSE = \text{SQRT}(\sum ((O_i - X_i) - \text{mean}(O_i - X_i)/N) \quad (\text{S.2})$$

Categorical index

The FAR (false alarm rate) indices and PC (percent correct or hit rate) are defined here in terms of two thresholds, low (FAR30, PC30) and high threshold (FAR80, PC80) corresponding to 30 grains/m³ and 80 grains/m³ respectively.

Let consider the following table:

	Observed	Non-observed
Forecast	A	B
Not forecast	C	D

False alarm rate = $B/(B+D)$ (event forecast but non-observed divided by total number of observation of non-event)

Percent correct = $(A+D)/(A+B+C+D)$ (number of forecast correct divided by total number of events).

More details and definition about verification metrics could be found at

http://www.swpc.noaa.gov/forecast_verification/Glossary.html.

The Spearmann coefficient of correlation is defined as

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}.$$

where d_i is the distance between rank which are paired and n the number of cases. Note that the Spearman correlation is less sensitive than the Pearson correlation to strong outliers and heavy tails in both samples. This is because Spearman's coefficient of correlation limits the outliers to the value of its rank [Maritz, 1981]. Finally, the mean absolute error (MAE in %) is also computed as Equation S.1 but using absolute values.

S6. Method for evaluating the contribution of different source regions

One of the goal of this study is to identify what is the most important source region for birch pollen arriving in Montreal; that is, to delineate the contribution of different regional source areas of TDMLD. Shutting off all pollen emissions sources except for a given region is a way to evaluate the contribution of a given region (source attribution). When a given source is activated, the vegetation distribution can emit pollen with the same configuration as for the normal simulation whereas outside the region selected, the pollen emission is shutoff to zero. Evaluation of the individual contributions are done for 6 regions (Figure 2 in the main text): 1) Labrador and Newfoundland region to the northeast (long-distance dispersal, hereafter, LDD), 2) Appalachian region to the east of Montreal (source of both medium and long-range distance dispersal, hereafter

MDD/LDD), 3) Hautes Laurentides to the north and northwest of Québec City (MDD/LDD), 4) Basses Laurentides (MDD) and 5) northern Ontario and Canadian Prairies to the far west (LDD). An additional experiment (region 6) deals with analyzing the impact of local sources in the Montreal Island and vicinity (by shutting off all remote MDD/LDD sources). A final experiment was also conducted to exclude all above sources (all regions 1-6 in Figure 4) to assess the impact of very remote sources elsewhere in North America.

S7. Methods for data validation

Validation of model pollen concentration will be made using observations collected at Université de Montréal (UdeM). This site has monitored pollen for many decades and this is one major reason why Montreal was chosen the centre of the focus for this study. Since pollen count measurements are rare in Canada, we will also use the statistical forecast issue by Weather Network for Montreal along with other sites in the province of Quebec and few in the neighbour province of Ontario as proxies for supplementary observations needed for validation. Figure S1 below shows the different locations where validation will be performed.

Pollen collected at UdeM station

Pollen data from the site at Université de Montréal is used to validate the results during the simulation period 11th April to 31st May 2012. Sampling is made by a volumetric spore-trap (LANZONI VPPS2000). Air is sucked in through a 2 mm by 14 mm nozzle at a rate of 10 litre per minute. Airborne particles are collected on a Melinex tape mounted on a drum which is slowly turned by clockwork and then analysed under a microscope using standard method of counting. *Mandrioli et al.* [1998] describes in more details the process of pollen measurement and monitoring. During the study period, *Betula* pollen and other biological aerosols (other allergenic pollen and spore particles) were monitored uninterruptedly. The data collected for a specific day corresponds to a period from noon the previous day to noon the actual day. Note that the methodology described above is limited by important potential bias due to human errors [*INSPQ*, 2013]. *Comtois et al.* [1999] have estimated that for birch pollen count errors from aerobiological samples are in the range 5-7%, i.e. lower than that for other taxons.

Statistical forecast produced by Weather Network

Weather Network (WXN) is a private firm that provides on a daily basis a short-term forecast across Canada during the pollen season. The ROTOROD[®] sampler is a rotating arm impactor that collects airborne particles on two rapidly spinning polystyrene rods. Airborne particulates impact on one side of a translucent 6 cm long square rods (1.6 by 1.6 mm) that whirl through the air at 2400 rpm which corresponds to a sampling rate of about 120 litres per minute. After collection, the rods are removed and examined under light microscopy [Scheifinger *et al.*, 2013]. Note that the year 2012 was selected for convenience (better availability of data required for this study) and because WXN was producing forecast pollen concentration (in grains per cubic meter) for the last time (only categorical forecast is publically available after 2012). Note also that the year 2012 does not correspond to a mast year and not to a low year either but to an average year for pollen production (i.e. within one standard deviation of typical climatological values, according to climatology).

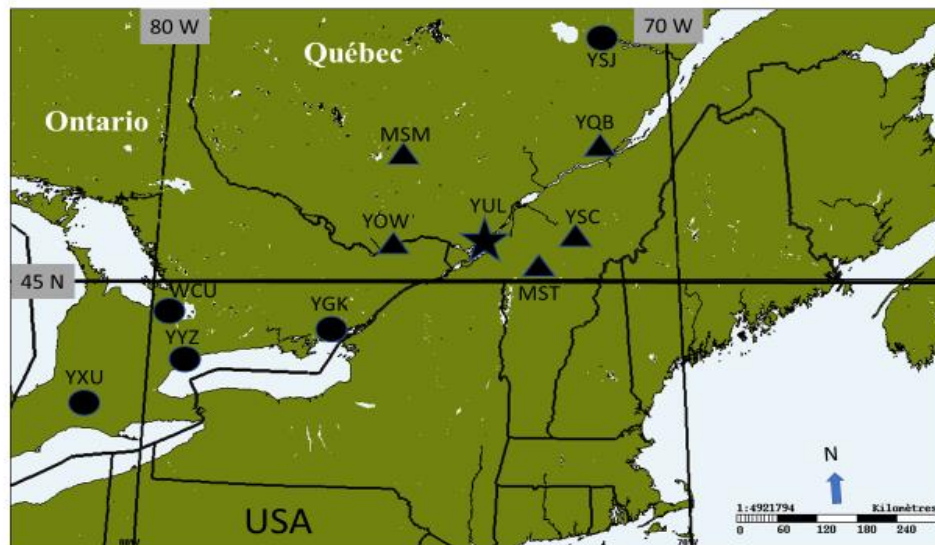


Figure S2. Location of sites used in this study for validation purposes: YUL (Montreal depicted by the star), YSC (Sherbrooke), YOW (Ottawa), YQB(Quebec city), YSJ (Lac St-Jean), MSM (Mont-St-Michel), MST (Mont-Sutton) in the province of Quebec and YYZ (Toronto), YGK (Kingston), WCU (Barrie), YXU (London) in Ontario.

Validation of phenology.

	Distance (km)	R	SPI (model)	SPI (WXN)
Montreal (YUL)	0	0.72	2506	1536
Sherbrooke (YSC)	150	0.685	2558	1681
Ottawa (YOW)	165	0.849	2395	2121
Quebec (YQB)	275	0.496	3339	1897
Kingston (YKG)	289	0.63	1441	1772
Chicoutimi (YSJ)	456	0.56	2866	1166
Toronto (YYZ)	545	0.43	2380	1910
Barrie (WCU)	621	0.30 ^a	2144	2158
London (YXU)	721	0.05 ^b	1854	1102

Note. $p < 0.01$ except if mentioned otherwise, ^a: $p < 0.05$, ^b non-significant (i.e. $p > 0.1$).

Table S3. Coefficient of correlation between model value and statistical forecast (WXX-YUL) as a function of distance. A linear fit of the data defines a correlation function (see Figure 8 of main text).

S8. Modelling and coding considerations

The basic configuration of the GEM-MACH aerosol module utilizes two bins (bin 1 for $PM_{2.5}$ and bin 2 for $PM_{2.5-10}$). Even though the pollen grain dimension is significantly larger than the maximum size of bin 2 (i.e. 10 μm), we will consider birch pollen as a coarse particulate in the model as far as the transport, washout due to precipitation and clouds are concerned. Several authors have adopted this strategy in the past for similar pollen studies [*Sofiev et al 2006a,b* and *Siljamo et al, 2008* in SILAM and *Efstathiou et al., 2011* as well as *Zhang et al., 2014* both in CMAQ]. On the other hand, boundary layer processes are very important for vertical transport and dispersion of particles. However, the meteorological models such as the GEM model does not seem to reproduce adequately the vertical transport of tracers and their mixing especially when there is so-called shallow convection (rather intermittent and localized) where LES (large-scale eddy simulation) would be needed and absent from the current version of GEM-MACH. Moreover, GEM-MACH model has a boundary layer which tends to be too stable. Morning and evening transition, and the development of low level jet have been shown to be poor for the version used (v1.3.8) of GEM-MACH [*Bosveld et al., 2014*] and this corresponds to the time that the model is re-initialized with observations of chemicals, meteorological observations or peak morning pollen emission. Moreover, there is no mechanism for free convection or deep convection vertical

transport of chemical or biological tracers within GEM-MACH. Finally, *Lanner* [1966] argues that the formation of small convective currents at the local scale (thermal shells) provides the mechanism for pollen to be lifted to high altitudes and thus dispersed long distances. Again, the numerical model does not simulate well these phenomena since it is considered as a sub-grid scale feature. To circumvent these problems, a simple redistribution of tracer mass was tested to project the ‘emitted’ pollen concentrations over the bottom 10 levels of the model with a linear weighting decreasing from 1.0 (at the surface) to 0.1 (at model level 10, the latter corresponding to the top of the warm season boundary layer, i.e. altitude ~ 1.5 km) while making sure that the total mass is conserved in the process. This is equivalent to define a correlation vertical length for vertical diffusion of height equivalent to the mean maximum boundary layer height. Supplementary material S9 describes the methodology used to convert surface pollen emissions to surface pollen concentration. The method of vertical projection for the tracer mass (the concept of vertical correlation length) is found in *Robichaud* [2017] and was used successfully in the data assimilation configuration for ozone, nitrogen dioxide and PM_{2.5}.

Many other complicated processes likely take place in the vertical transport of pollen. As identified for wind-blown, sand and dust, saltation, reptation¹, and airborne suspension often occur in the atmosphere [*Kok et al.*, 2012] at times amplified by the Magnus force² [*Huang et al.*, 2010]. At the current moment, it is not possible to resolve these in the context of our simulation platform (resolution of 10 km is not sufficient). Nevertheless, we believe that the concept of vertical correlation length as described above compensates for the lack of extensive modelling likely required to simulate vertical trajectories of airborne pollen (from local to regional scale).

S9. Converting emission flux into surface concentration

The emission obtained from the modified emission model described in *Helbig et al.* [2004] is translated into surface concentration here using basic principles. The concentration at a given time could be expressed as the mass of pollen emitted per square meter ET where E is the emission flux (grains/m²/sec) and T the characteristic time scale (sec). This characteristic time scale could be expressed as H^2/κ from non-dimensional analysis [*J. Pudykiewicz, pers. comm.*, 2010] where H is

¹ Saltation is a vertical motion having a trajectory such as a leap or a spring caused by turbulent fluctuations while reptation is a mode of transport similar to saltation (although over shorter distances) but where the impact with the soil is involved (bouncing).

² Magnus force is a lifting force due to particle spin.

the vertical characteristic height and κ the vertical diffusion. Therefore, the concentration $\langle c \rangle$ could then be expressed as:

$$\langle c \rangle = (E - \langle c \rangle v_d) / H$$

(i.e. emission minus deposition over the layer H) with V_d , the sedimentation velocity. Re-arranging and substituting for T gives:

$$\langle c \rangle = E / (\kappa/H + v_d)$$

Given the emission E from *Helbig et al.* [2004], κ from GEM model output and taking fixed values for other variables, the surface concentration $\langle c \rangle$ is then easily obtained. Typical values for H and κ are 40 m and $10 \text{ m}^2/\text{s}^2$ respectively.

S10. Animated pollen simulation over eastern Canada (11th April – 31st May 2012)

Horizontal distribution of pollen as a function of time (each frame is one hour starting 11th April 2012 at 12UTC, e.g. 8AM in the morning) during *Betula* (birch) pollen season of 2012. The units are in grains/ m^3 and values follow the color bar to the right. Notice that most of the pollen arriving in Montreal comes from the Laurentides (in agreement of results of Table 2).

References

- Chang, J.C. Hanna, S.R. 2004. Air quality model performance evaluation. *Meteorol. Atmos. Phys.* 87:167-196, doi 10.1007/s00703-003-0070-7.
- Comtois, P., Alcazar, P., Néron, D. 1999. Pollen counts statistics and its relevance to precision. *Aerobiologia* 15:19-28.
- Côté, J., Gravel, S., Méthot, A., Patoine, A., Roch, M., Staniforth, A.N. 1998a. The operational CMC-MRB Global Environmental Multiscale (GEM) model. Part I: design considerations and formulation. *Mon. Wea. Rev.* 126: 1373-1395.
- Côté, J., Desmarais, J.G., Gravel, S., Méthot, A., Patoine, A., Roch, M., Staniforth, A. 1998b. The operational CMC-MRB Global Environmental Multiscale (GEM) model. Part II: Results. *Mon. Wea. Rev.* 126:1397–1418.
- de Weger, L.A., Bergmann, K.C., Rantio-Lehtimäki, Dahl, A., Buters, J., Déchamp, J., Belmonte, J. 2013. *Impact of pollen*. In: Allergenic pollen, A review of the production, release, distribution and health impacts, Chap. 6. Eds: M. Sofiev and K.C. Bergmann. Springer Science+Business

Media, Dordrecht, 2013, 256 pp. ISBN 978-94-007-4880-4e-ISBN 978-94-94-007-4881-1. doi 10.1007/978-94-007-4881-1_6.

Emberlin, J. 1997. *Grass, tree and weed pollen*. In: Kay A.B. (Ed). Allergy and allergic diseases. National Heart and Lung Institute, London.

Erdtman, G. 1943. *An introduction to pollen analysis*. Chronica Botanica. Waltham, Mass., USA.

Efstathiou, C., Isukapalli, S., Georgopoulos, P. 2011. A mechanistic modeling system for estimating large scale emissions and transport of pollen and co-allergens. *Atmos. Environ.* 45(13):2260-2276, doi:10.1016/j.atmosenv.2010.12.008.

Gong, S. L., Barrie, L.A., Blanchet, J.P., von Salzen, K., Lohmann, U., Lesins, G., Spacek, L., Zhang, L., Girard, E., Lin, H., Leaitch, R., Leighton, H., Chylek, P., Hung, P., Jiang, J. 2003. Canadian Aerosol Module: A size-segregated simulation of atmospheric aerosol processes for climate and air quality models: 1. Module development. *J. Geophys. Res.* 108 (D1), 4007, doi:10.1029/2001JD002002.

Gregory, H.P. 1973. *The microbiology of the atmosphere*. Second edition. John Wiley & Sons, New-York, Toronto.

Helbig, N., Vogel, B., Vogel, H., Fiedler, F. 2004. Numerical modelling of pollen dispersion on the regional scale, *Aerobiologia* 3:3-19.

Institut national de santé publique du Québec (INSPQ) 2013. *États des connaissances sur le pollen et les allergies*. Les assises pour une gestion efficace, ISBN 978-2-550-68403-9 (available at <http://www.inspq.qc.ca>).

Jiang, W. 2003. Instantaneous secondary organic aerosol yields and their comparison with overall aerosol yields for aromatic and biogenic hydrocarbons. *Atmospheric Environment* 37:5439–5444.

Lurmann, F.W., Lloyd, A.C., Atkinson, R. 1986. A chemical mechanism for use in long-range transport/acid deposition computer modelling. *J. Geophys. Res.* 91:10905-10936.

Makar, P.A. 1995. *Fast use chemical numeric methods: the use of “vectorization by grid point”*. In: Moussiopoulos, H.N., Brebbia, C.A. (Eds.), Air Pollution III, Vol. 1. Computational Mechanics Publications, Southampton.

Maritz, J. S. 1981. *Distribution-Free Statistical Methods*. Chapman & Hall. 217 p. ISBN 0-412-15940-6.

Mandrioli, P., Comtois, P., Levizzani, V. 1998. *Methods in aerobiology*. Pitagora Editrice Bologna, ISBN 8-371-1043-X.

SAS®, 1989. Statistical Analysis Software. SAS/STAT User’s guide. SAS Institute, Cary, N.C. USA.

Scheifinger, H., Belmonte, J., Buters, J., Celenk, S., Damialis, A., Dechamp, C., García-Mozo, H., Gehrig, R., Grewling, H., Halley, J. M. 2013. *Monitoring, modelling and forecasting of the pollen season*. Chap 4. In: Allergenic pollen, A review of the production, release, distribution and health

impacts. Editors: M. Sofiev and K.C. Bergmann. Springer Science+Business Media, Dordrecht, 56 pp. ISBN 978-94-007-4880-4e-ISBN 978-94-94-007-4881-1.

Seinfeld, J.H. Pandis, S.N. 2006. *Atmospheric chemistry and physics: from air pollution to climate change*. Wiley-Interscience.

Shao, Y., Lu, H. 2000. A simple expression for wind erosion threshold friction velocity. *J. Geophys. Res.* 105(D17), 22437-22443.

Siljamo, P. 2013. *Numerical modelling of birch pollen emissions and dispersion on regional and continental scales*. Finnish Meteorological Institute Contributions no. 99, FMI-CONT-99. Helsinki, Finland. ISBN 978-951-697-795-2.

Siljamo, P., Sofiev, M., Severova, E., Ranta, H., Kukkonen, J. Polevova, S., Kubin, E., Minin, A., A. 2008. Sources, impact and exchange of early-spring birch pollen in the Moscow region and Finland. *Aerobiologia* 24:211-230, doi: 10.1007/s10453-008-9100-8.

Sofiev, M., Belmonte, J., Gehrig, R., Izquierdo, R., Smith, M., Dahl, A., Siljamo, P. 2013. *Airborne pollen transport*. Chap 5. In: Allergenic pollen, A review of the production, release, distribution and health impacts. Editors: M. Sofiev and K.C. Bergmann. Springer Science+Business Media, Dordrecht, 256 pp. ISBN 978-94-007-4880-4e-ISBN 978-94-94-007-4881-1.

Sofiev, M., Siljamo, P., Ranta, H., Rantio-Lehtimäki, A. 2006a. Towards numerical forecasting of long-range air transport of birch pollen: theoretical considerations and a feasibility study. *Int. J. Biometeorol.* 50:392-402, doi:10.1007/s00484-006-0027-X.

Sofiev M., Siljamo P., Valkama I., Ilvonen M., Kukkonen J. 2006b. A dispersion modelling system SILAM and its evaluation against ETEX data. *Atmos. Environ.* 40:674-685.

Wesely, M. L. 1989. Parameterization of surface resistance to gaseous dry deposition in regional-scale numerical models, *Atmos. Environ.* **23**:1293-1304.

Young, T.R., and Boris, J.P. 1977. A numerical technique for solving stiff ordinary differential equations associated with the chemical kinetics of reactive flow problems. *J. Phys. Chem.* 81:2424-2427.

Zhang, L., Gong, S., Padro, J., Barrie, L. 2001. A size-segregated particle dry deposition scheme for an atmospheric aerosol module. *Atmos. Environ.* 35:549–560.

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. and VanReken, T.M. 2014. 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, doi:10.5194/bg-11-1461-2014.
