I am grateful to the editors for giving me the opportunity to comment on the recent paper “Ensemble Averaging and the Curse of Dimensionality,” by Bo Christiansen (Christiansen 2018). I will limit my comments to the theoretical material in section 2.

It is intriguing that in ensembles of climate simulations, the multimodel mean (MMM) sometimes outperforms even the best ensemble member in root-mean-square error (RMSE), computed over the pixels of a spatial field; a formal statement is given below. This outperformance is present in several fields in Fig. 9.7 of chapter 9 of the Fifth Assessment Report (FAR) of the IPCC (IPCC 2013). Prompted by this figure, in Rougier (2016) I provided a set of conditions that imply this outcome. To paraphrase my interpretation of these conditions, if individual simulators are tuned more on their overall bias than their large pixel errors, then we would expect the MMM to outperform all ensemble members, when the number of pixels is large.

The challenge in the mathematics was the correlations between all of the RMSEs (more on this below). I tackled this challenge in a rather blunt way, by considering the asymptotic limit as the number of pixels grows without bound. In this case, the concentration of means around their expectations plus some additional asymptotic theory produced my key result, result 2. This made for rather a technical paper, and I was keen to read Dr. Christiansen’s geometric explanation.

Dr. Christiansen’s model is

\[ X_i \sim N_n(\mu, \sigma^2 I_n), \quad i = 1, \ldots, k \]

\[ Z \sim N_n(\mu, \sigma^2 I_n), \] (1)

where \( X_i \) is the \( n \)-dimensional field of outputs from simulator \( i \), \( Z \) is the corresponding vector of observations, and \( (X_1, \ldots, X_k, Z) \) are treated as mutually independent, and hence they are independent and identically distributed (IID). The MMM is denoted \( \bar{X} \). Being treated as known, the common expectation vector \( \mu \) can be set to zero without loss of generality; likewise, the scalar \( \sigma^2 \) can be set to 1. Dr. Christiansen, like me, considers the situation where \( n \to \infty \).

In my view (1) is not appropriate for ensembles of climate simulators [see Rougier et al. (2013) for a technical discussion]. In my analysis in Rougier (2016), I was careful not to require the assumption of zero bias for every member of the ensemble, let alone the stronger assumption that the ensemble members and the observations are jointly IID. IID is inappropriate both a priori, given what we know about how climate simulators are constructed and tuned, and a posteriori, in the light of the “genealogy” evidence of Knutti et al. (2013).

Putting these reservations aside, we need to be clear about what Dr. Christiansen must show with his geometric explanation:

Under the model in (1),

\[ \| \bar{X} - Z \| < \min\{\|X_1 - Z\|, \ldots, \|X_k - Z\|\} \]

as \( n \to \infty \) for fixed \( k \), where \( \| \cdot \| \) is Euclidean distance in \( n \) dimensions.

In this notation, \( \|X_i - Z\| \) is the RMSE of simulator \( i \). In fact I have already proved this result, which is a special case of my result 2 and its generalizations, in which all of my \( \mu_i \) (simulator biases) are set to zero, and my \( \sigma^2 = 2 \).

Now there is another difficulty with Dr. Christiansen’s model—a kind of logical trap. For his model, I have already proved that the MMM always has the smallest RMSE for sufficiently large \( n \). And yet this is not what...
we see in, say, the different output fields of Fig. 9.7 of the
FAR: in some output fields the MMM is best, whereas
in others it is little better than the median. To save his
model, Dr. Christiansen would have to argue that
in the FAR is not large enough to enforce convergence.
But if n is not large enough, then an explanation based
on large n is vacuous. This issue does not arise in my
model because each output field can have a different
configuration of biases across the simulators, and the
performance of the MMM depends on the configuration
of biases.

As I have already remarked, the challenge in the
mathematics is that the $k + 1$ RMSEs are all correlated,
because of the common term $Z$, and because $X$ involves
$X_1, \ldots, X_k$. Thus the joint distribution of all $k + 1$
RMSEs is very complicated, and the distribution of the
rank of the MMM RMSE among the ensemble RMSEs
even more so, the rank being a highly nonlinear func-
tion. Unfortunately, however, Dr. Christiansen’s geo-
metric explanation is oblivious to the presence and
effect of correlation, which indicates that he has made
an error.

The nature of Dr. Christiansen’s explanation is to
treat each of the terms $Z, X_1, \ldots, X_k, \text{and } X$ sepa-
ately, and then to conclude from their individual behaviors that the required ordering of the RMSEs
follows. This not a valid procedure. It violates the cal-
culus of probability, which asserts that the distribution
of a specific function of a vector of random variables
cannot, in general, be deduced from the marginal dis-
tributions alone.

Dr. Christiansen’s treatment of $X$ is particularly trou-
bling. He requires $X$ to be effectively zero, and he
produces a marginal argument to this effect without ref-
ence to the underlying ensemble values $(X_1, \ldots, X_k)$. But $X$ lies precisely at the centroid of the ensemble: it
goes where the ensemble is, and to detach it from the
ensemble values when comparing the RSME of $X$ with
the set of ensemble RSMEs is invalid.

Also in section 2, there is the striking claim on p. 1590
that Dr. Christiansen’s model and his geometric ex-
planation can resolve a long-standing regularity, which
is that the RMSE of the MMM is often about 30%
smaller than the median of the RMSEs of the $k$
ensemble members. Unfortunately Dr. Christiansen’s
explanation again relies on $X$ being effectively zero,
regardless of the values of $(X_1, \ldots, X_k)$. It also suffers
from the same logical trap identified above: empiri-
cally, the ratio of RMSEs varies substantially from one
output field to another, and yet Dr. Christiansen’s model
and argument asserts that the MMM RSME is always
about 30% smaller than the ensemble median, for suffi-
ciently large $n$.

In summary, section 2 of Dr. Christiansen’s paper aims
to give a geometric explanation of the outperformance of
the MMM. Unfortunately his explanation is not valid,
owing to its violation of the probability calculus. I also
maintain that Dr. Christiansen’s model is too restrictive
for the current ensemble of climate simulators, and its
strong conclusions are refuted empirically.

Acknowledgments. This research was supported by the
EPSRC SuSTaIn Grant EP/D063485/1.

REFERENCES
Christiansen, B., 2018: Ensemble averaging and the curse of di-
JCLI-D-17-0197.1.
Cambridge University Press, 1535 pp., https://doi.org/10.1017/
CBO9781107415324.
Knutti, R., D. Masson, and A. Gettelman, 2013: Climate model
genealogy: Generation CMIP5 and how we got there. Geophys.
Rougier, J., 2016: Ensemble averaging and mean squared
JCLI-D-16-0012.1.
——, M. Goldstein, and L. House, 2013: Second-order exchange-
ability analysis for multimodel ensembles. J. Amer. Stat. Assoc.,