Minimax filtering for sequential aggregation: Application to ensemble forecast of ozone analyses

Vivien Mallet,1,2 Alexander Nakonechny,3 and Sergiy Zhuk4

Received 13 May 2013; revised 18 July 2013; accepted 8 August 2013; published 2 October 2013.

[1] This paper presents a new algorithm for sequential aggregation of an ensemble of forecasts. At any forecasting step, the aggregation consists of (1) computing new weights for the ensemble members represented by different numerical models and (2) forecasting with a weighted linear combination of the ensemble members. We assume that the time evolution of the weights is described by a linear equation with uncertain parameters and apply a minimax filter (and also Kalman filter, for comparison) in order to estimate the vector of weights given “observations”. The “observation” equation for the filter compares the aggregated forecast with the analysis determined in a data assimilation cycle together with its variance. The minimax approach allows one to work with flexible uncertainty description: deterministic bounding sets for uncertain parameters in weight’s equation, and error covariance matrices for the “observational” errors. Our key contribution is an uncertainty estimate of the aggregated forecast, for which we introduce an evaluation test. The performance of the method is assessed for the forecast of ground-level ozone daily peaks over Europe, for the year 2001. Compared to forecasts generated by classical data assimilation, the root mean square error is decreased by 16% for prediction of the analyses and by 20% for prediction of the observations.


1. Introduction

[2] In the task of forecasting the atmospheric state, one can nowadays rely on different sources of information: forecasts provided by numerical models, field observations, and error statistics for both observations and model simulations. The availability of several numerical models reflects the fact that different physical parameterizations were derived to describe the same atmospheric phenomena. Also, the mathematical model could be turned into a numerical model by means of various numerical techniques. In addition, the input data can be provided by different sources and are usually uncertain to some extent. This suggests to consider an ensemble of forecasts: an ensemble brings together various sources of information and allows, therefore, to derive a new forecast which performs better than any individual ensemble member. The improved forecast is usually obtained by taking a linear or convex combination of ensemble members with some weights and is therefore called an aggregated forecast. Now, the aggregation problem may be formulated as follows: construct weights such that the corresponding aggregated forecast is close in terms of some performance measure to the given reference which is usually represented by observations. The weights of the combination are updated as soon as new observations become available. So the procedure is referred to as sequential aggregation.

[3] One of the main problems of forecasting algorithms based on ensemble aggregation is uncertainty estimation. Given various numerical models and related observations, with different uncertainty descriptions, one needs to combine these descriptions all together and transform them into an estimate of the uncertainty associated with the aggregated forecast. We stress that the weights of the aggregated forecast change when new observations become available; hence they evolve over time. Thus, the uncertainty transformation should be accomplished by the aggregation algorithm along with the evolution of the weights. In other words, the dynamics of the weights drives the aggregated forecast and its uncertainty estimate. Since this dynamics is uncertain, one needs to assume an appropriate uncertainty description for the weights. We stress that technically the error statistics for observations may be incomparable with uncertainty description for the weight’s evolution or even for individual ensemble members: for instance, the measurement error may be stochastic, and the error of the numerical models may be deterministic. The minimax filter can handle such case and provide an uncertainty estimation for the weights (hence for the aggregated forecast as well) in the form of a bounding set.

[4] Another important technical issue to consider is the sparsity (in space) of the observation’s network. It may
result in weights that are optimal only locally, i.e., at observed locations. This problem is solved using "ensemble forecast of analyses" (EFA) [Mallet, 2010]. In EFA, one first uses a data assimilation algorithm to generate an analysis. At a given date and under given assumptions, the analysis is the a posteriori estimate of the atmospheric state that optimally combines (in the least squares sense) observations, simulations, and error statistics. The main idea of EFA is to forecast forthcoming analyses instead of observations. The analyses are the preferred target because they include all a posteriori knowledge on the atmospheric state, they take into account observational errors, and they provide comprehensive information (i.e., concentrations for all pollutants in all model grid cells). The weights of the aggregated forecast are thus adapted to forecast the analyses instead of the observations, and the weights can depend on space (one weight per model and per state component).

[5] In air quality applications, the aggregation of ensemble simulations has been carried out by different strategies. Delle Monache and Stull [2003] relied on the ensemble mean, where all simulations were given the same weight. Analysis of ensemble mean variance was carried out in Potemski and Galmarini [2009] and Solazzo et al. [2012]. Multimodel forecast based on ensemble median was reported in Riccio et al. [2007]. Bias correction techniques were tested in Monteiro et al. [2013]. Nonstationary weighting procedure for ozone forecasting based on a dynamic linear regression was presented in Pagowski et al. [2006]. Our approach also assumes an equation for the dynamics of the aggregation weights and an observation equation. Note that the latter, in fact, allows one to compare observations (or analyses) and aggregated forecasts corresponding to given weights. However, our uncertainty description differs. In the dynamic linear regression, the errors—in the weights equation and observation equation—are Gaussian and the variance of the observational error is unknown. In contrast, the minimax approach assumes that uncertainty description is given in terms of bounding sets: the errors in the weight equation are elements of a prescribed set. Similarly, the variances of the observational errors may not be prescribed, but they must belong to the given set as well. As a result, the algorithm is robust with respect to uncertainty in observation error covariance matrix unlike Kalman filter: it is well known from the control literature (see, for instance, Shen and Deng [1997] and Başar and Bernhard [1995]), that $H_2$ or Kalman state estimators may be sensitive to uncertainty in the statistical error description. In other words, small perturbations in error covariance matrices may lead to significant deviations in estimates and/or error estimates. Such perturbations are often introduced in practice because the covariance parameters are usually estimated from data. Mallet and Sportisse [2006] used plain least squares methods on a large ensemble. Mallet et al. [2009] applied several machine learning algorithms on the same ensemble, especially a version of the ridge regression with discount in time. Machine learning algorithms are robust, adapted to operational forecasting and guarantee good performance in the long term. However, contrary to dynamic linear regression or minimax approach, they do not evaluate the uncertainty associated with their forecasts. Also, we prove that the weights obtained by means of the discounted ridge regression can be generated by our algorithm for a suitable choice of parameters.

[6] The main contribution of this paper is a minimax aggregation algorithm and uncertainty estimate associated with forecasts together with a simple method to check its reliability. Our aggregation method is based upon a minimax state estimation approach [see Bertsekas and Rhodes, 1971; Nakonechny, 1978; Milanese and Tempo, 1985; Milanese and Vicino, 1991; Kurzhanski and Valyi, 1997]. The estimation problem is defined as follows: given a state equation (the model, for the weights), an observation equation, and error descriptions, one needs to estimate the state of the model assuming that model errors belong to a given bounding set, and observational errors are realizations of random variables with given mean and unknown variance—which is in a given bounded set as well. To solve the estimation problem, we construct a worst-case error which selects the worst possible realization of uncertain parameters and leads to the maximal possible estimation error. The minimax estimate is chosen to have the least possible worst-case estimation error. This, in turn, allows us to construct a set of all possible estimates that are compatible with the model, observed data, and uncertainty description. This set represents, in fact, an uncertainty estimate.

[7] In this work, we construct the minimax estimate for aggregation weights in the form of a linear recursive filter. The so-called state equation (or model) defines the weights dynamics. The so-called observation equation actually involves the analysis, which is supposed to be equal to a linear combination of the ensemble of simulations, plus some unknown error. The filter estimates the aggregation weights which are in turn used to compute an aggregated forecast. The filter also provides an uncertainty estimation for the weights, from which the forecast uncertainty can be derived. In addition, we propose a method to check the reliability of the uncertainty estimation. We stress that the bounding set for the initial weight may be unbounded reflecting the fact that the initial weight may be chosen arbitrarily. This, in turn, proves that the algorithm’s convergence does not depend upon a choice of the initial weight. In addition, the algorithm is robust to the variation in observation error covariance matrices unlike Kalman filters as it was mentioned above.

[8] The paper is organized as follows. Section 2 introduces a version of the minimax filter dedicated to sequential aggregation. It explains how to compute and assess the uncertainty estimation and discusses the links to Kalman filter and discounted ridge regression. Section 3 introduces the application to air quality, with further explanations on the EFA strategy. It briefly describes the ensemble simulations and the generation of the analyses. It also gives the parameters related to the minimax filter. Section 4 evaluates the forecast performance and uncertainty estimations. We also address the sensitivity of the results.

2. Sequential Aggregation With Nonstationary Weights

2.1. Notation

[9] Let $\mathcal{M}_t$, be a model at time instant $t \in \{0, \ldots, T - 1\}$ that carries out the time integration from time $t$ to time $t + 1$. 

11,295
The model state vector \( \mathbf{x}_t \) (in our case, the vector of all pollutant's concentrations across the grid) is updated with \( \mathbf{x}_{t+1} = M_t(\mathbf{x}_t) \). Let \( \mathbf{x}_{true} \) denote the “true” state vector of the process. At time \( t \), the observation vector is denoted \( \mathbf{y}_t \). An observation operator \( \mathbf{H} \) maps the model state space into the observation space so that \( \mathbf{H} \mathbf{x}_{t} \) can be compared with \( \mathbf{y}_t \).

[16] Let \( (\mathbf{x}_1^t, \ldots, \mathbf{x}_M^t) \) denote an ensemble of forecasts generated by given models \( M_1^t, \ldots, M_M^t, M \in \mathbb{N} \), at time \( t \): \( \mathbf{x}_i^t = M_i^t(\mathbf{x}_{true}^t) \). The \( i \)-th components of the aforementioned vectors are denoted with an additional subscript: \( \mathbf{x}_{i,t}^t, \mathbf{y}_{i,t}^t \) and \( \mathbf{x}_{i,m}^t \). Let \( \mathbf{w} \) denote the vector of weights \( (w_1^t, \ldots, w_M^t)^T \). \( \mathbf{y}_t \) stands for the 2 norm of the vector \( \mathbf{v} \).

### 2.2. Minimax Weights

[11] Our main objective is to forecast the true process state \( \mathbf{x}_{true}^t \) by means of a linear combination \( \sum_{m=1}^M w_m^t \mathbf{x}_m^t \) of ensemble members \( (\mathbf{x}_1^t, \ldots, \mathbf{x}_M^t) \) with coefficients \( w_m^t = (w_1^t, \ldots, w_M^t)^T \). Intuitively, each ensemble member \( \mathbf{x}_m^t \) represents a direction in the state space and, thus, we seek a vector \( \mathbf{x}_t \) in the subspace spanned by \( (\mathbf{x}_1^t, \ldots, \mathbf{x}_M^t) \) giving the best (in the sense of the minimax criterion) approximation of \( \mathbf{x}_{true}^t \). Noting that \( \mathbf{x}_t \) is solely defined by the corresponding weights \( \mathbf{w}_t \), we will be looking for a vector of weights which is optimal in the minimax sense. We will refer to the optimal weights as minimax weights.

#### Model for Weights

[12] We introduce the true weights \( w_{true}^t \) that yield the best approximation to \( \mathbf{x}_{true}^t \) in the 2 norm sense. We assume that the true weights satisfy the following weights equation:

\[
\mathbf{w}_{true}^t = \mathbf{A} \mathbf{w}_{true}^t + \mathbf{e}_t, \quad \mathbf{e}_0 = \mathbf{e},
\]

where \( \mathbf{A} \) is an \( M \times M \) matrix, \( \mathbf{e}_t \) is a model error, and \( \mathbf{e} \) is an uncertain initial condition.

[13] If one assumes that \( \mathbf{e}_t \) and \( \mathbf{e}_s \) (\( s \neq t \)) are independent normal random variables, then the weights \( \mathbf{w}_{true}^t \) would represent a realization of a nonstationary Ornstein-Uhlenbeck process with discrete time. We stress that this is very well aligned with the nonstationary nature of the underlying physical process. However, the normal assumption is not adapted to physical processes that are always modeled with bounded error, while normal random variables can take values outside any bounded set with nonzero probability. Thus, we will assume that \( \mathbf{e}_t \) is bounded within a given ellipsoid (see below, inequality (4)). In what follows, we will not rely on any hypothesis about the distribution of the model error. This shows that the minimax weights are robust and the weights’ model can be adapted to any real-life application (based on any ensemble).

#### Observation for Weights

[14] We introduce a relation between the observations and the weights. The observation equation for the weights reads as follows:

\[
y_t = \mathbf{E}_t \mathbf{w}_{true}^t + \eta_t,
\]

where \( \mathbf{y}_t \) is the observation vector, \( \eta_t \) is the observational error (random vector), and the \( m \)-th column of \( \mathbf{E}_t \) is the forecast \( \mathbf{H}_t \mathbf{x}_m^t \) of \( \mathbf{y}_t \) by the \( m \)-th model:

\[
\mathbf{E}_t = [\mathbf{H}_t \mathbf{x}_1^t, \ldots, \mathbf{H}_t \mathbf{x}_m^t, \ldots, \mathbf{H}_t \mathbf{x}_M^t].
\]

[15] **Uncertainty Description:** In the minimax setting, the errors are assumed to satisfy the following constraints (also called uncertainty description):

\[
(e - \mathbf{w}_0)^T Q^{-1} (e - \mathbf{w}_0) + \sum_{r=0}^{T-1} q_{r}^T \mathbf{q}_{r} \leq 1, \quad \sum_{r=0}^{T} \mathbb{E} [\eta_r^T R_r \eta_r] \leq 1,
\]

with \( \mathbf{w}_0 \) being our initial guess of the weights, \( \mathbb{E} [\cdot] \) standing for the expectation, and \( \mathbf{Q}, \mathbf{Q}_r, \mathbf{R}_r \) being symmetric positive definite matrices. Each constraint defines an ellipsoid where \( \mathbf{e}, \mathbf{e}_r, \) or \( \eta_r \) lie. We stress that the second constraint is actually on the covariance matrix of \( \eta_r \). The best estimation of \( \mathbf{R}_r \) is proportional to the variance of \( \eta_r \), but any matrix \( \mathbf{R}_r \) satisfying the constraint can be taken here. Once \( \mathbf{R}_r \) is fixed, it determines the acceptable range for the variance of \( \eta_r \).

#### Minimax Approach

[16] The weights \( \hat{w}_t \) are said to be minimax if

\[
\sup_{\mathbf{w}_t} \mathbb{E} [\ell^T (\mathbf{w}_{true}^t - \hat{w}_t)] \leq \sup_{\mathbf{w}_t} \mathbb{E} [\ell^T (\mathbf{w}_{true}^t - \mathbf{w}_0)]
\]

for any weights \( \mathbf{w}_t \) that follow the model (1) and for any vector \( \ell \in \mathbb{R}^M \). In other words, the minimax weights \( \hat{w}_t \) have the minimal worst-case error (expressed by sup overall uncertain parameters satisfying (4)). Also they do not depend on a particular realization of the uncertain model errors or the observation error covariance matrices and are, therefore, robust to these parameters. In addition, all possible realizations of the weights \( \mathbf{w}_{true}^t \) form a set which is centered at the minimax weights \( \hat{w}_t \). This set allows to construct the uncertainty estimate for the minimax weights as well as for the aggregated forecast.

#### Forecasting

[17] We assume that the ensemble of forecasts \( (\mathbf{x}_{1,t}, \ldots, \mathbf{x}_{M,t}) \) is available at any time step \( t \). The aggregated forecast \( \mathbf{x}_{true,t+1} \) is produced with the forecast weight vector \( \mathbf{A}\hat{w}_t \) as follows:

\[
\mathbf{x}_{true,t+1} = \sum_{m=1}^M (\mathbf{A}_m \hat{w}_t) \mathbf{x}_{m,t+1},
\]

if \( \mathbf{A}_m \) is the \( m \)-th row of \( \mathbf{A} \). Note that \( \mathbf{A}\hat{w}_t \) corresponds to the expected weights at \( t + 1 \) according to equation (1).

#### Minimax Weights

[18] The minimax weights and their uncertainty description \( \mathbf{G}_t^{-1} \) are initialized with

\[
\mathbf{G}_0 = Q^{-1} + E_0 R_0^{-1} E_0^T,
\]

\[
\hat{w}_0 = \mathbf{G}_0^{-1} (Q^{-1} w_0 + E_0 R_0^{-1} y_0),
\]

and they are updated as follows:

\[
\mathbf{G}_t = (\mathbf{B} \mathbf{Q}_t \mathbf{B}^T + \mathbf{A} \mathbf{G}_t^{-1} \mathbf{A}^T)^{-1} + E_t R_t^{-1} E_t^T,
\]

\[
\hat{w}_t = \mathbf{A}\hat{w}_{t-1} + \mathbf{G}_t^{-1} E_t R_t^{-1} (y_t - E_t \hat{w}_{t-1}).
\]

The derivation of these formulae is sketched in Appendix A.

#### Uncertainty Estimation

\( \mathbf{G}_t^{-1} \) is the inverse of the minimax gain matrix, and it provides the uncertainty description for \( \hat{w}_t \). The forecast weights \( \mathbf{A}\hat{w}_t \) are associated with uncertainty description

\[
\mathbf{F}_{t+1} = \mathbf{B} \mathbf{Q}_t \mathbf{B}^T + \mathbf{A} \mathbf{G}_t^{-1} \mathbf{A}^T,
\]

which is the same as \( \mathbf{G}_t^{-1} \) (see equation (9)) with \( \mathbf{E}_{t+1} = 0 \) in order to reflect the absence of observations at the forecast time instant \( t + 1 \). Note that \( \mathbf{F}_0^{-1} = \mathbf{Q} \).
[30] Considering (6), we find that the uncertainty on the i-th component of the aggregated forecast \( \hat{x}_{t+1} \) is
\[
\gamma^2_{t+1} := (x_{t+1}^1, \ldots, x_{t+1}^M, \mathbf{E}_{t+1}^1, \ldots, \mathbf{E}_{t+1}^M) \mathbf{T},
\]
so that if the error descriptions (4) are correct, then on average, the truth \( x_{t+1}^{true} \) is guaranteed to satisfy
\[
\mathbb{E}[x_{t+1}^{true}] = \left[ \hat{x}_{t+1} - y_{t+1}, \hat{x}_{t+1} + y_{t+1} \right].
\]

2.2.1. A Posteriori Evaluation of the Uncertainty Estimation

[21] We propose a method to check the quality of the uncertainty \( y_{t+1} \). It plays the same role as the \( \chi^2 \) diagnosis in Kalman filtering [e.g., Ménard et al., 2000]. In our case, the uncertainty description is given by bounded sets, see inequalities (4) and (13), in which we will assume uniform distribution. We therefore assume that every single observation \( y \) is uniformly distributed around the unknown truth \( x^{true} \); \( y \sim U(x^{true} - \varepsilon, x^{true} + \varepsilon) \), where \( \varepsilon \) is related to the diagonal elements of \( \mathbf{R} \). In this paper, we assume \( \varepsilon \) is equal to the square root of the corresponding diagonal element of \( \mathbf{R} \). Similarly, as a consequence of (13), we assume that the truth is uniformly distributed around the minimax estimate \( \hat{x}^M \); \( x^{true} \sim U(\hat{x}^M - \gamma, \hat{x}^M + \gamma) \).

[22] Then we compute the probability that the observation \( y \) falls outside the interval \( [\hat{x}^M - \gamma, \hat{x}^M + \gamma] \) with probability 1. With observational errors and given the truth \( x^{true} \), the probability that \( y \in [x^{true} - \varepsilon, x^{true} + \varepsilon] \) falls outside \( [\hat{x}^M - \gamma, \hat{x}^M + \gamma] \) is equal to \( (y \notin [\hat{x}^M - \gamma, \hat{x}^M + \gamma]) = 1 - \frac{1}{\pi} \int_{[\hat{x}^M - \gamma, \hat{x}^M + \gamma]} |U(x^{true} - \varepsilon, x^{true} + \varepsilon) \cap [\hat{x}^M - \gamma, \hat{x}^M + \gamma]| \), where \( [a, b] = b - a \) (interval width). Considering that \( x^{true} \sim U(\hat{x}^M - \gamma, \hat{x}^M + \gamma) \), we find that
\[
\text{E}[P(y \notin [\hat{x}^M - \gamma, \hat{x}^M + \gamma]; x^{true})] = \begin{cases} \frac{1}{2} & \text{if } \varepsilon \leq 2 \gamma, \\ 1 - \frac{1}{\pi} & \text{if } \varepsilon > 2 \gamma. \end{cases}
\]

This expectation can be compared with the actual frequency in which the observations fall outside the predicted interval \( [\hat{x}^M - \gamma, \hat{x}^M + \gamma] \). They should be similar if the uncertainty description (4) is reliable.

2.2.2. Minimax Versus Kalman Filter

[25] It is possible to make links with Kalman filter. To do so, we would apply the minimax filter with this uncertainty description instead of inequality (4):
\[
(e - w_0) Q^{-1} (e - w_0) + \sum_{i=0}^{F-1} e_i' Q_i^{-1} e_i + \sum_{i=0}^{F-1} \eta_i' R_i^{-1} \eta_i \leq 1,
\]
assuming \( \eta_i \) is deterministic in this case.

[24] In the Kalman setting, we would interpret \( Q, Q_i, \) and \( R_i \) as covariance matrices for the initial weights error, for the weights model error, and for the observational error, respectively. Under the weights’ equation (1) and the observation equation (2), the Kalman filter and the minimin filter compute the same estimate
\[
\tilde{w}_{t+1} = A \tilde{w}_t + \mathbf{G}_t, \mathbf{E}_t^1, \mathbf{R}_t, (y_{t+1} - E_{t+1} A \tilde{w}_t), \quad \tilde{w}_0 = \tilde{w}_0.
\]

The minimax gain matrix \( \mathbf{G}_t \) reads
\[
\mathbf{G}_t = (\mathbf{BQ}_t \mathbf{B}^T + A \mathbf{G}_t A^T)^{-1} + \mathbf{E}_t^1 \mathbf{R}_t \mathbf{E}_t^1, \quad \mathbf{G}_0 = Q^{-1} + \mathbf{E}_0 \mathbf{R}_0 \mathbf{E}_0.
\]

This update is the same as in the Kalman filter, but in the Kalman filter, \( \mathbf{G}^{-1}_t \) is the trace of the error on \( \tilde{w}_{t+1} \).

[25] In this paper, we will report some of the results obtained with the Kalman filter.

2.2.3. Minimax versus Ridge Regression

[26] In the previous work [Mallet et al., 2009, 2010], the aggregation weights were computed using the discounted ridge regression:
\[
\tilde{w}_t = \arg \min_{w \in \mathbb{R}^M} \left[ \lambda \|u\|^2 + \sum_{i=0}^{F-1} (1 + \psi_i) |y_i - \sum_{m=1}^M u_m H_i x_m^i|^2 \right],
\]
where \( \lambda > 0 \) and \( \psi_i > 0 \) is a decreasing sequence of some form [Mallet et al., 2007a].

[27] In case uncertainties are described as inequality (15), both the minimax filter and the Kalman filter lead to the same estimate as the ridge regression at \( t \) if \( A = I, B = 0, w_0 = 0, Q = I, R_t = (1 + \psi_t) I \).

3. Experiment Setup for Ozone Ensemble Forecasting

3.1. Ensemble Simulations and Generation of Analyses

[28] This section is essentially a summary of section 3.1 from Mallet [2010] since the experiment setup is the same in this work.

[29] We aim at forecasting ground-level ozone at 15:00 UTC on the next day across western Europe. The ensemble simulations were generated within the Polyphemus platform [Mallet et al., 2007c] for the full year 2001. They were generated and analyzed by Garaud and Mallet [2010]. Each simulation was computed by a different 3-D Eulerian chemistry-transport models. By “model,” we refer to a unique description of the phenomena in terms of physical formulation, numerical discretization, and input data. All models share a horizontal resolution of 0.5°. The meteorological data are from the European Centre for Medium-Range Weather Forecasts. The differences between the models lie in the chemical mechanism (which can be Regional Acid Deposition Model 2 (RADM 2), [Stockwell et al., 1990] or Regional Atmospheric Chemistry Model (RACM) [Stockwell et al., 1997]), the computation of the photolysis rates, the parameterization for vertical diffusion (from Louis [1979] or Troen and Mahrt [1986]), the deposition velocities (Wesely [1989] or Zhang et al. [2003]), the evaluation of the cloud attenuation, the number of vertical layers, and so forth. Twelve input fields, such as boundary conditions, emissions or winds, were perturbed with homogeneous and constant perturbations following either normal or log-normal distributions. The ensemble is composed of 20 members in total. Most of these members were randomly generated using the alternatives and perturbations previously mentioned.

[30] One model (the first member) from the ensemble is used to generate the analyses. Of course, the analyses themselves are not inside the ensemble, but the first member relies on data assimilation until 19:00 UTC. Then a forecast for 15:00 UTC the next day is computed by the model starting from the analysis at 19:00 UTC. This procedure is repeated for every day, so as to mimic an operational cycle. The other members do not benefit from data assimilation.
The analyses are generated for ozone in the three first model layers with the optimal interpolation method. This method computes the so-called best linear unbiased estimator (BLUE) and updates the model state with BLUE whenever observations become available. We chose this method after the work by Wu et al. [2008] in which optimal interpolation gave good results compared to other data assimilation methods.

If we assume that the error on the model state $x^f$ before assimilation (the so-called background) has variance $P_f$ and the error on observations $y_t$ has variance $U_t$, then BLUE reads

\[ \overline{x}_t^a = x^f + P_f^T \left( H^T P_f^T H_f + U_t \right)^{-1} (y_t - H_f x^f). \]

The observational error variance is diagonal: $U_t = \sigma_i I_t$. Let $l_h$ and $l_v$ be respectively the horizontal and vertical distances between two grid cells (in the first three layers where the analysis is computed). The background state error covariance between these points is given in the form

\[ \text{cov} (x_{i+h}, x_{i+v}) = b \left( 1 + \frac{l_h}{L_h} \right) e^{-\frac{l_h}{L_h}} \left( 1 + \frac{l_v}{L_v} \right) e^{-\frac{l_v}{L_v}}, \]

where the decorrelation lengths are set to $L_h = 1^\circ$ and $L_v = 150$ m. After a $\chi^2$ diagnosis [e.g., Ménard et al., 2000] applied in Mallet [2010], the variances are set to $b = 190 \mu g m^{-6}$ and $r = 51 \mu g m^{-6}$. In this case, the standard deviation of the observational error is about $7.1 \mu g m^{-3}$ which is supposed to take into account measurement errors and representativeness errors.

The observations from the European Monitoring and Evaluation Programme monitoring network are assimilated. Every hour, there are available observations from about 90 active stations distributed across Europe—see Figure 4 for the locations of the stations.

The analysis error variance is $S_i = (P_i^{-1} + H_i^T U_i^{-1} H_i)^{-1}$. Inside the optimal interpolation algorithm, we compute the diagonal of $S_i$, which is necessary in the application of minimax-based EFA. In this paper, $P_i^T$ does not depend on time, but $U_t$ depends on time because of the availability of the monitoring stations.

### 3.2. Ensemble Forecast of Analyses

In this paper, we apply the ensemble forecast of analyses [Mallet, 2010] so that the aggregation is driven by analyses instead of observations. The aggregation weights are computed independently for each component of the state vector. For each state component $i$, we have an independent weight equation $w_{ij,t} = Aw_{ij,t} + B \eta_{ij,t}$, and the aggregation follows as $\overline{x}_{ij,t} = \sum_{m=1}^{M} w_{ij,t} x_{ij,t}^m$. The observations $y_t$ at time $t$ are replaced with the (scalar) analysis $\overline{x}_{ij,t}$ which is available at the same time as the observations (with a small delay due to the BLUE computations (19)). The observation operator from equation (2) reads $E_i = [x^f_{ij,t}, \ldots, x_{ij,t}^M]$. The “observation” equation (2) is therefore replaced with

\[ \overline{x}_{ij,t}^a = \sum_{m=1}^{M} w_{ij,t} x_{ij,t}^m + \eta_{ij,t}, \]
the Kalman filter for comparison. Applying the filters independently in every grid cell allows better performance, because the weights are adapted to the local situation. In theory, it is possible to use the same weights for part of the components or for all the components, though.

Note that we are driven by the analyses, but we take into account the uncertainties on the analyses (with \( \eta_{it} \)). In fact, we try to forecast the true state \( \bar{x}_{it}^{true} \) in every grid cell \( i \), using the ensemble of forecasts and the analysis which can be seen as some imperfect observation of the true state.

### 3.3. Aggregation Parameters

The analyses, their variances and the models simulations depend on the grid cell. The other parameters of the aggregation algorithm do not depend on the grid cell. Efficient grid-independent parameters were found after trials and tests in all grid cells at once. The parameters were essentially selected so that the diagnosis (14) is nearly satisfied. They were also tuned to minimize the errors with respect to the analysis and the observations. The robustness of these parameters was evaluated afterward (see section 4.3).

The model for the weights is trivial: \( A = I \) (identity matrix) and \( B = 0 \). In the initial weight vector \( \mathbf{w}_0 \), all components are zero except the component that corresponds to the model that benefits from data assimilation. This model, whose forecasts originate from analyses at 19:00 UTC, is given a weight 1. The first aggregated forecast therefore coincides with the forecast of this model. The value of \( \mathbf{Q} \) (see equation (4)) is \( 0.2I \) and \( \mathbf{Q}_0 \) is set to \( 0.015^2I \). The value for \( \mathbf{Q}_0 \) was selected based on the performance of the filter, but the filter shows moderate sensitivity to this parameter—see section 4.3. We stress that the minimax estimate is not sensitive to the initial weights as it is shown by Figure 2 and discussed in section 4.

### 4. Results

#### 4.1. Forecasting Performance

The forecast performance is evaluated from 1 February 2001 (at 15:00 UTC) to the end of the year (actually, 30 December at 15:00 UTC), hence over 333 days. The month of January serves as a spin-up period for the minimax filter. During the evaluation period, we first compare the individual models from the ensemble with the analyses. The root mean square error (RMSE) is computed with the ground-level ozone concentrations from all \( n = 3082 \) grid cells, so that the RMSE of the \( m \)-th model is \( \sqrt{\sum_{i=1}^{n} \sum_{t=2}^{364} (\bar{x}_{it}^{true} - \bar{x}_{it}^{m})^2} \). The RMSE varies greatly among the models of the ensemble: the highest RMSE is 51.2 \( \mu \text{g m}^{-3} \), and the lowest RMSE among the models without assimilation is 16.4 \( \mu \text{g m}^{-3} \). The best model is the first model as it benefits from data assimilation until 19:00 UTC in the previous day. Its RMSE is 13.5 \( \mu \text{g m}^{-3} \). Another reference RMSE is that of the ensemble mean: 18.4 \( \mu \text{g m}^{-3} \). With the parameters from section 3.3, the minimax aggregated forecast has a RMSE of 11.3 \( \mu \text{g m}^{-3} \). The Kalman filter performs equally well (11.3 \( \mu \text{g m}^{-3} \) too). Note that this is the same as the performance of the discounted ridge regression, whose RMSE is 11.3 \( \mu \text{g m}^{-3} \) as well.

In Figure 1, we compare the temporal mean of the analyses with (1) the aggregated forecasts, (2) the first model (hence with data assimilation), and (3) the first model without data assimilation. The latter is not part of the ensemble, but it is plotted to illustrate the difference between the model’s simulation with and without data assimilation. On average, the minimax forecasts capture all patterns from the analyses, with the right amplitude, while the reference model, with or without assimilation, fails in different regions, e.g., north to Spain.

Following the classical evaluation approach, we also compute the RMSE against observations, including all observations from the evaluation period. The RMSE of the best model, i.e., the first model, is 19.8 \( \mu \text{g m}^{-3} \). The RMSE of the ensemble mean is 22.5 \( \mu \text{g m}^{-3} \). The RMSE
Figure 3. Ensemble simulations (left) and minimax aggregated forecast (right), in g m\(^{-3}\), for one grid cell \((j = 18, i = 35)\) over about 110 days. The analysis is plotted in red on both figures. On the left, the 20 members of the ensemble are plotted in light blue. The blue background fills the space between the upper and lower envelopes of the ensemble (in black). The white line is the first model, which is the best model as it benefits from data assimilation until 19:00 UTC in the previous day. On the right, the minimax aggregated forecast is in white, and the blue background represents the uncertainty.

The uncertainty of the aggregated forecasts is 16.2 g m\(^{-3}\) for the minimax filter and 16.6 g m\(^{-3}\) for the Kalman filter. This is higher than the RMSE of the discounted ridge regression, which is 15.6 g m\(^{-3}\).

4.2. Uncertainty Estimation

A key contribution of the approach is the estimation of the uncertainty on the weights. Figure 2 shows the time evolution of the weights for the 20 models in a grid cell. The model that benefits from data assimilation receives the largest weight during the whole period. Note that the method is not sensitive to the initial weights. In the figure, we also show the evolution of the weights starting from \(Q^{-1} = 0\), i.e., with infinite error on the weights. At the end of the time period, a comparable distribution is reached and the overall performance (not shown in this paper) is essentially the same as with \(Q = 0.2I\). The uncertainty range contains the...
values where the true weights can lie according to the uncertainty description for the model errors and the observational errors. The uncertainty estimation corresponds to the first diagonal elements of $F_q^{-rac{1}{2}}$. The initial value is $F_q^{-rac{1}{2}} = Q^{-rac{1}{2}}$. It tends to grow in the first steps as the model error, quantified by $Q^{-1}$, accumulates in the weights. At the same time, the assimilation of the analyses tends to decrease the uncertainty on the weights. After about 50 or 70 steps, the amplitude of the uncertainty on the first weight reaches balanced values between the accumulation of model error and the corrections due to the assimilation.

[44] The uncertainty on the concentration can be derived from the uncertainty on the weights, following equation (13). Figure 3 shows the time evolution of the concentrations, for the ensemble models and for the aggregated forecast. They are compared with the analysis, and the uncertainty estimation is provided with the minimax forecast. The grid cell and the time period were selected so that they are representative of the best model RMSE, of the aggregated-forecast RMSE, and of the amount of analyses falling inside the uncertainty bounds. The uncertainty range from the minimax filter is much narrower than the ensemble envelope. Over the whole domain and the evaluation period, the analysis lies in the uncertainty bounds with frequency 0.751, which is similar to the expected value of 0.757 as computed according to the diagnosis (14).

[45] The uncertainty depends on the grid cell since the aggregation is carried out independently in all grid cells and the analysis variance depends on space. Indeed, the analysis variance is lower at observed locations and higher far off from the observation network. Figure 4 shows the uncertainty map for the aggregated forecast for 1 May. The map shares common features with classical uncertainty maps, e.g., the high uncertainty along the coasts, especially in the south [see Garaud and Mallet, 2011]. At the same time, the uncertainty is reduced around observed locations, as a result of the observations assimilation. This can be clearly seen in Spain where the blue spots correspond to observation stations. The blue spots can be shifted (from the exact station locations) because of the effect of transport. In Figure 5, we show the same uncertainty map, but for 14 February and along with the standard deviation of the ensemble. The standard deviation is only a rough uncertainty representation since the ensemble is not calibrated for uncertainty estimation. The uncertainty estimated by the filter is much lower on average. The patterns are significantly different for most days. Again, lower uncertainty is often found around observed locations.

**Table 1. Performance of the Aggregation Against the Weights’ Model Uncertainty $q$ So That $Q_q = q^2I^q$**

<table>
<thead>
<tr>
<th>$q$</th>
<th>0.0002</th>
<th>0.002</th>
<th>0.01</th>
<th>0.015</th>
<th>0.02</th>
<th>0.03</th>
<th>0.04</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Minimax</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE (analysis), $\mu$g m$^{-3}$</td>
<td>11.4</td>
<td>11.0</td>
<td>11.1</td>
<td>11.3</td>
<td>11.6</td>
<td>12.1</td>
<td>12.4</td>
</tr>
<tr>
<td>RMSE (observations), $\mu$g m$^{-3}$</td>
<td>16.4</td>
<td>15.8</td>
<td>15.9</td>
<td>16.2</td>
<td>16.4</td>
<td>16.8</td>
<td>17.1</td>
</tr>
<tr>
<td>Analysis outside $[\hat{x} - y, \hat{x} + y]$ (target)</td>
<td>70.9%</td>
<td>59.7%</td>
<td>31.4%</td>
<td>24.3%</td>
<td>19.6%</td>
<td>13.6%</td>
<td>9.7%</td>
</tr>
<tr>
<td>Analysis outside $[\hat{x} - y, \hat{x} + y]$ (actual)</td>
<td>66.0%</td>
<td>55.9%</td>
<td>33.1%</td>
<td>24.9%</td>
<td>18.7%</td>
<td>9.7%</td>
<td>4.1%</td>
</tr>
<tr>
<td><strong>Kalman</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE (analysis), $\mu$g m$^{-3}$</td>
<td>11.5</td>
<td>11.0</td>
<td>11.0</td>
<td>11.3</td>
<td>11.5</td>
<td>11.8</td>
<td>12.1</td>
</tr>
<tr>
<td>RMSE (observations), $\mu$g m$^{-3}$</td>
<td>17.0</td>
<td>16.3</td>
<td>16.4</td>
<td>16.6</td>
<td>16.9</td>
<td>17.2</td>
<td>17.4</td>
</tr>
</tbody>
</table>

*The best model in the ensemble, which benefits from data assimilation, has a RMSE of 13.5 $\mu$g m$^{-3}$ against analyses and of 19.8 $\mu$g m$^{-3}$ against the observations. The interval $[\hat{x} - y, \hat{x} + y]$ is defined in section 2.2, see equations (12) and (13). The lines “Analysis outside $[\hat{x} - y, \hat{x} + y]$” report with what frequency (1) the analyses are supposed to fall outside $[\hat{x} - y, \hat{x} + y]$ (line “target”) according to the diagnosis (14) from section 2.2.1, and (2) the analyses are actually found outside $[\hat{x} - y, \hat{x} + y]$ in the experiments (line “actual”). For a reliable uncertainty estimation, both values should coincide.*
The temporal mean of the forecasts lower bounds, i.e., the temporal mean of $\tilde{x}_{i,t} - y_{i,t}$ (see equations (12)-(13)), and the temporal mean of the forecasts upper bounds, i.e., the mean of $\tilde{x}_{i,t} + y_{i,t}$, are displayed in Figure 6.

4.3. Sensitivity to the Parameters

We computed the aggregation performance for different values of the weights model error, i.e., different values for $Q_t = q^t I$. Table 1 shows that the sensitivity of the RMSEs of minimax filter and Kalman filter with respect to $q$ is rather low. Despite the large variations in $q$, the performance in terms of errors remains significantly better than the performance of the best model. In terms of uncertainty estimation, the criterion (14) is approximately met with values from $q = 0.002$ to $q = 0.02$. As it was mentioned above, the weights are robust to the choice of the initial weight. In other words, the weights model and the observations contain enough information for the weights to converge to the optimal (in the minimax sense) weights over time.

5. Conclusions

We introduced a scheme for ensemble aggregation based on minimax filtering. The ensemble of forecasts is linearly combined with dynamical weights evolving over time in order to better forecast the true state. We assumed a linear uncertain model for the aggregation weights and applied a minimax filter on these weights to reduce the uncertainty. The filter computes the set of weights that are compatible with (1) the weight dynamics, (2) the weights’ model error that is supposed to belong to a prescribed ellipsoid, and (3) the observations and their errors which also belong to prescribed ellipsoids. Using a similar approach, a Kalman filter can be applied as well, but considering Gaussian errors instead of bounded errors.

A key point of the method is that it produces, along with an aggregated forecast, an estimation of the uncertainty on the weights and therefore of the uncertainty on the forecast concentrations. This uncertainty is described in terms of an ellipsoid where the true weights or concentrations are supposed to lie. We introduced a checking criterion based on the frequency with which the analysis falls outside the ellipsoid for the concentrations. The minimax uncertainty description assumes that all model errors (plus the error on the initial weights) fill the same ellipsoid: in other words, we assume one global bounding set. A more natural uncertainty description might rely on separate ellipsoids (for the model error) on every single time step, but some investigation is required so that the filter does not provide a strong overestimation of the uncertainty.

The method can be applied to forecast at individual observation stations, but in this paper, it is applied in the context of ensemble forecasting of analyses for ground-level ozone over Europe and during a full year. The performance of the filter is similar to that of the learning methods. Contrary to the previous work with learning methods, we have no comparison against all linear combinations constant over the full simulation period, but the filtering approach provides an uncertainty estimation. Furthermore, using appropriate parameters ($A = I$, $B = 0$), the filter can generate a constant weight vector which can be considered as a mean weight vector. One could think of using this mean weight, for instance, setting $A = a I$ and $B = I$, and introducing nontrivial systematic model error equal to the mean weight vector.

The analyses lie most of the time in the uncertainty range, and the uncertainty is much lower than the ensemble spread, especially around observed locations. Considering the performance and the availability of an uncertainty estimation, the method could be applied operationally just like ensemble forecast of analyses with machine learning is applied on the Prev’air operational forecasting platform operated by Institut National de l’Environnement Industriel et des Risques.

While the spatial patterns of the analyses are almost perfectly reproduced on average, the day-to-day forecasts of the analyses can show erroneous patterns. Future work should address the aggregation of spatial fields, which should involve weights depending on nonlocal analysis values. Since uncertainty estimation is available, additional information is provided and this may help future work on the forecast of threshold exceedance based on an ensemble. However, the final uncertainty estimations can be accurate only if the analyses error variances are accurate. In this paper, the estimations of the analyses error variances relied on a $\chi^2$ diagnosis, which is a clear limitation.

Appendix A: Construction of the Minimax Weights

We assume that the true weights $w^{true}_t$ are given in the following form:

$$w^{true}_t = Aw^{true}_{t-1} + Be + (e - w_0)Q^{-1}(e - w_0) + \sum_{r=0}^{T-1} e^r Q^{-1} e^r \leq 1,$$

$$y_{t+1} = E_{t+1} w^{true}_t + \eta_{t+1}, \quad E \left[ \sum_{r=0}^{T} \eta_r R_{t+1} \eta_{t+1} \right] \leq 1,$$

(41)

where the vectors $e$, $e_0$, and $\eta$ play the role of uncertain parameters which vary within the prescribed bounds. For simplicity, we derive the filter with $w_0 = 0$.

At some time $t \in \{0, \ldots, T\}$, we will be looking for the estimate of the projection of $w^{true}$ onto some $h$, i.e., $\hat{\ell}^T w^{true}_t$, in the class of linear functions $u$ defined on observations:

$$u(y_0, \ldots, y_T) = \sum_{s=0}^{T} u_s y_s.$$

The minimax estimate $\hat{u}$ minimizes the so-called worst-case error:

$$\hat{u} \in \text{Arginf} \sigma^2(u),$$

with $\sigma^2(u) = \sup_{e_0, \ldots, e_s, \eta_0, \ldots, \eta_1} \text{E}[(\ell^T w^{true}_t - u(y_0, \ldots, y_T))^2].$

In what follows, we sketch one way of constructing the minimax estimate. We refer the reader to Zhuk [2010] and Mallet and Zhuk [2010] for the details. Using Cauchy inequality for inner products weighted by $Q$, $Q_s$, and $R_s$, one can compute

$$\sigma^2(u) = z^T Q^{-1} z_0 + \sum_{s=0}^{T} z^T B^T Q^{-1} B z_s + \sum_{s=0}^{T} u_s^T R_s z_s,$$

(A2)

where $z_t$ is a solution of the so-called adjoint equation

$$z_t = -A^T z_{t+1} + E^T_{t+1} u_{t+1}, \quad z_0 = \ell.$$

(A3)
Now, we see that the minimax estimate $\hat{u}$ solves the so-called linear quadratic optimal control problem that is to minimize quadratic cost function (A2) (which is the worst-case error assigned to the estimate $u$) over the adjoint equation (A3). Solution to this problem has the following form (see Åström [2006] for details): $u_t = R(E,p)\sigma^2(u)$ where

$$z_t = A^T z_{t-1} + E^T R(E,p)$ \quad z_t = \ell$$

$$p_{t+1} = A_p + B^T Q^{-1} B z_t \quad p_0 = Q^{-1} z_0$$

(A4)

After simple algebra, we derive the following representation for the minimax estimate $\hat{u}$ and minimax error $\sigma^2(\hat{u})$:

$$\hat{u}(y_0, \ldots, y_t) = \ell^T \hat{w}_t \quad \sigma^2(\hat{u}) = \ell^T \hat{G}_t \ell$$

with $\hat{w}_t$ and $\hat{G}_t$ are defined in (7)–(10).

[36] Acknowledgments. The third author was partially supported by ERCIM “Alain Bensoussan” Fellowship Programme, which is supported by the Marie Curie cofunding by the European Commission.

References


