

Reply to “Comment on ‘Ensemble Averaging and the Curse of Dimensionality’”

BO CHRISTIANSEN

Danish Meteorological Institute, Copenhagen, Denmark

(Manuscript received 27 June 2018, in final form 27 August 2018)

In high dimensional spaces random vectors drawn from the same distribution have almost the same length, and independent vectors are almost always orthogonal. In [Christiansen \(2018a\)](#) I used these properties to explain the ubiquitous observation that the ensemble mean very often has a smaller error than all individual ensemble members when compared to observations. I also explained why the error of the ensemble mean often is 30% smaller than the typical error of the ensemble members. This is not only found in validation of climate models but also when dealing with ensemble weather forecasts.

Assuming the simple statistical model that the observations \mathbf{z} , and the K ensemble members \mathbf{x}_i , $i = 1, \dots, K$, are drawn from the same N -dimensional distribution ($N \gg 1$), then the properties of high dimensional spaces ensure that the ensemble mean, the observations, and an ensemble member form an isosceles right triangle with the ensemble mean at the right angle. The situation is illustrated in [Fig. 1](#). From this it follows that the error of the ensemble mean is $(1 - \sqrt{2})/\sqrt{2} \approx 29\%$ smaller than the error of the individual ensemble members. High dimensions are encountered when, for example, the error is measured as the root-mean-square over an extended spatial region. Apart from the high dimensions, the explanation also requires that the number of ensemble members is large so that the ensemble mean is located near the center.

In his comment [Rougier \(2018\)](#) claims there is a logical flaw in the argument above because the ensemble mean is not independent from the individual ensemble members. Actually, the argument does not require

that the ensemble mean is orthogonal on the individual ensemble members; it only requires that the ensemble mean is near the center. This is ensured by the law of large numbers as also discussed in [Christiansen \(2018a\)](#). Nevertheless, as the number of ensemble members increases, the contribution to the ensemble mean by each individual ensemble member vanishes. For example, considering the covariance between an ensemble member \mathbf{x}_i and the ensemble mean $\bar{\mathbf{x}} = \sum_{k=1}^K \mathbf{x}_k / K$, we get $\mathbf{x}_i \cdot \bar{\mathbf{x}} = \mathbf{x}_i \cdot \mathbf{x}_i / K + \sum_{i \neq j} \mathbf{x}_i \cdot \mathbf{x}_j / K = \|\mathbf{x}_i\|^2 / K + \sum \mathbf{x}_i \cdot \mathbf{x}_j / K \approx \sigma^2 N / K$, for large N . This term is a factor of $1/K$ smaller than the comparable terms $\|\mathbf{x}_i\|^2 \approx \sigma^2 N$. Also, the correlation between \mathbf{x}_i and $\bar{\mathbf{x}}$ becomes $\approx 1/\sqrt{K}$. Likewise, [Rougier \(2018\)](#) claims that the errors are correlated. But it is the central point in the argument that in the limit of large N we have $\|\mathbf{x}_i - \mathbf{z}\|^2 \approx \|\mathbf{x}_i\|^2 + \|\mathbf{z}\|^2 \approx 2\sigma^2 N$ as the cross-term vanishes because of the orthogonality and because the observations and the ensemble members have the same lengths. In the same way we get $\|\bar{\mathbf{x}} - \mathbf{z}\|^2 \approx N\sigma^2 / K + \sigma^2 N \approx \sigma^2 N$. In other words, it is the orthogonality that separates the joint distribution into the product of the marginal distributions.

The arguments up to this point are simple and flawless and, if in doubt, the results above can easily be confirmed with numerical experiments as they were in [Christiansen \(2018a\)](#).

[Rougier \(2018\)](#) also mentions that sometimes it is observed that some ensemble members have smaller errors than the ensemble mean and that this is in disagreement with the theory above. But the theory was derived in the limit $N \rightarrow \infty$. [Christiansen \(2018a\)](#) did include numerical analysis that showed that for smaller N the ensemble mean will often but not always be better than individual members (Figs. 3 and 5 in [Christiansen 2018a](#)). We can relax the requirement on N and still obtain analytical results. In a new paper

Corresponding author: Bo Christiansen, boc@dmi.dk

DOI: 10.1175/JCLI-D-18-0416.1

© 2018 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) (www.ametsoc.org/PUBSReuseLicenses).

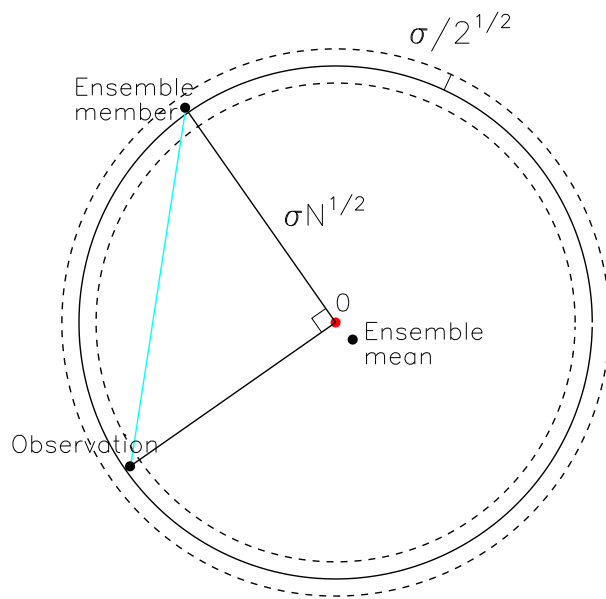


FIG. 1. Schematic diagram showing the situation in high-dimensional space when ensemble members and observations are drawn independently from an N -dimensional Gaussian distribution $\mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ with $N \gg 1$. The ensemble members and observations are located in a thin annulus with radius $\sigma\sqrt{N}$ and width (standard deviation) $\sigma/\sqrt{2}$. The ensemble mean is located close to the center $\mathbf{0}$ of the distribution (red dot). Independent vectors are almost always orthogonal. The ensemble mean, the observations, and an ensemble member therefore form an isosceles right triangle with the ensemble mean at the right angle. The distance between the observations and the ensemble member (the error of the ensemble member) is the length of the hypotenuse (cyan line). This is then a factor of $\sqrt{2}$ larger than the radius of the annulus (the error of the ensemble mean).

(Christiansen 2018b, manuscript submitted to *Mon. Wea. Rev.*) I show that for large N the fraction of ensemble members better than the ensemble mean (i.e., the normalized rank of the ensemble mean) is

$$\Phi\left(-\frac{\sqrt{N} K - 1}{2 K}\right),$$

where Φ is the cumulative probability density of the Gaussian. As mentioned in that paper, I have confirmed numerically that this is a very good approximation even for $N = 10$. Also, in good agreement with the results in Christiansen (2018a) we get the normalized ranks 1.3×10^{-6} , 0.005, and 0.07, for $N = 100$, 30, and 10.

Finally, Rougier (2018) questions the applicability of the arguments to the ensemble of climate models. It is true that there are reasons to believe that the simple statistical model above is not fully representative for

the ensemble of climate models. This model is often described as the “indistinguishable” interpretation and requires that ensemble members and observations are exchangeable. However, Christiansen (2018a) did test this based on 17 CMIP5 models by successively swapping the observations and a selected climate model. Again I found for each swap that the ensemble mean was better than almost all individual ensemble members and that the ensemble mean was around 30% better than the typical ensemble member. It is also easy to directly test the relevant properties of the ensemble. Calculating the lengths of the ensemble members we get values in the range of 390–666 K with a mean of 514 K.¹ For the observations we get 419 K and for the ensemble mean 25 K. For the angles between the ensemble members we get 1.63 ± 0.23 and for the angles between the ensemble members and observations 1.63 ± 0.18 ($\pi/2 \approx 1.57$). This demonstrates that we are very close to the situation illustrated in Fig. 1.

Of course the simple statistical model may not be correct from all perspectives and I did mention in Christiansen (2018a) that the swapping experiment may lack power as a test of the statistical model. But as the aphorism goes, “all models are wrong but some are useful,” and the explanation given in Christiansen (2018a) does succeed in explaining the ubiquitous observations of the ensemble mean in a very simple way.

In Christiansen (2018a) the simple statistical model was extended to allow for bias and different variances of observations and ensemble members. I noted that the results regarding the ensemble mean are robust and only change little for small biases and small differences between the variances. In Christiansen (2018b, manuscript submitted to *Mon. Wea. Rev.*) I analyze weather forecasts using this extended model. Because the bias and variances depend on lead time so does the behavior of the model mean. For low lead times we are in the “truth plus error” interpretation where ensemble members are sampled from a distribution centered about the observations. For large lead times we approach the “indistinguishable” interpretation, where model ensembles and observations are all considered exchangeable. Again, I find an excellent agreement with the analytical

¹ We are looking at the monthly climatology of near-surface temperature. The length is calculated in the space spanned by all 10 512 grid points and the 12 calendar months. The ensemble and the observations have been centered to their common mean. A length of 500 K corresponds to $\sigma = 1.4$ K in the simple statistical model.

results of the extended statistical model in high dimensions.

Acknowledgments. This work is supported by the NordForsk-funded Nordic Centre of Excellence project (Award 76654) Arctic Climate Predictions: Pathways to Resilient, Sustainable Societies (ARCPATH).

REFERENCES

- Christiansen, B., 2018a: Ensemble averaging and the curse of dimensionality. *J. Climate*, **31**, 1587–1596, <https://doi.org/10.1175/JCLI-D-17-0197.1>.
- , 2018b: Analysis of ensemble mean forecasts: The blessings of high dimensionality. *Mon. Wea. Rev.*, submitted.
- Rougier, J., 2018: Comment on “Ensemble averaging and the curse of dimensionality.” *J. Climate*, **31**, 9015–9016.

