

## Ensemble forecast of analyses: Coupling data assimilation and sequential aggregation

Vivien Mallet<sup>1,2</sup>

Received 25 March 2010; revised 12 September 2010; accepted 17 September 2010; published 16 December 2010.

[1] Sequential aggregation is an ensemble forecasting approach that weights each ensemble member based on past observations and past forecasts. This approach has several limitations: The weights are computed only at the locations and for the variables that are observed, and the observational errors are typically not accounted for. This paper introduces a way to address these limitations by coupling sequential aggregation and data assimilation. The leading idea of the proposed approach is to have the aggregation procedure forecast the forthcoming analyses, produced by a data assimilation method, instead of forecasting the observations. The approach is therefore referred to as ensemble forecasting of analyses. The analyses, which are supposed to be the best a posteriori knowledge of the model's state, adequately take into account the observational errors and they are naturally multivariable and distributed in space. The aggregation algorithm theoretically guarantees that, in the long run and for any component of the model's state, the ensemble forecasts approximate the analyses at least as well as the best constant (in time) linear combination of the ensemble members. In this sense, the ensemble forecasts of the analyses optimally exploit the information contained in the ensemble. The method is tested for ground-level ozone forecasting, over Europe during the full year 2001, with a 20-member ensemble. In this application, the method proves to perform well with 28% reduction in root-mean-square error compared to a reference simulation, to be robust in time and space, and to reproduce many spatial patterns found in the analyses only.

**Citation:** Mallet, V. (2010), Ensemble forecast of analyses: Coupling data assimilation and sequential aggregation, *J. Geophys. Res.*, 115, D24303, doi:10.1029/2010JD014259.

### 1. Introduction

[2] Data assimilation [e.g., Daley, 1993; Bouttier and Courtier, 1999; Cohn, 1997; Todling, 1999] is a well-known approach that combines information from different sources to improve an estimate of a system's state. The information sources are usually a numerical model, various observations, and error statistics. Under given assumptions, several widely used assimilation methods compute the so-called best linear unbiased estimator (BLUE), which minimizes the total variance of the state. Among the popular methods, one may cite the Kalman filters (e.g., the ensemble Kalman filter) [Evensen, 1994] or the four-dimensional variational assimilation (4D-Var) method [Le Dimet and Talagrand, 1986]. Data assimilation has been successful in many fields. It is especially successful in numerical weather forecasting where improved initial conditions can dramatically improve a forecast cycle [e.g., Buehner et al., 2010].

[3] Another approach for improving forecasts using observations is called sequential aggregation. It is employed to produce an improved forecast out of an ensemble of forecasts. Each forecast of the ensemble is given a weight that depends on past observations and past forecasts. An aggregated forecast is then formed by the weighted linear combination of the forecasts of the ensemble. The aggregation is sequential since it is repeated before each forecast step, with updated weights. The weights can be computed with machine learning algorithms so that appealing theoretical results may hold in practical applications [Mallet et al., 2009]. Other forms of aggregation have been applied in geophysical forecasts, for example, with the dynamic linear regression in air quality [Pagowski et al., 2006] or with least-squares methods in climatology [Krishnamurti et al., 2000] and in air quality [Mallet and Sportisse, 2006].

[4] In this paper, the focus is on machine learning algorithms because of their key theoretical properties. If the forecast performance is measured by a mean quadratic discrepancy with the observations, the learning algorithms guarantee that, in the long run, the aggregated forecast performs at least as well as the best linear combination of models that is constant in time. In other words, over a long-enough time period, the mean quadratic error of any constant combination (that is, any linear combination of models with

<sup>1</sup>INRIA, Paris-Rocquencourt Research Center, Le Chesnay, France.

<sup>2</sup>Also at CERE, Joint Laboratory École des Ponts ParisTech, EDF R&D, Université Paris-Est, Marne la Vallée, France.

weights that do not depend on time) tends to be greater than or equal to the mean quadratic error of the aggregated forecast. In particular, the aggregated forecast will perform better than any individual model (whose forecast is the linear combination with unitary weight on the model and null weights otherwise) and the ensemble mean (associated with uniform weights). Both data assimilation and sequential aggregation have advantages and drawbacks, in terms of theoretical framework, practical application, performance, and computational efficiency. An introduction to these techniques can be found in Appendix A.

[5] In this paper, a method called “ensemble forecast of analyses” (EFA) is designed to combine both approaches. A key motivation is to address two important limitations of sequential aggregation. The first limitation is that sequential aggregation may not take into account the observational errors. The linear combination of the ensemble forecasts is determined to minimize its discrepancy with the observations. Since the observations are not perfect, this approach is not entirely satisfactory. The second limitation is that the weights are computed only at the locations and for the variables that are observed. Computing weights for other locations and other variables is beyond the scope of the methods. It is possible to compute a single set of weights for all locations. In this case, it can be argued, and sometimes observed in applications, that the weights are reasonably spatially robust, but there is no theoretical framework to support this assumption. Further details on the motivation for the development of EFA are given in section 2.1.

[6] In EFA, the leading idea is to carry out sequential aggregation to forecast an analysis, instead of observations. The analysis is the result of a data assimilation step. In some sense, it is the best estimate of the true state that can be produced out of the available information. Analyses can be produced whenever observations become available. Therefore the sequence is similar to that of sequential aggregation. First, EFA produces a forecast of the forthcoming analysis. Second, when the date of the forecast is reached, the observations become available and the analysis can be computed. Third, this analysis is compared with the EFA output. After that step, the cycle goes on with EFA producing a forecast for the next analysis. In short, the EFA method tries to predict, using sequential aggregation, the analysis that *will* be computed with future observations.

[7] The observation errors are taken into account in the analysis, and EFA naturally computes a multivariate and multidimensional field in the same space as the model state. The method is described in section 2. This fairly general method is illustrated through application to ozone forecasting in section 3.

## 2. Ensemble Forecast of Analyses

### 2.1. Motivation

[8] Sequential aggregation is appealing because it is robust. The theoretical guarantee (to compete in the long run against the best constant linear combination) holds in practical applications because learning algorithms assume essentially nothing about the observations and the ensemble. The methods are pessimistic enough to perform well (in terms of minimization of forecast errors) in any situation. This robustness is an important feature in operational forecasts.

[9] A criticism one usually makes of statistical methods is that they may perform well on average but they miss the extreme events. Sequential aggregation relies on physical models (through a linear combination of them), which circumvents this limitation to the extent that the models are able to capture the events. This was shown in a practical application with ground-level ozone data [Mallet *et al.*, 2009].

[10] In data assimilation, when only the initial conditions are constrained, the sensitivity to these initial conditions is a key parameter. In meteorology, this sensitivity is so high that there is room for large improvements in the forecasts. This is not the case in air quality where the benefit of data assimilation is significant only for about a day after the assimilation period [e.g., Elbern and Schmidt, 2001]. Indeed a chemistry-transport model applied to photochemistry is so stable that changes in its initial concentrations are quickly dampened. Note that sequential aggregation is not subject to this limitation as its weights may be used for a long time period.

[11] One drawback of sequential aggregation is that it does not take into account observational errors. Its objective is to minimize the discrepancy between the forecasts and the observations, while data assimilation aims to estimate the true state of a system. This can make a clear difference when significant instrumental errors or representativeness errors are involved. In this case, the data assimilation approach makes more sense.

[12] Another drawback of sequential aggregation is that the weights are computed considering the observation locations only. It is possible to compute the same weights at all observed locations, and this is even recommended, so that the weights may reasonably be applied at nonobserved locations. If the weights were computed independently at the observation locations, it would be unlikely that they would be relevant elsewhere, even in the vicinity of the observed locations. In contrast, when the weights are not space dependent, they are likely to still perform well in some region enclosing the observation network. However, this region might not cover entirely the target region; that is, the region over which forecasts are to be computed. For ground-level ozone, tests not reported here showed that the weights can be applied relatively far from the observation stations (about 1.5° to 2° in latitude/longitude around the stations, at European scale). Nevertheless, no theoretical framework supports the weights being applied outside the observation network.

[13] The same is true for multivariate forecasts (e.g., in air quality for ozone and particulate matter at the same time). The quality of the weights is guaranteed for the variables that are observed, but the same weights may not perform well for the nonobserved variables. This contrasts with data assimilation that naturally corrects the state of a system based on the error covariances between the state components, and hence the spatial distribution of a multivariate field is corrected. However, in applications, this is not always a straightforward process because the error covariance matrix (see section A.1) must be properly estimated. It is often advocated not to include all nonobserved variables in the state vector or, equivalently, to assume no error correlation between observed and some nonobserved variables. Nonetheless, the improvements in the controlled variables will eventually propagate to the other variables during the simulation because of the coupling between the variables in the model.

## 2.2. Introduction to the Method

[14] The EFA is introduced to address the issues discussed in section 2.1. The leading idea is to carry out sequential aggregation in order to *forecast an analysis* instead of the observations.

[15] The analysis to be forecast will be computed with a background state, forecast by some model, and the observations as soon as they become available. This analysis will therefore be produced at the same time as the observations, maybe with a delay due to the computational cost of the assimilation method.

[16] In EFA, a sequential-aggregation algorithm is applied before each forecast time step. It computes *one weight per model and per state component* (and per forecast step). Unlike the usual aggregation procedure [Mallet et al., 2009] (see also section A.2), there is no need to compute a single weight per model (and per step) for all observed locations, because all state components have a value in the analysis. This a clear advantage compared to the sequential aggregation alone in that the weights are determined for a single target in the state vector, that is, a single grid cell and a single variable.

## 2.3. Notation

[17] Let  $\tilde{\mathbf{x}}_t^f$  be the forecast at time  $t \in \{1, \dots, T\}$  of a reference model  $\mathcal{M}_{t-1}$ . It is also referred to as the state vector of the model. The analysis state vector is denoted  $\tilde{\mathbf{x}}_t^a$ . At time  $t$ , the observation vector is denoted  $\mathbf{y}_t$ . An observation operator  $\mathbf{H}_t$  maps from the space of the model state to the observation space:  $\mathbf{H}_t \tilde{\mathbf{x}}_t^f$  can be compared to  $\mathbf{y}_t$ . Hereinafter, it is assumed that the observation operator is linear. Nevertheless, EFA can be applied with a nonlinear observation operator, provided that a proper data assimilation method is used.

[18] An ensemble of forecasts is available at time  $t$  with the sequence  $(\mathbf{x}_t^1, \dots, \mathbf{x}_t^M)$ , where  $M$  is the number of models. The forecast of a single model  $m$  is therefore denoted  $\mathbf{x}_t^m$ . The models are denoted  $\mathcal{M}_t^m$  at time  $t$ :  $\mathbf{x}_{t+1}^m = \mathcal{M}_t^m(\mathbf{x}_t^m)$ . The  $i$ th components of the aforementioned vectors are denoted with an additional subscript:  $\tilde{x}_{i,t}^f, y_{i,t}, x_{i,t}^m$ .

## 2.4. Calculations and Algorithm

[19] First, a simulation is carried out with assimilation. It produces a sequence of forecasts  $\tilde{\mathbf{x}}_t^f$  and a sequence of analyses  $\tilde{\mathbf{x}}_t^a$ . Henceforth, for the sake of clarity, we assume that the assimilation method is simply an optimal interpolation, computing the best linear unbiased estimator (BLUE; see section A.1 for its formula). Nevertheless, a Kalman filter, a 4D-Var, or any other method can be used.

[20] Furthermore, an ensemble of simulations  $(\mathbf{x}_t^1, \dots, \mathbf{x}_t^M)$  is available at any forecast time step  $t$ . This ensemble may or may not be related to the sequence  $\tilde{\mathbf{x}}_t^f$ . The only mandatory relation is that  $\tilde{\mathbf{x}}_t^f$  and all  $\mathbf{x}_t^m$  ( $m \in \{1, \dots, M\}$ ) have the same size and represent the same quantities.

[21] At time  $t-1$ , the observations  $\mathbf{y}_{t-1}$  become available and are then assimilated to provide the analysis. Thus  $\tilde{\mathbf{x}}_{t-1}^a$  becomes available. The ensemble of forecasts for  $t$ ,  $\mathbf{x}_t^m$  ( $m \in \{1, \dots, M\}$ ) is also generated. Based on all previous analyses  $\tilde{\mathbf{x}}_{t'}^a$  ( $t' < t-1$ ) and on all ensemble computations  $\mathbf{x}_{t'}^m$  ( $m \in \{1, \dots, M\}, t' \leq t$ ), an aggregated forecast  $\hat{\mathbf{x}}_t$  is produced with the weight vectors  $(\mathbf{w}_t^m)_m$ . Here,  $\mathbf{w}_t^m$  is the vector of

weights associated with model  $m$  and time  $t$ ; it contains one weight per component in the model state. Thus,

$$\hat{\mathbf{x}}_{i,t} = \sum_{m=1}^M w_{i,t}^m x_{i,t}^m. \quad (1)$$

This aggregated forecast should be as close as possible to the analysis to be computed at the next time  $t$ , thus  $\tilde{\mathbf{x}}_t^a$ .

[22] Many methods for sequential aggregation may be used to compute the weight vectors  $\mathbf{w}_t^m$ . If the discounted ridge regression is used (see section A.2 for further details), the weights will satisfy

$$\mathbf{w}_{i,t} = \operatorname{argmin}_{\mathbf{u} \in \mathbb{R}^M} \left[ \lambda \|\mathbf{u}\|_2^2 + \sum_{t'=1}^{t-1} (1 + \psi_{t-t'}) \left( \tilde{x}_{i,t'}^a - \sum_{m=1}^M u_m x_{i,t'}^m \right)^2 \right], \quad (2)$$

if  $\mathbf{w}_{i,t} = (w_{i,t}^1, \dots, w_{i,t}^M)^T$  is the vector of weights for the  $i$ th state component,  $\lambda > 0$  and  $\psi_t > 0$  is a decreasing sequence.

[23] In the case that the optimal interpolation is the assimilation algorithm used and that the discounted ridge regression is the aggregation method, the EFA algorithm is performed using the following steps:

[24] 1. Initialization

[25] (i) Initial conditions:  $\tilde{\mathbf{x}}_0^a, \mathbf{x}_0^m, \mathbf{x}_1^m = \mathcal{M}_0^m(\mathbf{x}_0^m)$ , for  $m \in \{1, \dots, M\}$

[26] (ii) Assimilation parameters:  $\mathbf{B}_t, \mathbf{R}_t$

[27] (iii) Aggregation parameters:  $\lambda, \psi_t$

[28] 2. Time loop; forecasting time  $t = 2, \dots, T$ :

[29] (i) Forecast of the reference model:

$$\tilde{\mathbf{x}}_{t-1}^f = \tilde{\mathcal{M}}_{t-2}(\tilde{\mathbf{x}}_{t-2}^a). \quad (3)$$

[30] (ii) Computing the analysis:

$$\tilde{\mathbf{x}}_{t-1}^a = \tilde{\mathbf{x}}_{t-1}^f + \mathbf{B}_{t-1} \mathbf{H}_{t-1}^T (\mathbf{H}_{t-1} \mathbf{B}_{t-1} \mathbf{H}_{t-1}^T + \mathbf{R}_{t-1})^{-1} \cdot (\mathbf{y}_{t-1} - \mathbf{H}_{t-1} \tilde{\mathbf{x}}_{t-1}^f). \quad (4)$$

[31] (iii) Ensemble of forecasts: for all  $m \in \{1, \dots, M\}$ ,

$$\mathbf{x}_t^m = \mathcal{M}_{t-1}^m(\mathbf{x}_{t-1}^m). \quad (5)$$

[32] (iv) Computing a weight vector for any state component  $i$ :

$$\mathbf{w}_{i,t} = \operatorname{argmin}_{\mathbf{u} \in \mathbb{R}^M} \left[ \lambda \|\mathbf{u}\|_2^2 + \sum_{t'=1}^{t-1} (1 + \psi_{t-t'}) \left( \tilde{x}_{i,t'}^a - \sum_{m=1}^M u_m x_{i,t'}^m \right)^2 \right]. \quad (6)$$

[33] (v) Computing the forecast:

$$\hat{\mathbf{x}}_{i,t} = \sum_{m=1}^M w_{i,t}^m x_{i,t}^m. \quad (7)$$

## 2.5. Theoretical Background

[34] It can be shown [Cesa-Bianchi and Lugosi, 2006; Mallet et al., 2007b] that, for every state component and, in the long run, the aggregated forecast will perform as well as the best constant linear combination of the ensemble's forecasts (this is the same guarantee as introduced in section A.2, equation (A3), except that the reference is the analysis, not the observations):

$$\sum_{t=1}^T \frac{1}{T} (\tilde{x}_{i,t}^a - \hat{x}_{i,t})^2 - \min_{\mathbf{u} \in \mathbb{R}^M} \sum_{t=1}^T \frac{1}{T} \left( \tilde{x}_{i,t}^a - \sum_{m=1}^M u_m x_{i,t}^m \right)^2 \lesssim \mathcal{O}\left(\frac{\ln T}{T}\right). \quad (8)$$

[35] In the long run, EFA produces a forecast that is at least as close to the best a posteriori estimate of the true state (BLUE) as the best constant (in time) combination is. EFA therefore achieves the potential of the ensemble in terms of constant linear combinations. Another formulation is:

$$\sum_{t=1}^T \frac{1}{T} \|\tilde{\mathbf{x}}_t^a - \hat{\mathbf{x}}_t\|_2^2 - \min_{\mathbf{u}_1, \dots, \mathbf{u}_N \in \mathbb{R}^M} \cdot \sum_{t=1}^T \frac{1}{T} \|\tilde{\mathbf{x}}_t^a - C(\mathbf{u}_1, \dots, \mathbf{u}_N, \mathbf{x}_t^1, \dots, \mathbf{x}_t^M)\|_2^2 \lesssim \mathcal{O}\left(\frac{\ln T}{T}\right), \quad (9)$$

where  $N$  is the size of a state vector and  $\mathbf{z}_t = C(\mathbf{u}_1, \dots, \mathbf{u}_N, \mathbf{x}_t^1, \dots, \mathbf{x}_t^M)$  is the constant (in time) linear combination with weights  $\mathbf{u}_1, \dots, \mathbf{u}_N$ :

$$z_{i,t} = \sum_{m=1}^M u_{m,i} x_{i,t}^m,$$

if  $u_{m,i}$  is the  $m$ th component of  $\mathbf{u}_i$ .

## 2.6. Comments

### 2.6.1. Aggregation Methods

[36] Other aggregation algorithms may be used with the same or similar theoretical results. An example is provided in Appendix B.

[37] The discrepancy between the analysis and the aggregated forecast is called the loss and it is measured with a quadratic difference in equation (8). Other losses can be considered; many aggregation methods can be applied to any loss that is convex in the weights. For example, while the loss in equation (8) focuses on a single state component, a more global loss could be introduced to take into account spatial structures in the fields.

### 2.6.2. Aggregation Target

[38] While the usual aggregation strategy (see section A.2) forecasts the (erroneous) observations, the EFA procedure aims to forecast the best estimate of the state that will be available in the near future. This is more satisfactory because the a posteriori estimate of the true state takes into account all error sources, it can be a multivariate spatial field, and it is more complete information about the observed system. Both strategies can be employed at the same time. In that case, EFA would produce the forecast of the complete state, on one the hand, and an aggregation can be carried out at each observation location, on the other hand. The latter would produce a forecast closer to the observation (at each station) than the aggregation applied to all stations together (section A.2) or EFA.

[39] Although this is not an objective, note that EFA may better forecast the observations than the usual aggregation procedure. Indeed, the aggregation alone forecasts the observations directly, but it is constrained by many locations at the same time. EFA forecasts the analyses, which generally have a significant discrepancy with the observations, but the procedure is much less constrained since it is applied per state component. In short, EFA may forecast the analyses well enough to better estimate the observations than the aggregation alone (when applied to all stations at the same time).

## 3. Application to Air Quality

### 3.1. Experiment Setup

[40] EFA is applied to forecast ground-level ozone concentrations over Europe on the next day at 15:00 UTC, when ozone concentrations often reach their daily peak. All simulations are carried out for the full year 2001 by Eulerian chemistry-transport models generated within the Polyphemus platform [Mallet et al., 2007b].

[41] The models employed in this application (including the model with which the analyses are generated) are part of the ensemble introduced in Garaud and Mallet [2010]. All models have a horizontal resolution of  $0.5^\circ$ . Meteorological inputs are from the European Centre for Medium-Range Weather Forecasts (12-hr forecast cycles starting from analyzed fields). The raw emissions are from the European Monitoring and Evaluation Programme (EMEP) database and they are chemically processed following Middleton et al. [1990]. The biogenic emissions are generated following Simpson et al. [1999]. The lateral and top boundary conditions are from simulations by MOZART 2 [Horowitz et al., 2003]. In the generation of the ensemble, the input fields may be in addition randomly perturbed.

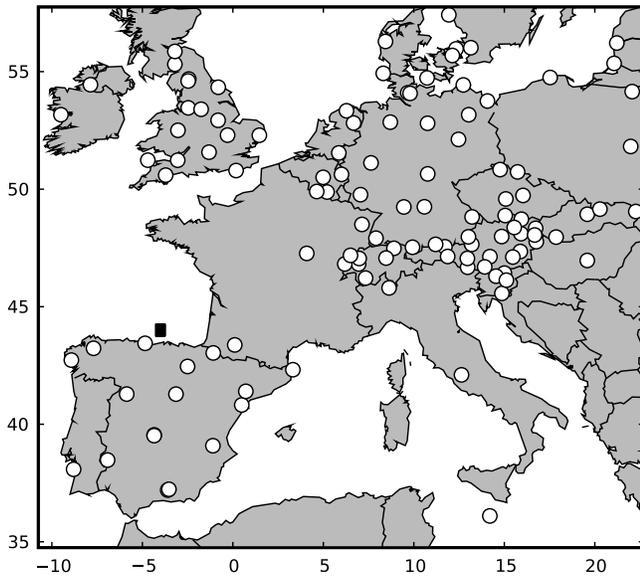
#### 3.1.1. Generation of the Analyses

[42] The model with which the analyses are generated (this model is referred to as the fourth reference model (R3) by Garaud and Mallet [2010]) uses the Regional Atmospheric Chemistry Mechanism (RACM) [Stockwell et al., 1997]. It models the vertical diffusion with the Troen and Mahrt [1986] parameterization within the unstable boundary layer and with the Louis [1979] parameterization otherwise.

[43] The analyses are produced with the optimal interpolation which demonstrated good performance in Wu et al. [2008] for ozone forecast. The controlled state is the ozone concentration in the three first model layers above the surface. The assimilated observations are hourly ground-level ozone concentrations from the EMEP network. Every hour, the network provides observations at about 90 active background stations over Europe (Figure 1).

[44] The observation error covariance matrix (see section A.1) is taken diagonal:  $\mathbf{R}_t = r\mathbf{I}_t$  where  $r$  is a scalar variance,  $\mathbf{I}_t$  is the  $O_t \times O_t$  identity matrix, and  $O_t$  is the number of observations at time  $t$ . The background error covariance matrix  $\mathbf{B}_t$  is taken in Balgovind form [Balgovind et al., 1983], that is, with covariance between two state components determined by their geographical distance:

$$\text{covar}(l_h, l_v) = b \left(1 + \frac{l_h}{L_h}\right) e^{-\frac{l_h}{L_h}} \left(1 + \frac{l_v}{L_v}\right) e^{-\frac{l_v}{L_v}}, \quad (10)$$



**Figure 1.** Simulation domain with the monitoring stations of EMEP network. The black rectangle, to the north of Spain, locates the grid cell in which the weights of Figure 5 are computed by EFA.

where  $l_h$  and  $l_v$  are the distances in the horizontal and in the vertical respectively,  $L_h$  and  $L_v$  are a characteristic lengths along the horizontal and the vertical respectively, and  $b$  is a variance. The characteristic lengths are similar to those of *Wu et al.* [2008]:  $L_h = 1^\circ$  and  $L_v = 150m$ . The variances  $b$  and  $r$  are determined so that the simulation passes the  $\chi^2$  diagnosis [e.g., *Ménard et al.*, 2000], that is, so that there is consistency between the error statistics and the innovations:

$$\sum_{t=1}^T \frac{1}{TO_t} (\mathbf{y}_t - \mathbf{H}_t \mathbf{x}_t^f)^T (\mathbf{R}_t + \mathbf{H}_t \mathbf{B}_t \mathbf{H}_t^T)^{-1} (\mathbf{y}_t - \mathbf{H}_t \mathbf{x}_t^f) \simeq 1. \quad (11)$$

Several simulations were carried out to find a reasonable variance pair that satisfies the relation (11). With  $(b, r) = (190 \text{ g}^2 \text{ m}^{-6}, 51 \text{ g}^2 \text{ m}^{-6})$ , the sum in equation (11) is 1.02. The ratio (The individual values of  $r$  and  $b$  have no impact on the assimilation procedure as long as the ratio between  $r$  and  $b$  remains the same.) between  $r$  and  $b$  is therefore about 0.27, which is also an appropriate ratio according to *Wu et al.* [2008]. Note that the individual values of  $r$  and  $b$  have no impact on the assimilation procedure as long as the ratio between  $r$  and  $b$  remains the same.

[45] It should be emphasized that checking the reliability of the assimilation procedure is an important step in EFA as the final target is the analysis. One mistake would be to underestimate the observation variance: Then the analysis and therefore the result of EFA would be artificially close to the observations at station locations. Roughly speaking, as the ratio  $r/b$  tends to 0, the procedure tends to sequential aggregation carried out at individual observed locations, which is certainly not the objective here. EFA is designed to forecast the best a posteriori knowledge of the state, not the observations.

### 3.1.2. Ensemble Simulations

[46] A model in the ensemble is uniquely defined by its physical formulation, its numerical discretization, and, for

convenience, its input data. Note that none of the models are simplified. Any of them could a priori be used for operational forecasting or for modeling activities; the comparison with observations, however, shows that several models have poor average performance. EFA is applied to an ensemble of 20 models. More members could possibly be added, but the number of models is kept low so the spin-up period in the learning procedure should not be too long. A 30-day spin-up period was found to be long enough for 20 models. The EFA algorithm is applied starting from 2 January 2001, and the performance is evaluated from 1 February 2001. For the first ensemble forecast, at 15:00 on 1 February 2001, 30 past analyses are available in each grid cell to compute the 20 weights.

[47] Out of the simulations of *Garaud and Mallet* [2010], 19 members are selected for inclusion in the ensemble. Five of them are very similar to the reference simulation, except that one or two parameterizations were changed (for vertical diffusion, and for vertical-wind diagnosis). Following the terminology of *Garaud and Mallet* [2010], these five models are the six “reference members”, except (R3). The other members were randomly generated based on 30 alternatives (with two, three, or four options) related to the physical parameterizations, the numerical discretization, and the input data. Tests show that the overall results of EFA are not sensitive to the choice of these 19 members. A fully random selection of these members gives very similar results.

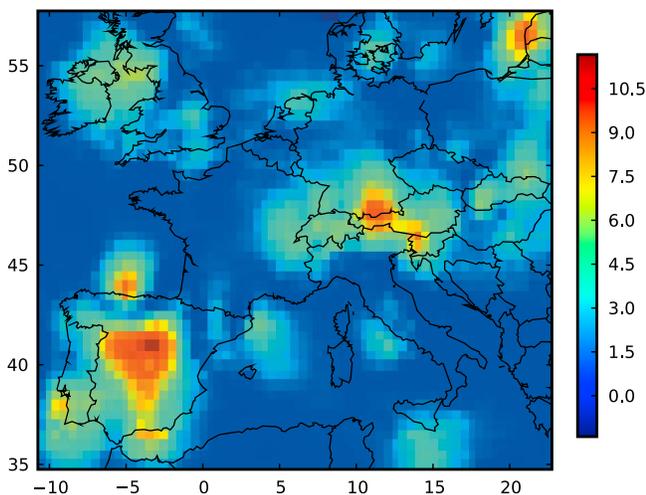
[48] The 20 member of the ensemble is based on the reference model involved in an assimilation cycle. Since sequential data assimilation is applied to that model, it is possible to compute an improved forecast starting from an earlier analysis. In that case, the optimal interpolation is carried out, on the reference model (from the beginning of 2001) until 19:00 in day  $D$ , and a subsequent forecast is carried out until 15:00 in day  $D+1$ . This is applied every day in a cycle supposed to mimic operational forecasting. Contrary to any of the 19 other members of the ensemble, the inclusion of this 20th member makes a noteworthy (though not essential) improvement to the EFA results.

### 3.1.3. Ensemble Forecast of the Analyses

[49] A slight adjustment to the algorithm is needed since the assimilation (equation (4)) is applied on an hourly basis, while the aggregation is carried out on a daily basis (to forecast the ozone field at 15:00). If the time index  $t$  refers to hours, starting from  $t = 1$  on the 2 January 2001 at 01:00, then the assimilation produces analyses  $\tilde{\mathbf{x}}_t^a$  for every hour (equation (4)). But the aggregation occurs only for any  $t = 24n + 15$  ( $n \geq 0$ ), so only  $\mathbf{w}_{i,24n+15}$  (equation (12)) and  $\tilde{\mathbf{x}}_{i,24n+15}^a$  (equation (7)) are computed. In addition, the weights are computed only on the basis of data at 15:00, for  $n \geq 0$ :

$$\mathbf{w}_{i,24n+15} = \operatorname{argmin}_{\mathbf{u} \in \mathbb{R}^M} \left[ \lambda \|\mathbf{u}\|_2^2 + \sum_{n'=1}^{n-1} (1 + \psi_{n-n'}) \left( \tilde{\mathbf{x}}_{i,24n'+15}^a - \sum_{m=1}^M u_m \mathbf{x}_{i,24n'+15}^m \right)^2 \right]. \quad (12)$$

These adjustments do not alter the theoretical guarantees. In the long run, the result of EFA will be at least as good as that of the best constant (in time) linear combination considering only the concentrations at 15:00.



**Figure 2.** Difference between the RMSE ( $\mu\text{g m}^{-3}$ ) for the reference with assimilation and the RMSE for EFA. In each grid cell, the RMSE is computed with all the 15:00 concentrations from 1 February to 30 December.

[50] In this application, only the ground-level ozone field (that is, ozone concentrations in the first layer of the models) is aggregated, not the whole state. Nevertheless it would be possible to aggregate the three-dimensional (3-D) ozone field and the 3-D concentration fields of other chemical species. Due to the assimilation procedure, the concentrations of all chemical species might improve as the corrections applied to ozone propagate into the model.

[51] The aggregation parameters were chosen according to previous studies [Mallet *et al.*, 2007a; Gerchinovitz *et al.*, 2008; Mallet *et al.*, 2009]. A few preliminary tests on the available data have also been conducted. The penalization is weighted with  $\lambda = 125$  and the discounting coefficients are  $\psi_n = \gamma/t^2$  with  $\gamma = 20$ . The results proved not to be very sensitive to these parameters. The preliminary tests explored values of  $\lambda$  in [75,125] and of  $\gamma$  in [10,30] without noteworthy changes in the performance. This is consistent with the low dependency of the aggregation results to the parameters, which was observed in Mallet *et al.* [2009].

### 3.2. Results and Discussion

[52] The reference model, to which the optimal interpolation is applied, shows good performance and could be used in operational forecasting. It is used to better quantify the improvements brought by the EFA procedure. The forecasts (starting from analyses at 19:00) of the assimilation cycle are also taken as a means to assess the ensemble forecasts. These forecasts will be referred to as those of the reference model with assimilation. Note that the reference model is not in the ensemble, but the reference model with assimilation is.

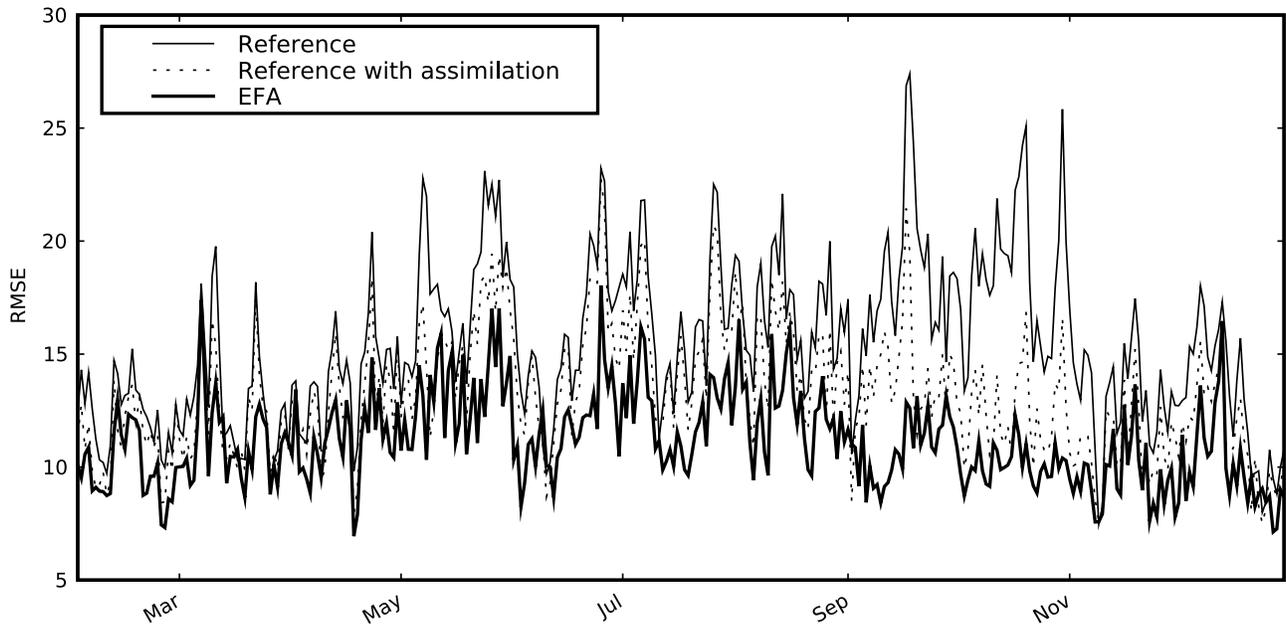
[53] The performance is first evaluated with the root-mean-square error (RMSE) between the forecasts and the analyses, from 1 February to 30 December. All grid cells in the first model layer are included in the calculation. The RMSE between the reference forecasts and the analyses is  $15.8 \mu\text{g m}^{-3}$ . The RMSE between the reference forecasts with assimilation and the analyses is the lowest among all members in the ensemble:  $13.5 \mu\text{g m}^{-3}$  which is a 15% reduction of the error. The RMSE between EFA and the analyses is  $11.3 \mu\text{g m}^{-3}$ ,

which is a 28% reduction of the error, compared to the reference simulation. Note that, in a few tests with ensembles composed of the reference model with assimilation and of 19 randomly generated models, the RMSE remained in the range of  $11.3\text{--}11.4 \mu\text{g m}^{-3}$ .

[54] To evaluate the spatial distribution of the improvements, the RMSE of time series may be computed on a grid cell basis. In Figure 2, the difference between the RMSE for the reference with assimilation and the RMSE for EFA demonstrates that EFA performs better than the reference with assimilation in most grid cells—its RMSE is lower in over 90% of the grid cells. The largest improvements are essentially found in regions that are observed. In the observed regions, the analyses differ the most from the ensemble simulations, and there is room for improvement. The RMSEs may also be computed on a daily basis (Figure 3). Again, EFA performs better than the reference, with or without assimilation, at over 90% of the dates. It is noteworthy that the peak RMSEs are significantly lower with EFA, even compared to the reference with assimilation, for example in September, October, late June, and late July. In this application, EFA is therefore not only efficient, with low global RMSE, but also robust in the sense that it does not introduce large local errors.

[55] In each grid cell, the weights of EFA are learned independently of the other grid cells. Yet EFA is able to reproduce spatial patterns. On average, EFA produces patterns that are almost identical to that of the analyses. In Figure 4, the ground-level ozone field at 15:00, averaged from February to December, is shown for the reference simulation with and without assimilation, for the EFA and for the analyses. The two reference forecasts show very similar patterns, and a number of patterns that appear in the analyses are either missing or very faint. On the contrary, EFA clearly captures almost all patterns, such as the low concentrations off the coasts of northern Spain; the lower concentrations (compared to the references) over Ireland, over southern Scotland, over Slovakia, along Portuguese coasts, or in the vicinity of Lyon (France); the higher concentrations (compared to the references) in the vicinity of Madrid or in northern Germany. On seasonal averages (not shown here), EFA also produces almost the same maps as the analyses. On a daily basis, EFA still better catches the patterns than the reference with assimilation, but there are clear misses.

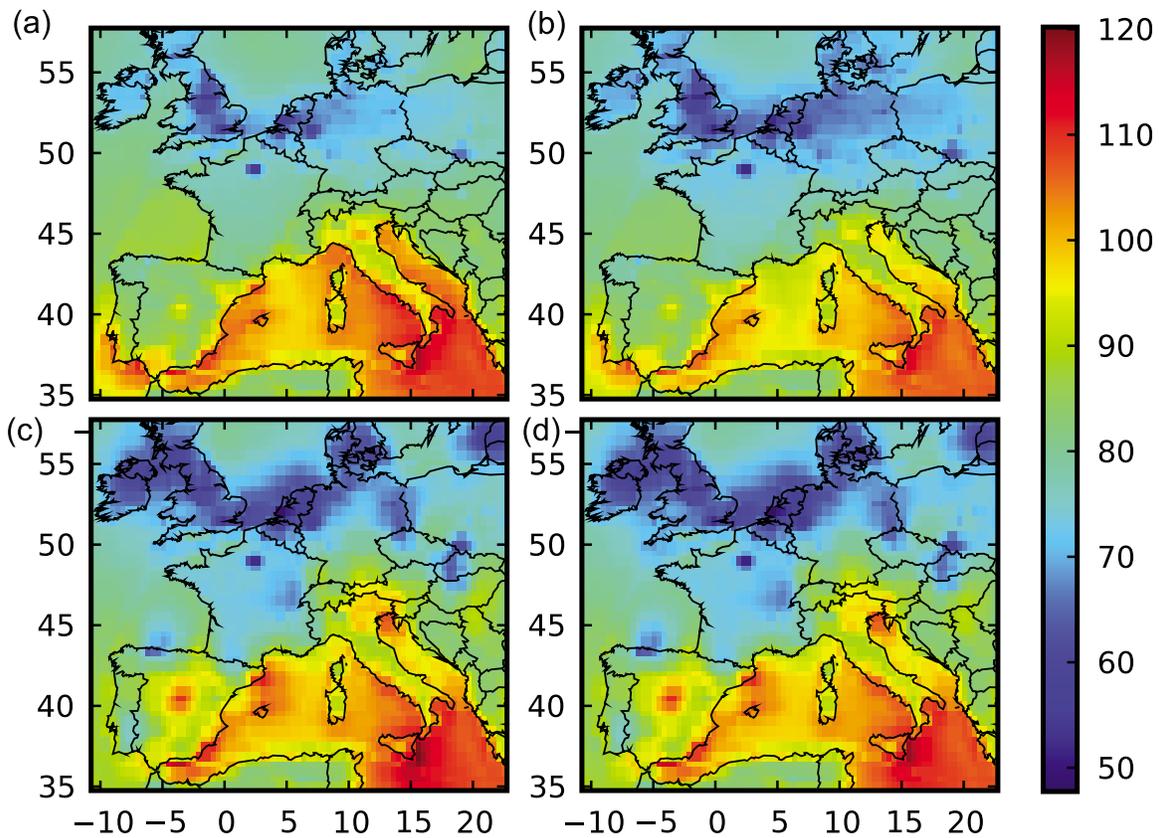
[56] In Figure 5, the time evolutions of the weights associated with the 20 members of the ensemble are shown in a grid cell to the north of Spain. This grid cell is marked with a black rectangle in Figure 1 and it is located in a region, with low concentrations, that is reconstituted only by EFA (see Figure 4). Many models are given a weight significantly different from zero. Hence the algorithm seems to extract information from the whole ensemble. The reference simulation with assimilation gets one of the highest weights, probably because it produces the closest concentrations to the analyses. At locations farther from the network, the reference model with assimilation can be clearly given the highest weights since it almost coincides with the analyses. It is difficult to further interpret weights that can be negative and of any magnitude. The member that has the best performance in a given grid cell is not necessarily given a strong weight, since the objective of the aggregation is to produce the best linear combination, not to select the best model.



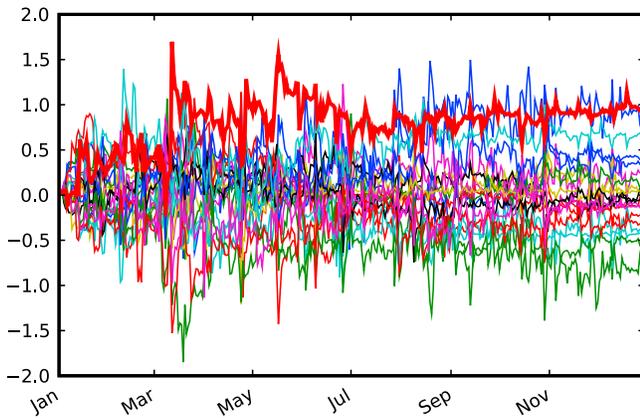
**Figure 3.** RMSE ( $\mu\text{g m}^{-3}$ ) against time for EFA and the reference simulation with and without assimilation. For each day, the RMSE is computed with all grid cells of the simulation domain.

[57] The time evolution of the weights is slightly erratic, which may bring into question the robustness of the method. The time evolution of the RMSE, as shown in Figure 3, demonstrates some robustness in the performance. Further-

more, the weights can be applied for  $D+2$  forecasts with a limited loss in performance. In this case, the weights originally computed for  $D+1$  are applied as such to the  $D+2$  forecasts of the ensemble members. The RMSE of EFA  $D+2$



**Figure 4.** Mean field of ground-level ozone concentration ( $\mu\text{g m}^{-3}$ ) at 15:00 from 1 February to 30 December. (a) Reference and (b) reference with assimilation. (c) EFA and (d) analyses.



**Figure 5.** Time evolution of the weights computed by EFA from 2 January to 31 December, in a grid cell to the north of Spain (black rectangle in Figure 1). The thick red line in the upper part is associated with the reference model with assimilation.

forecasts, from 1 February to 30 December, is  $11.9 \mu\text{g m}^{-3}$ . On the contrary, the  $D+2$  forecasts of the reference simulation with assimilation are close to those without assimilation, and their RMSE is  $15.0 \mu\text{g m}^{-3}$ . This low performance is essentially due to the low sensitivity of ozone forecasts to the initial conditions; the benefits of data assimilation vanish quickly. EFA overcomes this limitation. The  $D+2$  forecasts of EFA retain all the spatial patterns of the analysis: the mean ground-level ozone field is very similar to that of  $D+1$  forecasts and of the analyses (as shown in Figure 4).

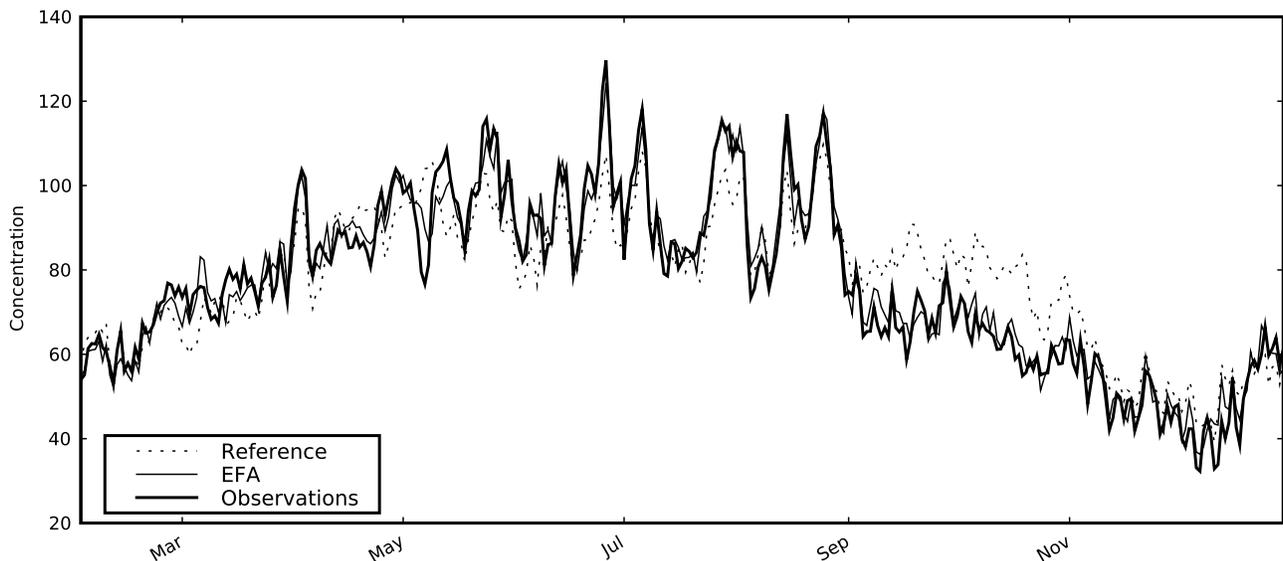
[58] Finally, the forecasts are compared to the observations (Figure 6). Note that the forecasts should not reproduce exactly the observations because there are measurement errors and representativeness errors. This is the reason why EFA does not aim to forecast the observations but the analyses. The distance to the observations is measured with a root-mean-square discrepancy (RMSD), whose formula is

the same as the RMSE but involves only observed locations. The RMSD between the forecasts and the observations is computed from 1 February to 30 December, using the 15:00 observations of EMEP network. The RMSD of the reference forecasts without assimilation is  $21.6 \mu\text{g m}^{-3}$ . The reference forecasts with the assimilation cycle perform better with a RMSD decreased by 9%:  $19.8 \mu\text{g m}^{-3}$ . The RMSD of EFA is 28% lower than that of the reference forecasts (without assimilation):  $15.6 \mu\text{g m}^{-3}$ . The  $D+2$  forecasts of EFA show again good performance, with a RMSD of  $16.4 \mu\text{g m}^{-3}$  (24% improvement compared to the reference without assimilation). In contrast, the reference with assimilation only slightly improves the  $D+2$  forecasts with a RMSD of  $21.0 \mu\text{g m}^{-3}$ .

#### 4. Conclusions

[59] This paper introduces an approach to couple data assimilation and sequential aggregation of ensemble simulations. Contrary to classical ensemble forecasting, the aim is not to forecast the observations but the best a posteriori knowledge of a model's state. In this paper, the latter is an analysis produced by a data assimilation method. The proposed approach therefore produces an ensemble forecast of the analysis. The ensemble aggregation is carried out by a learning algorithm, which produces a linear combination with weights associated with every model, time step and component of the model's state.

[60] The learning algorithm theoretically guarantees that, in the long run, the error between the aggregated forecasts and the analyses is at least as low as the error of the best constant combination of the ensemble members. This property is verified for every component of the state, independently of the other components, and without assumptions on the analyses and on the ensemble simulations. Thus the method competes in every grid cell and for every variable with the best constant linear combination of the ensemble members. In that sense, EFA is able to optimally exploit the information brought by the ensemble. Also, observation errors can be properly accounted for in the assimilation stage.



**Figure 6.** Concentration ( $\mu\text{g m}^{-3}$ ) averaged over all stations, against time, for the reference simulation, EFA and the observations.

[61] The method is applied to ground-level ozone forecasting with an ensemble of Eulerian chemistry-transport models. EFA forecasts the analyses significantly better than any single member in the ensemble, including the reference member with assimilation. In the forecast of the analyses, the RMSE of EFA is decreased by 28% compared to that of the reference simulation. The method is able to reproduce the average spatial patterns of the analyses. It also performs very well in comparisons against observations (with also a reduction of 28% of the RMSD, compared to the reference simulation). The improvements are still strong in the  $D+2$  forecasts, with 25% improvement (compared to the reference) in forecasting analyses and 24% in forecasting observations, which contrasts (in this application) with data assimilation alone.

[62] Ensemble forecasting of analyses provides a rigorous mathematical and algorithmic framework, but it opens a number of new issues at the same time. There may be further work on the aggregation algorithm. For instance, in regions where there are no observations, the reference forecasts and the analyses should be similar: The algorithm should quickly converge to the reference forecasts. The aggregation algorithm should also take into account the spatial distribution of the concentrations. Currently, the aggregation is performed independently for each element of the model's state. This strategy gives a strong theoretical guarantee since EFA and then competes against the best constant linear combination for every single state component, but it does not include any information about the spatial patterns. Another area for further research is how to extend the method to estimate the uncertainty associated with the forecasts.

## Appendix A: Data Assimilation and Sequential Aggregation

[63] Let  $\tilde{\mathbf{x}}_t^f$  be a model forecast at time  $t \in \{1, \dots, T\}$ . It is also referred to as the state vector of the model. The analysis state vector is denoted  $\tilde{\mathbf{x}}_t^a$ , and the true state is  $\tilde{\mathbf{x}}_t^t$ . At time  $t$ , the observation vector is denoted  $\mathbf{y}_t$ . A linear observation operator  $\mathbf{H}_t$  maps from the space of the model state to the observation space, so  $\mathbf{H}_t \tilde{\mathbf{x}}_t^f$  is the forecast of  $\mathbf{y}_t$ .

### A1. Data Assimilation: Generation of the Analyses

[64] In many data assimilation methods, the forecast error  $\mathbf{e}_t^f = \tilde{\mathbf{x}}_t^f - \tilde{\mathbf{x}}_t^t$  is modeled as a random variable, with zero mean and with variance  $\mathbf{B}_t$ . Similarly, the observational error  $\mathbf{e}_t^o = \mathbf{y}_t - \mathbf{H}_t \tilde{\mathbf{x}}_t^t$  is modeled as a random variable, with zero mean and with variance  $\mathbf{R}_t$ .

[65] The analysis  $\tilde{\mathbf{x}}_t^a$  may be taken as the best linear unbiased estimator. This estimator is defined so (1) it is a linear combination of  $\tilde{\mathbf{x}}_t^f$  and  $\mathbf{y}_t$ , (2) it is unbiased, and (3) its error  $\mathbf{e}_t^a = \tilde{\mathbf{x}}_t^a - \tilde{\mathbf{x}}_t^t$  has a variance with minimal trace; it minimizes  $\sum_i \text{var}(e_{i,t}^a)$ , where  $e_{i,t}^a$  is the  $i$ th component of  $\mathbf{e}_t^a$ . In that sense, the BLUE can be seen as the best estimate of the true state, based on the simulated state, the observations and their error statistics  $\mathbf{B}_t$  and  $\mathbf{R}_t$ .

[66] The BLUE has an explicit formulation:

$$\tilde{\mathbf{x}}_t^a = \tilde{\mathbf{x}}_t^f + \mathbf{B}_t \mathbf{H}_t^T (\mathbf{H}_t \mathbf{B}_t \mathbf{H}_t^T + \mathbf{R}_t)^{-1} (\mathbf{y}_t - \mathbf{H}_t \tilde{\mathbf{x}}_t^f). \quad (\text{A1})$$

[67] Other estimators can be considered in the EFA method. The key point is that the estimator should be the *best estimate* (in some sense) of the *true state*. The BLUE may be a good option, which can be computed by a data assimilation method like a Kalman filter.

### A2. Sequential Aggregation

[68] An ensemble of forecasts is available at time  $t$  with the sequence  $(\mathbf{x}_t^1, \dots, \mathbf{x}_t^M)$ , where  $M$  is the number of models. The forecast of a single model  $m$  is therefore denoted  $\mathbf{x}_t^m$ . Sequential aggregation consists of generating a linear combination of the ensemble forecasts. The weights are computed based on past observations and ensemble predictions. This procedure is repeated before each forecast time step. With proper weight computations, the aggregated forecasts are guaranteed to perform, in the long term, at least as well as the best constant (in time) linear combination of models. Other algorithms compete against the best constant convex combination (i.e., with positive weights summing to 1), or simply against the best model. Many aggregation algorithms have been derived in the machine learning community, see *Cesa-Bianchi and Lugosi* [2006] for a good introduction, and see *Mallet et al.* [2009] for their application in the context of atmospheric forecasting.

[69] Before every forecast step  $t$ , new weights  $\mathbf{v}_t$  are computed according to past observations  $\mathbf{y}_1, \dots, \mathbf{y}_{t-1}$ , the past predictions  $\mathbf{x}_1^m, \dots, \mathbf{x}_{t-1}^m$  and, in given methods, the predictions to be aggregated  $\mathbf{x}_t^m$  ( $m = 1, \dots, M$ ). For example, an aggregation algorithm can be given by the discounted ridge regression [*Mallet et al.*, 2007a, section 13]:

$$\mathbf{v}_t = (v_{1,t}, \dots, v_{M,t})^T = \underset{\mathbf{u} \in \mathbb{R}^M}{\text{argmin}} \left[ \lambda \|\mathbf{u}\|_2^2 + \sum_{t'=1}^{t-1} (1 + \psi_{t-t'}) \left\| \mathbf{y}_{t'} - \sum_{m=1}^M u_m \mathbf{H}_{t'} \mathbf{x}_{t'}^m \right\|_2^2 \right], \quad (\text{A2})$$

where  $\|\cdot\|_2$  is the Euclidean norm (2-norm),  $\lambda > 0$ , and  $\psi_t$  is a decreasing sequence that discounts the quadratic discrepancies of the faraway past; e.g.,  $\psi_t = \gamma/t^\beta$ , with  $\gamma, \beta > 0$ .

[70] It can be shown [*Cesa-Bianchi and Lugosi*, 2006] that, in the long run, the mean quadratic performance of the aggregated forecasts,  $\sum_{m=1}^M v_{m,t} \mathbf{H}_t \mathbf{x}_t^m$ , is at least as good as the mean quadratic performance of the best constant (in time) linear combination. Formally,

$$\sum_{t=1}^T \frac{1}{TO_t} \left\| \mathbf{y}_t - \sum_{m=1}^M v_{m,t} \mathbf{H}_t \mathbf{x}_t^m \right\|_2^2 - \min_{\mathbf{u} \in \mathbb{R}^M} \sum_{t=1}^T \frac{1}{TO_t} \left\| \mathbf{y}_t - \sum_{m=1}^M u_m \mathbf{H}_t \mathbf{x}_t^m \right\|_2^2 \lesssim \mathcal{O} \left( \frac{\ln T}{T} \right), \quad (\text{A3})$$

where  $O_t$  is the size of  $\mathbf{y}_t$ . On the left-hand side, the first term is the mean-square error of the aggregated forecast, and the second term is the mean-square error of the best constant linear combination.

[71] In other words, the aggregated forecast tends to have an RMSE at least as good as that of the best constant linear combination. This theoretical guarantee holds whatever the sequence of observations and predictions may be. No assumption is made, except that the quadratic error per time step is bounded (which is obviously satisfied in geophys-

ical applications); in particular, no stochastic assumption is made.

[72] Note that both the learned weights and the weights of the best constant combination are unconstrained. In particular, the weights can be negative. Also note that there are many other algorithms, several of which produce weights for convex combinations. See *Mallet et al.* [2007a] for an overview of the other algorithms.

## Appendix B: Convex EFA

[73] In the body of the paper, EFA is illustrated with the discounted ridge regression. Other methods can be applied in the EFA framework. The reader may refer to *Mallet et al.* [2009] for an introduction to machine learning methods with similar guarantees as that of section 2.5. These methods are further detailed in the technical report *Mallet et al.* [2007a].

[74] The weights produced by the discounted ridge regression are unconstrained. They can take any value, even negative. There are learning methods that produce weights for convex combinations in which the weights are all positive and they sum up to 1. One example is the exponentiated gradient [e.g., *Cesa-Bianchi*, 1999], which in the EFA context, reads:

$$w_{i,t}^m = \frac{\exp\left[-2\eta \sum_{t'=1}^{t-1} x_{i,t'}^m \left(\sum_{m'=1}^M w_{i,t'}^{m'} x_{i,t'}^{m'} - \tilde{x}_{i,t'}^a\right)\right]}{\sum_{m'=1}^M \exp\left[-2\eta \sum_{t'=1}^{t-1} x_{i,t'}^{m'} \left(\sum_{m'=1}^M w_{i,t'}^{m'} x_{i,t'}^{m'} - \tilde{x}_{i,t'}^a\right)\right]}, \quad (\text{B1})$$

starting with  $w_{i,t}^m = 1/M$ , for all  $m$ . The aggregated forecasts  $\hat{\mathbf{x}}_t$  will perform in the long run at least as well as the best convex combination:

$$\sum_{t=1}^T \frac{1}{T} \|\tilde{\mathbf{x}}_t^a - \hat{\mathbf{x}}_t\|_2^2 - \min_{\mathbf{u}_1, \dots, \mathbf{u}_N \in \mathcal{X}_M} \sum_{t=1}^T \frac{1}{T} \|\tilde{\mathbf{x}}_t^a - C(\mathbf{u}_1, \dots, \mathbf{u}_N, \mathbf{x}_t^1, \dots, \mathbf{x}_t^M)\|_2^2 \leq \mathcal{O}\left(\frac{1}{\sqrt{T}}\right), \quad (\text{B2})$$

for an optimal choice of  $\eta$ .  $\mathcal{X}_M$  is the set of all vectors whose components are positive and sum up to 1. The theoretical guarantee is weaker than that of the (discounted) ridge regression since only convex combinations are formed here. The method could anyway show interesting performance in special cases. Notice that there is also a discounted version of the exponentiated gradient, see *Mallet et al.* [2009] for further details.

## References

Balgovind, R., A. Dalcher, M. Ghil, and E. Kalnay (1983), A stochastic-dynamic model for the spatial structure of forecast error statistics, *Mon. Weather Rev.*, *111*(4), 701–722.  
 Bouttier, F., and P. Courtier (1999), Data assimilation concepts and methods, *Meteorological training course lecture series*, ECMWF.  
 Buehner, M., P. L. Houtekamer, C. Charette, H. L. Mitchell, and B. He (2010), Intercomparison of variational data assimilation and the ensemble

kalman filter for global deterministic NWP. Part II: One-month experiments with real observations, *Mon. Weather Rev.*, *138*(5), 1567–1586.  
 Cesa-Bianchi, N. (1999), Analysis of two gradient-based algorithms for on-line regression, *J. Comput. Syst. Sci.*, *59*, 392–411.  
 Cesa-Bianchi, N., and G. Lugosi (2006), *Prediction, Learning, and Games*, 394 pp., Cambridge Univ. Press, New York.  
 Cohn, S. E. (1997), An introduction to estimation theory, *J. Meteorol. Soc. Jpn.*, *75*(1B), 257–288.  
 Daley, R. (1993), *Atmospheric Data Analysis*, 457 pp., Cambridge Univ. Press, New York.  
 Elbern, H., and H. Schmidt (2001), Ozone episode analysis by four-dimensional variational chemistry data assimilation, *J. Geophys. Res.*, *106*(D4), 3569–3590.  
 Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, *99*(C5), 10,143–10,162.  
 Garaud, D., and V. Mallet (2010), Automatic generation of large ensembles for air quality forecasting using the Polyphemus system, *Geosci. Model Dev.*, *3*(1), 69–85.  
 Gerchinovitz, S., V. Mallet, and G. Stoltz (2008), *A Further Look at Sequential Aggregation Rules for Ozone Ensemble Forecasting*, Tech. Rep., INRIA and Ecole normale supérieure, Paris.  
 Horowitz, L. W., et al. (2003), A global simulation of tropospheric ozone and related tracers: Description and evaluation of MOZART, version 2, *J. Geophys. Res.*, *108*(D24), 4784, doi:10.1029/2002JD002853.  
 Krishnamurti, T. N., C. M. Kishtawal, Z. Zhang, D. B. T. LaRow, and E. Williford (2000), Multimodel ensemble forecasts for weather and seasonal climate, *J. Clim.*, *13*, 4196–4216.  
 Le Dimet, F.-X., and O. Talagrand (1986), Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects, *Tellus, Ser. A* *38*, 97–110.  
 Louis, J.-F. (1979), A parametric model of vertical eddy fluxes in the atmosphere, *Boundary Layer Meteorol.*, *17*, 187–202.  
 Mallet, V., and B. Sportisse (2006), Ensemble-based air quality forecasts: A multimodel approach applied to ozone, *J. Geophys. Res.*, *111*, D18302, doi:10.1029/2005JD006675.  
 Mallet, V., B. Mauricette, and G. Stoltz (2007a), *Description of Sequential Aggregation Methods and Their Performances for Ozone Ensemble Forecasting*, Tech. Rep. DMA-07-08, École normale supérieure de Paris.  
 Mallet, V., et al. (2007b), Technical Note: The air quality modeling system Polyphemus, *Atmos. Chem. Phys.*, *7*(20), 5479–5487.  
 Mallet, V., G. Stoltz, and B. Mauricette (2009), Ozone ensemble forecast with machine learning algorithms, *J. Geophys. Res.*, *114*, D05307, doi:10.1029/2008JD009978.  
 Ménard, R., S. E. Cohn, L.-P. Chang, and P. M. Lyster (2000), Assimilation of stratospheric chemical tracer observations using a Kalman filter: Part I. Formulation, *Mon. Weather Rev.*, *128*, 2654–2671.  
 Middleton, P., W. R. Stockwell, and W. P. L. Carter (1990), Aggregation and analysis of volatile organic compound emissions for regional modeling, *Atmos. Environ.*, *24A*(5), 1107–1133.  
 Pagowski, M., et al. (2006), Application of dynamic linear regression to improve the skill of ensemble-based deterministic ozone forecasts, *Atmos. Environ.*, *40*, 3240–3250.  
 Simpson, D., et al. (1999), Inventorying emissions from nature in Europe, *J. Geophys. Res.*, *104*(D7), 8113–8152.  
 Stockwell, W. R., F. Kirchner, M. Kuhn, and S. Seefeld (1997), A new mechanism for regional atmospheric chemistry modeling, *J. Geophys. Res.*, *102*(D22), 25,847–25,879.  
 Todling, R. (1999), Estimation theory and foundations of atmospheric data assimilation, in *Note Series on Global Modeling and Data Assimilation 1999-01*, Data Assimilation Office, Goddard Space Flight Center, Greenbelt, Md.  
 Troen, I., and L. Mahrt (1986), A simple model of the atmospheric boundary layer: Sensitivity to surface evaporation, *Boundary Layer Meteorol.*, *37*, 129–148.  
 Wu, L., V. Mallet, M. Bocquet, and B. Sportisse (2008), A comparison study of data assimilation algorithms for ozone forecasts, *J. Geophys. Res.*, *113*, D20310, doi:10.1029/2008JD009991.

V. Mallet, INRIA, Paris-Rocquencourt Research Center, BP 105, F-78153 Le Chesnay CEDEX, France. (vivien.mallet@inria.fr)