Reply to Anonymous Referee #1

First of all we would like to thank you very much for your comments and proposed improvements and corrections.

General comments

We have decided to maintain the demonstrations in the main text though only for a simpler cases only because we believe that a real understanding of a problem requires knowing whether something is true or not. However the reader can easily skip them as they are marked in the text in a clear way.

In the paper we have focused on the optimization of the representation of the ensemble by minimization of the mean square error, which is related to variance and bias. Removing the bias is a typical procedure and as you know, the variances of biased and non-biased random variables are the same: $V(\overline{X}) = V(\overline{X} - E(\overline{X} - Y))$. The question of uncertainty in a prediction you have raised, is a big issue that needs more investigation. In a sense, we consider this paper a starting point and an attempt to build a more robust mathematical background. We do hope to continue in this direction tackling also that question from the same angle. We regret having so rapidly deflated your interest in the paper by not tackling that issue upfront and we do realize we might have tried to cure your thirst with a glass of sea water. However, we still think that the paper raises issues that in the atmospheric dispersion and air quality communities have not been tackled yet nor presented.

We completely agree with your suggestion of including information on the use of ensemble in the weather forecasting which we realized was largely overlooked in the present version of the manuscript. We have complemented the introduction with a paragraph on that.

p. 14265 l. 2

Ensemble weather prediction should serve as an example of techniques build out of a robust theory that relates to predictability and uncertainty. Approaches based on either singular vector or bred vectors, have been developed from that theoretical framework and are used in operational activities such as ECMWF and NCEP (Atger 1997; Buizza and Palmer 1995; Buizza 1997; Buizza et al 1999a; Buizza, Miller, Palmer 1999b; Molteni 1996; Toth and Kalnay 1993; 1997; see Kalnay 2004 for more bibliography).

Specific comments

p. 14266 l. 10

The paragraph has been completely rewritten. We agree with you that the wording could be easily misunderstood.

Another important element motivating ensemble model practices and, in particular the multi-model ensemble, is the relationship between multi-model ensemble result and scientific consensus. Examples in this sense are the last IPCC reports or to a much smaller scale and different contest, the ENSEMBLE activity (Galmarini et al., 2004a). The concurrence of different results, originating from different sources, to the determination of an ensemble is an optimal method to represent all available scientific evidences (including their variability) and a way to facilitate agreement around a synthetic and relatively comprehensive result.

Whatever motivates the choice of ensemble modeling in atmospheric dispersion and air quality, we feel like pointing out that no investigation has ever been published on the fundamental elements that define an ensemble of atmospheric transport and dispersion model results and on the theoretical requirements that define it.

p. 14267 bullets 1-4. We've corrected it according to your suggestion and it now reads:

1. Is the ensemble average result always superior to that of individual members?
2. If one of the models has essentially a higher variance, should we remove it from the ensemble while calculating the ensemble average to minimize the ensemble variance? Under which conditions?

p. 14268. We included the definition of Talagrand diagram.

In a Talagrand diagram regular bins are created extending from the minimum to the maximum value predicted by the ensemble of model results. A normalised-frequency distribution is then obtained by counting the number of measured values that fall in the corresponding bin. An even distribution of the measurements guarantees that the ensemble covers the spectrum of measurement values. Any deviation from that structure is an indication of biased ensemble behaviour.

p. 14270 l15-20. Actually we think that the assumption that model results and measurements are independent is very reasonable. We have reformulated the paragraph to clarify that we use the notion of independence and correlation as it is used in probability theory. Below we have marked main changes in red.

…More formally, the independence of two systems can be expressed by the independence of random variables i.e.: two variables $z_1$ and $z_2$ representing two models are independent whenever their joint probability can be calculated as a product of individual ones i.e. $p(z_1, z_2)=p(z_1)p(z_2)$. This condition is reasonable in the case in which $z_1$ is the result of a model and $z_2$ is a measurement, but it is not difficult to imagine that it will not apply necessarily to two models. In fact in the latter case the condition applies to all results extracted from the two pdfs and implies that there is not possibility of a synergic contribution of the two models to the same result. In general this is a condition that is difficult to satisfy and verify for any atmospheric model. We will therefore relax the independence condition, thus requiring that the members of our ensemble are un-correlated in the sense of un-correlated random variables. This is a more realistic assumption to the extent that it applies to the average behavior rather than the intrinsic properties of each model. Formally the un-correlation is in fact defined as: $E\{z_1 z_2\}=E\{z_1\}E\{z_2\}$ where $E\{\}$ is the expectation value. The un-correlation includes the VL(2007) condition of independence of identical pdfs but also the condition of different pdfs or partly overlapping ones, as depicted schematically in Figure 2. Hence we have transferred the notion of correlation (or independency) from probability theory to the models. We are aware of the fact this is not commonly used term in ensemble systems but this allows us to use precise mathematical formulations.

p. 14271 Eq 2. In fact the variable *y* has been defined in the previous section, but we can recall it, of course.

Using the notation from Section 2 we can introduce the bias ($b$) and mean square error (shortly written as $S_2$) of the ensemble as:…

p. 14277 l8-12. We have a feeling that these statements, although obvious, should be clearly stated before we consider more complicated cases where variances are to be replaced by eigenvalues.

p. 14278. The expression "correlated models" derives from the assumption that the models are represented by random variables. Hence we transferred notion of correlation from random variables to the models (actually the same can be said about model independency). We are aware of the fact this is not commonly used term but we want to use precise mathematical definition. We have included a statement on this point on page 14270 where definitions of independency and correlation are given (see above).

p. 14282. In general there is no easy way to find analytical expression which is continuous relation between correlated and uncorrelated case. However if we consider an example for a simple two dimensional matrix from page 14284 assuming that p>1 we get

$$\frac{s_2}{s_1} = \frac{p^2+1+\sqrt{(p^2-1)^2+4a^2}}{p^2+1-\sqrt{(p^2-1)^2+4a^2}} \xrightarrow{a\to 0} p^2$$

while $V(\bar{x}_m) = \frac{p^2+1}{4} + \frac{a^2}{2} \xrightarrow{a\to 0} \frac{p^2+1}{4} \le 1 \; for \; p^2 \le 3$.

Hence one can conclude that slightly correlated models can have better bound than *m*. We put in the paper general estimation, but we have commented it in appropriate way.

After line 7 page 14283 we have added:

Modifying this example by letting $\varepsilon \to -\frac{1}{2}$ and putting 3 as the second variance (instead of 2) one can conclude that slightly correlated models can have better bound than *m*. However, there is no easy way to find a general analytical expression which would be a continuous relation between correlated and uncorrelated case.

p. 14284 l.4-13. We should make this clearer. Suppose that there are 2 different ensemble: the first one consists of 2 uncorrelated models, the second one consists of 2 correlated models but the variances of these models are the same as the variances of the models from the first system. Which ensemble would you expect to be better? We expect the first one to be better – however the example shows that in a specific situation the second one can produce lower mean square error. We have corrected the text by replacing lines 10-14 with the text below.

This example shows that if we consider two different ensembles: the first one consisting of two uncorrelated models with variances $\sigma_1^2$, $\sigma_2^2$, and the second with two correlated models with the same variances $\sigma_1^2$, $\sigma_2^2$, then there are conditions for which the second system can produce lower mean square error than the first one.

p. 14290 l.15-20. We also do not like the concept of removing the models, however this is a natural question arising when one of the models seems to be worse than others, or is an outlier, so we felt an answer ought to be included.

p. 14291 p.2. Of course there are different sources of errors, not only meteorological. However often in the simulation – for practical reasons – we use the same source term (or emission) or we make a number of simulations with different source terms, and then meteorological parameters are crucial. But you are right – we have commented it as shown below.

2. More advanced approaches could be based on meteorological data from Ensemble Prediction Systems (EPS), for example the ones available at ECMWF or NCEP. Additionally one can perturb initial data or model parameters.

p. 14293 l.10 We really do not want to say that atmospheric modelers are not good statisticians - we have deleted this sentence.

Updated reference


Atger F., 1999, The skill of ensemble prediction systems. *Mon. Wea. Rev*, **127**, 1941–1953

Buizza R and T. Palmer, 1995, The singular vector structure of the atmospheric general circulation, *J Atmos Sci*, 52, 1434-1456.

Buizza, R., 1997, Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF ensemble prediction system, *Mon. Wea. Rev.*, 125, 99-119.

Buizza, R., J. Barkmeijer, T.N. Palmer, and D. Richardson, 1999a: Current status and future developments of the ECMWF Ensemble Prediction System. *Meteorol. Appl*., 6, 1-14, January 1999.

Buizza R., M. Miller, T.N. Palmer, 1999b: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System, *Q. J. R. Met. Soc.* 125, 2887-2908.

Molteni, F., R. Buizza, T.N. Palmer, and T. Petroliagis, 1996: The ECMWF Ensemble Prediction System: Methodology and Validation. *Quart. J .Roy. Meteor. Soc.* 122, 73-119.

Toth,Z., and E. Kalnay, 1993: Ensemble Forecasting at the NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, 74, 2317-2330.

Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method, *Mon. Wea. Rev.* 125, 3297-3319.