

# Ensemble-based air quality forecasts: A multimodel approach applied to ozone

Vivien Mallet and Bruno Sportisse

Cerea: Teaching and Research Center in Atmospheric Environment  
 Joint laboratory École Nationale des Ponts et Chaussées / Électricité de France R&D  
<http://www.enpc.fr/cerea/>  
 Clime  
 Joint team Inria / École Nationale des Ponts et Chaussées  
<http://www-rocq.inria.fr/clime/>

**Abstract.** This paper investigates the potential of ensemble techniques to improve ozone forecasts. Ensembles with up to 48 members (models) are generated using the modeling system Polyphemus. Members differ in their physical parameterizations, their numerical approximations and their input data. Each model is evaluated during four months (summer 2001) over Europe with hundreds of stations from three ozone-monitoring networks. We found that several linear combinations of models have the potential to drastically increase the performances of model-to-data comparisons. Optimal weights associated with each model are not robust in time or space. Forecasting these weights therefore requires relevant methods, such as selection of adequate learning data sets, or specific learning algorithms. Significant performance improvements are accomplished by the resulting forecasted-combinations. A decrease of about 10% of the root-mean square error is obtained on ozone daily peaks. Ozone hourly concentrations show stronger improvements.

## 1. Introduction

Though sparsely evaluated, the uncertainty in chemistry-transport models is a major limitation of air-quality forecasting. The source of this uncertainty lies in input fields (emissions, deposition velocities, land data, meteorological fields, etc.), as detailed in *Hanna et al.* [1998, 2001], and in the models themselves [*Russell and Dennis*, 2000; *Mallet and Sportisse*, 2006]. The uncertainty is so high that the reliability of model results should be carefully assessed, and ensemble forecast is relevant to address this issue. *Straume et al.* [1998]; *Dabberdt and Miller* [2000]; *Galmarini et al.* [2004]; *Straume* [2001]; *Warner et al.* [2002] have estimated the uncertainty in dispersion modeling using ensemble forecasts. Dealing with ozone exposure, *Hanna et al.* [1998]; *Beekmann and Derognat* [2003] accounted for uncertainties in input fields with Monte Carlo simulations to check the efficiency of emission reductions. *Hanna et al.* [2001]; *Hanna and Davis* [2002]; *Mallet and Sportisse* [2006] estimated the uncertainty in photochemical forecasts based on Monte Carlo simulations and a multi-model approach respectively.

With respect to day-to-day photochemical forecasts, only few developments have been undertaken in order to associate uncertainties with the forecasts or in order to overtake the limitations of uncertain processes or data. Improvements in air-quality forecasts have been sought in modeling developments, input data refinements and increasing computational resources. Unfortunately the performances have only slightly increased [*Russell and Dennis*, 2000]. A reasonable explanation is that the high uncertainties hide modeling efforts and that models are usually tuned to deliver satisfactory forecasts (the latter is also suggested in *Russell and Dennis* [2000]). Taking into account the uncertainty could help in enhancing the forecasts. A promising technique is

to perform ensemble forecasts and to combine the ensemble members.

A brute-force approach is the use of ensemble mean *Delle Monache and Stull* [2003]; *McKeen et al.* [2005]. Underlying (and strong) assumptions are that the ensemble provides an accurate approximation of the output-concentrations probability density function and that the mean of this probability density function is close to the true state. Because of the limited number of models and the unsatisfactory description of the uncertainty, it is hard to satisfy the first assumption. Moreover there is no study supporting the second assumption. More sophisticated methods have been used, mainly in other fields, such as superensembles in meteorology [*Krishnamurti et al.*, 2000] or for ozone forecasts [*Pagowski et al.*, 2005], or Bayesian model averaging [*Hoeting et al.*, 1999] (in many fields).

In this paper, we investigate several methods to build optimal combinations of ensemble members. The objective is to increase day-to-day forecast performances (estimated through comparison against field measurements). The methods are applied to ozone hourly concentrations and daily peaks at European scale during summer 2001 and over hundreds of stations from three monitoring networks. The involved ensembles include up to 48 members, which allows us to study the characteristics of efficient ensembles. Section 2 gives further details about the ensemble members and the system used to generate these forecasts. In Section 3 we introduce the methods we investigated and we review their potential, that is the quality of their a posteriori (i.e. knowing all observations) combinations. In Section 4, methods to forecast optimal ensemble-combinations are described and tested. Selection of the best suited members is also addressed.

## 2. Ensemble Forecasts

### 2.1. Forecasting System Polyphemus

Polyphemus [*Mallet et al.*, 2005] is an air-quality modeling system with ensemble-forecasting abilities based on multiple configurations. These configurations define almost all

components of the modeling system so that each configuration should be viewed as a new model. Polyphemus is primarily composed of:

1. a library for physical parameterizations (and data processing), AtmoData [Mallet and Sportisse, 2005], that includes several parameterizations for major processes;
2. a chemistry-transport model, Polair3D [Boutahar *et al.*, 2004], whose gas-phase version is basically a numerical solver for the reactive-dispersion equation;
3. a set of programs that make calls to AtmoData in order to compute the input-data to the chemistry transport model.

Contrary to most modeling systems that rely on an “all-in-one chemistry-transport model”, Polyphemus splits the numerical solver from physical parameterizations and data management. The programs that compute input data to the chemistry-transport model provide flexibility. They propose several options supported by the multiple physical-parameterizations available in AtmoData. Polair3D is also versatile enough to propose several chemical mechanisms and numerical approximations. In addition independence of the components eases the workflow control, such as corrections in input fields to Polair3D or incorporation of new data sets. These features enable to build ensembles with a high number of members. Moreover very different models can be built so as to produce an ensemble with a wide spread in output concentrations (see Section 2.2).

In this paper, the system is run at European scale ( $[40.25^{\circ}\text{N}, 10.25^{\circ}\text{W}] \times [56.75^{\circ}\text{N}, 22.25^{\circ}\text{E}]$ ) during summer 2001 (27 April 2001 to 31 August 2001). It aims at forecasting ozone concentrations (hourly concentrations and daily peaks). We define a reference configuration (the reference model or reference ensemble-member, not necessary the best model when compared to observations) in the following way:

1. meteorological data: ECMWF<sup>1</sup> fields (resolution of  $0.36^{\circ} \times 0.36^{\circ}$ , TL511 spectral resolution in the horizontal, 60 levels, time step of 3 hours, 12 hours forecast-cycles starting from analyzed fields);
2. land use coverage: USGS<sup>2</sup> land cover map (24 categories, 1 km Lambert);
3. chemical mechanism: RACM [Stockwell *et al.*, 1997];
4. emissions: the EMEP<sup>3</sup> inventory, converted according to Middleton *et al.* [1990];
5. biogenic emissions: computed as proposed in Simpson *et al.* [1999];
6. deposition velocities: the revised parameterization from Zhang *et al.* [2003];
7. vertical diffusion: within the boundary layer, the Troen and Mahrt parameterization described in Troen and Mahrt [1986], with the boundary-layer height computed by ECMWF; above the boundary layer, the Louis parameterization found in Louis [1979];
8. boundary conditions: output of the global chemistry-transport model Mozart 2 [Horowitz *et al.*, 2003] run over a typical year;
9. numerical schemes: a first-order operator splitting, the sequence being advection–diffusion–chemistry; a direct space-time third-order advection scheme with a Koren flux-limiter; a second-order order Rosenbrock method for diffusion and chemistry [Verwer *et al.*, 2002].

Since ensemble forecasting is computationally consuming, we kept a low vertical resolution. The first layer is located between 0 m and 50 m. The thickness of the other layers is about 600 m with the top of the last layer at 3000 m.

## 2.2. Ensembles description

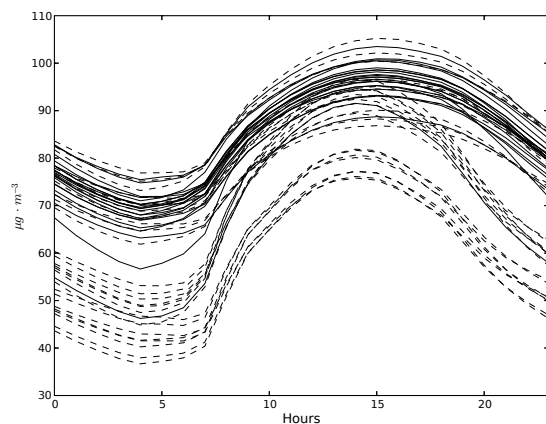
We introduce three ensembles:

– Ensemble 1 is composed of the reference simulation and 21 similar simulations but for one change in the physical parameterizations, in the raw input data (to Polyphemus), in the numerical approximations or in uncertain input-data computed in the system workflow. Table 1 lists all changes.

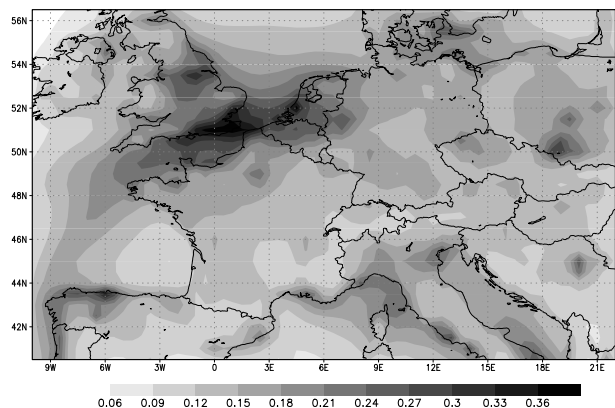
– Ensemble 2 is built with the changes involved in models 17, 8, 4, 2, 1 (numbers from Table 1). All possible combinations of these changes are included in the ensemble. There are therefore 32 members in ensemble 2.

– Ensemble 3 collects all members from ensembles 1 and 2. Ensembles 1 and 2 have six common members (0, 1, 2, 4, 8 and 17); hence there are 48 members in ensemble 3.

Ensembles similar to ensembles 1 and 2 were introduced in Mallet and Sportisse [2006] in order to estimate uncertainties in output ozone concentrations. One may refer to this paper for a detailed description of the ensembles and of their spread. A rough idea of the wide spread is given in Figure 1. The mean of hourly standard-deviations of en-



**Figure 1.** Ozone daily profiles of the 48 models (ensemble 3). The discontinuous lines correspond to the models that are in ensemble 2 and not in ensemble 1. The concentrations are averaged over the whole domain (excluding a three-cell band around the domain borders) and over the 127 simulated days.



**Figure 2.** Spatial distribution of ensemble 3 spread. The standard deviation of the ensemble is computed in each cell and for each hour. Then resulting standard deviations are averaged (time averages) in each cell and divided by the mean concentration of the cell, which gives a relative standard deviation.

**Table 1.** Physical parameterizations, raw input data (to Polyphemus), numerical approximations and perturbed input data involved in ensemble 1. Each model has the same configuration as the reference model but for one change (column “alternative”).

#	Model	Reference	Alternative	Comment
<i>Physical parameterizations</i>				
1. <sup>a</sup>	Chemistry	RACM	RADM 2 [ <i>Stockwell et al.</i> , 1990]	
2.	Vertical diffusion	Troen & Mahrt	Louis [ <i>Louis</i> , 1979]	
3.			Louis in stable conditions	Troen & Mahrt kept in unstable conditions
4.	Deposition velocities	Zhang [ <i>Zhang et al.</i> , 2003]	Wesely [ <i>Wesely</i> , 1989]	
5.	Surface flux	Heat flux <sup>b</sup>	Momentum flux <sup>b</sup>	For the aerodynamic resistance (in deposition velocities)
6.	Cloud attenuation	RADM method [ <i>Chang et al.</i> , 1987; <i>Madronich</i> , 1987]	Esquif ( <i>ESQUIF</i> [2001])	
7.	Critical relative humidity	Depends on $\sigma$	Two layers	Used in the RADM method to compute cloud attenuation
<i>Raw input data</i>				
8.	Emissions vertical distribution	All in the first cell	All in the two first cells	
9.	Land use coverage	USGS	GLCF	For deposition velocities
10.	Land use coverage	USGS	GLCF	For biogenic emissions
11.	Exponent $p$ in Troen & Mahrt	2	3	
12.	Photolytic constants	JPROC (from EPA Models-3)	Dependent on the zenith angle (only)	
<i>Numerical approximations</i>				
13.	Time Step	600 s	100 s	
14.			1800 s <sup>c</sup>	
15.	Vertical resolution	5 layers	9 layers	The first layer height remains 50 m
16.	First layer height	50 m	40 m	The top height of every other layer does not change
17.	Continuity equation	$\text{div}(V) = 0$	$\text{div}(\rho V) = 0$	
<i>Perturbed input data</i>				
18.	Boundary layer height	ECMWF	Increased by 10%	
19.	NO emissions	EMEP	Increased by 25%	Including biogenic emissions
20.	Biogenic emissions	<i>Simpson et al.</i> [1999]	Increased by 100%	Excluding NO biogenic emissions
21.	Ozone boundary conditions	Mozart 2	Decreased by 10%	

<sup>a</sup> The reference model is referred to as model #0.

<sup>b</sup> Computed using Louis formulae.

<sup>c</sup> The advection is integrated over submultiples of 1800 s so as to satisfy the CFL (Courant-Friedrichs-Lewy) condition.

**Table 2.** Performances of ensembles against field observations from the three networks. RMSE is in  $\mu\text{g} \cdot \text{m}^{-3}$ , correlation in %, and bias in %. Best results, for each network, are in bold. Mean statistics are averaged statistics of individual models.

Ensemble	Hourly concentrations			Daily peaks		
	RMSE	Correlation	Bias	RMSE	Correlation	Bias
<i>Network 1</i>						
Ensemble 1						
Best member	27.0	66.1	<b>1.8</b>	<b>22.7</b>	73.8	<b>0.1</b>
Mean statistics	29.0	63.8	11.3	24.2	71.1	2.9
Ensemble 2						
Best member	<b>26.7</b>	<b>67.9</b>	<b>1.8</b>	23.0	<b>74.8</b>	<b>0.1</b>
Mean statistics	29.1	64.8	13.4	26.4	69.1	6.2
Ensemble 3						
Mean statistics	29.0	64.4	12.6	25.6	69.8	4.9
Worst member	32.1	60.8	27.1	33.5	62.2	17.2
<i>Network 2</i>						
Ensemble 1						
Best member	<b>25.7</b>	63.6	0.5	<b>21.5</b>	69.7	<b>0.1</b>
Mean statistics	26.8	60.6	7.8	22.6	67.4	2.7
Ensemble 2						
Best member	26.3	<b>63.9</b>	<b>0.2</b>	21.6	<b>70.2</b>	0.4
Mean statistics	28.9	59.9	12.9	25.4	64.4	6.6
Ensemble 3						
Mean statistics	28.1	60.1	11.0	24.4	65.5	5.1
Worst member	35.1	54.4	28.7	32.1	56.7	17.3
<i>Network 3</i>						
Ensemble 1						
Best member	29.4	65.5	3.2	<b>24.9</b>	72.2	<b>0.2</b>
Mean statistics	32.5	61.6	15.3	26.5	67.8	2.9
Ensemble 2						
Best member	<b>29.0</b>	<b>67.8</b>	<b>0.2</b>	25.1	<b>74.4</b>	0.5
Mean statistics	31.2	62.9	12.8	29.1	65.4	6.8
Ensemble 3						
Mean statistics	31.7	62.4	13.8	28.2	66.2	5.4
Worst member	35.8	58.8	26.0	37.5	55.4	17.7

ensemble 3 profiles (shown in Figure 1) is  $10.4 \mu\text{g} \cdot \text{m}^{-3}$ . To show the spatial distribution of ensemble spread, the standard deviation of ensemble 3 is computed in each cell and for each hour, then averaged relative standard-deviations (time average) in each cell are plotted in Figure 2.

### 2.3. Comparisons with Observations

We use ozone measurements from three monitoring networks (described below). All stations in these networks have observations for at least 30% of possible measurements during the 127 simulated days (for both hourly concentrations and peaks). Here is a description of the networks:

- Network 1 is composed of 241 urban and regional stations over Europe. A large part of the stations are in France (116 stations) and in Germany (81 stations). It provides about 619,000 hourly concentrations and 27,500 peaks.

- Network 2 includes 85 EMEP stations, that is regional stations distributed over Europe, with about 240,000 hourly observations and 10,400 peaks.

- Network 3 includes 356 urban and regional stations in France from BDQA (“Banque de Données sur la Qualité de l’Air”, managed by ADEME<sup>4</sup> and gathering 40 approved associations for monitoring air quality). It provides 997,000 hourly measurements and 42,000 peaks. Note that it includes most French stations of network 1.

Network 1 and 2 have a large extent while network 3 provides a large amount of measurements in a single country. Network 2 allows us to test combining methods only with regional stations. Three statistical measures are introduced in order to estimate model performances: the root-mean square error (RMSE), the correlation and a bias factor, de-

finied as

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

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

$$\text{bias} = \left| 1 - \frac{1}{\tilde{n}} \sum_{i=1}^{\tilde{n}} \frac{\tilde{y}_i}{\tilde{o}_i} \right| \quad (3)$$

where  $y$  is the vector of model outputs,  $o$  the vector of the corresponding observations. Both vectors have  $n$  components. Their means are  $\bar{y}$  and  $\bar{o}$ .  $\tilde{o}$  is the vector of observations above  $40 \mu\text{g} \cdot \text{m}^{-3}$  and  $\tilde{y}$  is the corresponding computed concentrations. Both vectors have  $\tilde{n}$  components.

Table 2 shows the performances of the three ensembles against the measurements from the three networks.

## 3. Combining Forecasts: Methods and Potentialities

### 3.1. Introduction

For day-to-day forecasts, the modeler is usually able to choose a model with performances close to the best model. In particular, it means that the performances of the reference configuration (Section 2.1) are similar to those of the best model. The objective is therefore to deliver a forecast with higher performances than the best available model. In other words, a method to determine the best model would

not be of great help. Hence ensemble members should be combined. Knowing that ensemble forecasting is computationally consuming, a satisfactory model-combination has to bring significant improvements. We consider that a decrease by 10% of the root-mean square error of the best model (that is, about  $2\text{--}3\mu\text{g} \cdot \text{m}^{-3}$ ) is required for an ensemble method to be interesting. This threshold is arbitrary, but there is some background to support it. The best model is usually a tuned model, that is, a favorable configuration found by the modeler. Improving a well tuned model, so as to decrease the root-mean square error by 10%, is not an easy task, especially for day-to-day forecasts.

### 3.2. Notations

An ensemble is denoted  $\mathcal{E}$  or  $\mathcal{E}_i$ . For instance,  $\mathcal{E}_3 = \mathcal{E}_1 \cup \mathcal{E}_2$ . A network is denoted  $\mathcal{N}$  or  $\mathcal{N}_i$ . The cardinal of a network (number of stations) or of an ensemble (number of models) is denoted  $|\cdot|$ . Output concentrations of a model are denoted  $M_{t,x}$  or  $M_{m,t,x}$  (if the model is indexed by  $m$ ), where  $t$  is the time step and  $x$  denotes a station. Time and spatial averages are denoted  $\overline{M}_x^t$  and  $\overline{M}_t^x$  respectively. The mean over all stations and during the whole simulation period is  $\overline{M}^{t,x}$ . Observations are denoted  $O_{t,x}$  and  $C_{t,x}$  are combined concentrations.

### 3.3. Introduction to Combining Methods

#### 3.3.1. Ensemble Mean and Ensemble Median

The ensemble mean is defined as

$$\text{EM}_{t,x} = \frac{1}{|\mathcal{E}|} \sum_{M \in \mathcal{E}} M_{t,x} \quad (4)$$

The ensemble median is defined as

$$\text{EMD}_{t,x} = \text{median}(\{M_{t,x}\}_{M \in \mathcal{E}}) \quad (5)$$

If there is an even number of models, the mean of the two middle models is used.

#### 3.3.2. Models Selection

At each station, the best model is selected. The resulting model is denoted  $\text{EB}^s$  ('B' stands for "best" and 's' stands for "station"). In the same way, selecting the best model for each date (but for all stations) defines the meta-model  $\text{EB}^d$  ('d' stands for "date").

#### 3.3.3. Least-Squares Methods

The best linear combination, in the least-squares (LS) sense, is

$$\text{ELS}_{t,x} = \sum_m \alpha_m M_{m,t,x} \quad (6)$$

where  $\alpha$  is a vector of unconstrained weights that minimizes

$$\sum_{t,x} \left[ O_{t,x} - \sum_m \alpha_m M_{m,t,x} \right]^2 \quad (7)$$

An unbiased version is

$$\text{EULS}_{t,x} = \overline{O}^{t,x} + \sum_m \alpha_m \left( M_{m,t,x} - \overline{M}_m^{t,x} \right) \quad (8)$$

where  $\alpha$  (still unconstrained) minimizes

$$\sum_{t,x} \left[ O_{t,x} - \overline{O}^{t,x} - \sum_m \alpha_m \left( M_{m,t,x} - \overline{M}_m^{t,x} \right) \right]^2 \quad (9)$$

EULS may be referred as superensemble [following *Krishnamurti et al.*, 2000].

Weights ( $\alpha$ ) may be computed for each station or for each time step. The corresponding combinations are denoted with the superscripts 's' (station) and 'd' (date), e.g.,  $\text{ELS}^s$  and  $\text{ELS}^d$ . Averages are adapted to the new target; for instance:

$$\text{EULS}_{t,x}^s = \overline{O}_x^t + \sum_m \alpha_{m,x}^s \left( M_{m,t,x} - \overline{M}_{m,x}^t \right) \quad (10)$$

where the vector  $\alpha_x^s = (\alpha_{1,x}^s, \alpha_{2,x}^s, \alpha_{3,x}^s, \dots)$  minimizes

$$\sum_t \left[ O_{t,x} - \overline{O}_x^t - \sum_m \alpha_{m,x}^s \left( M_{m,t,x} - \overline{M}_{m,x}^t \right) \right]^2 \quad (11)$$

### 3.4. Potentials

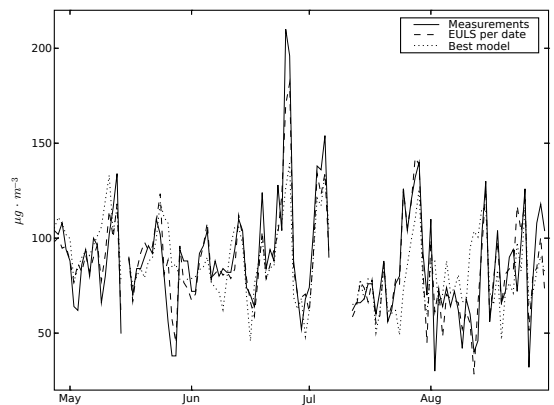
In previous formulae, weights are computed on the basis of *all* observations. In operational forecasts, these weights should be forecasted, that is, on the basis of *past* observations. However, in this subsection, methods are assessed through their a posteriori (i.e. with all observations known) performances. This gives the potential of all methods.

All statistical measures are provided in Table 3.

#### 3.4.1. Ensemble Mean and Ensemble Median

For every ensemble, results of EM and EMD are better than the averaged statistics of the ensemble. However, they often have lower performances than the best member. No ensemble mean or ensemble median has a RMSE below 90% of the best RMSE of the same ensemble. Ensemble mean and ensemble median therefore show poor performances. This is in contradiction with the results from *Delle Monache and Stull* [2003]; *McKeen et al.* [2005]. Nonetheless the former study implied only four models, during six days and with five stations, which limits the reliability of the conclusions, as also pointed out by the authors. The latter study is also limited with seven models.

#### 3.4.2. Models Selection



**Figure 3.** Ozone daily peaks at Harwell (station in network 2) over the 127 days (120 available measurements). The best model is extracted from ensemble 1. The combination  $\text{EULS}^d$  is based on ensemble 1. The best model is associated with a RMSE of  $22.0\mu\text{g} \cdot \text{m}^{-3}$  and a correlation of 63.4%.  $\text{EULS}^d$  is associated with a RMSE of  $12.1\mu\text{g} \cdot \text{m}^{-3}$  and a correlation of 90.6%.

**Table 3.** Potential performances of model combinations against field observations from the three networks. RMSE is in  $\mu\text{g} \cdot \text{m}^{-3}$ , correlation in %, and bias in %. For networks 1 and 3, only the best combinations with respect to RMSE are shown. Conclusions drawn from results over network 2 are very similar for networks 1 and 3.

Ensemble	Hourly concentrations			Daily peaks		
	RMSE	Correlation	Bias	RMSE	Correlation	Bias
<i>Network 1</i>						
Ensemble 1						
EULS <sup>d</sup>	16.7	87.3	2.6	13.5	91.6	1.4
Ensemble 2						
EULS <sup>d</sup>	16.3	87.9	2.5	13.3	91.9	1.4
Ensemble 3						
EULS <sup>s</sup>	16.5	87.7	2.0	10.9	94.5	1.0
EULS <sup>d</sup>	14.5	90.6	2.0	11.6	93.9	1.1
<i>Network 2</i>						
Ensemble 1						
EM	25.9	61.9	6.3	22.0	68.7	0.7
EMD	26.4	60.9	7.7	22.1	68.0	1.0
EB <sup>s</sup>	23.1	70.6	2.4	19.7	75.3	2.4
EB <sup>d</sup>	24.2	67.0	2.6	19.9	74.8	2.4
ELS	23.7	68.0	0.8	18.7	78.2	2.5
EULS	23.4	68.8	0.0	18.5	78.7	3.2
ELS <sup>s</sup>	16.4	86.3	0.7	12.9	90.3	1.2
EULS <sup>s</sup>	16.0	86.8	0.2	12.5	90.9	1.4
ELS <sup>d</sup>	17.1	84.8	0.5	12.5	90.9	1.3
EULS <sup>d</sup>	16.7	85.5	0.2	12.1	91.4	1.4
Ensemble 2						
EM	25.2	64.4	5.5	23.1	70.5	4.6
EMD	25.3	64.0	4.9	23.3	69.6	4.6
EB <sup>s</sup>	22.4	72.5	0.9	19.1	77.1	2.0
EB <sup>d</sup>	24.0	67.3	1.3	19.6	75.6	2.1
ELS	24.3	66.2	0.7	19.6	75.8	2.7
EULS	24.0	66.9	0.4	19.4	76.2	3.4
ELS <sup>s</sup>	17.3	84.6	0.9	12.8	90.4	1.0
EULS <sup>s</sup>	16.9	85.3	0.1	12.3	91.2	1.4
ELS <sup>d</sup>	15.9	87.1	0.3	11.4	92.4	1.1
EULS <sup>d</sup>	15.4	87.9	0.1	11.0	93.1	1.2
Ensemble 3						
EM	24.9	64.2	1.3	22.3	70.7	2.7
EMD	25.7	61.5	4.5	22.2	69.2	0.7
EB <sup>s</sup>	22.1	73.2	0.8	18.9	77.6	1.7
EB <sup>d</sup>	23.8	68.1	1.6	19.4	76.2	2.0
ELS	23.5	68.8	0.9	18.3	79.3	2.4
EULS	23.2	69.6	0.0	18.1	79.7	3.0
ELS <sup>s</sup>	15.5	87.8	0.6	10.5	93.7	0.8
EULS <sup>s</sup>	15.2	88.3	0.2	10.1	94.1	1.0
ELS <sup>d</sup>	11.9	93.0	0.1	8.3	96.1	0.6
EULS <sup>d</sup>	11.6	93.3	0.0	8.0	96.3	0.6
<i>Network 3</i>						
Ensemble 1						
EULS <sup>d</sup>	16.9	88.3	3.1	13.9	91.9	1.4
Ensemble 2						
EULS <sup>d</sup>	16.4	89.0	3.0	13.3	92.6	1.3
Ensemble 3						
EULS <sup>d</sup>	15.0	90.8	2.6	11.9	94.1	1.1

Performances of EB<sup>s</sup> and EB<sup>d</sup> are satisfactory, especially on the peaks. RMSE are then below 90% of the best-model RMSE.

### 3.4.3. Least-Squares Methods

All comments are made for both the regular least-squares version and the corresponding unbiased version since their performances are very similar.

Least-squares method applied with a single combination, over the whole network and at all dates, brings significant improvement in results. RMSE are usually well below 90% of the best-model RMSE.

Meanwhile, the best performances are reached by far with the least-squares methods per station and per date.

Over network 2, EULS<sup>d</sup> based on ensemble 3 even reaches a RMSE of  $8\mu\text{g} \cdot \text{m}^{-3}$  and a correlation of 96.3% for daily peaks.

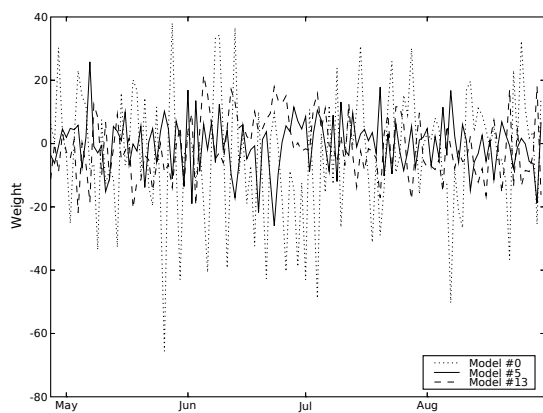
Combinations based on ensemble 3 logically show the best results since ensemble 3 includes all simulations. Least-squares combinations based on ensemble 2 are slightly better than ensemble 1, which may be due to the number of members, the wider spread or a favorable configuration. Least-squares combinations per date usually perform better than combinations per station. The ratio between the number of available stations per date and the number of measurements per station might be an explanation for hourly concentrations. However it is likely that the spatio-temporal structure of the computed fields plays an important role.

Ozone daily peaks at a representative station (with respect to the statistics) illustrate the improvements in Figure 3.

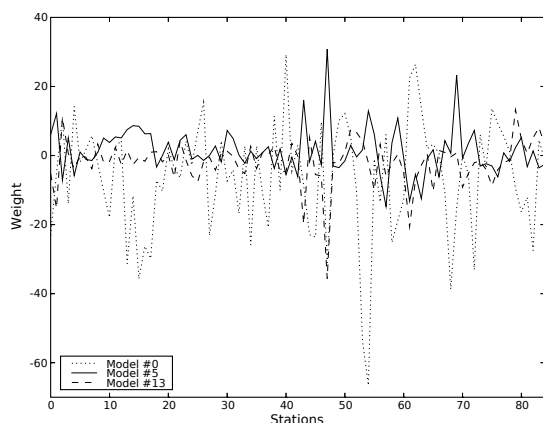
## 4. Forecasting Ensemble Combinations and Selecting Ensemble Members

The previous results show a strong potential for least-squares methods. The objective is to use them for forecasts, that is to forecast the weights associated with every model based on the weights computed in the past days. This may be viewed as a data assimilation procedure constrained by the ensemble structure.

Unless specified, the following tests are performed with ensemble 1, over network 2 and with ozone daily peaks.



**Figure 4.** Time evolution of the three less stable weights in  $ELS^d$  (ensemble 1, network 2), that is, weights associated with the highest standard-deviations. Weights of other models are also highly variable.



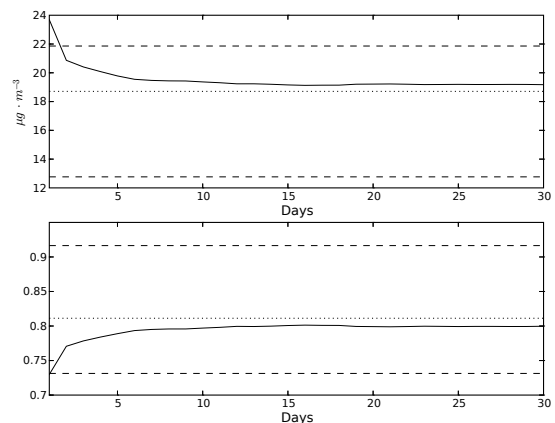
**Figure 5.** Distribution over the 85 stations of the three less stable weights in  $ELS^s$  (ensemble 1, network 2), that is, weights associated with the highest standard-deviations. Weights of other models are also highly variable.

### 4.1. Weights Stability

Since  $ELS^s$  and  $ELS^d$  both show promising performances, the combinations may be forecasted at each station (and over a given period) or for each time step (and for all stations). In order to ease weights forecasting, combinations with a low time-dependency are of high interest. It is also useful to have spatially-robust weights, that is, weights that may be applied to another network or to other grid cells. With such weights, the whole ground-field may be forecasted, which is a key feature of 3D chemistry-transport models.

It is noteworthy that (1) there exist constant weights over the whole period (127 days) for an efficient combination ( $ELS^s$ ), and (2) there also exist uniform weights (over a network) associated with high performances ( $ELS^d$ ). The question is primarily to know whether these coefficients can be forecasted.

#### 4.1.1. Least-Squares Method per Date



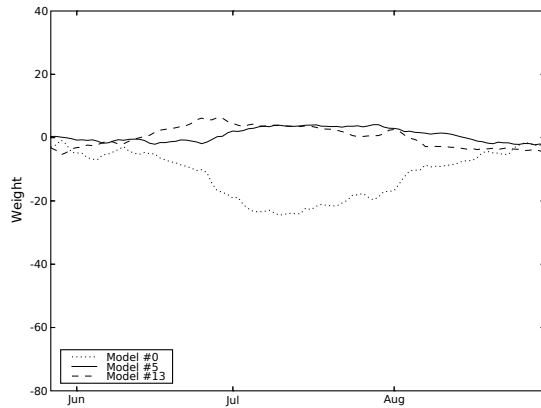
**Figure 6.** RMSE (upper graphic) and correlation (lower graphic) of a combination with weights (same at all stations) computed at each step with a least-squares optimization over a learning period of  $x$  days (abscissa). The discontiguous lines are the performances of the best model in the ensemble and of  $ELS^d$  (ensemble 1). The dotted line is the performance of  $ELS$ .

**Table 4.** Performances on ozone daily peaks over the last 96 simulated days for  $ELS^d$ ,  $ELS$ , the best model (in the ensemble) and the combination with “least-squares weights” computed with 30-day learning periods preceding each forecasted day. In each column, RMSE in  $\mu g \cdot m^{-3}$  is followed by correlation in %.

Ensemble	$ELS^d$	$ELS$	Best model	Forecast
<i>Network 1</i>				
Ensemble 1	14.1 – 91.7	19.6 – 83.3	22.4 – 78.0	20.5 – 81.7
Ensemble 2	13.9 – 92.0	20.5 – 81.5	22.4 – 78.1	21.3 – 80.0
Ensemble 3	12.0 – 94.1	19.2 – 84.0	22.4 – 78.1	20.2 – 82.2
<i>Network 2</i>				
Ensemble 1	12.8 – 91.6	18.7 – 81.1	21.9 – 73.1	19.2 – 80.0
Ensemble 2	11.6 – 93.1	19.6 – 79.0	21.9 – 73.8	20.4 – 77.2
Ensemble 3	8.4 – 96.4	18.2 – 82.3	21.9 – 73.8	19.0 – 80.4
<i>Network 3</i>				
Ensemble 1	14.6 – 91.8	21.1 – 81.8	24.0 – 76.4	21.8 – 80.6
Ensemble 2	13.9 – 92.5	21.1 – 81.9	23.9 – 76.6	22.1 – 80.0
Ensemble 3	12.4 – 94.1	20.2 – 83.5	23.9 – 76.6	21.2 – 81.6

Time evolution of weights for ELS<sup>d</sup>, for three models, is shown in Figure 4. These weights are highly variable. Even the highest weights (in absolute value) which constitute the main part of the combination are highly unstable. It makes the combination very hard to forecast.

Another property is that it is not easy to use these weights over another network or in other cells. Applying the weights computed for network 3 (and ensemble 1, ozone peaks) to network 2 leads to a RMSE of  $55.8 \mu\text{g} \cdot \text{m}^{-3}$ . This is the least favorable extension since the two networks contain stations of different nature and their spatial extent strongly differ. A more favorable experiment is to compute the weights over network 2 (Europe) and to apply them to network 3 (France). The resulting RMSE is  $24.6 \mu\text{g} \cdot \text{m}^{-3}$  (correlation of 74.7%) which is reasonable but similar to the RMSE of the best model ( $24.9 \mu\text{g} \cdot \text{m}^{-3}$ , correlation of 72.2%). There is a more favorable experiment. Like network 2, network 1 has stations over whole Europe but it has regional and urban stations, including stations from network 3. Network 3



**Figure 7.** Time evolution of three weights computed over a 30-day learning period (preceding each date), for ensemble 1 and network 2. This figure should be compared to Figure 4 which shows more variable weights. Both figures have the same range of values along y.

**Table 5.** Performances on ozone hourly concentrations over the last 96 simulated days for ELS<sup>d</sup>, ELS, the best model (in the ensemble) and the combination with “least-squares weights” computed with 30-day learning periods preceding each forecasted day. The weights associated with a given hour  $h$  are estimated with the computed and observed concentrations at hour  $h$  during the learning period. In each column, RMSE in  $\mu\text{g} \cdot \text{m}^{-3}$  is followed by correlation in %.

Ensemble	ELS <sup>d</sup>	ELS	Best model	Forecast ELS <sup>d</sup>
<i>Network 1</i>				
Ensemble 1	17.2 – 87.3	22.9 – 75.9	26.8 – 68.4	22.7 – 76.6
Ensemble 2	16.8 – 87.9	24.0 – 73.2	26.7 – 69.9	23.3 – 75.2
Ensemble 3	14.9 – 90.6	22.7 – 76.5	26.7 – 69.9	22.5 – 77.1
<i>Network 2</i>				
Ensemble 1	17.3 – 85.5	23.9 – 70.1	25.9 – 65.6	23.6 – 71.0
Ensemble 2	16.1 – 87.7	24.6 – 67.9	26.7 – 65.7	24.6 – 68.0
Ensemble 3	11.9 – 93.4	23.6 – 71.0	25.9 – 65.7	23.4 – 71.5
<i>Network 3</i>				
Ensemble 1	17.2 – 88.4	23.3 – 77.5	28.7 – 68.0	22.9 – 78.4
Ensemble 2	16.7 – 89.2	24.9 – 73.7	28.5 – 69.9	23.7 – 76.8
Ensemble 3	15.3 – 91.0	22.9 – 78.4	28.5 – 69.9	22.8 – 78.7

is therefore closer to network 1 than to network 2. Applying weights computed over network 1 to network 3 gives better performances, with a RMSE of  $17.4 \mu\text{g} \cdot \text{m}^{-3}$  and a correlation of 87.1%. This promising result tends to show that it is possible to apply suited weights to cells without observations.

#### 4.1.2. Least-Squares Method per Station

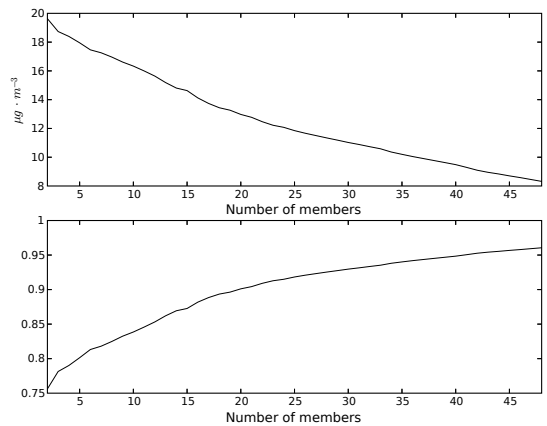
Weights (for ELS<sup>s</sup>) associated with each model are highly variable over the network, as shown in Figure 5. In addition, there is no subset of stations over which the weights are similar. This is not surprising because setting a single weight per model (for all stations and all dates) does not provide very strong improvements (Table 3, ELS).

#### 4.2. Using Weights from the Previous Days

An obvious method is to use weights computed in the previous days. In this section, the statistics are computed

**Table 6.** Performances on ozone daily peaks over the last 96 simulated days for ELS<sup>d</sup>, ELS<sup>d</sup> with forecasted weights (as in Section 4.2.2 and Table 4), the best model (in the ensemble) and the combination computed with the gradient-descent algorithm. The first 30 days constitute a minimum learning-period. All forecasted concentrations are preceded by at least 30 contiguous peak observations. This is the reason why the comparisons with observations slightly differ from Table 4 (whose fifth column corresponds the third column of this table). In each column, RMSE in  $\mu\text{g} \cdot \text{m}^{-3}$  is followed by correlation in %.

Ensemble	ELS <sup>d</sup>	Forecast ELS <sup>d</sup>	Best model	Gradient descent
<i>Network 1</i>				
Ensemble 1	13.8 – 91.9	20.3 – 81.5	22.4 – 77.5	20.1 – 82.1
Ensemble 2	14.2 – 91.4	21.0 – 80.0	22.4 – 77.7	19.5 – 83.0
Ensemble 3	11.2 – 94.7	20.0 – 82.1	22.4 – 77.7	19.6 – 83.0
<i>Network 2</i>				
Ensemble 1	13.0 – 91.1	18.8 – 80.0	21.8 – 72.5	18.8 – 80.6
Ensemble 2	13.1 – 90.9	20.2 – 76.8	21.8 – 73.5	18.2 – 81.7
Ensemble 3	10.6 – 94.2	18.8 – 80.2	21.8 – 73.5	18.2 – 81.6
<i>Network 3</i>				
Ensemble 1	14.7 – 91.7	21.8 – 80.6	24.2 – 76.2	21.7 – 81.0
Ensemble 2	15.0 – 91.4	22.1 – 80.1	24.1 – 76.4	22.7 – 82.9
Ensemble 3	12.6 – 94.0	21.3 – 81.6	24.1 – 76.4	21.0 – 82.3



**Figure 8.** Performances (RMSE in upper graphic and correlation in lower graphic) of ELS<sup>d</sup> against the number of models in the ensemble. Models are taken from ensemble 3.



in the last 96 days so that up to 30 days may be used as a learning period. A learning period of  $n$  days includes the  $n$  preceding days of the forecasted day. This is a “moving learning-period”.

#### 4.2.1. Least-Squares Method per Station

Computing weights per station over a learning period of 22 to 30 days (22 is a minimum because there are 22 weights) fails to improve the forecasts. Best results are obtained based on a 30-day learning period with a RMSE of  $40.7\mu\text{g} \cdot \text{m}^{-3}$ . Extending the learning period should help (ELS<sup>s</sup> performs well) but the test period would be too short to draw reliable conclusions. However we report that a 60-day learning period allows to reach, during the last 36 simulated days, a RMSE of  $22.3\mu\text{g} \cdot \text{m}^{-3}$  (best model:  $21.6\mu\text{g} \cdot \text{m}^{-3}$ ) and a correlation of 76.5% (best model: 74.6%). As a conclusion, the strategy is not satisfactory for this simulation. Nevertheless, further investigations, with a simulation during a longer period, are needed.

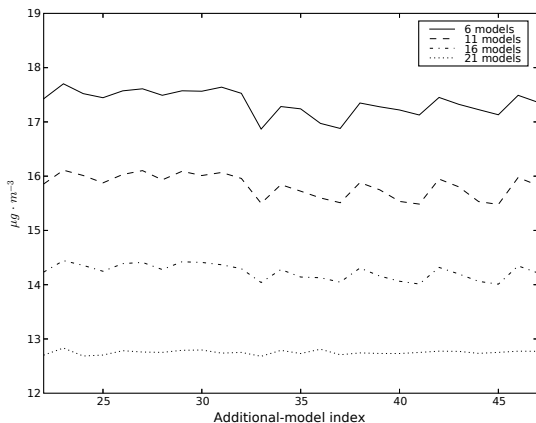
#### 4.2.2. Least-Squares Method per Date

At each date, weights are the same for all stations. They are computed on the basis of a learning period ranging from 1 day to 30 days. Figure 6 shows that this method performs well with a short learning period of about 5–7 days. Longer learning periods do not improve the results. Performances are close to the ones of ELS.

With a 30-day learning period, RMSE of the forecasted combination is  $19.2\mu\text{g} \cdot \text{m}^{-3}$  (best model:  $21.9\mu\text{g} \cdot \text{m}^{-3}$ ) and correlation is 80.0% (best model: 73.3%). The criteria on RMSE (below 90% of RMSE of the best model) is therefore fulfilled (for ensemble 1 and network 2). This is not the case with all ensembles and networks, as shown in Table 4. However there are always significant improvements.

A key point to explain these improvements is the time evolution of the weights. Figure 4 shows strong variations which explain that a one-day learning period has little chance to be suited. Indeed, as shown in Figure 6, a one-day learning period gives poor performances (RMSE of  $23.7\mu\text{g} \cdot \text{m}^{-3}$  and correlation of 73%). Coefficients computed over a 30-day learning period are more stable, see Figure 7.

#### 4.2.3. Hourly concentrations



**Figure 9.** Four ensembles are built with the first 5, 10, 15 and 20 members of ensemble 1. A single model from ensemble 2 (abscissa) is added to these ensembles and RMSE of ELS<sup>d</sup> (ordinate) is computed. 26 models from ensemble 2 (i.e., all models that are in ensemble 2 but not in ensemble 1) are included this way. It shows the contribution that each model can make to overall performances.

Hourly forecasts may also be improved using weights learned in the previous 30 days and estimated per date as in Section 4.2.2. In order to forecast the weights at a given hour  $h$ , only concentrations computed and observed at hour  $h$  during the learning period are included. Including all hourly concentrations lowers performances.

All results are collected in Table 5. Performances are significantly improved, especially over networks 1 and 2. Note that these performances are similar to the performances of ELS (for which one can say that the learning period is the whole simulation).

#### 4.3. Learning algorithms

Applying an optimal combination computed over a learning period may be efficient (Sections 4.2.2 and 4.2.3) but more sophisticated algorithms were designed in machine learning. A classical algorithm is, for instance, the gradient descent algorithm for on-line regression [Cesa-Bianchi et al., 1996]. In our case, this method is independently applied at each station and the objective is to minimize a loss function defined as

$$L_t(\alpha_t) = \left( \sum_m \alpha_{m,t} M_{m,t} - O_t \right)^2 \quad (12)$$

Weights  $\alpha_{t-1} = (\alpha_{1,t-1}, \alpha_{2,t-1}, \alpha_{3,t-1}, \dots)$  are updated according to:

$$\alpha_t = \alpha_{t-1} - \eta L'_{t-1}(\alpha_{t-1}) \quad (13)$$

$\eta$  is the learning rate. Results are sensitive to this parameter (tests not reported here). We chose  $\eta = 5 \cdot 10^{-7}$  for ensembles 1 and 2, and  $\eta = 2.5 \cdot 10^{-7}$  for ensemble 3. Results are stable in the vicinity of these parameters (e.g.,  $\pm 50\%$ ). Initial weights are set to  $\frac{1}{N}$  where  $N$  is the number of models (this corresponds to the ensemble mean).

Table 6 shows results of the gradient descent algorithm. Performances (fifth column) are slightly better than performances of the least-squares method with weights computed at each date (introduced in Section 4.2.2, third column). The learning algorithm succeeds while applying weights computed during the previous days fails (Section 4.2.1). Knowing that there are many variants of learning algorithms (with updates that differ from equation 13), this is certainly a promising direction for further improvements.

#### 4.4. Members Selection

In Table 4, ensemble 1 shows better performances than ensemble 2 even if ensemble 1 has less members (22 against 32) and is less spread. Because of computational costs, it is useful to reduce the number of models to be included. Figure 8 shows performances of ELS<sup>d</sup> against the number of models, where the models from ensemble 3 are included one by one in the optimization. Even if the impact of additional models decreases with the number of models, performances are still significantly improved.

Another question is whether there are models that contribute more than others to performance improvement. In Figure 9, contributions of several models from ensemble 2 (to four sub-ensembles based on ensemble 1 and of size 5, 10, 15 and 20) are shown. Contributions are primarily distinguishable in small ensembles. We also report that contributions of the models from ensemble 1 (to four sub-ensembles based on ensemble 2 and of size 5, 10, 15 and 20) are less distinguishable. In addition, correlation between the models RMSE and their contribution to the combined-model RMSE is below 30%. It seems that the best models do not necessarily bring the best contributions.

There is no clear reason why forecasted combinations based on ensemble 2 show lower performances than those based on ensemble 1. Ensemble 2 includes simulations with multiple changes (see Section 2.2) and only five choices are involved, which might be poorer than involving 21 different single changes.

## 5. Conclusion

The forecasting system Polyphemus has the ability to generate ensemble forecasts with a wide spread in output concentrations and with a high number of different members. Combining the models in an optimal way has a strong potential. While ensemble mean and ensemble median barely improve the performances, results may be dramatically enhanced by linear combinations with optimal weights in some sense.

It was shown that weights computed over a given network do not necessarily apply to another network and consequently to other grid cells. This low spatial robustness of the weights should be studied since gridded forecasts are an important feature of 3D chemistry-transport models.

Daily forecasts also require to forecast the weights of an optimal combination. Weights appear to be highly unstable from one day to another or from one station to another. More stable weights are found in combinations constant over a 30-day period and over a whole network. These weights can be reasonably forecasted and the associated combinations provide significant improvements on hourly concentrations and on daily peaks. A decrease of about 10% of the RMSE is achieved on daily peaks. Hourly concentrations even show better improvements.

In addition, there is a promising application of learning algorithms (machine learning) which do not need to introduce weights computed over the numerous stations of a monitoring network. The gradient-descent algorithm shows good performances when applied to each station, while applying weights computed over a 30-day learning period fails.

An ensemble with less members and less spread than another can lead to better combinations. Member selection was therefore discussed. Additional models always bring improvements, but slightly related to their individual performances.

Future work should address this issue. Additional sources of uncertainty could be introduced. Meteorological ensemble forecasts and Monte Carlo simulation on other input data are necessary steps to account for all uncertainties. The computational costs will be a crucial point. Relevant strategies are needed for the introduction of Monte Carlo methods together with discrete changes (in the model formulation, as performed in this paper through changes in physical parameterizations and numerical approximations).

An obvious future-work lies in forecasting the weights. As shown in this paper, the potential of model combination is very high and it is much higher than what is achieved with the forecasted combinations tested so far. Specific learning algorithms should be involved.

Ensemble forecasting may also deliver probabilistic forecasts. It would be an improvement of the forecasts through additional information. It would help in assessing uncertainties and it would allow reasonable integrated-uses of air-quality models, e.g. for risk assessment.

Finally an open question is the relations between ensemble forecast and classical data assimilation. Would sequential or variational data assimilation perform better than ensemble-based forecast? How could both strategies be combined? Data assimilation may be performed on each member of an ensemble, or only on a reference member (to be determined) whose updates (from the data assimilation procedure) would be applied to other members (e.g., optimized emissions or corrected initial conditions). In addition, the ensemble spread may be valuable information in the data assimilation procedure.

**Acknowledgments.** The first author is partially supported by the Île-de-France region. We thank the monitoring networks that provided the numerous observations. We also thank Gilles Stoltz (CNRS) for his introduction to machine learning algorithms.

## Notes

1. European Centre for Medium-Range Weather Forecasts
2. U.S. Geological Survey
3. Co-operative Programme for Monitoring and Evaluation of the Long-range Transmission of Air Pollutants in Europe
4. Agence de l'Environnement et de la Maîtrise de l'Énergie

## References

- Beekmann, M., and C. Derognat, 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(D17), 8,559, 2003.
- Boutahar, J., S. Lacour, V. Mallet, D. Quélo, Y. Roustan, and B. Sportisse, Development and validation of a fully modular platform for numerical modelling of air pollution: POLAIR, *Int. J. Env. and Pollution*, 22(1/2), 17–28, 2004.
- Cesa-Bianchi, N., P. M. Long, and M. K. Warmuth, Worst-case quadratic loss bounds for prediction using linear functions and gradient descent, *IEEE Trans. Neural Net.*, 7(3), 604–619, 1996.
- Chang, J., R. Brost, I. Isaken, S. Madronich, P. Middleton, W. Stockwell, and C. Walcek, A three-dimensional Eulerian acid deposition model: physical concepts and formulation, *J. Geophys. Res.*, 92(D12), 14,681–14,700, 1987.
- Dabberdt, W. F., and E. Miller, Uncertainty, ensembles and air quality dispersion modeling: applications and challenges, *Atmos. Env.*, 34(27), 4,667–4,673, 2000.
- Delle Monache, L., and R. B. Stull, An ensemble air-quality forecast over western Europe during an ozone episode, *Atmos. Env.*, 37, 3,469–3,474, 2003.
- ESQUIF, Étude et simulation de la qualité de l'air en île de France – rapport final, 2001.
- Galmarini, S., et al., Ensemble dispersion forecasting – part I: concept, approach and indicators, *Atmos. Env.*, 38(28), 4,607–4,617, 2004.
- Hanna, S. R., and J. M. Davis, Evaluation of a photochemical grid model using estimates of concentration probability density functions, *Atmos. Env.*, 36, 1,793–1,798, 2002.
- Hanna, S. R., J. C. Chang, and M. E. Fernau, Monte Carlo estimates of uncertainties in predictions by a photochemical grid model (UAM-IV) due to uncertainties in input variables, *Atmos. Env.*, 32(21), 3,619–3,628, 1998.
- Hanna, S. R., Z. Lu, H. C. Frey, N. Wheeler, J. Vukovich, S. Arunachalam, M. Fernau, and D. A. Hansen, Uncertainties in predicted ozone concentrations due to input uncertainties for the UAM-V photochemical grid model applied to the July 1995 OTAG domain, *Atmos. Env.*, 35(5), 891–903, 2001.
- Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky, Bayesian model averaging: a tutorial, *Stat. Sci.*, 14(4), 382–417, 1999.
- Horowitz, L. W., et al., A global simulation of tropospheric ozone and related tracers: description and evaluation of MOZART, version 2, *J. Geophys. Res.*, 108(D24), 2003.
- Krishnamurti, T. N., C. M. Kishtawal, Z. Zhang, D. B. T. LaRow, and E. Williford, Multimodel ensemble forecasts for weather and seasonal climate, *J. Climate*, 13, 4,196–4,216, 2000.
- Louis, J.-F., A parametric model of vertical eddy fluxes in the atmosphere, *Boundary-Layer Meteor.*, 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(D8), 9,740–9,752, 1987.

- Mallet, V., and B. Sportisse, Data processing and parameterizations in atmospheric chemistry and physics: the AtmoData library, *Tech. Rep. 12*, CERE, 2005.
- Mallet, V., and B. Sportisse, Uncertainty in a chemistry-transport model due to physical parameterizations and numerical approximations: An ensemble approach applied to ozone modeling, *J. Geophys. Res.*, *111*(D1), 2006.
- Mallet, V., D. Quélo, and B. Sportisse, Software architecture of an ideal modeling platform in air quality – A first step: Polyphemus, *Tech. Rep. 11*, CERE, 2005.
- McKeen, S., et al., Assessment of an ensemble of seven real-time ozone forecasts over eastern North America during the summer of 2004, *J. Geophys. Res.*, *110*(D21), 2005.
- Middleton, P., W. R. Stockwell, and W. P. L. Carter, Aggregation and analysis of volatile organic compound emissions for regional modeling, *Atmos. Env.*, *24A*(5), 1,107–1,133, 1990.
- Pagowski, M., et al., A simple method to improve ensemble-based ozone forecasts, *Geophys. Res. Lett.*, *32*, 2005.
- Russell, A., and R. Dennis, NARSTO critical review of photochemical models and modeling, *Atmos. Env.*, *34*, 2,283–2,234, 2000.
- Simpson, D., et al., Inventorying emissions from nature in Europe, *J. Geophys. Res.*, *104*(D7), 8,113–8,152, 1999.
- Stockwell, W. R., P. Middleton, J. S. Chang, and X. Tang, The second generation regional acid deposition model chemical mechanism for regional air quality modeling, *J. Geophys. Res.*, *95*(D10), 16,343–16,367, 1990.
- Stockwell, W. R., F. Kirchner, M. Kuhn, and S. Seefeld, A new mechanism for regional atmospheric chemistry modeling, *J. Geophys. Res.*, *102*(D22), 25,847–25,879, 1997.
- Straume, A. G., A more extensive investigation of the use of ensemble forecasts for dispersion model evaluation, *J. Applied Meteor.*, *40*, 425–445, 2001.
- Straume, A. G., E. N. Koffi, and K. Nodop, Dispersion modeling using ensemble forecasts compared to ETEX measurements, *J. Applied Meteor.*, *37*, 1,444–1,456, 1998.
- Troen, I., and L. Mahrt, A simple model of the atmospheric boundary layer; sensitivity to surface evaporation, *Boundary-Layer Meteor.*, *37*, 129–148, 1986.
- Verwer, J. G., W. Hundsdorfer, and J. G. Blom, Numerical time integration for air pollution models, *Surveys on Math. for Indus.*, *10*, 107–174, 2002.
- Warner, T. T., R.-S. Sheu, J. F. Bowers, R. I. Sykes, G. C. Dodd, and D. S. Henn, Ensemble simulations with coupled atmospheric dynamic and dispersion models: illustrating uncertainties in dosage simulations, *J. Applied Meteor.*, *41*, 488–504, 2002.
- Wesely, M. L., Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical models, *Atmos. Env.*, *23*, 1,293–1,304, 1989.
- Zhang, L., J. R. Brook, and R. Vet, A revised parameterization for gaseous dry deposition in air-quality models, *Atmos. Chem. Phys.*, *3*, 2,067–2,082, 2003.

---

Vivien.Mallet@cerea.enpc.fr, Bruno.Sportisse@cerea.enpc.fr