Atmos. Chem. Phys. Discuss., 9, 14263–14314, 2009
www.atmos-chem-phys-discuss.net/9/14263/2009/

Atmospheric Chemistry and Physics Discussions

This discussion paper is/has been under review for the journal *Atmospheric Chemistry and Physics (ACP)*. Please refer to the corresponding final paper in *ACP* if available.

Est modus in rebus: analytical properties of multi-model ensembles

S. Potempski[1,2] **and S. Galmarini**[1]

[1]European Commission – DG Joint Research Centre, Institute for Environment and Sustainability, 21020 Ispra VA, Italy
[2]Institute of Atomic Energy, 05-400 Otwock-Swierk, Poland

Received: 3 March 2009 – Accepted: 23 May 2009 – Published: 1 July 2009

Correspondence to: S. Galmarini (stefano.galmarini@jrc.it)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

In this paper we investigate some basic properties of the multi-model ensemble systems, which can be deduced from a general characteristic of statistical distributions of the ensemble members with the help of mathematical tools. In particular we show how to find optimal linear combination of model results, which minimizes the mean square error both in the case of uncorrelated and correlated models. By proving basic estimations we try to deduce general properties describing multi-model ensemble systems. We show also how mathematical formalism can be used for investigation of the characteristics of such systems.

1 Introduction

The use of ensemble techniques in atmospheric dispersion is becoming more and more a popular research topic as well as application. A large number of modeling communities opt for joining forces in a common multi-model effort to improve their results, thus moving from the "deterministic" approach typical of the 80ies and 90ies to the statistically based ensemble approach of the last decade.

As described by Galmarini et al., 2004a there are several ways in which an ensemble can be constructed: either as a group of model runs produced by different modeling systems, or with one model and different input data or model settings. In this work we will mainly focus on multi-model ensembles in which models in principle have "nothing more" in common than the modeled case. Examples of these kind can be found in Stull et al. (1997), Krishnamurti et al. (2000), Dabbert and Miller (2000), Ziehmann (2000), Galmarini et al. (2001, 2004a, 2004b, 2008), Delle Monache and Stull (2003), Killip et al. (2003), Vijaya Kumar et al. (2003), Williford et al. (2003), McKeen et al. (2004), Mutemi et al. (2004), Yun et al. (2004), Delle Monache et al. (2006a), Mallet and Sportisse (2006), Van Loon et al. (2007), Vautard et al. (2008), Wang et al. (2008), Potempski et al. (2008). The multi-model ensemble approach constitutes

ACPD

9, 14263–14314, 2009

yet the extreme application case of the ensemble techniques and therefore it is worth attention.

Regardless of the methodology selected it is rather indisputably recognized that the treatment of several model results produces an overall improvement of the quality of the model simulations when compared with measurements. It is our opinion, however, that the technique deserves a yet renewed attention by the scientific community that should attempt to define its boundaries of applicability by formal methods other than through examples and applications. An attempt in this direction has been made, recently, in Riccio et al. (2007) where the ensemble results of Galmarini et al. (2004b) where reproduced in a Bayesian context and a formal explanation was presented.

To date, several questions remain unanswered. Among them one that is constantly present in the case of ensemble dispersion modeling relates to the way in which the ensemble should be set up. In other words, which criteria should be adopted to guarantee that the ensemble results will always be superior to those of any individual member? Moreover, how should the members be selected?

It is extremely interesting to note that all the multi-model ensembles presented in the last decade or so in the literature, found their reasons for existence in the opportunity of joining forces rather than in the analysis of the model characteristics, the specificity and peculiarity of their results, and the way that might have affected the ensemble. This interesting aspect of science sociology can be easily verified. The ensembles are made out of existing groups of scientists that get together offering their model results and having in return the added value of the ensemble. It feels good to be on one side or the other of the ensemble mean since it does not make a difference as long as the latter compares well with the measurements and no one is neither too near nor too far from it[1]. No a priori investigation is known to have been conducted on the degree of kinship each model has in the constituted community, on the models independence and complementarity. Models are not selected for a specific common

[1] We take the liberty to be so critical since we have been responsible for setting up one of these communities and we have the insider's perspective.

or exclusive characteristic they show, their results are just assumed to be appropriate for that specific application. In this context, the ensemble practice relies on what we could define *phenotypical model distinction*. Models differ for a limited number of modules, characteristics, or simply for the data used. The differentiation of a common model genotype occurs whenever the model is adopted or used in a specific modeling environment (modeling group or modeling application). In most of the cases the ensemble practice brings together models that show no substantial differences and that are available since they survived the *natural selection* of individual model evaluation. Another important element motivating ensemble model practices and, in particular the multi-model ensemble, is the relationship between multi-model ensemble result and political 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, from different sources, to the determination of an ensemble is an optimal method to represent all available scientific evidences (including its variability) and an efficient way to drive political consensus around it.

It should be clear that we are not at all against this process of natural aggregation of scientists and we do know and understand that most of the time, realism, pragmatism and practical constraints rule people's actions. However, 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. A clear and robust theory should guide the practice rather than heuristic approaches. In this respect the ensemble generation and treatment of weather forecast can be of inspiration to the atmospheric transport and dispersion modelling community.

This paper is a humble attempt to give formulas and ways to identify a priori the ensemble characteristics and to try to give a formal definition to a number of aspects that have never been discussed in the context of atmospheric dispersion modeling. The main aim of this paper is to introduce some basic properties of multi-model ensemble systems, which can be deduced from general characteristics of statistical distributions

ACPD
9, 14263–14314, 2009

of the ensemble members with the help of mathematical analysis. We have identified a precise number of questions that we wish to address in this work, namely:

1. Is the ensemble result always superior to that of individual members?
2. If one of the models has essentially a higher variance (and mean square error), should we remove it from the ensemble? Under which conditions?
3. How should an average of the ensemble be modified in order to extract an optimal representation from all its members?
4. Is there a condition which guarantees that the variance (and mean square error) of the ensemble mean is less than that of any individual model?

The paper is structured as follows: in the second section we present basic assumptions, and then the analysis for uncorrelated multi-model ensemble follows. In the fourth section we generalize the results to the case where the models should be considered as correlated ones. In Sect. 5 we summarize multi-dimensional case. The last section contains the conclusions. In the Appendix we include some technicalities related to multi-dimensional case.

An explanation is probably due on the title of this paper. The expression: "*Est modus in rebus*" is a sentence from the Satire (1, 1, 106–107) by the Latin poet Quintus Horatius Flaccus (aka Horace). The sentence should be translated as: "There is an optimal condition in all things" which in the original text is followed by the verse: "There are therefore precise boundaries beyond which one cannot find the right thing" (*sunt certi denique fines* | *quos ultra citraque nequit consistere rectum*). We think that these expressions summarize quite well the central topic of this paper and we hope it will result clearly in the proceedings.

ACPD
9, 14263–14314, 2009

2 The starting point

Our work starts from the article of Van Loon et al., 2007 (VL2007) in which a simple formula was presented to explain the advantage of the multi-model ensemble system for long-term ozone simulations in comparison with a single model approach. The formula proposed was obtained under the assumption that statistical distributions of model results and observations were identical and independent. More precisely, they considered daily ozone maxima in summertime determined by seven models and observations made in a number of stations. After bias corrections, the Talagrand diagram created from model results became flatter, which allowed them to conclude that the actual ozone concentration has similar statistical properties as the ensemble members. This justified the assumption that both simulated and observed concentrations could be sampled randomly from identical distributions (i.e. their variances are all equal), which in turn led to the following formula for the mean square error (msqe):

$$\text{msqe} = (1 + \frac{1}{m})\sigma^2 + b^2 \quad (1)$$

In (1) σ^2 and b stand for the variance and ensemble bias, respectively, while m represents the number of ensemble members. Expression (1) shows the advantage of ensemble approach as it demonstrates that the msqe for the ensemble (after bias correction) is always less than that of any individual model (m=1).

In general the assumption on identical statistical distribution is obviously a simplification as the models differ in terms of used parameterizations or numerical concept not mentioning application of distinct meteorological data. Also in the case considered in VL2007, the Talagrand diagram although improved after bias correction still was not perfect. An analysis we conducted on 204 different realizations produced by four atmospheric dispersion models applied to the ETEX-1 case (Girardi et al., 1998), using ECMWF-Ensemble Prediction System (EPS) weather fields, demonstrates that the spreads of the results sometimes can be significantly different among the models (see Fig. 1), and as a consequence also the variances.

ACPD
9, 14263–14314, 2009

Regardless of these additional considerations, the simple formulation given in (1) illustrates very well and quite synthetically the basic idea of the multi-model ensemble approach. Starting from this, we would like to give a slightly deeper look at the analysis based on statistical characteristics of the mean square error and present more general mathematical formalism. We will start our consideration with the simplest one-dimensional case i.e. when the simulation results can be described by a *scalar-valued random variable* for each model (for instance when multi-model ensemble is applied at a single point in space and time), then we will provide analogous formulas for the multi-dimensional case i.e. when the problem is described by a *vector-valued random variable*.

First we introduce the notation and some relevant assumptions. The model results we want to focus on are those of a prediction of the concentration levels of an unspecified substance at a single point in space and time. Let us assume that by using a set of m atmospheric dispersion models we obtain m simulations of the concentration evolution at that point. We assume that at the same point measurements were collected for the same variable. Both model data and observations are characterized by some estimation of variability or the error. For the models this can be done for example, by perturbing model parameters and some input data using Monte Carlo technique or through sensitivity analysis (e.g. Saltelli et al., 2006). The formal description of this situation implies that the values predicted by the models are represented by random variables x_j, $j = 1,\ldots,m$, where x_j corresponds to data produced by model j; each x_j has statistical distributions characterized by probability density functions (pdf) with bias and variance, which we denote as b_j and σ_j^2, respectively. Analogously we also assume that the observed values have some uncertainty characterized by random variable y with a pdf, which is described by the variance σ_o^2 (we can treat the measurements as not biased). In this context we do not need to specify any particular form of the pdf neither for the models nor for the observations. What we need to know however, are biases and variances.

ACPD
9, 14263–14314, 2009

An important aspect that requires clarification from the beginning is the level of diversity, independence, or correlation that we expect each of the m models of the ensemble to exhibit as a priori condition. These concepts or definitions are most of the time ignored, or given for granted thus leaving every reader with his own interpretation and sometime misunderstanding. To avoid that, we define explicitly the conditions we want the ensemble members to satisfy. The first condition we could impose on the members is that the individual models are independent[2]. In this sense we will need to specify whether we intend independence of the systems or of the results. Two models could be defined independent if they are structurally different, in other words if they are based on different modeling approaches or philosophies or if they are based on different parameterization of physical processes. At the same time they might be considered independent (or partially so) because they calculate atmospheric dispersion starting from meteorological fields originating from different weather models. In this context we will not distinguish between these two instances and we will refer to both cases indistinctively as independent models. More formally, the independence of two systems can be expressed as: 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. $p(z_1, z_2) = p(z_1)p(z_2)$. While this condition is automatically valid in the case in which $z_1$is the result of a model and z_2 is a measurement, it is not difficult to imagine that it will not apply necessarily to two models. In fact in this 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 very difficult to satisfy for any atmospheric model and if it applies it will be by all means too restrictive. We will therefore relax the independence condition, thus requiring that the members of our ensemble are *un-correlated*. This is indeed a more realistic assumption to the extent that it applies to the expectation value rather than the intrinsic properties of the pdf of each model. The un-correlation is in fact expressed

[2]This concept is most of the time confused with model difference which is just a qualitative and unrealistic definition.

ACPD
9, 14263–14314, 2009

as: $E\{z_1 z_2\}=E\{\ z_1\}E\{\ z_2\}$ where $E\{\}$ is the expectation value. In other words there is no need to know the pdf of the models which could be or not independent in this case. 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 Fig. 2.

Any consideration that will be derived for un-correlated models case, will then be extended to the correlated-models case, thus producing as result that whole model space will be covered. In fact by looking at Fig. 2 it is also clear that when we cover the correlated and uncorrelated model spaces we will include automatically also the dependent model class (dashed area) which is complementary to independent model class.

3 The case of an ensemble of uncorrelated model results

We assume that the results predicted by the m models can be represented in the ensemble form. We define the ensemble representative as any combination of model results in the form of an average or a generic linear combination of model results or median or any other percentile and that we denote by $\overline{x}$. The ensemble value is also supposed to be independent of measurements as all the members of the ensemble.

Finally we use standard notation $E(z)$ and $V(z)$ for the expectation operator and variance one, respectively, of any random variable z.

Now we can introduce the bias (b) and mean square error (shortly written as S_2) of the ensemble as:

$$b = E(\overline{x}-y) \text{ and } S_2 = E(\overline{x}-y)^2 \tag{2}$$

Under these assumptions the following formula holds:

$$S_2 = V(\overline{x}-y) + b^2 \tag{3}$$

ACPD
9, 14263–14314, 2009

Formula (3) is a direct consequence of the definition of bias and msqe, namely:

$$S_2 = E(\overline{x}-y)^2 = E(\overline{x}-y)^2-(E(\overline{x}-y))^2 + (E(\overline{x}-y))^2 = V(\overline{x}\text{-}y) + b^2 \ .$$

Using well known properties of the variance and under the assumed condition of the models independence of observations, from (3) we obtain:

$$S_2 = V(\overline{x}) + V(y) + b^2 \quad . \tag{4}$$

At this stage we define $\overline{x}$ explicitly as a linear combination of ensemble members i.e.:

$$\overline{x} = \sum_j \alpha_j \times_j \quad . \tag{5}$$

It is also reasonable to assume that the coefficients α_j are normalized i.e.

$$\sum_j \alpha_j = 1 \quad . \tag{6}$$

Since the models are uncorrelated expression (4) leads to the following equation:

$$S_2 = \sum_j \alpha_j^2 V(x_j) + V(y) + b^2 = \sum_j \alpha_j^2 \sigma_j^2 + \sigma_o^2 + \left(\sum_j \alpha_j b_j \right)^2 \tag{7}$$

Please note that formula (1) derived in VL2007, is a very special case of (7), as it can be obtained from the latter by taking the mean (i.e. all $\alpha_j = 1/m$) and assuming that all the variances to be equal, i.e.: $\sigma_j^2 = \sigma_o^2 = \sigma^2$.
More in general formula (7) offers us the possibility of finding optimal coefficients α_j so that they minimize msqe. This can be transformed into the optimization problem:
Find $\overline{\alpha}_1, \ldots, \overline{\alpha}_m$ such that:

$$S_2(\overline{\alpha}_1, \ldots, \overline{\alpha}_m) = \min_{\alpha_1,\ldots,\alpha_m} \left\{ \sum_j \alpha_j^2 \sigma_j{}^2 + \sigma_o{}^2 + \left(\sum_j \alpha_j b_j \right)^2 \right\} \text{ and } \sum_j \alpha_j = 1 \tag{8}$$

ACPD
9, 14263–14314, 2009

We then proceed by removing the bias from the model results first (i.e. introducing new variables: $\tilde{x}_j = x_j - b_j$) or by seeking for the solution of minimization problem in the form: $x' = \sum_j \alpha_j x_j - \beta = \overline{x} - \beta$, which introduces a new parameter β representing the bias of the ensemble. In the second case we would need an additional equation for β, which can be simply $E(x')=0$ (i.e. we postulate that x' is not biased), and leads to the following solution for β: $\beta = \sum_j \alpha_j b_j$. In one way or the other the bias can be easily removed. Therefore from now on we will consider that this operation was done a priori and that the models are all unbiased. The solution of problem (8) is equivalent to minimizing the following Lagrangian function: $L(\alpha_1, \ldots, \alpha_m, \lambda) = \sum_j \alpha_j^2 \sigma_j^2 + \sigma_o^2 + \lambda \left(1 - \sum_j \alpha_j\right)$, which leads to linear system of equations:

$$2\sigma_j^2 \alpha_j - \lambda = 0 \text{ for } j = 1, \ldots, m \tag{9}$$

$$\sum_j \alpha_j = 1$$

System (9) can be solved explicitly. By determining α_j from the first m equations and applying the normalization condition (6) we get:

$$\overline{\alpha}_k = \frac{\frac{1}{\sigma_k^2}}{\sum_j \frac{1}{\sigma_j^2}} \quad \text{for } k = 1, \ldots, m \tag{10}$$

As a matter of fact Eq. (10) corresponds to the minimization of the variance $V(\overline{x})$, which is widely used in a number of different applications for a large spectrum of problems, like the optimal interpolation in meteorology or Kalman filter (Gandin, 1964; Talagrand

ACPD
9, 14263–14314, 2009

1997; Kalnay, 2003 and references there). Formula (10) allows us to determine the optimal variance of the ensemble as the variance of the optimal linear combination i.e.:

$$V(\overline{x}_{\text{opt}}) = \sum_j \overline{\alpha}_j^2 \sigma_j^2 = \frac{1}{\sum_j \frac{1}{\sigma_j^2}} \tag{11}$$

From (11), because of the minimization of the msqe, the optimal variance is always less than any individual model variance. In fact for any $k=1,\ldots,m$ we have:

$$\frac{1}{\sigma_k^2} < \sum_j \frac{1}{\sigma_j^2},$$

which immediately implies that:

$$\frac{1}{\sum_j \frac{1}{\sigma_j^2}} < \sigma_k^2.$$

For the case where all the individual variances are equal, Eq. (11) leads to $V(\overline{x})=\frac{\sigma^2}{m}$, corresponding to the formula used in VL2007.

In summary formulas (10) and (11) can be applied to multi-model ensemble systems to produce optimal linear combination of model results, which minimizes msqe. This answers questions (a) and (c) set in the introduction.

In the remaining part of this section we try to demonstrate some generally valid properties of multi-model ensemble systems based on the formalism we have introduced and derived.

The first question that can be raised is about other than optimal linear combinations of ensemble which would have the property that the variance of the combination is always less than the variance of any individual models. This is not necessarily guaranteed by any linear combination as shown by the following example. Let us consider a

simple case with two models of variances σ_1^2, σ_2^2 such that $\sigma_2^2 = p\sigma_1^2$ for some $p > 1$. The combined variance is equal to:

$$\left(\alpha_1^2 + (1-\alpha_1)^2 p\right)\sigma_1^2$$

which is less then σ_1^2 only if $p < \frac{1+\alpha_1}{1-\alpha_1}$. For example for $\alpha_1 = \alpha_2 = 1/2$ (i.e. the mean of ensemble) this produces the condition $p < 3$. This suggests that in case of models with different variances it would be better if the difference between them is not too large, otherwise the combination described above would not produce an ensemble variance smaller than that of the individual models. Things would not be much different if we would assume that the linear combination coincides with the average. As shown above, even in this case, there is a possibility that an individual model variance (and consequently the msqe) can be less than the variance of the ensemble mean. We can therefore ask what conditions can guarantee that ensemble variance produced from the ensemble mean ($\alpha_1 = \ldots = \alpha_m = 1/m$) is always less than any individual model variance? The condition which should be fulfilled is:

$$V(\overline{x}_m) = \frac{\sigma_1^2 + \ldots + \sigma_m^2}{m^2} < \sigma_1^2,$$

where $\overline{x}_m$ represents the ensemble mean and we assume that the models variances are ordered such as $\sigma_1^2 \leq \sigma_2^2 \leq \ldots \leq \sigma_m^2$.
Then the following implication holds:

$$\text{if } \frac{\sigma_m^2}{\sigma_1^2} \leq m+1 \text{ then } V(\overline{x}_m) \leq \sigma_1^2 \leq \sigma_2^2 \leq \ldots \leq \sigma_m^2. \tag{12}$$

Condition (12) defines the relation between the maximal and minimal variances of ensemble datasets to assure that the variance of the ensemble mean is always not greater than any individual model variances, which answers question (d) in the introduction.

ACPD
9, 14263–14314, 2009

Proof: Indeed condition (12) implies that: $\frac{(m-1)\sigma_m^2}{m^2} \leq \frac{(m^2-1)\sigma_1^2}{m^2}$ and thus:

$(1-\frac{1}{m^2})\sigma_1^2 \geq \frac{m-1}{m^2}\sigma_m^2 \geq \frac{1}{m^2}(\sigma_2^2+\ldots+\sigma_m^2)$ which gives $\sigma_1^2 \geq \frac{\sigma_1^2+\ldots+\sigma_m^2}{m^2} = V(\overline{x}_m)$. ■

A question that follows from (12) is whether it is possible to define a priori limits of variability of the ensemble variance. Since $\sigma_1^2 \leq \sigma_j^2 \leq \sigma_m^2$ for $j=1,\ldots,m$ we have $\frac{1}{\sigma_1^2} \geq \frac{1}{\sigma_j^2} \geq \frac{1}{\sigma_m^2}$ and therefore the optimal ensemble variance will range as follows:

$$\frac{\sigma_1^2}{m} \leq V(\overline{x}_{\text{opt}}) \leq \frac{\sigma_m^2}{m} \tag{13}$$

The estimation shows that in principle it would be better to aggregate models whose individual variances are not very different (i.e. their relative ratio is close to 1). One could therefore conclude that when one model has a much bigger variance than the others, it is better to disregard it and exclude it from the ensemble. Unfortunately (or fortunately), this is not the case since the optimal variance obtained from m models is always less than the one calculated from $m-1$ ensemble members as shown by the following inequality:

$$V_m(\overline{x}_{\text{opt}}) = \frac{1}{\sum_j^m \frac{1}{\sigma_j^2}} < \frac{1}{\sum_j^{m-1} \frac{1}{\sigma_j^2}} = V_{m-1}(\overline{x}_{\text{opt}}).$$

The latter shows that σ_m^2 can be large but its contribution to the optimal representation will be very small (i.e. $\alpha_m \to 0$ as $\sigma_m^2 \to \infty$). In other words even a model (or models) with a huge variance cannot deteriorate the ensemble result if an optimal combination of model results is taken as ensemble representative. To corroborate this conclusion we can see that by combining inequality (13) with the fact that the optimal variance is always less than any individual model variance we get finally the following estimation:

ACPD
9, 14263–14314, 2009

$$\frac{\sigma_1^2}{m} \leq V(\overline{x}_{\text{opt}}) \leq \min\{\sigma_1^2, \frac{\sigma_m^2}{m}\}.$$

This shows that adding "a bad model" (i.e. the model with big variance) does not necessarily makes the estimation worse as msqe is bounded by the smallest individual variance anyway – hence it answers question (b) in the introduction. By the way (13) is also valid for the variance of the ensemble mean, but as shown above there is no guarantee that the individual model does not produce smaller msqe than the mean of the ensemble.

On the other hand a big difference between highest and lowest variances indicates that there is no agreement among the models in estimating uncertainty, so the ratio $\frac{\sigma_m^2}{\sigma_1^2}$ can be used as an indicator of the coherence of the multi-model ensemble simulations. If this ratio is close to 1 and the model predicted values are also close then there is a very good agreement within the ensemble.

At the end of this section we would like to add some comments related to the other possible way for obtaining weights (10). Namely this can be achieved by using the maximum likelihood principle (see for example Kalnay, 2003; Sasaki 1969; Parrish, Derber, 1992; Lorenc, 1986). Let us assume that x represents the truth and that conditional probability distributions are given by Gaussian pdf i.e.: $f_{\sigma_j}(x_j|x) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{(x_j-x)^2}{2\sigma_j^2}\right)$ for $j=1,..,m$. Then the likelihood of x being the truth is given by the following formula:

$$L(x|x_1, \ldots, x_m) = \prod_j f_{\sigma_j}(x_j|x) = \frac{1}{\left(\sqrt{2\pi}\right)^m \sigma_1 \ldots \sigma_m} \exp\left(-\sum_j \frac{(x_j-x)^2}{2\sigma_j^2}\right)$$

Hence the most likely value of x can be found by maximization of the likelihood function $x \rightarrow L(x|x_1,\ldots,x_m)$, which after taking logarithm and neglecting constant terms

ACPD
9, 14263–14314, 2009

leads to the minimization of the so called cost function:

$$J(x)=\frac{1}{2}\left[\frac{(x-x_1)^2}{\sigma_1^2}+\ldots+\frac{(x-x_m)^2}{\sigma_m^2}\right]$$

Then the solution of this problem is given by formulas (5) and (10).

The difference between this approach and preceding one is that here we assumed explicitly Gaussian distribution, while previously we did not take any particular assumption on pdf. On the other hand the minimization of the cost function is with respect to x not to parameters α_j, hence it shows that for Gaussian pdf appropriate linear combination produces an optimal solution. It can be also mentioned that the same cost function can be obtained using Bayesian interpretation (Edwards 1992; Kalnay, 2003).

4 What if the models are correlated?

While the formulas of the optimal combination for independent models correspond to the optimal interpolation in meteorology and therefore are generally well known, and have been already applied in a number of completely different areas (anywhere where independent measurements are considered), in this section we intend to extend the results to a more complicated situation, where the models cannot be no longer treated as uncorrelated ones. In particular we would like to derive analogous formulas for the optimal linear combination of multi-model results and variance. While this case might be perceived as an academic exercise, it is however a more realistic representation of the behavior of atmospheric dispersion models ensembles. We want to verify if properties analogous to those derived for the uncorrelated multi-model ensemble can be also obtained for the correlated case.

Let us consider the problem of minimizing msqe under the assumption that the statistical distributions of model results are not necessarily uncorrelated. If the ensemble is represented by a linear combination of the models results (according to the Eqs. 5–6)

ACPD
9, 14263–14314, 2009

then the formula for the ensemble variance is as follows:

$$V(\overline{x}) = \sum_i \alpha_i^2 \sigma_i^2 + \sum_{i,j,i\neq j} \alpha_i \alpha_j \mathrm{Cov}(x_i, x_j) = \sum_{i,j} \alpha_i \alpha_j \mathrm{Cov}(x_i, x_j),$$

where $\mathrm{Cov}(x_i, x_j)$ stands for the covariance of random variables x_i and x_j; we use the notation that $\sigma_i^2 = \mathrm{Cov}(x_i, x_i)$, where indices i, j correspond to model numbers.

Then the minimization problem (8) can be reformulated as follows where, as in the previous case, we consider the unbiased case:

Find $\overline{\alpha}_1, \ldots \overline{\alpha}_m$ such that:

$$S_2(\overline{\alpha}_1, \ldots, \overline{\alpha}_m) = \min_{\alpha_1,\ldots,\alpha_m} \left\{ \sum_{i,j} \alpha_i \alpha_j \mathrm{Cov}(x_i, x_j) + \sigma_o^2 \right\} \text{ and } \sum_j \alpha_j = 1. \quad (14)$$

This leads to the system of linear equations that can be written in a block form as:

$$\begin{bmatrix} 2\mathbf{K} & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

where $\mathbf{K}$ is the covariance matrix of dimension $m \times m$ (its ij-th element is $\mathrm{Cov}(x_i, x_j)$) and $\boldsymbol{\alpha}$ is a vector of coefficients (α_i). If we use notation $(\cdot, \cdot)$ for a dot product and define vector $\boldsymbol{I}$ as $\boldsymbol{I} = [1, \ldots, 1]^T$ this gives the equations:

$$\begin{aligned} &2\mathbf{K}\boldsymbol{\alpha} - \lambda \boldsymbol{I} = 0 \\ &(\boldsymbol{\alpha}, \boldsymbol{I}) = 1 \end{aligned} \quad (15)$$

Actually this system solves the minimization problem for the quadratic form $(\mathbf{K}\boldsymbol{\alpha}, \boldsymbol{\alpha})$ with the normalization condition (6) – in fact $V(\overline{x}) = (\mathbf{K}\boldsymbol{\alpha}, \boldsymbol{\alpha})$. This is also true for uncorrelated models case, in which the matrix $\mathbf{K}$ contains only variances as diagonal elements i.e.: $\mathbf{K} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_m^2)$.

To facilitate the treatment of the equations we take advantage of the fact that the covariance matrix is symmetric thus allowing us to apply the spectral theorem to write it

ACPD
9, 14263–14314, 2009

in the following form: $\mathbf{K} = \mathbf{USU}^*$ (Strang, 2003), where $\mathbf{U}$ is a unitary matrix (i.e. the columns are orthonormal vectors, which means that they are mutually orthogonal and have norm 1 i.e. $\|\cdot\|^2 = (\cdot,\cdot) = 1$; the same is true for the rows) and $\mathbf{S}$ is a diagonal matrix containing eigenvalues of the matrix $\mathbf{K}$. It should be added that the columns of $\mathbf{U}$ are eigenvectors of the matrix $\mathbf{K}$ and $\mathbf{U}^{-1} = \mathbf{U}^*$. The asterisk usually denotes generally adjoint operator i.e. complex conjugate and transposition defined by the following relation: $(\mathbf{U}\boldsymbol{u}, \boldsymbol{v}) = (\boldsymbol{u}, \mathbf{U}^*\boldsymbol{v})$ – since our case is real, $\mathbf{U}^* = \mathbf{U}^T$. The unitary matrix preserves also dot product i.e. $(\mathbf{U}\boldsymbol{u}, \mathbf{U}\boldsymbol{v}) = (\boldsymbol{u}, \boldsymbol{v})$.

System (15) can be also solved explicitly – we can repeat a similar procedure as used before to find that:

$$\boldsymbol{\alpha} = \tfrac{\lambda}{2}\mathbf{K}^{-1}\boldsymbol{I} = \tfrac{\lambda}{2}\mathbf{US}^{-1}\mathbf{U}^*\boldsymbol{I}, \text{ and}$$

$$1 = (\boldsymbol{\alpha}, \boldsymbol{I}) = \tfrac{\lambda}{2}(\mathbf{K}^{-1}\boldsymbol{I}, \boldsymbol{I}) = \tfrac{\lambda}{2}(\mathbf{US}^{-1}\mathbf{U}^*\boldsymbol{I}, \boldsymbol{I}) = \tfrac{\lambda}{2}(\mathbf{S}^{-1}\mathbf{U}^*\boldsymbol{I}, \mathbf{U}^*\boldsymbol{I}), \text{ hence}$$

$$\overline{\boldsymbol{\alpha}} = \frac{\mathbf{US}^{-1}\mathbf{U}^*\boldsymbol{I}}{(\mathbf{S}^{-1}\mathbf{U}^*\boldsymbol{I}, \mathbf{U}^*\boldsymbol{I})} = \frac{\mathbf{K}^{-1}\boldsymbol{I}}{(\mathbf{K}^{-1}\boldsymbol{I}, \boldsymbol{I})} \tag{16}$$

where we have assumed that there is no zero eigenvalue of the matrix $\mathbf{K}$. It is known that the covariance matrix $\mathbf{K}$ is nonnegative (Feller, 1968), so this also implies that all eigenvalues must be positive. In fact the case of zero eigenvalue would correspond to the situation where the model has zero variance, which means that the non-biased model is ideal (i.e. it does not produce errors).

If we denote the eigenvalues of $\mathbf{K}$ by $s_1,\ldots,s_m$, then obviously the diagonal matrix $\mathbf{S}^{-1}$ contains their inverses i.e. $1/s_1,\ldots,1/s_m$ (i.e. $\mathbf{S}^{-1}$=diag$(1/s_1,\ldots,1/s_m)$). By $(\boldsymbol{u}, \boldsymbol{v})_{\mathbf{S}^{-1}}$ we denote dot product generated by the operator $\mathbf{S}^{-1}$, which corresponds to the appropriate norm $((\boldsymbol{u}, \boldsymbol{v})_{\mathbf{S}^{-1}} = (\mathbf{S}^{-1}\boldsymbol{u}, \boldsymbol{v})$, $\|\boldsymbol{u}\|^2_{\mathbf{S}^{-1}} = (\boldsymbol{u}, \boldsymbol{u})_{\mathbf{S}^{-1}}$; analogously for $\mathbf{K}^{-1}$).

ACPD
9, 14263–14314, 2009

Then using formula (16) we can calculate the optimal variance:

$$V(\overline{x}_{\mathrm{opt}}) = (\mathbf{K}\overline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\alpha}}) = \frac{(\mathbf{K}\mathbf{K}^{-1}\boldsymbol{I}, \mathbf{K}^{-1}\boldsymbol{I})}{(\mathbf{K}^{-1}\boldsymbol{I}, \boldsymbol{I})^2} = \frac{1}{(\mathbf{K}^{-1}\boldsymbol{I}, \boldsymbol{I})} = \frac{1}{\|\boldsymbol{I}\|^2_{\mathbf{K}^{-1}}}$$

or

$$V(\overline{x}_{\mathrm{opt}}) = (\mathbf{U}\mathbf{S}\mathbf{U}^*\overline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\alpha}}) = \frac{1}{\|\mathbf{U}^*\boldsymbol{I}\|^4_{\mathbf{S}^{-1}}}(\mathbf{U}\mathbf{S}\mathbf{U}^*\mathbf{U}\mathbf{S}^{-1}\mathbf{U}^*\boldsymbol{I}, \mathbf{U}\mathbf{S}^{-1}\mathbf{U}^*\boldsymbol{I}) = \frac{1}{\|\mathbf{U}^*\boldsymbol{I}\|^4_{\mathbf{S}^{-1}}}(\mathbf{U}^*\boldsymbol{I}, \mathbf{S}^{-1}\mathbf{U}^*\boldsymbol{I}),$$

which reduces to:

$$V(\overline{x}_{\mathrm{opt}}) = \frac{1}{\|\mathbf{U}^*\boldsymbol{I}\|^2_{\mathbf{S}^{-1}}} = \frac{1}{\|\boldsymbol{I}\|^2_{\mathbf{K}^{-1}}} \tag{17}$$

If the matrix $\mathbf{K}$ is diagonal (i.e. the models are uncorrelated) formula (17) becomes (11) (as well as formulas (10) and (16) become identical). Please note that the denominator is simply the sum over the all elements of the matrix $\mathbf{K}^{-1}$.

Using the fact that $\boldsymbol{I} = \sum_j \boldsymbol{e}_j$, where $\boldsymbol{e}_j$ is the j-th versor ($\boldsymbol{e}_j = [0, \ldots, 1, \ldots 0]^T$ with 1 only on the j-th position) the denominator of Eq. (17) can be rewritten in terms of the elements of the matrix $\mathbf{U} = (u_{ij})$ as follows:

$$\|\mathbf{U}^*\boldsymbol{I}\|^2_{\mathbf{S}^{-1}} = \sum_j \frac{1}{s_j}\left(\sum_i u_{ij}\right)^2 = \sum_j \frac{1}{s_j}\sum_{i,k} u_{ij}u_{kj} = \sum_j \frac{1}{s_j}\Big(1 + \sum_{\substack{i,k \\ i \neq k}} u_{ij}u_{kj}\Big) \quad .$$

In comparison with the formula (11) an additional term appears (as an effect of correlation), while eigenvalues s_j play the role of variances. Thus in the remaining part of this section we investigate properties of multi-model ensemble similarly as for uncorrelated models case to verify whether already obtained results can be extended by transforming the variance into the eigenvalues.

First let us assume that all the eigenvalues of the covariance matrix are equal ($s_1 = \ldots = s_m = s$), which for uncorrelated models case corresponds to the situation with all the variances being the same. From formula (17) we can easily conclude that:

$$V(\overline{\boldsymbol{x}}_{\mathbf{opt}}) = \frac{1}{\frac{1}{s}\|\mathbf{U}^{*}\boldsymbol{I}\|^2} = \frac{1}{\frac{1}{s}\sum_i\sum_j(\mathbf{U}^{*}\boldsymbol{e}_j, \mathbf{U}^{*}\boldsymbol{e}_i)} = \frac{s}{\sum_i\sum_j(\boldsymbol{e}_j, \boldsymbol{e}_i)} = \frac{s}{m},$$

which is in accordance with the uncorrelated models case (of course all the weights α_k will be all equal).

Consider now the case of the ensemble mean (i.e. $\boldsymbol{\alpha} = \frac{1}{m}\boldsymbol{I}$). The implication analogous to (12) can be formulated as follows:

If $\frac{s_m}{s_1} \leq m$, then $V(\overline{x}_m) \leq s_1 \leq \ldots \leq s_m$,

where $\overline{x}_m$ represents the mean of ensemble (we assume that eigenvalues are ordered such that $s_1 \leq \ldots \leq s_m$).

Proof. This can be easily verified because of:

$$V(\overline{x}_m) = \frac{1}{m^2}(\mathbf{U}\mathbf{S}\mathbf{U}^{*}\boldsymbol{I}, \boldsymbol{I}) = \frac{1}{m^2}\|\mathbf{U}^{*}\boldsymbol{I}\|^2_{\mathbf{S}} = \frac{1}{m^2}\sum_j s_j\left(\sum_i u_{ij}\right)^2 \leq \frac{s_m}{m^2}\sum_j\left(\sum_i u_{ij}\right)^2 = \frac{s_m}{m^2}\|\mathbf{U}^{*}\boldsymbol{I}\|^2 =$$

$$= \frac{s_m}{m^2}\|\boldsymbol{I}\|^2 = \frac{s_m}{m} \leq s_1$$

as $\|\boldsymbol{I}\|^2 = m$ ■

In comparison with uncorrelated models case we see that the condition above is slightly more restrictive ($\frac{s_m}{s_1} \leq m$ versus $\frac{\sigma_m^2}{\sigma_1^2} \leq m + 1$), which is the effect of taking into account correlation terms. This condition cannot be changed to $\frac{s_m}{s_1} \leq m + 1$ as the following example shows. Consider the covariance matrix:

ACPD
9, 14263–14314, 2009

$\mathbf{K} = \begin{bmatrix} 1 & \frac{1}{2}+\varepsilon \\ \frac{1}{2}+\varepsilon & 2 \end{bmatrix}$, where $\varepsilon > 0$ is a parameter. Then one can calculate that:

$V(\overline{x}_m) = 1 + \frac{\varepsilon}{2} > 1 > s_1 = \frac{3-\sqrt{1+4(\frac{1}{2}+\varepsilon)^2}}{2}$, $\frac{s_2}{s_1} = \frac{3+\sqrt{1+4(\frac{1}{2}+\varepsilon)^2}}{3-\sqrt{1+4(\frac{1}{2}+\varepsilon)^2}}$. For $\varepsilon \leq \frac{\sqrt{5}-2}{4}$ we have $2 < \frac{s_2}{s_1} \leq 3$ i.e. $\frac{s_m}{s_1} \leq m+1$ but the condition $\frac{s_m}{s_1} \leq m$ is not satisfied (by the way $V(\overline{x}_{\text{opt}}) \leq 1$).

Finally also in the correlated case general bounds for the optimal variance can be obtained. Namely the following estimations hold:

$$\frac{s_1}{m} \leq V(\overline{x})_{\text{opt}} \leq \frac{s_m}{m}$$

Proof. $\|\mathbf{U}^* \boldsymbol{I}\|^2_{\mathbf{S}^{-1}} = \sum_j \frac{1}{s_j} \left(\sum_i u_{ij}\right)^2 \leq \frac{1}{s_1} \sum_j \left(\sum_i u_{ij}\right)^2 = \frac{1}{s_1}\|\mathbf{U}^* \boldsymbol{I}\|^2 = \frac{1}{s_1}\|\boldsymbol{I}\|^2 = \frac{m}{s_1}$,

and analogously: $\sum_j \frac{1}{s_j} \left(\sum_i u_{ij}\right)^2 \geq \frac{1}{s_m} \sum_j \left(\sum_i u_{ij}\right)^2 = \frac{m}{s_m}$. ■

Hence we get equivalent estimations for the optimal variance as for uncorrelated models case.

As optimal variance minimizes quadratic form $(\mathbf{K}\boldsymbol{\alpha}, \boldsymbol{\alpha})$ for any vector $\boldsymbol{\alpha}$ satisfying formula (6) then by taking versor $\boldsymbol{e_1}$ as vector $\boldsymbol{\alpha}$ we get: $V(\overline{x}_{\text{opt}}) \leq \sigma_1^2$, where σ_1^2 is the smallest model variance.

In such a way we can conclude that the following estimations for the optimal variance are true:

$$\frac{s_1}{m} \leq V(\overline{x}_{\text{opt}}) \leq \min\{\sigma_1^2, \frac{s_m}{m}\}.$$

ACPD
9, 14263–14314, 2009

In summary we can say that we have obtained similar results as for uncorrelated models case but the variances of models have been replaced by the eigenvalues of the covariance matrix **K**.

An interesting point relates to the fact that taking into account correlation one may improve msqe. As an example let us consider the following covariance matrix:

$\mathbf{K} = \begin{bmatrix} 1 & a \\ a & p^2 \end{bmatrix}$, (p>$a$≥0) then $V_{\text{opt}}^{\text{corr}} = \frac{p^2 - a^2}{p^2 - 2a + 1}$, while for a=0 (i.e. uncorrelated models case) we have: $V_{\text{opt}}^{\text{uncor}} = \frac{p^2}{1+p^2}$. It can be easily checked that $V_{\text{opt}}^{\text{corr}} < V_{\text{opt}}^{\text{uncor}}$ for $a \in (\frac{2p^2}{1+p^2}, p)$.

In other words the inclusion of the correlation term has improved msqe. This example shows also that the ensemble consisting of correlated models in some very specific situations can be more accurate than the ensemble (with the same pdfs) created from uncorrelated models.

At the end of this section we want to add that there is another way to obtain the formulas for the solution of the minimization problem formula (14), namely, by applying the spectral decomposition of the matrix **K** (Strang, 2003). This means that by considering **K** as a linear operator we have:

$\mathbf{K}\boldsymbol{\alpha} = \sum_j s_j(\boldsymbol{\varphi}_j, \boldsymbol{\alpha})\boldsymbol{\varphi}_j$, where ϕ_j are eigenvectors of matrix **K** (hence $\phi_j = \mathbf{U}\boldsymbol{e}_j$) forming the orthonormal basis. Using the same method as previously one can obtain the formula equivalent to (17):

$$V(\overline{x}_{\text{opt}}) = \frac{1}{\sum_j \frac{(\boldsymbol{\varphi}_j, \boldsymbol{I})^2}{s_j}}$$

ACPD
9, 14263–14314, 2009

5 Summary of multi-dimensional case

In this section we provide a summary on multi-dimensional case – all the technical details are given in the Appendix.

By multi-dimensional (or multivariate in general) case we consider the situation when the results of the simulations can be described by a *vector-valued random variable*. Typically this is the situation when we have a simulation domain with a number of spatial-temporal points and we want to include correlation between them. In fact the main difference between one- and multi-dimensional cases lies in taking into consideration correlation among different points. We are therefore adding an additional level of complexity to the cases analysed so far.

Analogously to the one-dimensional case we take *random vector* $\overline{\boldsymbol{X}}$ as the representative of the ensemble defined as a linear combination of *random vectors* $\boldsymbol{X_j}=[x_{1j}, \ldots, x_{nj}]^T$ representing multi-dimensional distribution for the model j:

$$\overline{\boldsymbol{X}}=[\overline{x_1}, \ldots, \overline{x_n}]^T = [\sum_{j=1}^{m} \alpha_{1j} x_{1j}, \ldots, \sum_{j=1}^{m} \alpha_{nj} x_{nj}]^T, \tag{18}$$

where m is the number of models and n denotes the dimension of the *random vectors* (for example number of points in some area). As previously we assume that the weights α_{ij} are normalized i.e.

$$\sum_{j=1}^{m} \alpha_{ij} = 1 \text{ for each } i = 1, \ldots, n. \tag{19}$$

We assume that the models are independent of the observations represented by *random vector* $\boldsymbol{Y}=[y_1, \ldots, y_n]^T$. We consider non-biased case as a similar procedure as for one-dimensional case can be also applied.

ACPD
9, 14263–14314, 2009

The first question is how to extend the definition of the square error. It seems that a natural way is the following one:

$$S_2^{(1)} = \sum_{i=1}^{n} E(\overline{x_i} - y_i)^2 \tag{20}$$

The mean square error can be expressed then as the average over all points i.e.: msqe= $S_2^{(1)}/n$.

First we assume that the models are mutually uncorrelated.

Then the problem of finding optimal coefficients which minimizes (20) leads to the minimization of the following Lagrange function (we can omit observation term):

$$L(\alpha_{11}, \ldots, \alpha_{1m}, \ldots, \alpha_{n1}, \ldots, \alpha_{nm}, \lambda_1, \ldots, \lambda_n) = \sum_i \sum_j \alpha_{ij}\sigma_{ij}^2 + \sum_i \lambda_i \left(1 - \sum_j \alpha_{ij}\right)$$

It can be easily seen that in such a way we obtain n separated systems of linear equations of the form (9) and therefore the formulas (10) and (11) can be applied for each point separately (for $i = 1, .., n$).

However, if we want to include also correlations between points (i.e. to consider the situation when the distributions $x_{1j}, \ldots, x_{nj}$ are correlated for any j), it seems that we should use the expression $E\{(\overline{\boldsymbol{X}} - \boldsymbol{Y})(\overline{\boldsymbol{X}} - \boldsymbol{Y})^T\}$ representing the covariance matrix. This upon the assumptions on independence between models and measurements leads to $E\{\overline{\boldsymbol{X}}\overline{\boldsymbol{X}}^T\}$ i.e. $n \times n$ matrix:

$$\mathrm{Cov}(\overline{\boldsymbol{X}}) = E\{\overline{\boldsymbol{X}}\overline{\boldsymbol{X}}^T\} = \begin{bmatrix} \sum_j \alpha_{1j}\alpha_{1j}E\{x_{1j}x_{1j}\} & \ldots\ldots\ldots & \sum_j \alpha_{1j}\alpha_{nj}E\{x_{1j}x_{nj}\} \\ \ldots\ldots\ldots & \ldots\ldots\ldots & \ldots\ldots\ldots \\ \sum_j \alpha_{nj}\alpha_{1j}E\{x_{nj}x_{1j}\} & \ldots\ldots\ldots & \sum_j \alpha_{nj}\alpha_{nj}E\{x_{nj}x_{nj}\} \end{bmatrix} \tag{21}$$

ACPD
9, 14263–14314, 2009

Then as a generalization of formula (20) we use the following formula:

$$S_2^{(2)} = \sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{j=1}^{m}\alpha_{kj}\alpha_{lj}E\{x_{kj}x_{lj}\} + \sum_{k,l}E\{y_k y_l\} = V(\overline{\boldsymbol{X}}) + V(\boldsymbol{Y}), \quad (22)$$

which corresponds to taking into account all the elements of the matrix (21). By msqe we put the average over all points i.e.: msqe= $S_2^{(2)}/n$.We split the formula into two terms: the first one related to the covariance matrix for the models ($V(\overline{\boldsymbol{X}})$) and the second one to the observations ($V(\boldsymbol{Y})$). It can be easily noticed that the previous formula (20) corresponds to taking into account only diagonal elements of the matrix (21). The matrix (21) should be distinguished from the covariance matrix introduced in Sect. 4. The latter one describes correlation between different models at one point while (21) defines the correlation between number of points for a linear combination of model results.

It should be also added that due to well known property of covariance (Feller, 1968):

$$E\{\overline{x}_i\overline{x}j\} \leq \sqrt{E\{\overline{x}_i^2\}E\{\overline{x}_j^2\}} \leq \frac{1}{2}\left(\{E\{\overline{x}_i^2\} + E\{\overline{x}_j^2\}\right) \quad ,$$

saying that off-diagonal elements are bounded by diagonal ones, formula (20) can be used to estimate upper bound of (22).

An extension to the correlated case is straightforward – we generalize formula (22) by including also terms related to the correlations between the models apart from already considered the correlations between the points. Then by the generalized mean square error we put msqe= $S_2^{(2)}/n$, where $S_2^{(2)}$ is expressed as follows:

$$S_2^{(2)} = \sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{j=1}^{m}\sum_{i=1}^{m}\alpha_{ki}\alpha_{lj}E\{x_{ki}x_{lj}\} + V(\boldsymbol{Y}) = V(\overline{\boldsymbol{X}}) + V(\boldsymbol{Y}) \quad (23)$$

It can be observed that in both cases (22, 23) the first term of the formulas can be expressed as: $V(\overline{\boldsymbol{X}}) = (\mathbf{K}\boldsymbol{\alpha}, \boldsymbol{\alpha})$, where:

ACPD
9, 14263–14314, 2009

– for uncorrelated case $\mathbf{K}$ is a block diagonal matrix: $\mathbf{K} = \mathrm{diag}(\mathbf{C_1}, \ldots, \mathbf{C_m})$,

$$\text{where } \mathbf{C_j} = \begin{bmatrix} E\{x_{1j}x_{1j}\} \ldots\ldots E\{x_{1j}x_{nj}\} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ E\{x_{nj}x_{1j}\} \ldots\ldots E\{x_{nj}x_{nj}\} \end{bmatrix},$$

– for correlated case $\mathbf{K} = \begin{bmatrix} \mathbf{C_{11}} \ldots\ldots\ldots\ldots \mathbf{C_{1m}} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \mathbf{C_{m1}} \ldots\ldots\ldots\ldots \mathbf{C_{mm}} \end{bmatrix}$

$$\text{where } \mathbf{C_{ij}} = \begin{bmatrix} E\{x_{1i}x_{1j}\} \ldots\ldots E\{x_{1i}x_{nj}\} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ E\{x_{ni}x_{1j}\} \ldots\ldots E\{x_{ni}x_{nj}\} \end{bmatrix},$$

and $\boldsymbol{\alpha} = [\boldsymbol{\alpha_1}, \ldots, \boldsymbol{\alpha_m}]^T$, $\boldsymbol{\alpha_j} = [\alpha_{1j}, \ldots, \alpha_{nj}]^T$.

Using this notation we can obtain the formulas for optimal weights and covariance matrix shown in Table 1.

For both cases we can obtain similar as in one-dimensional case ensemble properties, namely:

1. $\frac{\min \sigma(\mathbf{K})}{m} \leq \frac{V(\overline{X}_{\mathbf{opt}})}{n} \leq \frac{\max \sigma(\mathbf{K})}{m}$,

where by $\min \sigma(\mathbf{K})$ and $\max \sigma(\mathbf{K})$ we denote minimal and maximal elements of the spectrum $\sigma(\mathbf{K})$ of the covariance matrix $\mathbf{K}$, respectively i.e. the minimal and maximal eigenvalues.

2. $\frac{V(\overline{X}_{\mathbf{opt}})}{n} \leq \min_j s_n^{(j)} = \min_j \max_i s_i^{(j)}$,

ACPD
9, 14263–14314, 2009

where $s_i^{(j)}$, $i=1,\ldots,n$; $j=1,\ldots,m$ are the eigenvalues of **K** according to the block notation and put in the increasing order i.e.: $s_i^{(j)} \leq s_{i+1}^{(j)}$ for any i, j.

3. If $\frac{\max\sigma(\mathbf{K})}{\min\sigma(\mathbf{K})} \leq m$, then $\frac{V(\overline{X}_m)}{n} \leq \min\sigma(\mathbf{K})$, where $\sigma(\mathbf{K})$ is the spectrum of the matrix **K** and $\overline{\boldsymbol{X}}_{\boldsymbol{m}}$ is the ensemble mean.

We can see that we have analogous a priori estimations as in one dimensional case – optimal msqe is always bounded by minimal and maximal eigenvalues divided by the number of models. Similarly optimal msqe is always bounded by the maximal eigenvalue of the best individual model. And finally the ratio between highest and lowest eigenvalues can be used to find the condition when msqe for the ensemble mean is less than the one produced by any individual model.

6 Conclusions

In this study, by means of analytical formulation we have tried to fix some aspects never presented before, regarding the relationship between statistical behaviour of ensemble members and related expectations of the ensemble. The considerations presented here have been deduced having in mind the well known and extensively applied practice of ensemble dispersion modeling.

The results obtained show the importance of the knowledge of bias and variance of the statistical distributions for the models used in multi-model ensemble systems and how useful this information can be in the definition of the ensemble characteristics and in guaranteeing that the behaviour of the ensemble will fulfill the expectations. The results apply to both categories of correlated and uncorrelated models (or model results) filling a whole model space and can be summarized as follows:

– By choosing appropriate combination of model results we can find an optimal representative of the ensemble that after bias correction minimizes the mean square

ACPD
9, 14263–14314, 2009

error. This is equivalent to the minimization of the quadratic form defined by the covariance matrix with normalization condition. In fact the mean square error is expressed in terms of quadratic form determined by the covariance matrix.

– Some general a priori estimations for the optimal variance and msqe have been obtained, which show that multi-model ensemble has clear advantages in comparison to one model approach. This is expressed by the analytical formula demonstrating that msqe is bounded by the maximal variance or eigenvalue of the covariance matrix divided by the number of models. Similarly the lower bound of msqe is determined by the minimal variance or eigenvalue also divided by the number of models.

– We have devised the condition under which the mean of the ensemble still gives more accurate results in the sense of the minimization of msqe, than any individual model. This condition is expressed in the terms of the ratio between highest and lowest variances or eigenvalues of the appropriate covariance matrix. When the condition is not fulfilled, one can consider removing the responsible member from the ensemble. However in the sense of msqe, the ensemble results in principle cannot be deteriorated even by a model with a big variance if the optimal combination of models results is taken as a representative of the ensemble. It should be considered that eliminating a result from an ensemble is not as easy practice (especially for predictions) since there is no way to recognize when a single model is wrong or the ensemble is wrong or the case has a low predictability.

– If there is nothing wrong with any model then the ratio between the highest and lowest variances or eigenvalues can be considered as an indicator of the coherence of the multi-model ensemble. The biggest the ratio is, the highest disagreement among the models in estimating the uncertainty. In particular when this ratio is greater than the number of models this indicates that the ensemble mean may be worse than the best single models and special attention should be paid to take optimal representative of the ensemble. In this sense there is a relation

ACPD
9, 14263–14314, 2009

between the coherence of multi-model ensemble and the applicability of the ensemble mean, and this relation can be expressed simply as the ratio between the biggest and smallest eigenvalues of the appropriate covariance matrix. Thus the knowledge of the maximal and minimum eigenvalues has also practical meaning. When there is no big difference between them one can guarantee good coherence among models.

– We have also demonstrated that the same properties of multi-model ensemble are valid in the most general case in which both correlations between the models and between the points are taken into account.

Although it is out of the scope of the paper, we would like to add that it can be easily proved that the same formulas for the optimal covariance can also be applied in Kalman filter procedure to find optimal solution both for the gain matrix and ensemble representation at a time. This means that instead of using ensemble mean it is better to take optimal combination of models results accordingly to formulas shown in the paper.

The analysis produced points quite clearly toward the fact that one should acquire both the bias and variance of each ensemble member (see also Delle Monache et al., 2006b, Pagowski et al., 2006). To estimate the variance three approaches can be used:

1. Some of the models have built-in features to calculate variability of their results (e.g. Dabbert and Miller, 2000; Draxler, 2001; Stohl, 2005). It can be done by incorporating a kind of Monte Carlo simulations into the models for example by perturbing some crucial parameters.

2. More advanced approaches could be based by using meteorological data from Ensemble Prediction Systems (EPS), for example the ones available at ECMWF or NCEP.

ACPD
9, 14263–14314, 2009

3. A general approach based on sensitivity analysis in principle can be also applied (e.g. Saltelli, 2002; Hanson and Hemez, 2004; Saltelli et al., 2006; Campolongo et al., 2007)

All these approaches may not cover all possibly varying aspects of the model. However they are existing and applicable methods that can be used to produce yet useful information on the model variance. Of course the estimation of the variance can make simulation times much longer, in particular if it is to be based on EPS data. Bias correction methods have been already applied, in particular in air quality problems where there is enough amount of measurement data (Delle Monache et al., 2006, Delle Monache et al., 2008, Wilczak et al., 2006; Zupanski et al., 2007).

The estimation of correlation between the models can be based on various statistical tests (Lehman and Romano, 2008) for example, using Pearson correlation coefficient (see Rodgers and Nicewander, 1988 for an overview of different approaches). It may require however representative statistical material, which can be acquired from long term studies.

It should be also added that if individual models pdfs are known we can combine them using optimal weights to calculate ensemble pdf. This is in accordance with the general concept of applying the ensemble approach (Dabbert and Miller, 2000; Galmarini et al., 2004; Stull et al., 1997; Riccio et al., 2007; Potempski et al., 2009) to perform predictions in order to rely on stochastic paradigm rather than deterministic one.

Although mathematical framework used in the paper is not very sophisticated it shows how some basic results can be obtained. We wanted also to demonstrate that a formal mathematical approach can be useful to obtain general properties of the multi-model ensemble systems. It should be also added that a similar kind of analysis can be made for other than msqe metrics like the maximum norm. Obviously the estimations we presented can be applied to any multi-model ensemble system, not necessarily related to atmospheric dispersion models.

The real open issue still remains however, namely the connection between ensemble

ACPD
9, 14263–14314, 2009

coherence or agreement among the models which as we have seen can be predicted quite nicely with the analysis above, and uncertainty dealing with the relationship between the ensemble and observational data. The two are not coinciding concepts, and they are extremely relevant especially in the case of atmospheric dispersion ensemble forecast. In some cases the confusion between the two and an excessive confidence in the ensemble coherence as proxy of good result can lead to unwanted consequences, nicely summarized already in 1568 by P. Bruegel The Elder in the painting presented in Fig. 3. We will try to address also this aspect in the formal way in the future.

We are sure that the analysis presented here may result obvious to statisticians; we are not so sure the same effect will be produced to the eyes of atmospheric modellers who are extensive users of ensemble techniques.

Theoretical framing of practices is what we feel mostly needed at this stage of development of the ensemble dispersion technique and activities.

Appendix A

Details on multi-dimensional case

A1 Uncorrelated case

First we consider the problem of finding coefficients α_{ij} which minimize expression formula (22) similarly as in one-dimensional case. This leads to the minimization of the following Lagrange function (observation term can be omitted):

$$L(\alpha_{11},\ldots,\alpha_{1m},\ldots,\alpha_{n1},\ldots,\alpha_{nm},\lambda_1,\ldots,\lambda_n)=\sum_{k,l}\mathrm{Cov}_{kl}(\overline{\boldsymbol{X}})+\sum_i\lambda_i\left(1-\sum_j\alpha_{ij}\right), \qquad \text{(A1)}$$

ACPD
9, 14263–14314, 2009

where $\mathrm{Cov}_{kl}(\overline{\boldsymbol{X}})$ denotes (k, l) element of the covariance matrix (21) i.e.:

$$\mathrm{Cov}_{kl}(\overline{\boldsymbol{X}}) = \sum_{j=1}^{m} \alpha_{kj}\alpha_{lj} E\{x_{kj}x_{lj}\}$$

Then after differentiating formula (A1) with respect to all coefficients λ_i and α_{rs} we get the normalization formula (19) and:

$$\frac{\partial L}{\partial \alpha_{rs}} = 2\sum_{k=1}^{n} \alpha_{ks} E\{x_{rs}x_{ks}\} - \lambda_r \text{ for } r=1,..,n \text{ and } s=1,..,m, \tag{A2}$$

where the first index shows point number and the second one model number.

This system is not separable for each point but it can be written in a more convenient way, namely by grouping Eq. (A2) for each model we get:

$$\begin{aligned} &2\mathbf{C_1}\boldsymbol{\alpha_1} - \boldsymbol{\lambda} = 0 \\ &2\mathbf{C_2}\boldsymbol{\alpha_2} - \boldsymbol{\lambda} = 0 \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ &2\mathbf{C_m}\boldsymbol{\alpha_m} - \boldsymbol{\lambda} = 0 \\ &\sum_{j=1}^{m} \boldsymbol{\alpha_j} = \boldsymbol{I} \end{aligned} \tag{A3}$$

where $\boldsymbol{\alpha_j} = [\alpha_{1j}, \ldots, \alpha_{nj}]^T$ for $j=1,\ldots,m$, $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_n]^T$, $\boldsymbol{I} = [1, \ldots, 1]^T$ and $\mathbf{C_j}$ is the covariance matrix of $n \times n$ dimension for the model j:

$$\mathbf{C_j} = \begin{bmatrix} E\{x_{1j}x_{1j}\} \ldots\ldots E\{x_{1j}x_{nj}\} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ E\{x_{nj}x_{1j}\} \ldots\ldots E\{x_{nj}x_{nj}\} \end{bmatrix} \tag{A4}$$

In fact the system (A3) solves the minimization problem for the quadratic form $(\mathbf{K}\boldsymbol{\alpha}, \boldsymbol{\alpha})$ with the condition (19) (analogous to formula (6)), where $\mathbf{K}$ is a block diagonal matrix:

ACPD
9, 14263–14314, 2009

$\mathbf{K}=\mathrm{diag}(\mathbf{C_1}, \ldots, \mathbf{C_m})$ and $\boldsymbol{\alpha}=[\boldsymbol{\alpha_1}, \ldots, \boldsymbol{\alpha_m}]^T$. One can note that $V(\overline{\boldsymbol{X}}) = (\mathbf{K}\boldsymbol{\alpha}, \boldsymbol{\alpha})$, and this relation justifies the generalized formula of msqe (22) as it is natural extension of one-dimensional case.

From formula (A3) we can observe that the equations are coupled only via Lagrangian multipliers and normalization equation. Then again we can find explicit formulas for the optimal weights – namely as:

$$\boldsymbol{\alpha}_j=\frac{1}{2}\mathbf{C}_\mathbf{j}^{-1}\boldsymbol{\lambda} \text{ for } j=1,\ldots,m, \text{ from the last equation of (A3) we have } \boldsymbol{\lambda}=2\left(\sum_{j=1}^{m}\mathbf{C}_\mathbf{j}^{-1}\right)^{-1}\boldsymbol{I},$$

$$\text{hence } \boldsymbol{\alpha}_k=\mathbf{C}_\mathbf{k}^{-1}\left(\sum_{j=1}^{m}\mathbf{C}_\mathbf{j}^{-1}\right)^{-1}\boldsymbol{I}, \text{ for } k=1,..,m \tag{A5}$$

This formula is a natural extension of formula (10) for multi-dimensional case.

In order to obtain the optimal formula for the first term of the square error we rewrite it in the following form:

$$V(\overline{\boldsymbol{X}})=\sum_{j=1}^{m}\left(\mathbf{C}_\mathbf{j}\boldsymbol{\alpha}_j,\boldsymbol{\alpha}_j\right) \tag{A6}$$

Then by applying (A5) we get:

$$\sum_{j=1}^{m}\left(\mathbf{C}_\mathbf{j}\boldsymbol{\alpha}_j,\boldsymbol{\alpha}_j\right)=\sum_{j=1}^{m}\left(\mathbf{C}_\mathbf{j}\mathbf{C}_\mathbf{j}^{-1}\mathbf{C}\boldsymbol{I},\mathbf{C}_\mathbf{j}^{-1}\mathbf{C}\boldsymbol{I}\right)=\left(\mathbf{C}\boldsymbol{I},\sum_{j=1}^{m}\mathbf{C}_\mathbf{j}^{-1}\mathbf{C}\boldsymbol{I}\right), \text{ where} \tag{A7}$$

$$\mathbf{C}=\left(\sum_{j=1}^{m}\mathbf{C}_\mathbf{j}^{-1}\right)^{-1}$$

ACPD
9, 14263–14314, 2009

which leads finally to the following formula:

$$V(\overline{\boldsymbol{X}}_{\mathbf{opt}})=(\mathbf{C}\boldsymbol{I},\boldsymbol{I})=\left(\left(\sum_{j=1}^{m}\mathbf{C}_{\mathbf{j}}^{-1}\right)^{-1}\boldsymbol{I},\boldsymbol{I}\right) \tag{A8}$$

Hence we have obtained similar expression as for correlated models case (17), however with different covariance matrix. The matrix $\mathbf{C}$ defined by formula (A7) can be considered as the optimal covariance matrix.

Before starting examination of the properties of the optimal msqe we would like to add that formula (A7) has also Bayesian interpretation. Namely, if we consider for example case m=2 and assume that $\mathbf{X_1}$, $\mathbf{X_2}$ have Gaussian pdfs provided that $\mathbf{X_t}$ is the truth (i.e. $\boldsymbol{X_1}|X_t\sim N(\boldsymbol{X_t},\mathbf{C_1})$ and $\boldsymbol{X_2}|X_t\sim N(\mathbf{X_t},\mathbf{C_2})$), then the pdf of $\boldsymbol{X_t}|\ \boldsymbol{X_1},\boldsymbol{X_2}$ is also Gaussian of the distribution $N(\overline{X},\mathbf{C})$, where $\mathbf{C}^{-1}=\mathbf{C_1}^{-1}+\mathbf{C_2}^{-1}$ (Riccio, Giunta, Galmarini, 2007).

We start investigating the properties of the optimal combination of model results from the simple situation where all sub-matrices $\mathbf{C_j}$ are the same (i.e. all the models have the same distributions). Hence $\mathbf{C_j}=\mathbf{C}'$ for all j, and from spectral representation we get:

$$\mathbf{C}=\left(\sum_{j}\mathbf{C}_{\mathbf{j}}^{-1}\right)^{-1}=(m\mathbf{C}'^{-1})^{-1}=\frac{1}{m}\mathbf{C}' \text{ and } (\mathbf{C}\boldsymbol{I},\boldsymbol{I})=\frac{1}{m}(\mathbf{C}'\boldsymbol{I},\boldsymbol{I})=\frac{1}{m}(\mathbf{S}\mathbf{U}^{*}\boldsymbol{I},\mathbf{U}^{*}\boldsymbol{I})=\frac{\|\mathbf{U}^{*}\boldsymbol{I}\|_{\mathbf{S}}^{2}}{m}$$

If we denote eigenvalues of $\mathbf{C}'$ as $s_1\le\ldots\le s_n$ then by applying the same technique as in Sect. 3 (note that $\|\boldsymbol{I}\|^2=n$) we get similar estimations: $\frac{s_1}{m}\le\frac{V(\overline{\boldsymbol{X}})}{n}\le\frac{s_n}{m}$.

Let us consider now the general case and use the spectral representation for each $\mathbf{C_j}$: $\mathbf{C_j}=\mathbf{U_j}\,\mathbf{S_j}\mathbf{U_j}^{*}$, where $\mathbf{U_j}$ and $\mathbf{S_j}$ are the unitary and diagonal matrices, respectively. For the sake of convenience we put eigenvalues of $\mathbf{C_j}$ (i.e. elements of $\mathbf{S_j}$) in the increasing

ACPD
9, 14263–14314, 2009

order i.e.: $s_i^{(j)} \le s_{i+1}^{(j)}$ for any j.

The following estimations hold:

$$\frac{\min \sigma(\mathbf{K})}{m} = \frac{\min_j s_1^{(j)}}{m} \le \frac{V(\overline{\boldsymbol{X}}_{\mathbf{opt}})}{n} \le \frac{\max_j s_n^{(j)}}{m} = \frac{\max\ \sigma(\mathbf{K})}{m},$$

where by $\min\ \sigma(\mathbf{K})$ and $\max\ \sigma(\mathbf{K})$ we denote minimal and maximal elements of the spectrum $\sigma(\mathbf{K})$ of the covariance matrix $\mathbf{K} = \mathrm{diag}(\mathbf{C_1}, \ldots, \mathbf{C_m})$ i.e. the minimal and maximal eigenvalues.

Proof. Let put $\boldsymbol{w} = \mathbf{C}\boldsymbol{I} = \left(\sum_j \mathbf{C_j}^{-1} \right)^{-1} \boldsymbol{I}$,

then (denoting by $(\boldsymbol{v})_i$ the i-th element of any vector $\boldsymbol{v}$) we have:

$$(\mathbf{C}\boldsymbol{I}, \boldsymbol{I}) = (\boldsymbol{w}, \sum_j \mathbf{C_j}^{-1}\boldsymbol{w}) = \sum_j (\mathbf{C_j}^{-1}\boldsymbol{w}, \boldsymbol{w}) = \sum_j (\mathbf{S_j}^{-1}\mathbf{U_j}^*\boldsymbol{w}, \mathbf{U_j}^*\boldsymbol{w}) = \sum_j \sum_i \frac{1}{s_i^{(j)}} (\mathbf{U_j}^*\boldsymbol{w})_i^2 \ge$$

$$\sum_j \frac{1}{s_n^{(j)}} \|\mathbf{U_j}^*\boldsymbol{w}\|^2 = \sum_j \frac{1}{s_n^{(j)}} \|\boldsymbol{w}\|^2 = \sum_j \frac{1}{s_n^{(j)}} \|\mathbf{C}\boldsymbol{I}\|^2$$

Hence by Schwarz inequality:

$$\sum_j \frac{1}{s_n^{(j)}} \|\mathbf{C}\boldsymbol{I}\|^2 \le (\mathbf{C}\boldsymbol{I}, \boldsymbol{I}) \le \|\mathbf{C}\boldsymbol{I}\| \|\boldsymbol{I}\|, \quad \text{so} \quad \|\mathbf{C}\boldsymbol{I}\| \le \frac{\|\boldsymbol{I}\|}{\sum_j \frac{1}{s_n^{(j)}}},$$

and finally we get the following estimation: $\frac{V(\overline{\boldsymbol{X}}_{\mathbf{opt}})}{n} = \frac{(\mathbf{C}\boldsymbol{I}, \boldsymbol{I})}{n} \le \frac{\|\mathbf{C}\boldsymbol{I}\| \|\boldsymbol{I}\|}{n} \le \frac{1}{\sum_j \frac{1}{s_n^{(j)}}} \le \frac{\max_j s_n^{(j)}}{m}$.

To prove lower bound for optimal msqe we proceed in the following way – for any

$\boldsymbol{\alpha}=[\boldsymbol{\alpha_1}, \ldots, \boldsymbol{\alpha_m}]^T$ we have:

$$\sum_j (\mathbf{C_j}\boldsymbol{\alpha}_j, \boldsymbol{\alpha}_j) = \sum_j (\mathbf{S_j}\mathbf{U_j^*}\boldsymbol{\alpha}_j, \mathbf{U_j^*}\boldsymbol{\alpha}_j) = \sum_j \sum_i s_i^{(j)} (\mathbf{U_j^*}\boldsymbol{\alpha}_j)_i^2 \geq \sum_j s_1^{(j)} \|\boldsymbol{\alpha}_j\|^2 \geq \min_j s_1^{(j)} \|\boldsymbol{\alpha}\|^2$$

where the vector $\boldsymbol{\alpha}$ fulfils normalization condition (19). In order to find lower bound we minimize the norm i.e. $\|\boldsymbol{\alpha}\|^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_{ij}^2$. Taking into account condition (19) this can be done by minimizing inner sum separately for each $i=1, .., n$. It can be easily found that the minimum is reached at point $\alpha_{ij}=\frac{1}{m}$ for any i, j; which leads to the following inequality: $\|\boldsymbol{\alpha}\|^2 \geq \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{1}{m^2} = \frac{n}{m}$, true for any vector $\boldsymbol{\alpha}$ satisfying formula (19). Hence finally we obtain:

$$(\mathbf{C}\boldsymbol{I}, \boldsymbol{I}) = \min_{\alpha} (\mathbf{K}\boldsymbol{\alpha}, \boldsymbol{\alpha}) = \min_{\alpha} \sum_j (\mathbf{C_j}\boldsymbol{\alpha}_j, \boldsymbol{\alpha}_j) \geq \frac{n}{m} \min_j s_1^{(j)}, \quad \text{which gives :}$$

$$\frac{V(\overline{\boldsymbol{X}}_{\mathbf{opt}})}{n} \geq \frac{\min_j s_1^{(j)}}{m}. \quad \blacksquare$$

In such a way we have shown that also for multi-dimensional case the same estimation is valid as for one-dimensional case.

The other generally valid estimation is: $\frac{V(\overline{\boldsymbol{X}}_{\mathbf{opt}})}{n} \leq \min_j s_n^{(j)} = \min_j \max_i s_i^{(j)}$.

Proof. This can be proved in the following way: as the weights α are chosen to minimize lagrangian function then for any j the following inequality holds:

$(\mathbf{C}\boldsymbol{I}, \boldsymbol{I}) = \min_{\alpha} (\mathbf{K}\boldsymbol{\alpha}, \boldsymbol{\alpha}) \leq (\mathbf{C_j}\boldsymbol{I}, \boldsymbol{I}) = \sum_i s_i^{(j)} (\mathbf{U}^*\boldsymbol{I})_i^2 \leq \max_i s_i^{(j)} \|\boldsymbol{I}\|^2 = n \max_i s_i^{(j)}$, so taking minimum over j we get the estimation.

ACPD
9, 14263–14314, 2009

Combining this with previous estimation we have: $\frac{V(\overline{X}_{opt})}{n} \leq \min\left(\min_j s_n^{(j)}, \frac{\max_j s_n^{(j)}}{m}\right)$, saying that even if one model produces high eigenvalue msqe is still bounded by the maximal eigenvalue obtained from the best single model.

Let us finally consider the case where the mean is taken as the representative of the ensemble, which means that for any i, j, $\mathbf{K}$ we have: $\alpha_{ij}=\alpha_{ik}=\frac{1}{m}$, hence $\boldsymbol{\alpha}'=\boldsymbol{\alpha}_1=\ldots=\boldsymbol{\alpha}_m=[\frac{1}{m}, \ldots, \frac{1}{m}]^T$ (so $\|\boldsymbol{\alpha}'\|^2=\frac{n}{m^2}$). Then the following implication holds:

If $\frac{\max\sigma(\mathbf{K})}{\min\sigma(\mathbf{K})}=\frac{\max_j s_n^{(j)}}{\min_j s_1^{(j)}} \leq m$, then $\frac{V(\overline{X}_m)}{n} \leq \min_j s_1^{(j)}$, where $\overline{\boldsymbol{X}}_m$ is the ensemble mean.

Proof. This is a simple consequence of the following estimation:

$$V(\overline{\boldsymbol{X}}_m) = \sum_j (\mathbf{C_j}\boldsymbol{\alpha}', \boldsymbol{\alpha}') = \sum_j (\mathbf{S_j}\mathbf{U_j^*}\boldsymbol{\alpha}', \mathbf{U_j^*}\boldsymbol{\alpha}') = \sum_j \sum_i s_i^{(j)}(\mathbf{U_j^*}\boldsymbol{\alpha}')_i^2 \leq \sum_j s_n^{(j)}\|\mathbf{U_j^*}\boldsymbol{\alpha}'\|^2 \leq$$

$$\leq m \max_j s_n^{(j)}\|\boldsymbol{\alpha}'\|^2 = \frac{n}{m}\max_j s_n^{(j)}. \quad \blacksquare$$

Hence as for one-dimensional case the condition that ensemble mean produces lower msqe than any single model can be expressed by the ratio between highest and lowest eingevalues of the covariance matrix.

Finally we can add that some other estimations can be proved in a similar way like the following one: If $\sum_j s_n^{(j)} \leq m^2 \sum_j s_1^{(j)}$, then $\frac{V(\overline{X}_m)}{n} \leq \sum_j s_1^{(j)}$, where $\overline{\boldsymbol{X}}_m$ is the ensemble mean.

A2 Correlated case

For correlated case in order to simplify calculations we will use block matrix notation. Let $\mathbf{C_{ij}}$ ($i,j=1,\ldots,$m) be $n\times n$ matrix expressing dependence between two models i and

ACPD
9, 14263–14314, 2009

j at all n points (strictly – how model i is correlated to model j):

$$\mathbf{C_{ij}} = \begin{bmatrix} E\{x_{1i}x_{1j}\} \ldots\ldots E\{x_{1i}x_{nj}\} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ E\{x_{ni}x_{1j}\} \ldots\ldots E\{x_{ni}x_{nj}\} \end{bmatrix} \tag{A9}$$

Obviously we have $\mathbf{C_{ii}^T}=\mathbf{C_{ii}}$ and $\mathbf{C_{ij}^T}=\mathbf{C_{ji}}$ for $j \neq i$. Please note that strictly speaking the matrix $\mathbf{C_{ij}}$ (except of $\mathbf{C_{ii}}$) should not be treated as the covariance (or correlation) one (as it is not symmetric). In fact we ought to speak rather about the pair $\mathbf{C_{ij}}$ and $\mathbf{C_{ji}}$ as these both matrices describe fully mutual correlation between two models i and j.

If we introduce the matrix $\mathbf{K}$ of the dimension $mn \times mn$ as a block matrix:

$$\mathbf{K} = \begin{bmatrix} \mathbf{C_{11}} \ldots\ldots\ldots\ldots \mathbf{C_{1m}} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \mathbf{C_{m1}} \ldots\ldots\ldots\ldots \mathbf{C_{mm}} \end{bmatrix} \tag{A10}$$

and the vectors: $\alpha_j=[\alpha_{1j}, \ldots, \alpha_{nj}]^T$, for $j=1,\ldots,m$, $\boldsymbol{\alpha}=[\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m]^T$ (of dimension mn) then for the first term of formula (23) we get:

$$V(\overline{\boldsymbol{X}})=(\mathbf{K}\boldsymbol{\alpha}, \boldsymbol{\alpha})=\sum_{j=1}^{m}\sum_{i=1}^{m}(\mathbf{C_{ij}}\boldsymbol{\alpha}_j, \boldsymbol{\alpha}_i) \tag{A11}$$

The matrix $\mathbf{K}$ fully describes two types of correlations: between points and models, so it generalizes previously used covariance matrices – it has all the required properties.

First of all due to the mentioned above properties of matrices $\mathbf{C_{ij}}$ the matrix $\mathbf{K}$ is symmetric.

Secondly the matrix $\mathbf{K}$ is also positive semi-definite i.e. $(\mathbf{K}\boldsymbol{v},\boldsymbol{v}) \geq 0$ for any $\boldsymbol{v}$.

Proof. To prove it, consider auxiliary random variables:

$$Z_i=\alpha_{i1}x_{i1}+\ldots+\alpha_{im}x_{im}=\sum_{j=1}^{m}\alpha_{ij}x_{ij} \text{ for } i=1,\ldots,n, \text{ and their sum } Z=\sum_{i=1}^{n}Z_i=\sum_{i=1}^{n}\sum_{j=1}^{m}\alpha_{ij}x_{ij}.$$

ACPD
9, 14263–14314, 2009

Then the variance of Z is equal to the first term of the formula (23) i.e. $V(\overline{\boldsymbol{X}})$. According to (A11) for any $\boldsymbol{\alpha}$ we have: $(\mathbf{K}\boldsymbol{\alpha}, \boldsymbol{\alpha})$=V(Z)≥0 and as the variance is always nonnegative, this completes the proof.

This proof also shows that the first term of generalised msqe (i.e. $V(\overline{\boldsymbol{X}})$) can be defined as an averaged variance of the sum of linear combinations of multi-model results, taken over all the points (this to some extent justifies also used notation). On the other hand it is represented by the quadratic form $(\mathbf{K}\boldsymbol{\alpha}, \boldsymbol{\alpha})$, which corresponds to the previously considered cases.

If we use as previously vectors: $\boldsymbol{\lambda}=[\lambda_1, \ldots, \lambda_n]^T$ and $\boldsymbol{I}=[1, \ldots, 1]^T$ then the problem of finding coefficients minimizing (23) (or (A11)) can be solved by using the lagrangian function of the following form:

$$L=\sum_{i,j}(\mathbf{C_{ij}}\boldsymbol{\alpha}_j, \boldsymbol{\alpha}_i)+(\boldsymbol{\lambda}, \boldsymbol{I}-\sum_j \boldsymbol{\alpha}_j) \tag{A12}$$

Minimization with respect to all the elements of the vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\lambda}$ would lead to $nm+n$ equations. We simplify it by operating on the vectors $\boldsymbol{\alpha}_j$ and $\boldsymbol{\lambda}$ and by calculating Gaetaux derivative with respect to all $\boldsymbol{\alpha}_k$ along some vector $\boldsymbol{\delta}$:

$$\left.\frac{dL(\boldsymbol{\alpha}_k+t\boldsymbol{\delta})}{dt}\right|_{t=0}=\left.\frac{dL}{dt}\right|_{t=0}$$

$$\left(\sum_{i\neq k}(\mathbf{C_{ik}}(\boldsymbol{\alpha}_k+t\boldsymbol{\delta}), \boldsymbol{\alpha}_i)+\sum_{j\neq k}(\mathbf{C_{kj}}\boldsymbol{\alpha}_j, \boldsymbol{\alpha}_k+t\boldsymbol{\delta})+(\mathbf{C_{kk}}(\boldsymbol{\alpha}_k+t\boldsymbol{\delta}), \boldsymbol{\alpha}_k+t\boldsymbol{\delta})-(\boldsymbol{\lambda}, \boldsymbol{\alpha}_k+t\boldsymbol{\delta})\right)$$

$$=2(\mathbf{C_{kk}}\boldsymbol{\alpha}_k, \boldsymbol{\delta})+\sum_{i\neq k}(\mathbf{C_{ik}}\boldsymbol{\delta}, \boldsymbol{\alpha}_i)+\sum_{j\neq k}(\mathbf{C_{kj}}\boldsymbol{\alpha}_j, \boldsymbol{\delta})-(\boldsymbol{\lambda}, \boldsymbol{\delta})=2\sum_j(\mathbf{C_{kj}}\boldsymbol{\alpha}_j, \boldsymbol{\delta})-(\boldsymbol{\lambda}, \boldsymbol{\delta})=0$$

As the above equation is valid for any vector $\boldsymbol{\delta}$ then we get the following system of the equations:

$2\sum_{j=1}^{m}\mathbf{C_{kj}}\boldsymbol{\alpha}_j-\boldsymbol{\lambda}=0$ for $k=1,\ldots,m$ and $\sum_{j-1}^{m}\boldsymbol{\alpha}_j=\boldsymbol{I}$.

This system can be written in the following form:

$$\begin{bmatrix} 2\mathbf{C_{11}}\ldots..2\mathbf{C_{1m}} & -\mathbf{I} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ 2\mathbf{C_{m1}}\ldots..2\mathbf{C_{mm}} & -\mathbf{I} \\ \mathbf{I}\ldots\ldots\ldots\ldots\ldots\boldsymbol{I} & 0 \end{bmatrix}\begin{bmatrix}\boldsymbol{\alpha}_1\\ \ldots.\\ \boldsymbol{\alpha}_m\\ \boldsymbol{\lambda}\end{bmatrix}=\begin{bmatrix}0\\ \ldots\\ 0\\ \boldsymbol{I}\end{bmatrix},$$

which can be rewritten in a more concise form as:

$$\begin{bmatrix} 2\mathbf{K} & -\mathbf{I_R} \\ \mathbf{I_C} & 0 \end{bmatrix}\begin{bmatrix}\boldsymbol{\alpha}\\ \boldsymbol{\lambda}\end{bmatrix}=\begin{bmatrix}0\\ \boldsymbol{I}\end{bmatrix},$$

where $\mathbf{I_C}=[\mathbf{I},\ldots,\mathbf{I}]$ and $\mathbf{I_R}=[\mathbf{I},\ldots,\mathbf{I}]^T$ (i.e. $\mathbf{I_C^T}=\mathbf{I_R}$) and $\mathbf{I}$ is identity matrix of $n\times n$ dimension. Then the system of equations can be solved as previously as:

$\boldsymbol{\alpha}=\frac{1}{2}\mathbf{K^{-1}I_R}\boldsymbol{\lambda}$ and $\frac{1}{2}\mathbf{I_C}\,\mathbf{K^{-1}I_R}\boldsymbol{\lambda}=\boldsymbol{I}$, hence $\boldsymbol{\lambda}=2(\mathbf{I_C}\mathbf{K^{-1}I_R})^{-1}\boldsymbol{I}$, and finally

$$\boldsymbol{\alpha}=\mathbf{K^{-1}I_R}(\mathbf{I_R^T}\,\mathbf{K^{-1}I_R})^{-1}\boldsymbol{I}=\mathbf{K^{-1}}\,\mathbf{I_R}\,\mathbf{C}\boldsymbol{I} \tag{A13}$$

Please note that the matrix $\mathbf{I_R^T}\,\mathbf{K^{-1}I_R}$ has dimension $n\times n$ (so the inverse operator has sense), but the matrix $\mathbf{K}$ is of $mn\times mn$ dimension. By $\mathbf{C}$ we denote $(\mathbf{I_R^T}\,\mathbf{K^{-1}I_R})^{-1}$.

Then we can calculate the optimal first term of formula (23) $V(\overline{\boldsymbol{X}}_{\mathbf{opt}})$ as:

$$(\mathbf{K}\boldsymbol{\alpha},\boldsymbol{\alpha})=(\mathbf{K}\mathbf{K^{-1}I_R}\,(\mathbf{I_R^T}\mathbf{K^{-1}I_R})^{-1}\boldsymbol{I},\mathbf{K^{-1}I_R}(\mathbf{I_R^T}\mathbf{K^{-1}I_R})^{-1}\boldsymbol{I})=$$
$$((\mathbf{I_R^T}\mathbf{K^{-1}I_R})^{-1}\boldsymbol{I},\mathbf{I_R^T}\mathbf{K^{-1}I_R}(\mathbf{I_R^T}\mathbf{K^{-1}I_R})^{-1}\boldsymbol{I}) \text{ hence}$$

$$V(\overline{\boldsymbol{X}}_{\mathbf{opt}})=(\mathbf{C}\boldsymbol{I},\boldsymbol{I})=((\mathbf{I_R^T}\mathbf{K^{-1}I_R})^{-1}\boldsymbol{I},\boldsymbol{I}) \tag{A14}$$

The formula is similar to the previous ones – if we write the inverse matrix $\mathbf{K^{-1}}$ in the block form as:

ACPD
9, 14263–14314, 2009

$$\mathbf{K}^{-1} = \begin{bmatrix} \mathbf{C}_{11}^{(-1)} & \cdots & \mathbf{C}_{1m}^{(-1)} \\ \cdots & \cdots & \cdots \\ \mathbf{C}_{m1}^{(-1)} & \cdots & \mathbf{C}_{mm}^{(-1)} \end{bmatrix}, \text{ where } \mathbf{C}_{kj}^{(-1)} \text{ are sub-matrices of } n \times n \text{ dimension, then:}$$

$$\mathbf{C} = \left(\mathbf{I}_R^T \mathbf{K}^{-1} \mathbf{I}_R\right)^{-1} = \left(\sum_{k=1}^{m} \sum_{j=1}^{m} \mathbf{C}_{kj}^{(-1)}\right)^{-1} \tag{A15}$$

Thus the formula (A15) generalizes previously obtained expressions for the optimal covariance. In fact the multi-dimensional case with uncorrelated models corresponds to the situation where all the off-diagonal sub-matrices $\mathbf{C}_{ij}$ of the matrix $\mathbf{K}$ vanish. Then the inverse of the matrix $\mathbf{K}$ can be done block by block for matrices $\mathbf{C}_{ii}$, which corresponds to the formula (A7). On the other hand if we consider one-dimensional case with correlated models then the dimension of matrices $\mathbf{C}_{ij}$ is 1×1, so they become single elements of the matrix $\mathbf{K}^{-1}$ from Sect. 3, and the Eq. (A14) leads to formula (17).

The rest of the section will be devoted to prove similar estimations as previously for simpler cases.

We start with a general estimation for $V(\overline{\mathbf{X}}_{\text{opt}})$. The analogous estimations as before is valid:

$$\frac{\min \sigma(\mathbf{K})}{m} \leq \frac{V(\overline{\boldsymbol{X}}_{\mathbf{opt}})}{n} \leq \frac{\max \sigma(\mathbf{K})}{m},$$

where $\sigma(\mathbf{K})$ is the spectrum of the matrix $\mathbf{K}$ defined by (33) while $\min\sigma(\mathbf{K})$ and $\max \sigma(\mathbf{K})$ represent minimal and maximal eigenvalues, respectively.

Proof. First we write the matrix $\mathbf{I}_R^T \mathbf{K}^{-1} \mathbf{I}_R$ using spectral decomposition of the matrix $\mathbf{K} = \overline{\mathbf{USU}}^*$. We apply block notation i.e.

$$\mathbf{U}=\begin{bmatrix}\mathbf{U}_{11}\ldots\ldots\mathbf{U}_{1m}\\ \ldots\ldots\ldots\ldots\\ \mathbf{U}_{m1}\ldots\ldots\mathbf{U}_{mm}\end{bmatrix}\quad \mathbf{U}^{*}=\begin{bmatrix}\mathbf{U}^{*}_{11}\ldots\ldots\mathbf{U}^{*}_{1m}\\ \ldots\ldots\ldots\ldots\\ \mathbf{U}^{*}_{1m}\ldots\ldots\mathbf{U}^{*}_{mm}\end{bmatrix}\quad \mathbf{S}=\begin{bmatrix}\mathbf{S}_{1}\ldots\ldots\ldots 0\\ \ldots\ldots\ldots\ldots\\ 0\ldots\ldots\ldots\mathbf{S}_{m}\end{bmatrix},$$

where all the blocks have dimension $n\times n$. Due to the properties of the unitary matrix $\mathbf{U}$ we have: $\mathbf{U}^{*}_{ij}=\mathbf{U}^{T}_{ji}$. Then we get:

$$\mathbf{K}^{-1}=\mathbf{U}\mathbf{S}^{-1}\mathbf{U}^{*}=\begin{bmatrix}\sum_j \mathbf{U}_{1j}\mathbf{S}_j^{-1}\mathbf{U}^{*}_{j1}\ldots\ldots\ldots\sum_j \mathbf{U}_{1j}\mathbf{S}_j^{-1}\mathbf{U}^{*}_{jm}\\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ \sum_j \mathbf{U}_{mj}\mathbf{S}_j^{-1}\mathbf{U}^{*}_{j1}\ldots\ldots\ldots\sum_j \mathbf{U}_{mj}\mathbf{S}_j^{-1}\mathbf{U}^{*}_{jm}\end{bmatrix}\quad\text{and}$$

$$\mathbf{I}_R^{T}\mathbf{K}^{-1}\mathbf{I}_R=\sum_l\sum_k\sum_j \mathbf{U}_{lj}\mathbf{S}_j^{-1}\mathbf{U}^{T}_{kj}.$$

Let $\boldsymbol{w}=\mathbf{C}\boldsymbol{I}$ and denote eigenvalues of $\mathbf{S}$ according to block notation and put into the such an order that $s_k^{(j)}\le s_{k+1}^{(j)}$ for $j=1,\ldots,m$ and $k=1,.=..,n-1$. This leads to the following expressions:

$$(\mathbf{C}\boldsymbol{I},\boldsymbol{I})=(\boldsymbol{w},\mathbf{I}_R^{T}\mathbf{K}^{-1}\mathbf{I}_R\boldsymbol{w})=\sum_l\sum_k\sum_j(\mathbf{S}_j^{-1}\mathbf{U}_{lj}{}^{T}\boldsymbol{w},\mathbf{U}_{kj}^{T}\boldsymbol{w})=\sum_j(\mathbf{S}_j^{-1}\sum_l\mathbf{U}_{lj}^{T}\boldsymbol{w},\sum_k\mathbf{U}_{kj}^{T}\boldsymbol{w})=$$

$$=\sum_j\sum_i\frac{1}{s_i^{(j)}}\left(\sum_k\mathbf{U}_{kj}^{T}\boldsymbol{w}\right)_i^2\ge\sum_j\frac{1}{s_n^{(j)}}\sum_i\left(\sum_k\mathbf{U}_{kj}^{T}\boldsymbol{w}\right)_i^2\ge\frac{1}{\max_j s_n^{(j)}}\sum_j(\sum_l\mathbf{U}_{lj}^{T}\boldsymbol{w},\sum_k\mathbf{U}_{kj}^{T}\boldsymbol{w})=$$

$$=\frac{1}{\max_j s_n^{(j)}}\sum_l\sum_k\sum_j(\mathbf{U}_{lj}^{T}\boldsymbol{w},\mathbf{U}_{kj}^{T}\boldsymbol{w})=\frac{1}{\max_j s_n^{(j)}}(\boldsymbol{w},\mathbf{I}_R^{T}\mathbf{U}\mathbf{U}^{*}\mathbf{I}_R\boldsymbol{w})=\frac{m}{\max_j s_n^{(j)}}\|\boldsymbol{w}\|^2=\frac{m}{\max_j s_n^{(j)}}\|\mathbf{C}\boldsymbol{I}\|^2$$

because of $\mathbf{U}\mathbf{U}^{*}=\mathbf{I}$ and $\mathbf{I}_R^{T}\mathbf{I}_R=m\mathbf{I}$.

Then as previously using Schwarz inequality we can get similar estimation:

ACPD
9, 14263–14314, 2009

$\frac{m}{\max_j s_n^{(j)}}\|\mathbf{C}\boldsymbol{I}\|^2 \le \|\mathbf{C}\boldsymbol{I}\|\|\boldsymbol{I}\|$, hence $|\mathbf{C}\boldsymbol{I}\| \le \|\boldsymbol{I}\|\frac{\max_j s_n^{(j)}}{m}$, and finally (as $\|\boldsymbol{I}\|^2=n$)

$$\frac{V(\overline{\boldsymbol{X}}_{\mathbf{opt}})}{n}=\frac{(\mathbf{C}\boldsymbol{I},\boldsymbol{I})}{n} \le \frac{\max_j s_n^{(j)}}{m}.$$

Similarly as for uncorrelated models case one can also find the lower bound i.e.: $\frac{V(\overline{\boldsymbol{X}}_{\mathbf{opt}})}{n} \ge \frac{\min_j s_1^{(j)}}{m}$.

This can be proved in the same way as before since the following inequality holds for any vector $\boldsymbol{\alpha}$:

$(\mathbf{K}\boldsymbol{\alpha},\boldsymbol{\alpha})=(\mathbf{S}\mathbf{U}^*\boldsymbol{\alpha},\mathbf{U}^*\boldsymbol{\alpha})=\sum_j\sum_i s_i^{(j)}(\mathbf{U}^*\boldsymbol{\alpha})_{i,j}^2 \ge \min_j s_1^{(j)}\|\boldsymbol{\alpha}\|^2$, and the rest of the proof is identical.■

Identically as for uncorrelated models case we can also show that the following inequality holds: $\frac{V(\overline{\boldsymbol{X}}_{\mathbf{opt}})}{n} \le \min\left(\min_j s_n^{(j)}, \frac{\max_j s_n^{(j)}}{m}\right)$, saying that msqe is bounded also by the maximal eigenvalue obtained from the best single model.

Now let us consider the case where the average is taken as the representative of the ensemble, which means that we have: $\boldsymbol{\alpha}'=\boldsymbol{\alpha}_1=\ldots=\boldsymbol{\alpha}_m=[\frac{1}{m},\ldots,\frac{1}{m}]^T$. Let us assume that the matrix $\mathbf{K}$ has eigenvalues $s_1 \le \ldots \le s_N$, where $N=mn$. Then we have: $\frac{s_1}{m} \le \frac{V(\overline{\boldsymbol{X}}_m)}{n} \le \frac{s_N}{m}$, where $\overline{\boldsymbol{X}}_m$ is the ensemble mean.

Proof. Let us put $\boldsymbol{\alpha}''=[\boldsymbol{\alpha}',\ldots,\boldsymbol{\alpha}']$, so $\|\boldsymbol{\alpha}''\|^2=\frac{n}{m}$. Then we get:

$$V(\overline{\boldsymbol{X}}_m)=(\mathbf{K}\boldsymbol{\alpha}'',\boldsymbol{\alpha}'')=(\mathbf{S}\mathbf{U}^*\boldsymbol{\alpha}'',\mathbf{U}^*\boldsymbol{\alpha}'')=\sum_{i=1}^{N} s_i(\mathbf{U}^*\boldsymbol{\alpha}'')_i^2 \begin{cases} \le s_N n/m \\ \ge s_1 n/m \end{cases}.$$

ACPD
9, 14263–14314, 2009

From the above inequality the following implication follows: If $\frac{\max \sigma(\mathbf{K})}{\min \sigma(\mathbf{K})} \leq m$, then $\frac{V(\overline{X}_m)}{n} \leq \min \sigma(\mathbf{K})$, where $\sigma(\mathbf{K})$ is the spectrum of the matrix $\mathbf{K}$ defined by (A10). In such a way we have obtained analogous results as for uncorrelated models case.

Acknowledgements. H. "Chip" Levy II (GFDL/NOAA) is thanked for the interesting discussion had with SG on ensembles and political consensus.

References

Campolongo, F., Cariboni, J., and Saltelli, A.: An effective screening design for sensitivity analysis of large models, Environ. Model. Softw. 22(10), 1509–1518, 2007.

Dabbert W. F. and Miller E.: Uncertainty, ensembles and air quality dispersion modelling: applications and challenges, Atmos. Environ., 34, 4667–4673, 2000.

Delle Monache L. and Stull R. B.: An ensemble air-quality forecast over western Europe during an ozone episode, Atmos. Environ., 37(25), 3469–3474, 2003.

Delle Monache, L., Deng, X., Zhou Y., and Stull R.: Ozone ensemble forecasts: 1. A new ensemble design, J. Geophysi. Res., 111(D5), D05307,10.1029/2005JD006310, 2006a.

Delle Monache, L., Nipen, T., Deng, X., Zhou, Y., and Stull, R.: Ozone ensemble forecasts: 2. A Kalman filter predictor bias correction, J. Geophys. Res., 111, D05308, doi:10.1029/2005JD006311, 2006b.

Delle Monache, L., Wilczak, J., Mckeen, S., Grell, G., Pagowski, M., Peckham, S., Stull, R., McHenry, J., and McQueen, J.: A Kalman-filter bias correction method applied to deterministic, ensemble averaged and probabilistic forecasts of surface ozone, Tellus B, 60B(2), 238–249, 2008.

Edwards, A. W. F.: Likelihood, 2nd edn., The Johns Hopkins University Press, USA, 275 pp., 1992

Draxler, R. R.: Evaluation of an ensemble dispersion calculation, J. Appl. Meteorol., 42, 308–317, 2001.

Feller W.: Introduction to probability theory and its application, vol. 1, 3rd edn., vol. 2, 2nd edn., Wiley, 509 pp., 704 pp., 1968.

Gandin, L. S.: Objective analysis of meteorological field, Gidro-meteorologicheskoe Izdate'stvo, Leningrad, USSR, 286 pp., 1963.

Galmarini S., Bianconi R., Bellasio R., and Graziani G.: Forecasting consequences of accidental releases from ensemble dispersion modelling, J. Environ. Radioactiv., 57, 203–219, 2001.

Galmarini S., Bianconi, R., Klug, W., Mikkelsen, T., Addis, R., Andronopoulos, S., Astrup, P., Baklanov, A., Bartniki, J., Bartzis, J. C., Bellasio, R., Bompay, F., Buckley, R., Bouzom, M., Champion, H., D'Amours, R., Davakis, E., Eleveld, H., Geertsema, G. T., Glaab, H., Kollax, M., Ilvonen, M., Manning, A., Pechinger, U., Persson, C., Polreich, E., Potemski, S., Prodanova, M., Saltbones, J., Slaper, H., Sofiev, M. A., Syrakov, D., Sørensen, J. H., Van der Auwera, L., Valkama, I., and Zelazny, R.: Ensemble dispersion forecasting, Part 1: Concept, Approach and indicators, Atmos. Environ., 38(28), 4607–4617, 2004a.

Galmarini S., Bianconi., R., Addis, R., Andronopoulos, S., Astrup, P., Bartzis, J. C., Bellasio, R., Buckley, R., Champion, H., Chino, M., D'Amours, R., Davakis, E., Eleveld, H., Glaab, H., Manning, A., Mikkelsen, T., Pechinger, U., Polreich, E., Prodanova, M., Slaper, H., Syrakov, D., Terada, H., and Van der Auwera, L.: Ensemble dispersion forecasting, Part II: Application and evaluation, Atmos. Environ., 38(28), 4619–4632, 2004b.

Galmarini S., Bianconi, R., de Vries, G., and Bellasio, R.: Real-time monitorino data for real-time multi-model validation: coupling ENSEMBLE and EURDEP, J. Environ. Radioactiv., 99, 1233–1241, 2008.

Girardi, F., Graziani, G., van Veltzen, D., Galmarini, S., Mosca, S., Bianconi R., Bellasio, R., Klug, W.: The ETEX project, Office for official publication of the European Communities, Luxembourg, EUR Report 181–43 EN, 106 pp., 1998.

Hanson K. M. and Hemez, F. M.: Sensitivity Analysis of Model Output Proceedings of the 4th International Conference on Sensitivity Analysis of Model Output (SAMO 2004), Santa Fe, New Mexico, 8–11 March, 519 pp. 2004.

Kalnay, E.: Atmospheric Modeling, Data Assimilation and Predictability, Cambridge University Press, Cambridge, UK, 341 pp., 2004.

Killip, C. A., Brooke, A. S., Best, P. R., Application of various dispersion models to management of local and regional air quality in Australia, Guideline on Air Quality Models: The Path Forward, pp. 355-371, 2003.

Krishnamurti, T. N., Kishtawal, C.M., Zhang, Z., LaRow, T., Bachiochi, D., Williford, E., Gadgil, S., and Surendran, S.: Multimodel ensemble forecasts for weather and seasonal climate, J.

Climate, 13(23), 4196–4216, 2000.
Lehman, E.L. and Romano J. P.: Testing Statistical Hypotheses, 3rd edn., Springer, New York, usa, 768 pp., 2008
Lorenc A. C.: Analysis methods for NWP, QJRMS, 112, 1177–1194, 1986.
Mallet, V. and Sportisse, B.: Ensemble-based air quality forecasts: A multimodel approach applied to ozone, J. Geophys. Res., 111, D18302, doi:10.1029/2005JD006675, 2006.
McKeen S., Wilczak J., Grell G., Djalalova I., Peckham S., Hsie E.-Y., Gong W., BouchetV., Menard S., Moffet R., McHenry J., McQueen J., Tang Y., Carmichael G. R., Pagowski M., Chan A., Dye T., Frost G., Lee P., and Mathur R.: Assessment of an ensemble of seven real-time ozone forecasts over eastern North America during the summer of 2004, J. Geophys. Res.-Atmos., 110(21), 1–16, 2005.
Mutemi, J. N., Ogallo, L. A., Krishnamurti, T. N., Mishra, A. K., and Vijaya Kumar, T. S. V.: Multimodel based superensemble forecasts for short and medium range NWP over various regions of Africa, Meteorol. Atmos. Phys., 95(1–2), 87–113, 2007.
Pagowski M., Grell G. A., McKeen S. A., Devenyi D., Wilczak J. M., Bouchet V., Gong W., McHenry J., Peckham S., McQueen J., Moffet R., and Tang, Y.: A simple method to improve ensemble-based ozone forecasts, Geophys. Res. Lett., 32, L07814, doi:10.1029/2004GL022305, 2005.
Parrish, D. F. and Derber, J. D.: The National Meteorological Center spectral statistical interpolation analysis system, Mon. Weather. Rev., 120, 1747–1763, 1992.
Riccio, A., Giunta, G., and Galmarini, S.: Seeking for the rational basis of the Median Model: the optimal combination of multi-model ensemble results, Atmos. Chem. Phys., 7, 6085–6098, 2007,
http://www.atmos-chem-phys.net/7/6085/2007/.
Potempski S., Galmarini, S., Addis, R., Astrup, P., Bader, S., Bellasio, R., Bianconi, R., Bonnardot, F., Buckley, R., D'Amours, R., van Dijk, A., Geertsema, G., Jones, A., Kaufmann, P., Pechinger, U., Persson, C., Polreich, E., Prodanova, M., Robertson, L., Sørensen, J., and Syrakov, D.: Multi-model ensemble analysis of the ETEX-2 experiment, Atmos. Environ., 42, 7250–7265, 2008.
Potempski S., S. Galmarini, A. Riccio, G. Giunta, On the optimal combination of multi-model ensemble predictions and its portability, Atmos. Environ., submitted, 2009.
Rodgers J. L. and Nicewander W. A.: Thirteen Ways to Look at the Correlation Coefficient, The American Statistician, 42(1), 59–66, 1988.

ACPD
9, 14263–14314, 2009

Sasaki, Y.: Proposed inclusion of time variation terms, observational and theoretical, in numerical variational objective analysis, J. Meteorol. Soc. Jpn., 47, 115–124, 1969.
Saltelli, A.: Sensitivity analysis for importance assessment, Risk Anal., 22(3), 579–590, 2002.
Saltelli, A., Ratto, M., Tarantola, S., and Campolongo, F.: Sensitivity analysis practices: Strategies for model-based inference, Reliab, Eng. Syst. Saf., 91(10–11), 1109–1125, 2006.
Stohl, A., Forster, C., Frank, A., Seibert, P., and Wotawa, G.: Technical note: The Lagrangian particle dispersion model FLEXPART version 6.2, Atmos. Chem. Phys., 5, 2461–2474, 2005, http://www.atmos-chem-phys.net/5/2461/2005/.
Strang G.: Introduction to Linear Algebra, 4th edn., Wellesley-Cambridge Press, Wellesley MA, USA, 568 pp., 2003
Stull, R. B., Hacker, J., and Modzelewski, H.: Ensemble prediction of air-pollutant transport. Measurements and Modelling in Environmental Pollution in: Measurements and modelling in environmental pollution, edited by: San José, R. and Brebbia, C. A., Computational Mechanics Publications, Southampton, UK, ISBN 1-85312-461-3, 161–167, 1997.
Talagrand O.: Assimilation of observations, an introduction, J. Meteorol. Soc. Jpn., 75(1B), 191–209, 1997.
Van Loon, M., Vautard, R., Schaap, M., Bergström, R., Bessagnet, B., Brandt, J., Builtjes, P. J. H., Christensen, J. H., Cuvelier, C., Graff, A., Jonson, J. E., Krol, M., Langner, J., Roberts, L. Rouil, R., Stern, L., Tarrasón, P., Thunis, E., Vignati, L., and White, P.: Wind, Evaluation of long-term ozone simulations from seven regional air quality models and their ensemble, Atmos. Environ., 41, 2083–2097, 2007.
Vautard, R., Schaap, M., Bergstroöm, R., Bessagnet, B., Brandt, J., Builtjes, P. J. H., Christensen, J. H., Cuvelier, C., Foltescu, V., Graff, A., Kerschbaumer, A., Krol, M., Roberts, P., Rouïl, L., Stern, R., Tarrason, L., Thunis, P., Vignati, E., and Wind, P.: Skill and uncertainty of a regional air quality model ensemble, in press, doi:10.1016/j.atmosenv.2008.09.083, 2008.
Vijaya Kumar, T. S. V., Krishnamurti, T. N., Fiorino, M., and Nagata, M.: Multimodel superensemble forecasting of tropical cyclones in the Pacific, Mon. Weather Rev., 131(3), 574–583, 2003.
Wang, B., Lee, J.-Y., Kang, I.-S., Shukla, J., Park, C.-K., Kumar, A., Schemm, J., Cocke, S., Kug, J.-S., Luo, J.-J., Zhou, T., Wang, B., Fu, X., Yun, W.-T., Alves, O., Jin, E. K., Kinter, J., Kirtman, B., Krishnamurti, T., Lau, N. C., Lau, W., Liu, P., Pegion, P., Rosati, T., Schubert, S., Stern, W., Suarez, M., and Yamagata, T.: Advance and prospectus of seasonal prediction: assessment of the APCC/CliPAS 14-model ensemble retrospective seasonal prediction

(1980–2004), Clim. Dynam., 1–25, 2008.
Wilczak, J., McKeen S., Djalalova I., Grell G., Peckham S., Gong W., Bouchet V., Moffet R., McHenry J., McQueen J., Lee P., Tang Y., and Carmichael, G. R.: Bias-corrected ensemble and probabilistic forecasts of surface ozone over eastern North America during the summer of 2004, J. Geophys. Res., 111, D23S28, doi:10.1029/2006JD007598, 2006.
Williford, C. E., Krishnamurti, T. N., Torres, R. C., Cocke, S., Christidis, Z., and Kumar, T. S. V.: Real-time multimodel superensemble forecasts of Atlantic tropical systems in 1999, Mon. Weather Rev., Part 2, 131(8), 1878–1894, 2003.
Yun, W. T., Stefanova, L., Mitra, A. K., and Krishnamurti, T. N.: Multi-model synthetic superensemble prediction system, Bull. of Amer. Met. Soc., 4119–4122, 2004.
Ziehmann, C.: Comparison of a single model EPS with a multi-model ensemble consisting of a few operational models, Tellus A, 52, 280–299, 2000.
Zupanski, D., Denning, A. S., Uliasz, M., Zupanski, M., Schuh, A. E., Rayner, P. J., Peters, W., and Corbin, K. D.: Carbon flux bias estimation employing Maximum Likelihood Ensemble Filter (MLEF), J. Geophys. Res.-Atmos., 112, D17107, doi:10.1029/2006JD008371, 2007.

Table 1. Formulas in multi-dimensional case.

	Uncorrelated case	Correlated case
Optimal weights	$\boldsymbol{\alpha}_k = \mathbf{C}_k^{-1} \left(\sum_{j=1}^{m} \mathbf{C}_j^{-1} \right)^{-1} \boldsymbol{I}$ where $\boldsymbol{I}=[1, \ldots, 1]^T$, $k=1,\ldots,m$	$\boldsymbol{\alpha} = \mathbf{K}^{-1}\mathbf{I}_R(\mathbf{I}_R^T\mathbf{K}^{-1}\mathbf{I}_R)^{-1}\boldsymbol{I}$ where $\mathbf{I}_R=[\mathbf{I},\ldots,\mathbf{I}]^T$, $\boldsymbol{I}=[1,\ldots,1]^T$
Optimal covariance	$\mathbf{C} = \left(\sum_{j=1}^{m} \mathbf{C}_j^{-1} \right)^{-1}$	$\mathbf{C} = \left(\mathbf{I}_R^T\mathbf{K}^{-1}\mathbf{I}_R\right)^{-1} = \left(\sum_{k=1}^{m}\sum_{j=1}^{m} \mathbf{C}_{kj}^{(-1)} \right)^{-1}$ where $\mathbf{C}_{ij}^{(-1)}$ are sub-matrices of $\mathbf{K}^{-1} = \begin{bmatrix} \mathbf{C}_{11}^{(-1)} \ldots\ldots \mathbf{C}_{1m}^{(-1)} \\ \ldots\ldots\ldots\ldots \\ \mathbf{C}_{m1}^{(-1)} \ldots\ldots \mathbf{C}_{mm}^{(-1)} \end{bmatrix}$
Optimal $V(\overline{X})$	$V(\overline{X}_{opt}) = (\mathbf{C}\boldsymbol{I}, \boldsymbol{I}) = \left(\left(\sum_{j=1}^{m} \mathbf{C}_j^{-1} \right)^{-1} \boldsymbol{I}, \boldsymbol{I} \right)$	$V(\overline{X}_{opt}) = (\mathbf{C}\boldsymbol{I}, \boldsymbol{I}) = ((\mathbf{I}_R^T\mathbf{K}^{-1}\mathbf{I}_R)^{-1}\boldsymbol{I}, \boldsymbol{I})$

ACPD
9, 14263–14314, 2009

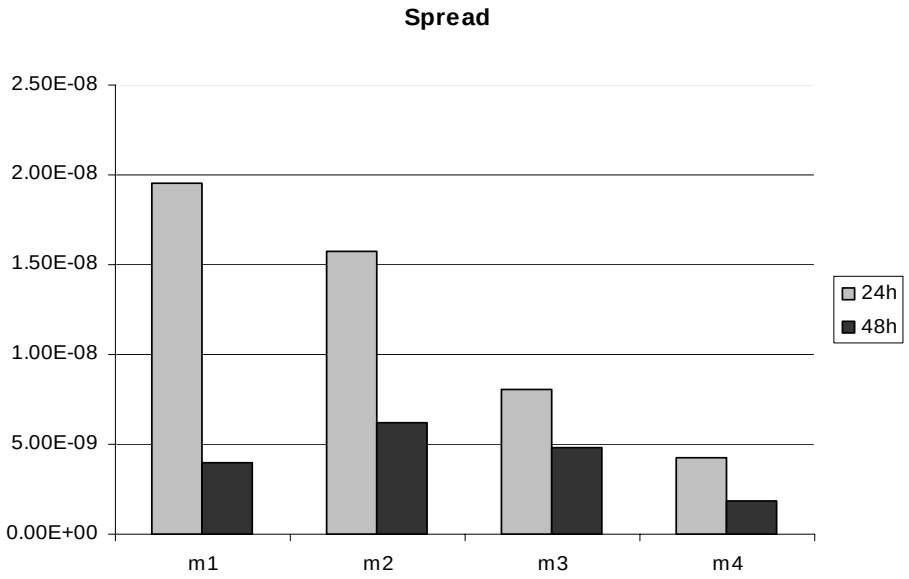

Fig. 1. ETEX-1 case: spread for 4 different models using the same meteo data (ECMWF-EPS).

ACPD
9, 14263–14314, 2009

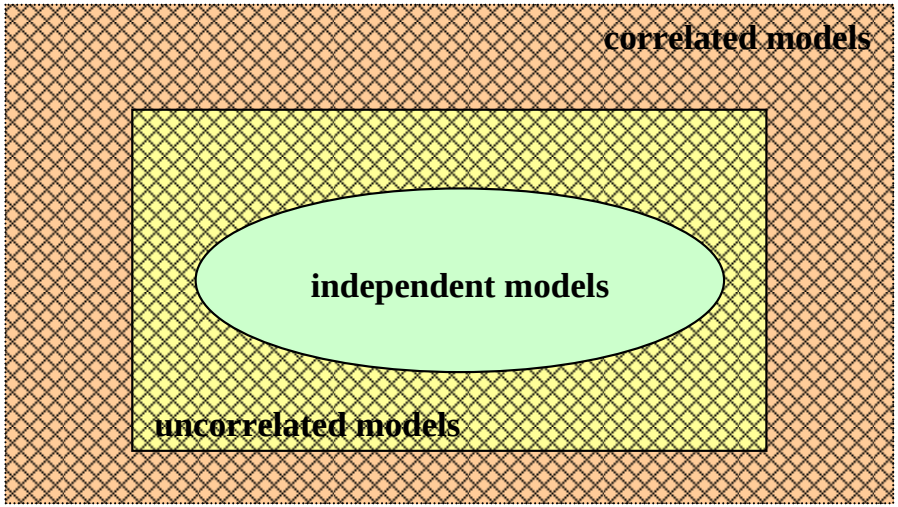

Fig. 2. Classes of models.

ACPD
9, 14263–14314, 2009

Fig. 3. "The parable of the blind leading the blind", by Peter Bruegle The Elder, (1568), courtesy of the Museo e Gallerie Nazionali di Capodimonte, Naples.

ACPD

9, 14263–14314, 2009