

[Interactive
Comment](#)

Interactive comment on “De praeceptis ferendis: good practice in multi-model ensembles” by I. Kioutsioukis and S. Galmarini

Anonymous Referee #3

Received and published: 14 July 2014

1 Summary

This article presents a detailed analysis of statistical properties of ensembles, their decompositions and demonstrates methods for constructing optimal ensembles in forecast/hindcast purposes modes.

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)

[Discussion Paper](#)

2 Verdict

The article appears to be an important contribution to an area where ensemble methods are becoming increasingly important. The article is well-written and I believe it should be published. However there are a small number of major issues, alongside a large number of minor points on which the manuscript could be improved. If these points are met, then I can recommend the article for publication. I congratulate the authors on their efforts - it is a good piece of work, and I am glad I had the chance to review it.

3 General, major points

- While I enjoyed reading the manuscript, I think it was unnecessarily long. The authors are encouraged to lay out a clear "take-home" message for the whole paper, and the individual sections. The connections between successive sections should be clear, and so should their importance to the main message of the article. Once this is done, it may be obvious that certain sections can be abbreviated, merged or removed. This may improve the overall readability of the manuscript.
- Related to the above point, the overall goal of the exercise should be emphasised, both in the introduction and the conclusions. Is it to obtain better predictions at non-observed sites, better predictions at future times, or a better historical reanalysis?
- The manuscript includes general findings (mathematically derived), which hold for any ensemble data-set, as well as findings specific to particular data-sets. I suggest clarifying the extent to which the findings from the particular data-sets

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



- examined are general properties of geophysical ensemble modelling, rather than specific to these data-sets.
- The authors need to argue the case that the AQMEII data-set is an ideal data-set for use in this context. The data-set covers one year only, so important cyclical (i.e. annual) features cannot be examined. It may have been preferable to use a longer time-series (e.g. an ensemble of multi-year climate simulations). The reasoning would certainly be stronger if additional data-sets were examined alongside the AQMEII data-set, or if there were extra references to similar analyses for longer time-series (e.g. as arising in the climate literature).
 - If I understand correctly, the mme< and mmW require observations to choose optimal weights/members. This means that the data must be divided somehow into a "training set" and a "test set". In the manuscript, it was often unclear how the data-sets were partitioned between testing and training. Also, for a fair comparison, the same test set should be used for all methods, including the mme, even though this does not involve a training set.
 - Please clarify where bias-correction has been applied, and where it has not. Is it applied everywhere? If so, how? For each individual time-series?

4 General, minor points

- I recommend the use of upper case acronyms (e.g. MME instead of mme, ID instead of id, KZ instead of kz, I.I.D. instead of i.i.d.). Upper and lower-case acronyms are used inconsistently (e.g. MSE and RMSE were generally capitalised in the manuscript).
- I recommend finding another name for the mme< ensemble. The < in mme< can be misleading, because the term mme is also used, and since < has a well-

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

understood meaning. See pp. 15825, line 25 for an example of how the term $mme<$ can appear confusing - its placement alongside / makes it look like a typing error. Perhaps MES for "mean of the ensemble subset" could be used instead of $mme<$.

- In your multi-plot figures, if the range of the x- or y-axes differ between the different panels (e.g. in Figure 1), make a note in the caption that the axes differ.

5 Specific, major points

- pp. 15811, l. 14-16: There is an important statement missing here, and in the rest of the article, as far as I can see: mme is a special case of $mm<$ (since it uses the full subset), and $mm<$ is a special case of mmW (since in $mm<$ some weights are zero and the others are equal and sum to 1).
- pp. 15812, l. 5-6: If the optimal weights can be negative, then the ensemble estimator is no longer bound by the ensemble. In the case of the AQMEII data-sets, the ensemble data-set used pertains to ozone concentrations. In theory, negative weights could result in negative concentration estimates. Did this occur? And if so, how was this dealt with? Were they truncated at zero or was the negative value simply used in the analyses?
- pp. 15829, l. 11-12: "On the other hand, static weights outscore all other products". I think this is one of the most interesting findings of the paper. I think it deserves emphasis elsewhere in the manuscript. Table 3 shows that the mmW gives by far the best results of the F(s) column. This is a very important finding. I would like to know if this is a common feature, or specific to the data-sets considered.

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

6 Specific, minor points

As a native English speaker, I have noted a range of minor grammatical or spelling errors. These are noted, for example, as:

- pp. 15820, l. 28: "the the" -> "the"

However in cases where the authors have used a phrase that I think sounds odd, it is noted as a suggestion, for example:

- pp. 15805, l. 26: "object of" -> "subject to"

So here are the comments:

- pp. 15804, l. 7: suggestion: cut "for which one cannot be gained without expense of the other"
- pp. 15804, l. 10: the -> these
- pp. 15805, l. 2: suggestion: "The availability of computing means in recent ..." -> "The availability of increasingly powerful computing in recent ..."
- pp. 15805, l. 3: suggestion: "application" -> "feasibility and use"
- pp. 15805, l. 11: what is meant by "mathematical bibliography"?
- pp. 15805, l. 14: suggestion: "driven" by the initial conditions uncertainty" -> "driven by uncertainty in the initial conditions"
- pp. 15805, l. 19: "condition" -> "conditions"

C4810

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



- pp. 15805, l. 23: suggestion: "the one of" -> "the error of"
- pp. 15805, l. 26: "object of" -> "subject to"
- pp. 15806, l. 1: suggestion: "risky" -> "less reliable"
- pp. 15806, l. 14: suggestion: "independent members." -> "independent members only."
- pp. 15806, l. 15-16: suggestion: add to this sentence a note that this will be demonstrated later in the article. Otherwise, the reader may want a reference.
- pp. 15806, l. 17: IID around observations - do you simply mean unbiased?
- pp. 15806, l. 17-19: I suggest justifying this statement
- pp. 15806, l. 17: "this property could not be" -> "this property can not be"
- pp. 15806, l. 22: suggestion: delete "to exploit ways"
- pp. 15808, l. 14: one of the termse between = signs is repeated and hence unnecessary
- pp. 15809, l. 7-8: I am not convinced by the statement that "the covariance term indicates the diversity or disparity". Covariance is a joint metric of variance and correlation. Correlation is a better measure of diversity/disparity than covariance, per se.
- pp. 15809, l. 12: "as little as possible" -> "as low as possible"
- pp. 15809, l. 10-12: This claim is not self-evident. Even though two of the terms are necessarily positive, it does not follow that one should focus only on the term which may be negative. It depends on the scale of the individual terms.

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

- If the necessarily positive terms dominate, then error-minimisation may be more effective by focussing on these terms.
- pp. 15810, l. 1: suggestion: "we have no criterion" -> "we have no a priori criterion"
 - pp. 15810, l. 1: It is claimed that "we have no criterion for identifying the best individual", there are the observations and a wealth of literature on verification - or is something else meant here?
 - pp. 15810, l. 14: suggestion: "we need to" -> "it is necessary to"
 - pp. 15811, l. 1: Please define omega
 - pp. 15811, l. 10: "presented decompositions" -> "decompositions presented"
 - pp. 15811, l. 24: "the models are assumed as random variables (i.e. their distribution is identical)". The statement in parentheses does not follow.
 - pp. 15812, l. 9: "as more models as possible" -> "as many models as possible"
 - pp. 15812, l. 11: suggestion: "it provided" -> "it provides"
 - pp. 15812, l. 11: "At the same time, it provided" - what does "it" refer to? This section? This method? The arithmetic mean? Something else?
 - pp. 15812, l. 15: suggestion: "through ..., points" -> "points, through ..., "
 - pp. 15812, l. 18: suggestion: "through ..., relies" -> "relies, through ..., "
 - pp. 15812, l. 20: suggestion: "through ..., provides" -> "provides, through ..., "
 - pp. 15813, l. 5: Suggestion: add line "Note that 2 is a general case of 1, and 3 is a general case of 2.", as noted above.

- pp. 15813, l. 18: The factorial term should not have a horizontal line separating the M and the k, otherwise this may be read as a fraction.
- pp. 15813, l. 21: suggestion: "the optimal weights do not deviate" -> "the optimal weights show little deviation"
- pp. 15814, l. 1: "two-third" -> "two-thirds"
- pp. 15814, l. 5: suggestion: "particular" -> "notable"
- pp. 15814, l. 6-7: Rewrite "the sentence "This upper bound ... equal sign" for clarity
- pp. 15814, l. 10: "the optimal combination" - meaning with the lowest RMSE?
- pp. 15814, l. 11: What is meant by "no clue"?
- pp. 15815, l. 6: suggestion: remove "well", or replace "well" with "largely"
- pp. 15815, l. 16: "was organized which consisted in having the two communities" -> "was organized, involving the two communities"
- pp. 15815, l. 20-21: "meteorological driver, air quality model, emission" -> "meteorological drivers, air quality models, emissions"
- pp. 15815, l. 25: "JJA" -> "JJA (the period of June-July-August)"
- pp. 15815, l. 26: "thirteen models that give rise to" -> "thirteen models, which give rise to " or "thirteen models, giving rise to "
- pp. 15816, l. 15-16: "The RMSE of each possible combination obtained theoretically" - this part of the sentence was unclear and could be rewritten for clarity

- pp. 15816, l. 19: "EU4r". This acronym needs to be introduced properly. It appears later, and the reader can figure it out from the context, but I would advise explaining these terms at this point in the manuscript.
- pp. 15817, l. 5: "sub-regions" - this terms should be introduced along with the above acronym
- pp. 15817, l. 4: suggestion: "quasi constant" -> "roughly constant"
- pp. 15817, l. 7: suggestion: "This number is small" -> "This fraction is small"
- pp. 15817, l. 11: "normalization" - how?
- pp. 15817, l. 15: Model 4 has negative weights - what does this mean?
- pp. 15817, l. 18: Suggestion: "Definitely" -> "Clearly"
- pp. 15817, l. 22: "Low skill cluster" -> "A low skill cluster"
- pp. 15817, l. 22: "1, 2 and 10 that" -> "1, 2 and 10, which"
- pp. 15817, l. 24: "improved variance" - improved how? closer to the variance of the obserations?
- pp. 15817, l. 25: suggestion: "error, correlation" -> "error, and correlation"
- pp. 15817, l. 23: "light" -> "slight"
- pp. 15817, l. 23: Suggestion: "Compared with" -> "Considering"
- pp. 15817, l. 23: "participation statistics" - are these shown?
- pp. 15818, l. 2: suggestion: "to form good ensemble groups" -> "to form part of skilful ensemble groups"

- pp. 15818, l. 4: suggestion: "under proper combination scheme" -> "in the right combination"
- pp. 15818, l. 13: "will not produce the best ensemble output" - best how? as measured by the overall MSE?
- pp. 15818, l. 19: "form skilled ensemble products" -> "form skilful ensemble products"
- pp. 15818, l. 22: "independent to" -> "independent of"
- pp. 15818, l. 23: "part implying" -> "part, implying"
- pp. 15819, l. 1-3: This refers to the variance-covariance plot in Fig 2d. I would suggest scaling the covariance term somehow by the variance, so that the horizontal and vertical axes display vertical information.
- pp. 15819, l. 5-6: "error formula is becoming lower but the covariance term is deteriorated" -> "error formula falls while the covariance term deteriorates"
- pp. 15819, l. 7: suggestion: "highly correlated" -> "strongly positively correlated"
- pp. 15819, l. 8: suggestion: "bigger" -> "larger" or "greater"
- pp. 15819, l. 10: "granting" -> "leading to"
- pp. 15819, l. 10: suggestion: "attempted" -> "illustrated"
- pp. 15819, l. 13: suggestion: "conditions for being lower than the one of" -> "conditions for the MSE being lower than that of"
- pp. 15819, l. 14: "constrain" -> "constraint"

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

- pp. 15819, l. 16-17: "the ensemble, i.e. the error dependence" -> "the ensemble (i.e. the error dependence)"
- pp. 15819, l. 17: "explained variation" -> "variation explained"
- pp. 15819, l. 18: "The pairwise plot" - Is this part of figure 2? If so, which plot?
- pp. 15819, l. 25: suggestion: remove "thought", or replace with "concept", "principle", "idea" or similar
- pp. 15819, l. 28: suggestion: replace "candidate for the interquartile range" -> "estimator for commonly observed values". This is because "interquartile range" means something specific in statistics, namely the 75th percentile minus the 25th percentile.
- pp. 15820, l. 1: Figure 3 shows that the ensemble mean for the optimal combination performs much better at the extremes compared to the mean of the full ensemble. This could be noted here.
- pp. 15820, l. 10: suggestion: "behaving as an i.i.d. sample" -> "under the i.i.d. assumption"
- pp. 15820, l. 26: suggestion: "case by case" -> "for the particular data-set"
- pp. 15820, l. 28: "the the" -> "the"
- pp. 15821, l. 2: At least give a basic explanation of what is meant here. e_i and d_i are represented in table 2 – are these the same as presented here? Also, explain which of the two correlation matrices is most relevant in the context of...
- pp. 15821, l. 8: suggestion: "the above produced" -> "the method described above produced" or "the above procedure yielded"

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

- pp. 15821, l. 11: "inspection at" -> "inspection of"
- pp. 15821, l. 13: suggestion: "the visual clusters of" -> "the clusters visible in"
- pp. 15821, l. 14: "plor" - "plot"
- pp. 15821, l. 16: suggestion: "marked" -> "noted" or "clear" or "obvious"
- pp. 15821, l. 19: "ID, DU, SY, LT" - at least give a basic explanation of these acronym. Details can be left to the references.
- pp. 15821, l. 24: "not fundamentally correct" - do you mean that it may lead to a higher error in the sub-ensemble
- pp. 15822, l. 2: "trace" - what is meant here? The sum of the diagonal of a matrix?
- pp. 15822, l. 11, 13: "uncertainty" - what is meant here? error? if so, they have different meanings.
- pp. 15822, l. 19: "in principle uncorrelated errors" - is this really to be expected? They are based on the same limited set of ensemble members.
- pp. 15823, l. 8: "diagnostic" - meaning?
- pp. 15823, l. 20: "medium-range" - what time-scales?
- pp. 15823, l. 22: "evidences" -> "evidence"
- pp. 15824, l. 10: "if models were uncorrelated" - is this so? has this been tested?
- pp. 15824, l. 17: "As the window size decreases" - do you mean "As the window size increases"?

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

- pp. 15824, l. 17: "RMSE" - since the ensemble members have been bias corrected, should this metric be the bias-corrected RMSE? If so, why not call it BCRMSE to distinguish it from the normal use of RMSE, which has non-zero bias.
- pp. 15824, l. 19-20: "inversely proportional lower variance" - clarify this statement
- pp. 15824, l. 22: "The cases with high variability, where the majority of models fail to simulate well" - note that this is a fundamental problem with high-resolution forecasting, relating to phase vs. amplitude accuracy.
- pp. 15825, l. 1: "it is variable as it" - what is "it"?
- pp. 15825, l. 6: "ratios" - what is meant here?
- pp. 15825, l. 10: "joint restrictions" - what is meant here?
- pp. 15825, l. 13: what are the values in the round braces?
- pp. 15825, l. 24: what are the training/testing sets used here?
- pp. 15825, l. 25: this illustrates why the term $mme<$ is potentially confusing
- pp. 15825, l. 28: "error minimisation through $mme<$ " - what is meant here?
- pp. 15826, l. 2: "distorted" - what is meant here?
- pp. 15826, l. 7: "variance" - measured/defined how?
- pp. 15826, l. 20: "matrix whose skeleton" - what is meant here?
- pp. 15826, l. 22: "normalized" - how?
- pp. 15827, l. 2: "conceived" -> "interpreted as meaning"

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

- pp. 15827, l. 9: "real" - do you mean non-negative? In mathematics, a distinction is made between real and imaginary/complex numbers, although I think something else is meant here.
- pp. 15827, l. 21: "coherent" - what is meant here?
- pp. 15828, l. 17: "remaining" - do you mean "subsequent"?
- pp. 15828, l. 22: "real" - do you mean non-negative or not with a zero imaginary component?
- pp. 15828, l. 27: I have never seen this symbol before (looks like a percentage sign, but with an extra o) - what does it mean?
- pp. 15829, l. 23: "phenomenological" is probably not the right word here. Perhaps "fundamentally" is better.
- pp. 15829, l. 16, 18: "diagnostic mode", "prognostic mode" - do you mean training and testing?
- pp. 15836, l. 6-7: "The learning algorithms ... of it (e.g. diversity)" - please clarify
- pp. 15836, l. 11: "can be seen as an application of flow dependent error covariance" - I disagree. This has nothing to do with flow, which is only a feature of temporally dependent 1-, 2-, and 3-dimensional models. There is a substantial literature on flow dependent error covariances in the field of data assimilation, and this is quite a distinct problem.
- pp. 15836, l. 26: suggestion: "era" -> "range"
- pp. 15836, l. 29: suggestion: "threshold point ... progress further" -> "benchmark against which all other weighting schemes should be evaluated"

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

- pp. 15837, l. 9: "extent" - what is meant here?
- pp. 15837, l. 9: "confined from" - do you mean "distinct from"?
- pp. 15842, table 1: reference for the derivation
- pp. 15843, table 2: what do the stars mean in e_m, d_m
- pp. 15844, table 3: relate "hindcast" and "forecast" to "training" and "testing"
- pp. 15847, l. 3: "Multi aspects" -> "Multiple aspects"
- pp. 15837: "sm" - do you mean single model?
- pp. 15849, bottom-right panel: what is point R? Reference?
- pp. 15850: what are "dm" and "em"?
- pp. 15850: The caption needs further work to clarify all the terms. Please review.
- pp. 15851: Would these curves appear linear if the x-axis were plotted on a logarithmic scale?

Interactive comment on Atmos. Chem. Phys. Discuss., 14, 15803, 2014.

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)