

Automatic generation of large ensembles for air quality forecasting using the Polyphemus system

D. Garaud^{1,2} and V. Mallet^{2,1}

¹CEREA, joint research laboratory ENPC – EDF R&D, Université Paris-Est, Marne-La-Vallée, France

²INRIA, Rocquencourt, France

Received: 19 June 2009 – Published in Geosci. Model Dev. Discuss.: 9 July 2009

Revised: 15 December 2009 – Accepted: 18 December 2009 – Published: 18 January 2010

Abstract. This paper describes a method to automatically generate a large ensemble of air quality simulations. Such an ensemble may be useful for quantifying uncertainty, improving forecasts, evaluating risks, identifying process weaknesses, etc. The objective is to take into account all sources of uncertainty: input data, physical formulation and numerical formulation. The leading idea is to build different chemistry-transport models in the same framework, so that the ensemble generation can be fully controlled. Large ensembles can be generated with a Monte Carlo simulations that address at the same time the uncertainties in the input data and in the model formulation. This is achieved using the Polyphemus system, which is flexible enough to build various different models. The system offers a wide range of options in the construction of a model: many physical parameterizations, several numerical schemes and different input data can be combined. In addition, input data can be perturbed. In this paper, some 30 alternatives are available for the generation of a model. For each alternative, the options are given a probability, based on how reliable they are supposed to be. Each model of the ensemble is defined by randomly selecting one option per alternative. In order to decrease the computational load, as many computations as possible are shared by the models of the ensemble. As an example, an ensemble of 101 photochemical models is generated and run for the year 2001 over Europe. The models' performance is quickly reviewed, and the ensemble structure is analyzed. We found a strong diversity in the results of the models and a wide spread of the ensemble. It is noteworthy that many models turn out to be the best model in some regions and some dates.

1 Introduction

Due to the great uncertainties that arise in air quality modeling, ensembles of simulations are now considered in a wide range of applications. They are primarily built for uncertainty estimation. They can therefore evaluate the reliability of exposure studies based on model simulations. In the context of short-term forecasts, they can be used to evaluate risks, with probabilistic forecasts (e.g., threshold exceedence). Uncertainty estimation may also be useful for screening studies in which the impact of emission abatement, as predicted by numerical simulations, should be compared with the uncertainties. Data assimilation is another application where ensembles are often used: e.g., in the popular ensemble Kalman filter, the background-error covariance matrix is derived from them. For operational forecasts, an ensemble simulation may be sequentially aggregated so as to form forecasts better than the individual models.

A key step is the generation of the ensembles. They may be built (1) with perturbations in the input data to a single model or with an ensemble of input data (Straume, 2001), (2) with models that share little or no computer code (Galmarini et al., 2004; McKeen et al., 2005), or with models built on the same modeling platform. Uncertainty estimation, for instance, has been conducted with Monte Carlo simulations, thus with perturbations in the input data to a given model (Hanna et al., 1998; Beekmann and Derognat, 2003), and with different models built on the same platform (Mallet and Sportisse, 2006b). In data assimilation, the ensemble Kalman filter (Evensen, 1994) approximates background-error covariance matrices using an ensemble of simulations generated with perturbations in the input data (in air quality, e.g., Segers, 2002). A few studies make use of ensembles composed of models developed in different teams (for long-term simulations, see van Loon et al., 2007). For operational



Correspondence to: D. Garaud
(damien.garaud@cerea.enpc.fr)

forecasting, a weighted linear combination of models can form an improved forecast, as has been shown with an ensemble of models from different teams (Pagowski et al., 2006) and with an ensemble built on the same modeling platform (Mallet and Sportisse, 2006a; Mallet et al., 2009).

Whatever the application may be, a key step is the generation of the ensemble. In an ideal setting, one should take into account all uncertainty sources based on the best description available. Essentially, this would mean relying on Monte Carlo perturbations for uncertain input data like emissions, on the alternative descriptions available for data like land use cover, on calibrated ensemble weather forecasts, on different formulations for the subgrid parameterizations in the chemistry-transport models, on different numerical schemes in the chemistry-transport models. In this paper, we tend to this ideal setting with a simplified approach: we do not use a meteorological ensemble (the meteorological inputs are treated like other input data), and we rely on an alternative sampling approach to full Monte Carlo simulations. Nevertheless all uncertainty sources can be considered, and they are all taken into account at the numerical-simulation stage: no statistical correction is applied in a postprocessing. The approach described in this paper may be seen as a three-fold extension that of Mallet and Sportisse (2006b): new uncertainty sources are included, the uncertainty in input data is specifically taken into account, and the ensemble generation is entirely automatic.

From a technical point of view, building an ensemble of simulations is rather straightforward in the case of Monte Carlo simulations: one simply applies random perturbations to the input data of a single model. The perturbation scheme may be complex if it takes into account spatial and temporal correlations in the input fields and if advanced Monte Carlo variants are implemented. However this involves little complexity compared to building an ensemble composed of different models, e.g. of models based on various chemical mechanisms. There are essentially two ways (that may be combined) to form an ensemble with different models. One is to use existing models, usually developed in research groups. The resulting ensemble then includes a small number of models, say about ten. Another way is to generate different models within the same modeling platform: the models are assembled using basic components such as the chemical mechanism or the deposition module. Building such a platform is a tedious task, but it makes the generation of ensembles, even very large ones, practicable. In addition, the structure of the ensemble is fully controlled, which eases the scientific interpretation. This approach has been implemented in the modeling system Polyphemus (Mallet et al., 2007b), and it is described in this paper.

All the models considered in the platform assume that the concentrations of pollutants satisfy a system of partial differential equations and they approximate their solutions by discretizing the equations in an Eulerian framework. Each

equation of the system is an advection–diffusion–reaction equation of the form:

$$\frac{\partial c_i}{\partial t} = -\text{div}(\mathbf{V}c_i) + \text{div}\left(\rho\mathbf{K}\nabla\frac{c_i}{\rho}\right) + \chi_i(\mathbf{c},t) + E_i - \Lambda c_i, \quad (1)$$

where c_i is the concentration of the i th species, $\mathbf{c} = (c_1, \dots, c_S)$ is the vector of all concentrations, \mathbf{V} the wind vector, \mathbf{K} is the turbulent diffusion matrix, ρ the air density, χ_i the production term due to chemical reactions involving species i , E_i represents the emissions and Λc_i accounts for losses due to scavenging. The boundary conditions at ground level involve the surface emissions S_i and the deposition velocity v_i :

$$\mathbf{K}\nabla c_i \cdot \mathbf{n} = S_i - v_i c_i, \quad (2)$$

if \mathbf{n} is the upward-oriented normal to the ground.

All models solve a system of reactive transport equations like Eq. (1), but they rely on different coefficients in the equations (e.g., in the chemistry χ_i) and on different numerical schemes. The coefficients in the equations are estimated according to data from many sources (emission inventory, meteorological model, etc.) and many physical parameterizations (vertical diffusion, photolysis attenuation, etc.). We therefore uniquely define a model with (1) the input data and the physical parameterizations it uses and (2) its numerical schemes. Many alternative parameterizations, data sources and numerical schemes are available in Polyphemus – this flexibility is part of Polyphemus design principles. Most options are described in Sect. 2 which identifies the models that can be built on the platform.

In Sect. 3, the actual generation of the ensemble is addressed. This means selecting the models, also called ensemble members, which in turn means selecting the components (input data, physical parameterization, numerical options) for every model. One model is actually a set of programs that are launched in a given order. The simulation chain should be properly established to take into account the dependencies (e.g., the deposition velocities depend on the land use cover). It should also share the common computations among groups of models and distribute the computations over several computer processors in order to minimize the overall computational time. In addition to the changes in the physical and numerical formulations, several input fields that appear in the reactive transport equation are perturbed. It is assumed that the fields have a normal or a log-normal distribution, and they are perturbed accordingly.

In Sect. 4, the method is illustrated with a 101-member ensemble with gas-phase chemistry only.

2 Building one model

In this section, many options available in Polyphemus 1.5 (released 20 May 2009) for photochemical simulations are introduced. A summary of these options is given in Table 1.

Table 1. Alternatives for the physical parameterizations and numerical options. The numbers enclosed in brackets correspond to the occurrence probability of an option.

#	Parameterization	First option	Other option(s)	Comment
Physical parameterizations				
1.	Land use cover	USGS (0.5)	GLCF (0.5)	
2.	Chemistry	RACM (0.6)	RADM 2 (0.4)	
3.	Cloud attenuation	RADM method (0.6)	ESQUIF (0.4)	
4.	Critical relative humidity	Depends on σ (0.7)	Twolayers (0.3)	Used in the RADM method to compute cloud attenuation
5.	Vertical diffusion (K_z)	Troen & Mahrt (0.35)	Louis (0.3) Louis stable (0.35)	Troen & Mahrt kept in unstable conditions
6.	Deposition velocity	Zhang (0.5)	Wesely (0.5)	
7.	Coefficient R_a	Heat flux (0.7)	Moment flux (0.3)	For aerodynamic resistance (in deposition velocities)
8.	Emissions vertical distribution	Low (0.5)	Medium (0.5)	
9.	Photolysis rates	JPROC (0.7)	Zenith angle (0.3)	
Numerical issues				
10.	Time step	600 s (0.9)	1200 s (0.1)	
11.	Vertical resolution	5 layers (0.5)	9 layers (0.5)	The first layer height can be 50 m or 40 m
12.	First layer height	50 m (0.5)	40 m (0.5)	The top of every other layer does not change
13.	Vertical-wind diagnosis	$\text{div}(\rho V) = 0$ (0.5)	$\text{div}(V) = 0$ (0.5)	
14.	Minimal K_z	$0.2 \text{ m}^2 \text{ s}^{-1}$ (0.7)	$0.5 \text{ m}^2 \text{ s}^{-1}$ (0.3)	
15.	Minimal K_z in urban area	$0.2 \text{ m}^2 \text{ s}^{-1}$ (0.3)	$0.5 \text{ m}^2 \text{ s}^{-1}$ (0.3) $1.0 \text{ m}^2 \text{ s}^{-1}$ (0.4)	
16.	Vertical application of minimal K_z	Yes (0.8)	No (0.2)	If <i>no</i> , the lowest threshold is applied only to the top of the first layer, otherwise it is applied to all levels
17.	Exponent p	2 (0.7)	3 (0.3)	The value of the p exponent to compute the vertical diffusion coefficient (T&M only)
18.	Boundary layer height	raw value (0.6)	+10% (0.2) -10% (0.1) +20% (0.1)	Used to compute K_z (T&M only)

2.1 Physical formulation (parameterizations and input data)

2.1.1 Land use cover

The land use cover (LUC) describes the material covering the ground with a few categories. Polyphemus supports the USGS (U.S. Geological Survey) LUC with its 24 categories and the GLCF (Global Land Cover Facility) LUC that includes 14 categories. Both LUC have a $1 \times 1 \text{ km}^2$ resolution, with categories such as grassland, cropland, deciduous forest, urban areas, etc.

2.1.2 Chemistry

The chemical mechanism is a simplified representation of atmospheric chemistry, here related to photochemical activity. The mechanism includes species that may or may not exist as such, since many (real) chemical species are lumped into a few (model) species (e.g., the terminal alkenes are lumped into “OLT” in RACM). The mechanism describes the chem-

ical reactions between these species. Here, we consider two chemical mechanisms: RADM 2 (Stockwell et al., 1990) with 61 species and 157 reactions, and RACM (Stockwell et al., 1997) with 72 species and 237 reactions.

2.1.3 Critical relative humidity

The critical relative humidity is used to compute the cloud fraction, the cloudiness and the attenuation. One option is to compute the critical relative humidity q_c as a function of σ :

$$q_c = 1 - \alpha \sigma^a (1 - \sigma)^b \left(1 + \beta \left(\sigma - \frac{1}{2} \right) \right), \quad (3)$$

where $\sigma = \frac{P}{P_s}$, P is the pressure, P_s is the surface pressure, $\alpha = 1.1$, $\beta = \sqrt{1.3}$, $a = 0$ and $b = 1.1$. In another option (*two layers*), the critical relative humidity is simply constant in two distinct layers: $q_c = 0.75$ below 700 hPa and $q_c = 0.95$ above.

2.1.4 Photolysis

Two options are considered. Clear sky photolysis rates J_{clear} can be those computed by the JPROC software which is part of the Community Multiscale Air Quality (CMAQ) Modeling System (Byun and Ching, 1999), or they can be computed based on the zenith angle alone. The photolysis rates are of the form $J = \mathcal{A}J_{\text{clear}}$ where \mathcal{A} is the attenuation.

2.1.5 Attenuation

The cloud attenuation \mathcal{A} measures the decrease in the rates of photolysis reactions when solar radiation is partially absorbed or reflected by clouds. It can be computed using the RADM method (Madronich, 1987; Chang et al., 1987):

$$\begin{cases} \mathcal{A}_b = 1 - \min(1, \mathcal{N}_m + \mathcal{N}_h)(1 - 1.6Tr \cos Z) \\ \mathcal{A}_a = 1 + \min(1, \mathcal{N}_m + \mathcal{N}_h)(1 + (1 - Tr) \cos Z) \end{cases} \quad (4)$$

where \mathcal{A}_b and \mathcal{A}_a are the attenuations below and above the clouds, \mathcal{N}_m and \mathcal{N}_h are the medium cloudiness and the high cloudiness, Tr is the cloud transmissivity and Z is the zenith angle. The photolysis rates below and above the clouds are respectively $J_b = \mathcal{A}_b J_{\text{clear}}$ and $J_a = \mathcal{A}_a J_{\text{clear}}$.

A second parameterization was developed after the ESQUIF campaign (ESQUIF, 2001), using measurements of the photolysis rates for NO_2 . The attenuation is approximated by

$$\mathcal{A} = (1 - a\mathcal{N}_h)(1 - b\mathcal{N}_m)e^{-cB}, \quad (5)$$

where a , b , c and B are constants.

2.1.6 Vertical diffusion

The vertical diffusion coefficient K_z ($\text{m}^2 \text{s}^{-1}$) is the third diagonal term of the turbulent diffusion matrix \mathbf{K} (Eq. 1). This coefficient is computed at the interfaces of the model layers and can be estimated with two parameterizations. K_z may be computed with the Louis parameterization (Louis, 1979) at interface k :

$$K_{z,k} = L_k^2 F(Ri_k) \left[\left(\frac{\Delta U_k}{\Delta z_k} \right)^2 + \left(\frac{\Delta V_k}{\Delta z_k} \right)^2 \right], \quad (6)$$

where L_k is the mixing length at level k , Ri is the Richardson number and F is the stability function. Alternatively, K_z can be computed with the Troen&Mahrt parameterization (Troen and Mahrt, 1986):

$$K_{z,k} = u_* \kappa z_k \Phi_{m,k}^{-1} \left(1 - \frac{z_k}{\text{PBLH}} \right)^p, \quad (7)$$

where u_* is the friction velocity, κ is the von Kármán constant, $\Phi_{m,k}$ is the non-dimensional shear and PBLH is the planetary boundary layer height. This parameterization is more parametric and more robust than the Louis parameterization. A third option is a combination of both parameterizations: the Louis parameterization used in stable conditions and the Troen&Mahrt parameterization in unstable conditions.

In the Troen&Mahrt parameterization (7), the exponent p may be 2 or 3. In the ensemble generation, the boundary layer height PBLH may be perturbed at that stage.

In addition to the selected parameterization, a few options remain with the minimum value for K_z , the minimum value of K_z over urban areas, and whether the minimum values for K_z are applied only in the first layer or in all layers.

2.1.7 Deposition velocities

The deposition velocities (m s^{-1}) are assumed to be in the form

$$V_d = \frac{1}{R_a + R_b + R_c}, \quad (8)$$

where R_a is the aerodynamic resistance, R_b is the quasi-laminar sublayer resistance and R_c is the canopy resistance. R_a can be computed with the heat flux or the momentum flux. R_c can be computed by the Zhang parameterization (Zhang et al., 2003) or the Wesely parameterization (Wesely, 1989). It depends on the LUC.

2.1.8 Emissions

Pollutant emissions are usually divided into two parts: biogenic emissions emitted by vegetation and anthropogenic emissions originating from human activities (transport, industries, etc.). The biogenic emissions are surface emissions computed following Simpson et al. (1999). They depend on LUC. At the European scale, anthropogenic emissions are estimated by EMEP (European Monitoring and Evaluation Programme). EMEP provides annual quantities for a few pollutants (NO_x , VOC, SO_2 , CO and aerosols) and for 10 different sectors called SNAP (Selected Nomenclature for Air Pollution). These annual emissions are multiplied by monthly, daily (Saturday, Sunday, week days) and hourly factors which depend on the country and SNAP. Finally the emissions are split into surface and volume emissions, according to SNAP. The vertical distribution of the volume emissions is subject to a choice; here, we consider two options: a *low* distribution and a *medium* distribution – the former distribution assumes that the pollutants are released closer to the ground than with the latter distribution. Table ?? describes the 10 different SNAP and the emission vertical distribution for the options *low* and *medium*.

2.2 Numerical issues

In Polyphemus, three numerical schemes (for advection, diffusion and chemistry) are assembled to form a numerical integrator, called Polair3D (Boutahar et al., 2004), whatever schemes are used. The numerical integrators share the coordinate system: regular horizontal grid in latitude/longitude, vertical levels with fixed altitudes (in meters from the ground). The integration makes use of operator splitting: in

one time step, the advection is integrated first, then the diffusion and finally the chemistry.

Very few numerical options are considered here, because the uncertainty sources were mainly found in the physical formulation and in the input data (Mallet and Sportisse, 2006b). In Mallet et al. (2007a), a detailed study of many numerical options shows that the splitting time step and the advection scheme may have a significant impact. In the present study, the advection scheme is not an option: a third-order direct-space-time scheme with flux limiting (Verwer et al., 2002) is used in all the models. On the other hand, the splitting time step is an option (see below). Both diffusion and chemistry are integrated using a second-order Rosenbrock scheme (Verwer et al., 2002).

2.2.1 Time step

The (splitting) time step is set to 600 s (the usual time step) or 1200 s.

2.2.2 Simulation grid

The horizontal resolution is set to 0.5° in all simulations.

Along the vertical, the grid is made up of 5 layers or 9 layers, up to 3000 m. The height of the first layer may be 40 m or 50 m. Consequently, there are 4 possible vertical grids. Note that a change in the vertical grid has consequences in almost all computations.

2.2.3 Vertical-wind diagnosis

The vertical wind may be reconstructed from the horizontal-wind components by solving the equation $\text{div}(\rho \mathbf{V}) = 0$ where ρ is the air density and \mathbf{V} the wind vector.

It may also be estimated with the simplified equation $\text{div}(\mathbf{V}) = 0$. In this case, the diffusion term in Eq. 1 is changed, for consistency, to $\text{div}(\mathbf{K}\nabla c)$.

This diagnosis is carried out after the horizontal winds have been perturbed.

2.3 Other options

The options previously mentioned are summarized in Table 1. Other options are available in Polyphemus. They are not reported in this paper because they are not used in the illustrative example (Sect. 4).

Many of the other options are related to aerosols. Polyphemus includes a size-resolved aerosol module called SIREAM (Debry et al., 2007) and related preprocessing (anthropogenic emissions, sea salt emissions, deposition, boundary conditions). The aerosol module offers numerous options: choice of the aqueous module, nucleation model (binary, ternary), heterogeneous reactions, calculation of the wet diameter, aerosol density, thermodynamics module, etc. This ability was used in the sensitivity study by Sartelet et al.

(2008), and it should be used in the generation of an ensemble. In the preprocessing steps, several options also relate to aerosols, e.g., the parameterization for estimating the emissions of sea salt could that of Smith and Harrison (1998) or that of Monahan et al. (1986).

3 Ensemble generation

In order to build a large ensemble and to take into account all possible options, an automatic procedure is necessary. In addition to the changes in the model formulation, the procedure includes a perturbation step (Sect. 3.1): the input data of the numerical model are perturbed so as to take into account additional uncertainty sources. After that step, all simulations are completely defined and launched (Sect. 3.1.1 and 3.2).

3.1 Input data perturbation

In the final stage of a simulation, the numerical integration of Eq. (1) is carried out with the selected numerical scheme. At this stage, the fields that appear in the equation are also perturbed.

Estimations of the uncertainties were established by experts and reported in Hanna et al. (1998, 2001), for 18 km and 12 km resolutions, in regions of eastern USA, and for a few days. These estimations should be seen as guidelines to be adapted to the simulation region, to the resolution of the simulation, to the time span, and to other considerations on the quality of the fields. For instance, the uncertainty in the values of a field should decrease when the resolution gets higher. In addition, a few ensembles were generated in order to roughly calibrate the uncertainty parameters, based on comparisons with observations (not reported here). Several fields are given a distribution, normal or log-normal, and an uncertainty range determined by a parameter α . It is assumed that any value of the field is the random variable \hat{p} that satisfies

$$- \hat{p} = p + \frac{\gamma}{2}\alpha \text{ for a normal distribution,}$$

$$- \hat{p} = p\sqrt{\alpha}^\gamma \text{ for a log-normal distribution,}$$

where γ is distributed according to $\mathcal{N}(0, 1)$, and p is a (deterministic and known) value which is assumed to be the median of \hat{p} .

For a normal distribution, $\hat{p} \in [p - \alpha, p + \alpha]$ has a probability of 95%. Thus $\pm\alpha$ is an uncertainty range, around the mean (or median) p , associated with a probability of 0.95. α is twice the standard deviation of \hat{p} . For a log-normal distribution, the same applies to $\ln \hat{p}$, with an uncertainty range of width $\pm\frac{1}{2}\ln\alpha$. The probability that $\hat{p} \in [p/\alpha, \alpha p]$ is 0.95.

Note that the perturbation will not depend on the date or on the position. We simply assume that $\hat{p}(t, x) = p(t, x)\frac{\gamma}{2}\alpha$ (or $\hat{p}(t, x) = p(t, x)\sqrt{\alpha}^\gamma$) for any date t and position x . Since γ does not depend on t and x , two values $\hat{p}(t, x)$ and $\hat{p}(t', x')$

Table 2. Perturbation of input data. The last column is the parameter α that defines the uncertainty range. If the median value of a normally-distributed random variable \hat{p} is p , the probability that $\hat{p} \in [p - 2\alpha, p + 2\alpha]$ is 0.95. If \hat{p} is log-normally distributed, the probability that $\hat{p} \in [p/\alpha, \alpha p]$ is 0.95.

#	Field	Source	Distribution	Uncertainty range
1.	Horizontal-wind module	ECMWF	Log-normal	1.5
2.	Horizontal-wind angle	ECMWF	Normal	$\pm 40^\circ$
3.	Temperature	ECMWF	Normal	± 3 K
4.	O ₃ boundary conditions	Mozart 2	Log-normal	2.0
5.	NO _x boundary conditions	Mozart 2	Log-normal	3.0
6.	VOCs (Volatile Organic Compounds) boundary conditions	Mozart 2	Log-normal	2.0
7.	NO _x anthropogenic emissions	EMEP	Log-normal	1.5
8.	VOCs anthropogenic emissions	EMEP	Log-normal	1.5
9.	Biogenic emissions	Computed	Log-normal	2.0
10.	Vertical diffusion	Computed	Log-normal	1.7
11.	Deposition velocities	Computed	Log-normal	1.5
12.	Photolysis rates	Computed	Log-normal	1.4

are fully correlated (correlation equals 1) for a normal distribution. In the log-normal case, $\ln \hat{p}(t, x)$ and $\ln \hat{p}(t', x')$ are fully correlated.

The list of the perturbed fields and the corresponding values of α are shown in Table 2. These values were used in the example (Sect. 4). The list of input fields includes meteorological variables, the boundary conditions, the emissions for different species and variables related to chemical species such as the deposition velocities or the photolysis rates. These input data come from different models (ECMWF, Mozart 2, EMEP) or are processed during the pre-processing.

Once the distribution and the parameter α are determined, the actual perturbation is not given by a random sampling of γ . The actual perturbed value is randomly and uniformly selected in a set of three values: the unperturbed value (i.e., the median) $\tilde{p}_0 = p$, and two other points \tilde{p}_1 and \tilde{p}_2 defined below. The points are chosen so that the empirical mean and the empirical standard deviation are the same as the mean and the standard deviation of \hat{p} , thus

$$- E(\hat{p}) = \frac{\tilde{p}_0 + \tilde{p}_1 + \tilde{p}_2}{3};$$

$$- \text{Var}(\hat{p}) = \frac{(\tilde{p}_0 - \bar{p})^2 + (\tilde{p}_1 - \bar{p})^2 + (\tilde{p}_2 - \bar{p})^2}{2}.$$

In the case where $\hat{p} \sim \mathcal{N}(p, \frac{1}{4}\alpha^2)$:

$$\tilde{p}_0 = p;$$

$$\tilde{p}_1 = p - \frac{1}{2}\alpha;$$

$$\tilde{p}_2 = p + \frac{1}{2}\alpha.$$

When \hat{p} is log-normally distributed:

$$\tilde{p}_0 = p;$$

$$\tilde{p}_1 = p\sqrt{\alpha}^{\gamma_1};$$

$$\tilde{p}_2 = p\sqrt{\alpha}^{\gamma_2},$$

with

$$\gamma_1 = 2\log\left(\frac{3\beta-1-\sqrt{\Delta}}{2}\right)/\log\alpha;$$

$$\gamma_2 = 2\log\left(\frac{3\beta-1+\sqrt{\Delta}}{2}\right)/\log\alpha,$$

and

$$\beta = \exp\left(\frac{1}{8}\log^2\alpha\right);$$

$$\Delta = 4\beta^4 - 7\beta^2 + 6\beta - 3.$$

3.1.1 Models automatic selection

The selection of the models to be included in the ensemble is carried out randomly. A probability is given to each option. The numbers between brackets in Table 1 are these probabilities. In any alternative, the sum of the probabilities equals 1. Each perturbation in the input data (Table 2) is uniformly selected from three possible values (Sect. 3.1). A model is defined once an option has been selected for any alternative (18 alternatives are shown in Table 1, 12 perturbations are listed in Table 2).

Except for the perturbations in the input data, the probabilities are chosen according to the confidence put in each option. There is no direct indicator to determine these probabilities. If two parameterizations are available for a given option, the choice lies between giving a probability one to a parameterization (no uncertainty), and giving 0.5 to both parameterizations (which leads to the largest uncertainty). If one option is supposed to be more accurate (a priori quality of a parameterization, finer grid resolution, etc.) or if it is usually associated with better model results (comparison with observations), its weight should be higher than that of alternative choices. For example, a time step equal to 600 s is supposed to give more accurate results than 1200 s – the numerical solution converges to the exact solution as the time

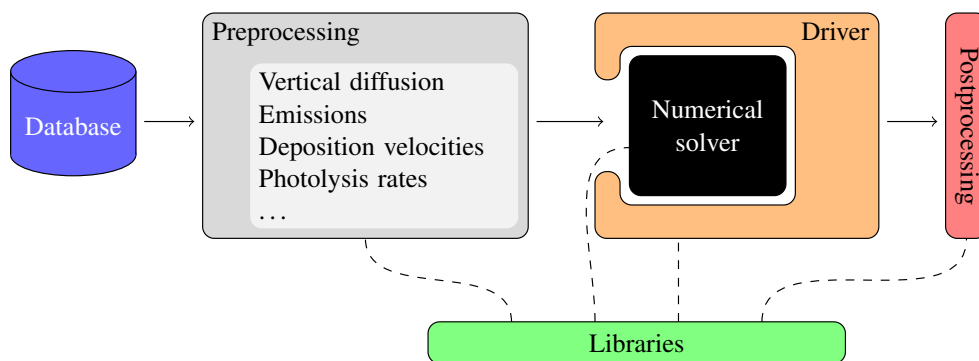


Fig. 1. A view on Polyphemus design: database storing all raw data, preprocessing stages for most physical computations, drivers in which the numerical solver is embedded, postprocessing and libraries that may be called at any time.

step tends to 0. Therefore, a higher probability is associated with the time step fixed to 600 s. Another example is the chemical mechanism RACM which is more detailed than RADM 2, and which has shown slightly better results in several studies (Gross and Stockwell, 2003).

3.2 Technical aspects

The structure of the Polyphemus system contains four (mostly) independent levels: data management, physical parameterizations, numerical solvers and high-level methods such as data assimilation. Figure 1 illustrates the structure of the modeling platform.

During the first stage, several C++ programs carry out the preprocessing. This is the most important part of the simulation process, both in terms of simulation definition (the physics is set there) and computer code. Almost all terms of the reactive transport Eq. (1) are computed at this stage. The computations are split into several programs to ensure flexibility. For instance, there is one program to process land use cover (actually two programs: one for USGS data and another for GLCF data), one program for the main meteorological fields, one program to compute biogenic emissions, another program for anthropogenic emissions, etc. Another example is the vertical diffusion coefficient: one program computes it with Louis parameterization and another with the Troen&Mahrt parameterization. In addition, these programs have several options (e.g., the parameter p in the Troen&Mahrt parameterization, see Eq. 7). The use of multiple programs makes it an efficient system to build an ensemble. Adding new options is easy since one may simply add a new program (or add the option into an existing program). Moreover the computations are well managed. For example, if two models have the same options except the deposition velocity, all computations except those depending on deposition (i.e., the computation of the deposition velocities, and the numerical integration of the reactive transport equation) will be shared.

In the second stage, the numerical solver carries out the time integration of the reactive transport equation. The numerical solver is actually embedded in a structure called the “driver”. The driver is primarily in charge of perturbing the input data as detailed in Sect. 3.1.

At a postprocessing stage, the ensemble is completely generated and the results are analyzed. At all stages, a few libraries, mainly in C++ and Python, offer support, especially for data manipulation.

Disk space usage is optimized since the models can share part of their preprocessing. Moreover, the perturbed input fields (Table 2) are not stored; only the unperturbed fields (medians) are stored, and the driver applies the perturbations during the simulation.

Python scripts generate the identities (i.e., the set of options and perturbations) of all models to be launched. The corresponding configuration files are created. The scripts then launch the preprocessing programs and the simulations. The simulations can obviously be run in parallel, so the scripts can launch the programs over SSH on different machines and processors. The only constraint lies in the dependencies between the programs: e.g., the deposition velocities must be computed after the meteorological fields because they depend on winds (among other fields). Groups of programs are defined with different priorities, and the scripts launch one group after the other. It is possible to relaunch parts of the ensemble computations. It is also possible to add new models (new simulations) after an initial ensemble has been generated. The Python code is available in the module EnsembleGeneration, from Polyphemus 1.5.

The same approach may be applied to another modeling system providing enough options (in the model formulation) are available. This requires that significant diversity is maintained in the system. In particular, when a new formulation (e.g., a more accurate chemistry) is developed, the previous formulation should remain available to the user. The rationale is that, while a formulation may seem better from a

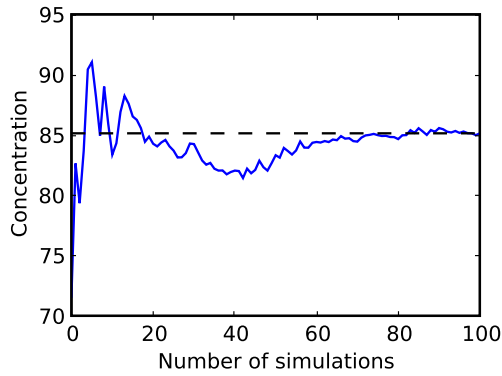


Fig. 2. The empirical mean of the ozone peaks averaged over all stations of network 1 and during 2001. It seems to have converged after about 70 simulations.

deterministic point of view (based on a priori considerations or on performance analysis), the previous formulation still has a significant probability (though lower than that of the new formulation) from a stochastic point of view.

4 An example of 101-member ensemble

With the previous method, about 620 billion models can be generated. An ensemble of 101 models is built and run throughout the year 2001 over Europe ($[10.75^\circ \text{W}, 22.75^\circ \text{E}] \times [34.75^\circ \text{N}, 57.75^\circ \text{N}]$). The models are not simplified to reduce the computational costs. All models have a 0.5° horizontal resolution, which is a usual resolution. Because the total computational cost is high, the ensemble size is limited to 101 simulations. This size is enough at least for the spatio-temporal empirical mean (of ozone peaks) to converge, as shown in Fig. 2.

Six reference models are included in the ensemble. These models are not generated automatically, but each of them corresponds to a possible combination of options in that they could have been selected by the automatic procedure.

Aerosols are not taken into account in these simulations. The output stored on disk are the hourly concentrations in the first layer for O_3 , NO , NO_2 and SO_2 – which already amounts to 45 Gb of data.

Section 4.1 briefly summarizes which members are included in the ensemble. Although this paper is a technical description of the ensemble generation procedure, we aim to provide insight into the ensemble structure. We review the performance of the models, compared to ground observations, in Sect. 4.2. We analyze the spread of the ensemble in Sect. 4.3. We do not address more complex issues like probabilistic forecasts, uncertainty estimation or sequential aggregation.

4.1 Experiment setup

In Table 1, a probability is associated with every option. The models are built according to these probabilities, but the actual frequency of an option in the 101-member ensemble may differ slightly because of the random sampling. The occurrence frequency (in percentages) of each parameterization, numerical option and field perturbation in the 101-member ensemble are shown in Table 3. For the field perturbations, there are three options: no perturbation (raw data), increased values in the field ($p\alpha$ if $p \geq 0$, or $p + \alpha$) and decreased values (Sect. 3.1).

The six additional models can be seen as reference models. They are built with the parameterizations that we trust the most, and without any perturbation in the input field. If we had to build a model for forecast, we would a priori choose one of them. They are formed with the parameterizations and numerical options from the first column of Table 1 but for the vertical diffusion parameterization and the mass conservation. Considering the three options for vertical diffusion (line 5) and the two options for vertical-wind diagnosis (line 13), six models may be constructed. These are listed in Table 4.

4.2 Evaluation of the ensemble members

4.2.1 Performance measures

In order to evaluate a model performance, n available observations o_i from different ground stations are compared with the corresponding simulated concentrations y_i , using

1. the root mean square error:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - o_i)^2};$$

2. the correlation:

$$\text{corr} = \frac{\sum_{i=1}^n (y_i - \bar{y})(o_i - \bar{o})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^n (o_i - \bar{o})^2}},$$

where $\bar{o} = \sum_{i=1}^n o_i$ and $\bar{y} = \sum_{i=1}^n y_i$;

3. the bias factor:

$$\text{BF} = \frac{1}{n} \sum_{i=1}^n \frac{y_i}{o_i}.$$

In practice, not all observations are retained. Stations that fail to provide observations at over 10% of all considered dates are discarded as these stations may not be reliable.

Table 3. Occurrence frequency of each parameterization, numerical option and field perturbation for the 101-member ensemble. As for the perturbations, “raw” means no perturbation, “raw⁻” means lower value after perturbation (p/α or $p-\alpha$) and “raw⁺” means higher value after perturbation ($p\alpha$ or $p+\alpha$).

#	Option			
Physical parameterizations				
1.	Land use cover	USGS (50)	GLCF (50)	
2.	Chemistry	RACM (61)	RADM 2 (39)	
3.	Attenuation	RADM method (50)	ESQUIF (50)	
4.	CRH	Depends on σ (75)	Two layers (25)	
5.	K_z	T&M (44)	Louis (30)	Louis stable (26)
6.	Deposition velocity	Zhang (55)	Wesely (45)	
7.	Coefficient Ra	Heat flux (74)	Moment flux (36)	
8.	Emissions vertical distribution	Low (54)	Medium (46)	
9.	Photolysis rates	JPROC (88)	Zenith angle (12)	
Numerical issues				
10.	Time step	600 s (91)	1200 s (9)	
11.	Vertical resolution	5 layers (48)	9 layers (52)	
12.	First layer height	50 m (40)	40 m (60)	
13.	Vertical-wind diagnosis	$\text{div}(\rho V) = 0$ (52)	$\text{div}(V) = 0$ (48)	
14.	Minimal K_z ($\text{m}^2 \text{s}^{-1}$)	0.2 (66)	0.5 (34)	
15.	Minimal K_z in urban area ($\text{m}^2 \text{s}^{-1}$)	0.2 (30)	0.5 (35)	1.0 (35)
16.	Vertical application of minimal K_z	Yes (84)	No (16)	
17.	Exponent p	2 (75)	3 (25)	
18.	Boundary layer height	raw value (61)	+10% (18)	-10% (7) +20% (14)
Input data				
19.	Temperature (K)	raw (39)	raw ⁻ (34)	raw ⁺ (27)
20.	Horizontal-wind angle (degrees)	raw (35)	raw ⁻ (31)	raw ⁺ (34)
21.	Horizontal-wind velocity (m s^{-1})	raw (36)	raw ⁻ (40)	raw ⁺ (24)
22.	K_z ($\text{m}^2 \text{s}^{-1}$)	raw (33)	raw ⁻ (32)	raw ⁺ (35)
23.	O_3 boundary conditions ($\mu\text{g m}^{-3}$)	raw (33)	raw ⁻ (36)	raw ⁺ (31)
24.	NO_x boundary conditions ($\mu\text{g m}^{-3}$)	raw (29)	raw ⁻ (35)	raw ⁺ (36)
25.	VOCs boundary conditions ($\mu\text{g m}^{-3}$)	raw (35)	raw ⁻ (37)	raw ⁺ (28)
26.	Biogenic emissions	raw (34)	raw ⁻ (28)	raw ⁺ (38)
27.	NO_x emissions	raw (34)	raw ⁻ (35)	raw ⁺ (31)
28.	VOCs emissions	raw (27)	raw ⁻ (25)	raw ⁺ (48)
29.	Deposition velocities (cm s^{-1})	raw (35)	raw ⁻ (32)	raw ⁺ (33)
30.	Photolysis rates	raw (34)	raw ⁻ (37)	raw ⁺ (29)

For ozone, the observations from three networks are considered:

- Network 1 is composed of 243 urban and regional stations, primarily in France and Germany (116 and 81 stations respectively). It provides about 1 365 000 hourly concentrations and 61 000 peaks.
- Network 2 includes 96 EMEP stations (regional stations distributed over Europe), with about 776 700 hourly observations and 33 300 peaks.

- Network 3 includes 371 urban and regional stations in France. It provides 2 800 000 hourly measurements and 122 000 peaks. Note that it includes most of the French stations of network 1.

4.2.2 The models' performance on ozone

Table 5 shows the performance of the six reference models for ozone and of the best model in the ensemble. The best model is selected with respect to the RMSE for the

Table 4. Description of the 6 reference models.

#	Vertical diffusion	Vertical-wind diagnosis
R0.	T&M	$\text{div}(\rho \mathbf{V}) = 0$
R1.	T&M	$\text{div}(\mathbf{V}) = 0$
R2.	Louis stable – T&M unstable	$\text{div}(\rho \mathbf{V}) = 0$
R3.	Louis stable – T&M unstable	$\text{div}(\mathbf{V}) = 0$
R4.	Louis	$\text{div}(\rho \mathbf{V}) = 0$
R5.	Louis	$\text{div}(\mathbf{V}) = 0$

considered network and target (ozone peaks or ozone hourly concentrations). It is noteworthy that, except for network 2 and for hourly concentrations, there is always at least one model in the 101-member ensemble which is better than the six reference models (according to the RMSE and the correlation). The automatic generation of 101 models therefore created models that are as good as or better than the models derived from experience.

It also generated models with poor performance. Figure 3 shows the performance of the 101 models sorted according to the mean, bias factor, correlation and RMSE. The performance can obviously vary greatly.

4.2.3 The best model

In this section, we define the “best model” as the model with the lowest RMSE. Of course, it depends on the target (the network, the pollutant, the time period), and considering the RMSE only is not enough to identify the best model, if any can be identified, as a modeler would do. Still this gives insights on the performance of models automatically generated.

Model 98 in the 101-member ensemble is the best model for ozone peaks on network 1, for ozone hourly concentrations and ozone peaks on network 2 (Table 5). For these targets, it beats the reference models. Several parameterizations and numerical options of model 98 are the same as those of the reference models (photolysis rates, deposition velocities, time step, etc.), but several selected options are unexpected. For instance, its chemical mechanism is RADM 2, and four fields are perturbed. See Table 6 for the complete description of model 98. It is interesting to note that (1) the random sampling generates several models with good performance (compared to the observations, with the RMSE), (2) the random sampling generates a model with lower square errors (over a long time period) than the models tuned by the modelers.

Among the 101 simulations, the median RMSE is about $27 \mu\text{g m}^{-3}$ and the median correlation is close to 0.73.

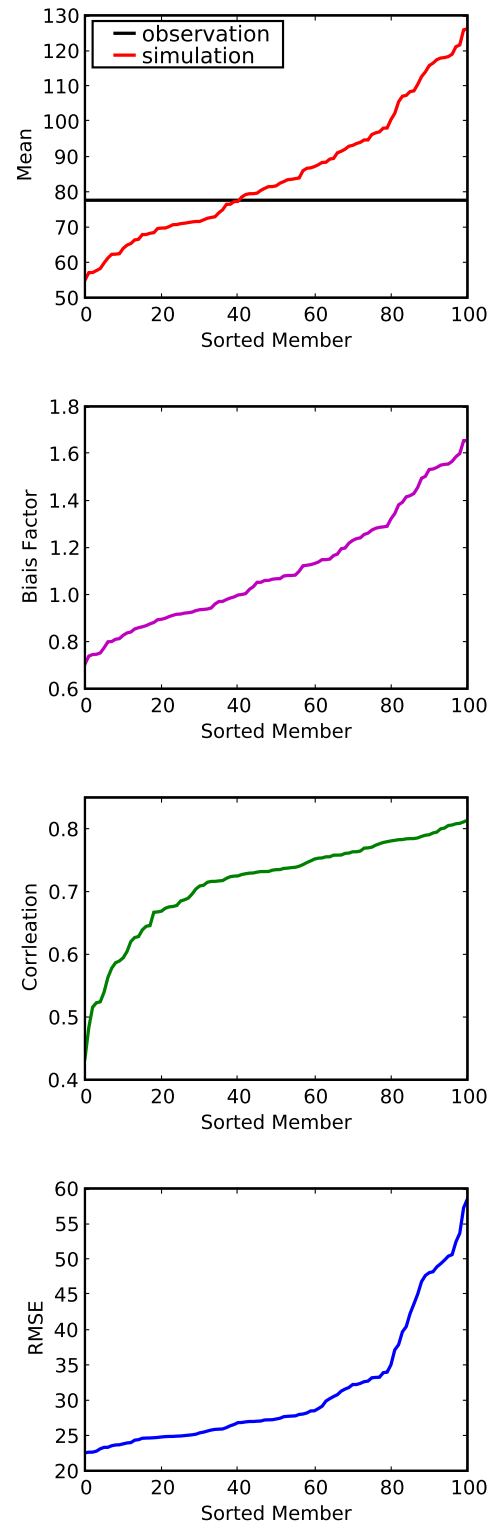


Fig. 3. Mean ($\mu\text{g m}^{-3}$), bias factor, correlation and RMSE ($\mu\text{g m}^{-3}$) for ozone peaks on network 1. In each plot, the models are sorted according to the indicator.

Table 5. Statistical measures for the 6 reference models and the best model from the 101-member ensemble, for hourly ozone concentrations and hourly ozone peaks. R0–5 refer to the 6 reference models.

#	BF	mean	corr	RMSE	#	BF	mean	corr	RMSE
Network 1 – Hourly					Network 1 – Peak				
R0.	1.06	62.0	0.67	28.09	R0.	1.10	85.1	0.76	24.54
R1.	0.96	55.5	0.68	25.55	R1.	1.00	76.9	0.77	23.19
R2.	1.21	72.3	0.68	31.19	R2.	1.13	87.1	0.78	24.49
R3.	1.10	65.1	0.68	26.85	R3.	1.03	78.8	0.78	22.82
R4.	0.89	51.1	0.69	25.87	R4.	0.98	75.6	0.79	23.30
R5.	0.82	46.9	0.70	25.74	R5.	0.91	70.1	0.78	23.95
48.	0.9	51.81	0.73	22.42	98.	1.08	83.6	0.80	22.54
R0.	0.99	65.2	0.64	25.28	R0.	1.06	84.2	0.73	21.66
R1.	0.90	59.0	0.64	24.90	R1.	0.97	76.7	0.73	21.51
R2.	1.12	74.1	0.65	25.74	R2.	1.09	86.0	0.74	21.22
R3.	1.02	67.3	0.65	23.52	R3.	0.99	78.4	0.74	20.74
R4.	0.83	54.2	0.66	26.47	R4.	0.93	74.4	0.74	23.36
R5.	0.77	50.1	0.66	27.75	R5.	0.87	69.5	0.73	24.80
98.	1.05	69.1	0.67	24.02	98.	1.04	82.6	0.76	20.24
R0.	1.12	65.1	0.64	31.18	R0.	1.15	86.0	0.76	26.59
R1.	1.01	58.0	0.66	27.21	R1.	1.04	77.3	0.76	23.98
R2.	1.27	75.2	0.66	35.98	R2.	1.18	88.0	0.77	27.09
R3.	1.15	67.4	0.67	30.44	R3.	1.06	79.2	0.77	24.15
R4.	0.96	54.3	0.68	26.34	R4.	1.02	76.9	0.79	23.09
R5.	0.89	49.7	0.69	25.13	R5.	0.95	71.1	0.79	22.79
48.	0.93	52.8	0.72	23.29	99.	0.91	68.5	0.81	22.41

4.3 Ensemble variability

Every model in the ensemble is unique, but one may ask whether the ensemble contains enough information and has a rich structure. For example, the ensemble should not be clustered into distinct groups of similar models. One measure of the difference between two models is the number of options that differ between them. Interestingly enough, two models with a similar RMSE can be made with many different options: for example models 98 and 58, which have close RMSEs (22.54 and 23.65 respectively, ozone peak, network 1), are generated with 17 different options (out of 30) shown in Table 7. This fact can be observed with the whole ensemble. In Fig. 4, the models are sorted according to their RMSE for ozone peaks on network 1 (model 0 has the lowest RMSE, and model 100 has the highest RMSE), and the matrix of the differences between the models (measured with the number of differing options) is shown. No overall structure can be identified. This tends to show that quite different models can achieve similar performance. The RMSE, seen as a function of the parameters, seems to have many local minima.

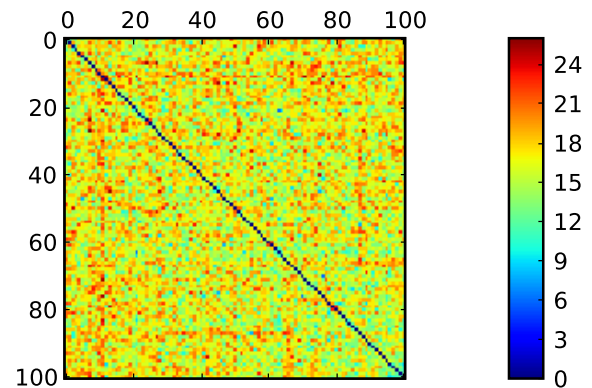


Fig. 4. Matrix of the number of different options between two models. The models are sorted according to the RMSE (from the best to the worst value).

On the other hand, the output of the best models are correlated. This is shown in Fig. 5 with the correlation computed with all ozone peaks observed in network 1. Two skillful models therefore have a similar spatio-temporal variability.

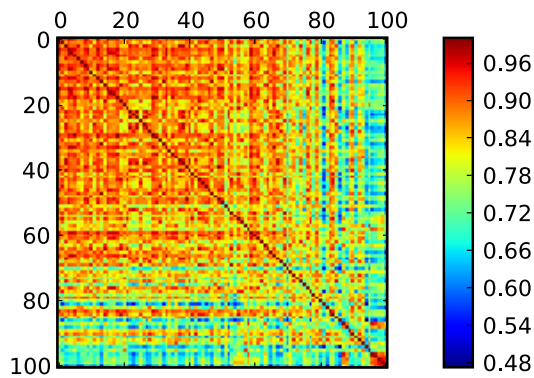


Fig. 5. Matrix of correlation between all observed ozone peaks (on network 1) and the corresponding model-concentrations. The models are sorted according to the RMSE (from the best to the worst value).

Table 6. Description of the best model.

#	Name	
Physical parameterizations		
1.	Land use cover	GLCF
2.	Chemistry	RADM 2
3.	Attenuation	ESQUIF
4.	CRH	Two layers
5.	K_z	T&M unstable – Louis stable
6.	Deposition velocity	Zhang
7.	Coefficient R_a	Heat flux
8.	Emissions vertical distribution	Low
9.	Photolysis rates	JPROC
Numerical issues		
10.	Time step	600 s
11.	Vertical resolution	5 layers
12.	First layer height	40 m
13.	Mass conservation	$\text{div}(\rho V) = 0$
14.	Minimal K_z ($\text{m}^2 \text{s}^{-1}$)	0.2
15.	Minimal K_z in urban area ($\text{m}^2 \text{s}^{-1}$)	1.0
16.	Vertical application for minimal K_z	Yes
17.	Exponent p	3
18.	Boundary layer height	raw value
Input data		
19.	Temperature (K)	raw ⁺
20.	Horizontal-wind angle (degrees)	raw
21.	Horizontal-wind velocity (m s^{-1})	raw
22.	K_z ($\text{m}^2 \text{s}^{-1}$)	raw ⁻
23.	O_3 boundary conditions ($\mu\text{g m}^{-3}$)	raw
24.	NO_x boundary conditions ($\mu\text{g m}^{-3}$)	raw
25.	VOCs boundary conditions ($\mu\text{g m}^{-3}$)	raw
26.	Biogenic emissions	raw ⁺
27.	NO_x emissions	raw
28.	VOCs emissions	raw ⁻
29.	Deposition velocities (m s^{-1})	raw
30.	Photolysis rates	raw

Table 7. A comparison between the model 98 and 58.

Name	Model 98	Model 58
Chemical Mechanism	RADM 2	RACM
Cloud attenuation	ESQUIF	RADM
Critical relative humidity	on 2 layers	with σ
Vertical diffusion	Troen & Mahrt	Louis
	unstable –	
	Louis stable	
Coefficient R_a	Heat flux	Moment flux
Vertical resolution	5 levels	9 levels
Time step	600 s	1200 s
Exponent p for K_z	3	2
First layer height	40 m	50 m
Minimal K_z in urban area	1.0	0.5
Temperature	raw ⁺	raw
NO_x boundary conditions	raw	raw ⁺
VOCs boundary conditions	raw	raw ⁺
Biogenic emissions	raw ⁺	raw
NO_x emissions	raw	raw ⁺
VOCs emissions	raw ⁻	raw
Deposition velocities	raw	raw ⁺

These high correlations are partly due to the structure of ozone fields. Because of the physical constraints, two reasonable ozone fields necessarily share a set of common features, such as higher concentrations in the south compared to the north, or low concentrations at high NO emission sources. However, two skillful models can show significant differences in their spatial patterns, as Fig. 6 demonstrates.

Figures 7, 8, 9, and 10 show the temporal mean of the concentration map of the fifth reference model and of a model from the 101-member ensemble, for O_3 , NO , NO_2 and SO_2 respectively. Again, the physical constraints make the models reproduce specific features, like high NO concentrations only at emission locations, but significant differences are found.

Figures 11 shows the mean daily profiles of all models from the 101-member ensemble, for O_3 , NO , NO_2 and SO_2 respectively. For the species O_3 and NO_2 , the daily profiles are computed on network 3 whereas the daily profiles for the species NO and SO_2 are computed with all cells. All models produce a similar profile shape, which is due to the physical phenomena accounted for in every model and the fact that these profiles are highly averaged (whole year, and full domain or all stations). The means can differ a lot, and, obviously, not all models are equally likely.

Nevertheless, even if the average performance of a model is very low, it may produce the best forecast at some location or some date. In other words, from a stochastic viewpoint, the model may have a very low probability, but it is

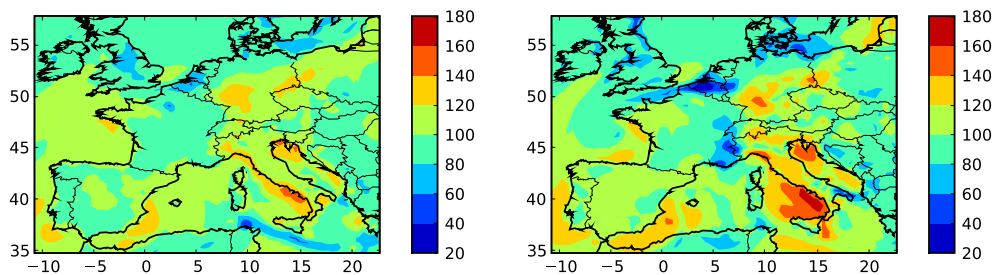


Fig. 6. Ozone map of model 98 (left) and model 58 (right), on 5 May 2001 at 17:00 UT. Both models show good performance, but they can produce ozone fields that differ significantly.

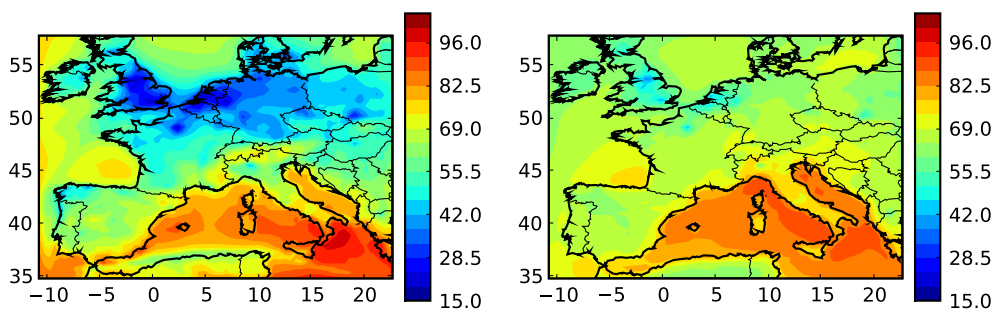


Fig. 7. Temporal average of ozone map for reference model 5 (left) and for model 76 of the 101-member ensemble (right).

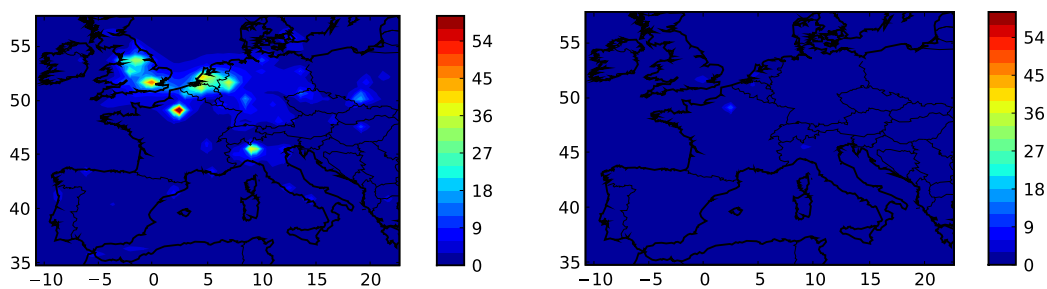


Fig. 8. Temporal average of NO map for reference model 5 (left) and for model 52 of the 101-member ensemble (right).

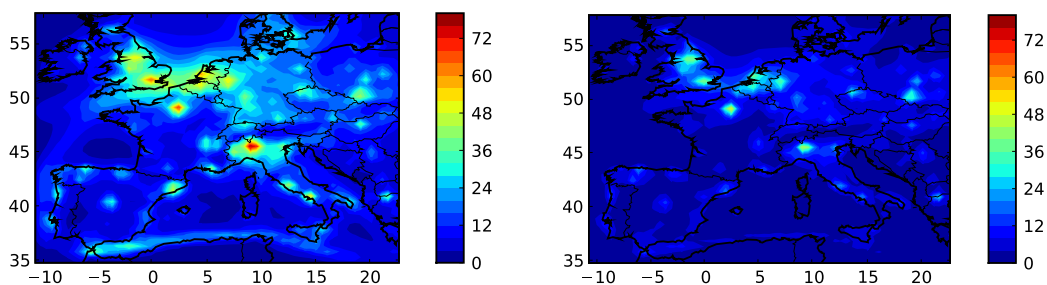


Fig. 9. Temporal average of NO₂ map for reference model 5 (left) and for model 90 of the 101-member ensemble (right).

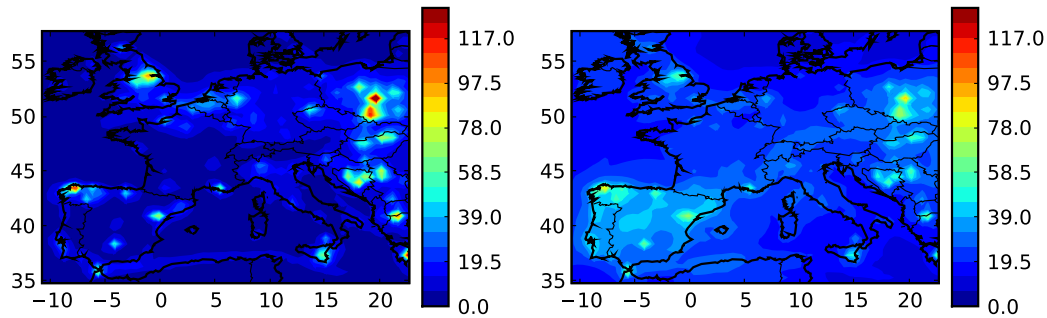


Fig. 10. Temporal average of SO_2 map for reference model 5 (left) and for model 88 of the 101-member ensemble (right).

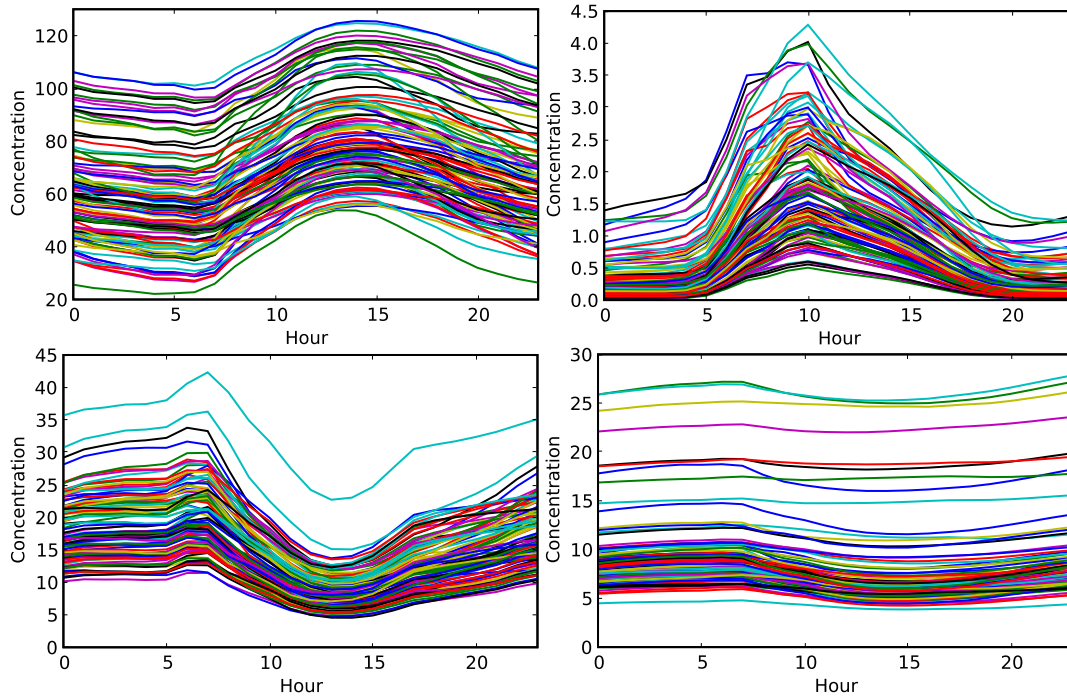


Fig. 11. Daily profile for ozone (*network 3*), NO, NO_2 (*network 3*) and SO_2 . The profile is computed at observation stations for O_3 and NO_2 . It is computed with all computed values (that is, from all grid cells) for NO and SO_2 .

still likely to produce the best forecast. This can be verified with a “map of the best-model index”. At a given date, the best model in each grid cell is determined as follows. The concentrations of the models and the observed concentration at the closest station (to the grid cell) are compared. The model that produces the closest concentration to the observed concentration is considered as the best model in the grid cell. Hence, in every grid cell, one “best model” is determined. A color is associated to each model (actually each model index) to generate the maps in Fig. 12. These maps show the best model for three different dates in June 2001. The best model varies frequently from one grid cell to another, and from one date to another. This shows that many models bring useful information, at least in some regions or on given dates.

5 Conclusions

This paper describes how a large ensemble may be automatically generated using the Polyphemus system. Contrary to most traditional approaches, which are based on perturbations of input data only, or on small ensembles of models from different teams, our approach takes into account all sources of uncertainties at once: input data, physical formulation and numerical formulation. Each member of the ensemble is a complete chemistry-transport model whose contents are clearly defined within the modeling platform. In this context, the ensemble and the differences between its members can be rigorously analyzed, and also controlled through the probabilities associated with every option. Our approach tries to combine the flexibility of Monte Carlo sim-

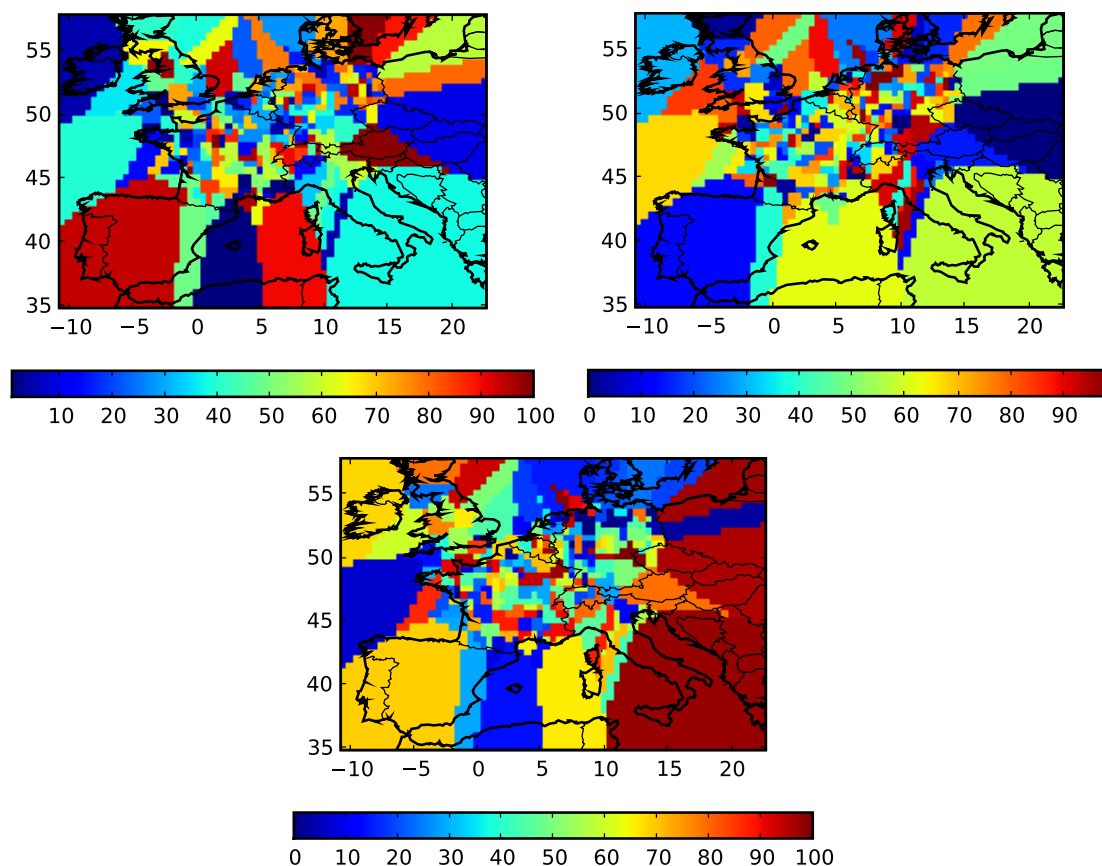


Fig. 12. Maps of best-model indexes. In each grid cell of the domain, the color shows which model (marked with its index, in $\llbracket 0, 100 \rrbracket$) gives the best ozone peak forecast on 1, 11 and 13 June 2001 at the closest station to the cell center. It shows that many models can deliver the best forecast at some point. Stations of network 1 are used. Of course, the colors are only reliable in regions that contain stations.

ulations (large ensembles of simulations with perturbed input data) and the completeness of a multimodel ensemble (models with alternative physical parameterizations, like in ensembles made of a few models from different teams).

In the ensemble, each model is defined by a unique set of physical parameterizations, numerical schemes and input data. Hence building a model means picking an option for every alternative that the system provides. The options are associated with probabilities – depending on how reliable the option is supposed to be – and they are randomly selected. In addition, input data is sampled from normal or log-normal distributions.

The computations are carried out, from the preprocessing to the actual simulation, using small programs whose output results may be shared by different models. This minimizes computational costs and increases flexibility. Thanks to the automatic procedure, the configuration and the generation of an arbitrarily-large ensemble is straightforward. The method can be applied to any simulation with Eulerian models in Polyphemus, such as simulations over a smaller region, or simulations with aerosols.

The ensemble given as example includes 101 photochemical models generated and run for the year 2001, over Europe. The ensemble has a wide spread for all chemical species. The models show a strong diversity both in their formulation and their performance. Many of them appear to be the best in many different regions and periods.

Many research issues are related to this procedure. One relates to the choice of the models to be included in the ensemble. How many models should be included for the ensemble to properly represent the uncertainties? Which models should be included? What probabilities should be associated with the options, and what distributions should be given to the input data? How should meteorological ensembles be integrated? Other research issues may deal with the best structure for an ensemble. How does our procedure compare with other approaches, such as Monte Carlo simulations or small ensembles based on models from different teams? How much information on the uncertainties is provided by the different approaches?

Table A1. Emission distribution in percentages for each level and for each SNAP category. Combustion in energy and transformation industries (S1); non-industrial combustion plants (S2); combustion in manufacturing industry (S3); production processes (S4); extraction and distribution of fossil fuels and geothermal (S5); solvent use and other product use (S6); road transport (S7); other mobile sources machinery (S8); waste treatment and disposal (S9), and agriculture (S10).

SNAP	Low					Medium				
	ground	0–50 m	50–150 m	150–300 m	>300 m	ground	0–50 m	50–150 m	150–300 m	>300 m
S1	0	28.6	71.4	0	0	0	13.8	34.5	51.7	0
S2	12.5	50	37.5	0	0	6.6	26.7	66.7	0	0
S3	0	28.6	71.4	0	0	0	13.8	34.5	51.7	0
S4	25	75	0	0	0	22.2	77.8	0	0	0
S5	25	75	0	0	0	22.2	77.8	0	0	0
S6	100	0	0	0	0	100	0	0	0	0
S7	100	0	0	0	0	100	0	0	0	0
S8	100	0	0	0	0	100	0	0	0	0
S9	0	28.6	71.4	0	0	0	13.8	34.5	51.7	0
S10	100	0	0	0	0	100	0	0	0	0

Appendix A

Emissions from EMEP

As described in Sect. 2, anthropogenic emissions are provided by EMEP. The vertical distribution of the pollutants depends on SNAP category. Two vertical distributions are used in this paper: a “low distribution” and a “medium distribution” – see Table A1.

Acknowledgements. We thank Richard James for proofreading the paper.

Edited by: J. Annan

References

- Beekmann, M. and Derognat, C.: Monte Carlo uncertainty analysis of a regional-scale transport chemistry model constrained by measurements from the atmospheric pollution over the Paris area (ESQUIF) campaign, *J. Geophys. Res.*, 108(8), 559, doi:10.1029/2003JD003391, 2003.
- Boutahar, J., Lacour, S., Mallet, V., Quélo, D., Roustan, Y., and Sportisse, B.: Development and validation of a fully modular platform for numerical modelling of air pollution: POLAIR, *Int. J. Environ. Pollut.*, 22, 17–28, 2004.
- Byun, D. W. and Ching, J. K. S., eds.: Science algorithms of the EPA models-3 community multiscale air quality (CMAQ) modeling system, US Environmental Protection Agency, Washington, 1999.
- Chang, J., Brost, R., Isaken, I., Madronich, S., Middleton, P., Stockwell, W., and Walcek, C.: A three-dimensional Eulerian acid deposition model: physical concepts and formulation, *J. Geophys. Res.*, 92, 14681–14700, 1987.
- Debry, E., Fahey, K., Sartelet, K., Sportisse, B., and Tombette, M.: Technical Note: A new Size Resolved Aerosol Model (SIREAM), *Atmos. Chem. Phys.*, 7, 1537–1547, 2007, <http://www.atmos-chem-phys.net/7/1537/2007/>.
- ESQUIF: Étude et simulation de la qualité de l’air en Île de France – rapport final, 2001.
- Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, 99, 10143–10162, 1994.
- Galmarini, S., Bianconi, R., Klug, W., Mikkelsen, T., Addis, R., Andronopoulos, S., Astrup, P., Baklanov, A., Bartniki, J., Bartzis, J. C., et al.: Ensemble dispersion forecasting – Part I: concept, approach and indicators, *Atmos. Environ.*, 38, 4607–4617, 2004.
- Gross, A. and Stockwell, W. R.: Comparison of the EMEP, RADM2 and RACM mechanisms, *J. Atmos. Chem.*, 44, 151–170, 2003.
- Hanna, S. R., Chang, J. C., and Fernau, M. E.: Monte Carlo estimates of uncertainties in predictions by a photochemical grid model (UAM-IV) due to uncertainties in input variables, *Atmos. Environ.*, 32, 3619–3628, 1998.
- Hanna, S. R., Lu, Z., Frey, H. C., Wheeler, N., Vukovich, J., Arunachalam, S., Fernau, M., and Hansen, D. A.: Uncertainties in predicted ozone concentrations due to input uncertainties for the UAM-V photochemical grid model applied to the July 1995 OTAG domain, *Atmos. Environ.*, 35, 891–903, 2001.
- Louis, J.-F.: A parametric model of vertical eddy fluxes in the atmosphere, *Bound.-Lay. Meteorol.*, 17, 187–202, 1979.
- Madronich, S.: Photodissociation in the atmosphere: 1. actinic flux and the effects of ground reflections and clouds, *J. Geophys. Res.*, 92, 9740–9752, 1987.
- Mallet, V. and Sportisse, B.: Ensemble-based air quality forecasts: A multimodel approach applied to ozone, *J. Geophys. Res.*, 111, D18302, doi:10.1029/2005JD006675, 2006a.
- Mallet, V. and Sportisse, B.: Uncertainty in a chemistry-transport model due to physical parameterizations and numerical approximations: An ensemble approach applied to ozone modeling, *J. Geophys. Res.*, 111, D01302, doi:10.1029/2005JD006149, 2006b.

- Mallet, V., Pourchet, A., Quélo, D., and Sportisse, B.: Investigation of some numerical issues in a chemistry-transport model: Gas-phase simulations, *Journal of Geophysical Research*, 112, D15301, doi:10.1029/2006JD008373, 2007a.
- Mallet, V., Quélo, D., Sportisse, B., Ahmed de Biasi, M., Debry, É., Korsakissok, I., Wu, L., Roustan, Y., Sartelet, K., Tombette, M., and Foudhil, H.: Technical Note: The air quality modeling system Polyphemus, *Atmos. Chem. Phys.*, 7, 5479–5487, 2007b.
- Mallet, V., Stoltz, G., and Mauricette, B.: Ozone ensemble forecast with machine learning algorithms, *J. Geophys. Res.*, 114, D05307, doi:10.1029/2008JD009978, 2009.
- McKeen, S., Wilczak, J., Grell, G., Djalalova, I., Peckham, S., Hsie, E.-Y., Gong, W., Bouchet, V., Menard, S., Moffet, R., McHenry, J., McQueen, J., Tang, Y., Carmichael, G. R., Pagowski, M., Chan, A., Dye, T., Frost, G., Lee, P., and Mathur, R.: Assessment of an ensemble of seven real-time ozone forecasts over eastern North America during the summer of 2004, *J. Geophys. Res.*, 110, D21307, doi:10.1029/2005JD005858, 2005.
- Monahan, E. C., Spiel, D. E., and Davidson, K. L.: *Oceanic Whitecaps – and Their Role in Air-Sea Exchange Processes*, chap. A model of marine aerosol generation via whitecaps and wave disruption, Kluwer Academic, Dordrecht Holland: Reidel, 167–174, 1986.
- Pagowski, M., Grell, G., Devenyi, D., Peckham, S., McKeen, S., Gong, W., Delle Monache, L., McHenry, J., McQueen, J., and Lee, P.: Application of dynamic linear regression to improve the skill of ensemble-based deterministic ozone forecasts, *Atmos. Environ.*, 40, 3240–3250, 2006.
- Sartelet, K., Hayami, H., and Sportisse, B.: MICS Asia Phase II–Sensitivity to the aerosol module, *Atmospheric Environment*, 42, 3562–3570, doi:10.1016/j.atmosenv.2007.03.005, 2008.
- Segers, A.: *Data assimilation in atmospheric chemistry models using Kalman filtering*, Ph.D. thesis, Delft University, 2002.
- Simpson, D., Winiwarter, W., Börjesson, G., Cinderby, S., Ferreira, A., Guenther, A., Hewitt, C. N., Janson, R., Khalil, M. A. K., Owen, S., Pierce, T. E., Puxbaum, H., Shearer, M., Skiba, U., Steinbrecher, R., Tarrasón, L., and Öquist, M. G.: Inventorying emissions from nature in Europe, *J. Geophys. Res.*, 104, 8113–8152, 1999.
- Smith, M. and Harrison, N.: The sea spray generation function, *J. Aerosol Sci.*, 29, 189–190, 1998.
- Stockwell, W. R., Middleton, P., Chang, J. S., and Tang, X.: The second generation regional acid deposition model chemical mechanism for regional air quality modeling, *J. Geophys. Res.*, 95, 16343–16367, 1990.
- Stockwell, W. R., Kirchner, F., Kuhn, M., and Seefeld, S.: A new mechanism for regional atmospheric chemistry modeling, *J. Geophys. Res.*, 102, 25847–25879, 1997.
- Straume, A. G.: A more extensive investigation of the use of ensemble forecasts for dispersion model evaluation, *J. Appl. Meteorol.*, 40, 425–445, 2001.
- Troen, I. and Mahrt, L.: A simple model of the atmospheric boundary layer; sensitivity to surface evaporation, *Bound.-Lay. Meteorol.*, 37, 129–148, 1986.
- van Loon, M., Vautard, R., Schaap, M., Bergström, R., Bessagnet, B., Brandt, J., Builtjes, P., Christensen, J., Cuvelier, C., Graff, A., Jonson, J., Krol, M., Langner, J., Roberts, P., Rouil, L., Stern, R., Tarrasón, L., Thunis, P., Vignati, E., White, L., and Wind, P.: Evaluation of long-term ozone simulations from seven regional air quality models and their ensemble, *Atmos. Environ.*, 41, 2083–2097, 2007.
- Verwer, J. G., Hundsdorfer, W., and Blom, J. G.: Numerical time integration for air pollution models, *Surveys on Mathematics for Industry*, 10, 107–174, 2002.
- Wesely, M. L.: Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical models, *Atmos. Environ.*, 23, 1293–1304, 1989.
- Zhang, L., Brook, J. R., and Vet, R.: A revised parameterization for gaseous dry deposition in air-quality models, *Atmos. Chem. Phys.*, 3, 2067–2082, 2003, <http://www.atmos-chem-phys.net/3/2067/2003/>.