Dear Dr. Neale

I have carefully reviewed their response and am deeply disappointed. Not only did their reply fail to address the key points, but more importantly, the authors appear to be responding to non-existent arguments (as noted in my comments below). I believe this study is methodologically flawed and conceptually confusing. Therefore, I am convinced that this work should not be published.

Below is my response to their replies. The red text represents their original comments,the bold black text is my previous feedback, and the blue text contains my current responses.

Reply: Our latest responses are provided in green (earlier replies in red). We respectfully disagree with the comment referring to our argument as "non-existent." Unfortunately, some of the reviewer's remarks appear to be misleading, and it seems that the reviewer has been inclined toward rejecting the manuscript from the outset, which we believe is unfair.

Comment-1: The article only explains how signal and noise variance are defined and calculated. Since variance itself is not the actual component, it is unclear how the signal and noise are extracted from the data. The concept and defintion are totally different between the variance and the variable itself.

Reply: There are numerous paper on how signal and noise components are extracted from model data and some of them are cited here (e.g. Kang and Shukla, 2006; Scaife et al., 2014; Saha et al., 2016a; Scaife and Smith, 2018; Weisheimer et al., 2018 and many more). While inter-ensemble spread is considered as noise/internal component, the ensemble mean is the signal/external component (equation 1 and 2 respectively in our manuscript). How signal and noise are extracted from data is clearly mentioned in lines 104-108 of the manuscript, section 2.3.1.

Comment: The articles cited by the authors only discuss signal variance and noise variance. It is problematic to treat the ensemble mean directly as the signal/external component. As a measure of variability, the signal should not be constrained by sign—how does one interpret a "positive signal" versus a "negative signal"? Therefore, it is more appropriate to use the square of the ensemble mean to represent the signal.

In the author's statements in lines 104–109 as below, I could not find a clear definition of either the signal or the noise.

respectively. Here, predictable and unpredictable components are termed external/signal and internal/noise components, respectively. The ratio of external to internal variance is known as the signal-to-noise ratio (SNR). If x is the precipitation field of the model, i is the year of the model integration (total year 'N'), and j is the number of ensemble simulations (total ensemble n = 52), then internal variance following Rowell et al. (1995), can be expressed as

Reply: The ensemble mean is always treated as the more reliable predictor; it is not a matter of a "positive" versus "negative" signal. This interpretation is incorrect. In our analysis, the *signal* (or external variance) represents the seasonal anomaly in the ensemble mean, with a correction term as defined in Equation (3). Our approach follows the methodology established by *Rowell et al.* (1995) and subsequently applied in numerous studies (e.g., Kang & Shukla, 2006; Scaife et al., 2014; Saha et al., 2016a; Scaife & Smith, 2018; Weisheimer et al., 2018). To avoid confusion, we have included excerpts below, from several of these references. Regarding the suggestion to use the "square of the ensemble mean" to represent the signal, we are not aware of any such approach in the existing literature. We have followed the standard and widely accepted method for estimating signal and noise components, as described in our manuscript (beginning at line 104).

We regret that the reviewer was unable to locate the definitions of *signal* and *noise* in the manuscript, which were clearly stated and also mentioned in our earlier response. Here is the cut-pest of our manuscript, describing signal and noise (beginning line 104).

2.3.1 ANalysis Of VAriance (ANOVA) method

In this method, the total variance is split into signal and noise components, i.e. external (σ^2_{EXV}) and internal (σ^2_{IV}) variances, respectively. Here, predictable and unpredictable components are termed external/signal and internal/noise components, respectively. The ratio of external to internal variance is known as the signal-to-noise ratio (SNR). If x is the precipitation field of the model, i is the year of the model integration (total year 'N'), and j is the number of ensemble simulations (total ensemble n = 52), then internal variance following Rowell et al. (1995), can be expressed as

$$\sigma^2_{IV} = \frac{1}{N(n-1)} \sum_{j=1}^{n} \sum_{i=1}^{N} (x_{ij} - \overline{x_i})^2 \tag{1}$$

where $\overline{x_i} = \frac{1}{n} \sum_{j=1}^{n} x_{ij}$ is the ensemble mean of the model for a year and the degrees of freedom is N(n – 1). The variance of ensemble mean (σ^2_{EV}) can be estimated as

$$\sigma^2_{EV} = \frac{1}{(N-1)} \sum_{i=1}^{N} (\overline{x_i} - \overline{\overline{x}})^2 \tag{2}$$

where $\overline{\overline{x}} = \frac{1}{Nn} \sum_{j=1}^{n} \sum_{i=1}^{N} x_{ij}$ is the average over all year and all ensemble. However, the variance of the ensemble mean is a biased estimate of external variance (Scheffe, 1959). As the number of ensemble members is not very large (here 52), the ensemble mean contains residual internal variability. Therefore, the external variance may be estimated following Scheffe

4

Paper by Rowell et al. (1995), describe how signal/variability due to SST/external and noise/internal components, which are based on 'analysis of variance' methodology, are calculated.

The next stage of research will be to explore the physical mechanisms which link the SST patterns to seasonal rainfall variability. Circulation changes over north Africa will be examined in a later publication, and some global-scale circulation patterns associated with Sahelian rainfall anomalies are presented by Ward *et al.* (1994).

Given that SST patterns are often predictable at least a few months in advance, this offers hope for the production of skilful forecasts of seasonal JAS rainfall anomalies averaged over the Sahel, Soudan and Guinea Coast. Indeed, such forecasts have now been issued by the UK Meteorological Office for the Sahel region since 1986, and for the Soudan and Guinea Coast regions since 1992, on an experimental basis (see Ward *et al.* (1993) for details). In order that such forecasts achieve maximum utility, further research is required on the variations of rainfall–SST relationships *within* the large regions used here and *within* the July to September season.

ACKNOWLEDGEMENTS

We are very grateful for the hard work of John Owen who set up and monitored many of the GCM experiments described here, and to David Parker who was involved in the early part of this work. We are also indebted to Mike Hulme (Climatic Research Unit, Norwich), who provided us with the observed rainfall data on the same grid as the GCM (under UK DoE contract PECD 7/12/78), and to John Rowell, whose statistical advice led to a much improved appendix.

APPENDIX

Here we reproduce the statistics required to separate the estimated total variance of simulated rainfall amounts (or any other parameter) (σ^2_{TOT}), into an SST-forced component (σ^2_{SST}) and an internal variability component (σ^2_{INT}). Although in this paper the technique is specifically applied to tropical north African seasonal and monthly rainfall totals, it will also prove useful as a general analysis tool for understanding natural climate variability and potential atmospheric predictability on any chosen time- or space-scale.

We consider a generalized case of N ensembles (years of SST forcing), each with n members. The simulated rainfall amounts for each experiment are 'modelled' as the sum of two independent components:

$$x_{ij} = \mu_i + \varepsilon_{ij} \tag{A.1}$$

where: x_{ij} = simulated rainfall, $i = 1, \ldots N$ is associated with a particular year (i.e. an ensemble of experiments with the same SST forcing), $j = 1, \ldots n$ is associated with a particular member of the ensemble (i.e. identifies the initial atmospheric conditions), μ_i = component of rainfall due to SST forcing, and ε_{ij} = anomalous rainfall due to internal variability.

The approach we use to estimate the components of variance closely follows an 'analysis of variance' methodology, using a so-called 'random-effects' model (see, for example, Scheffe (1959) section 7.2). The only assumptions required are that the internal variability is the same for each year, and that the effect of altering the initial atmospheric conditions has a purely random impact on the simulated seasonal-rainfall anomalies.

First, the internal variability is easily estimated by computing the variance of each datum's deviation from its ensemble mean:

$$\hat{\sigma}^2_{\mathrm{INT}} = \frac{1}{N(n-1)} \sum_{i=1}^{N} \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2 \qquad (N(n-1) \text{ being the degrees of freedom}) \tag{A.2}$$

where $\hat{y}$ denotes an estimate of y, and:

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^{n} x_{ij} \qquad \text{(the ensemble mean for the } i\text{th year).} \tag{A.3}$$

In order to estimate the variability due to SST forcing, we must first estimate the variance of the ensemble means:

$$\hat{\sigma}^2_{\mathrm{EM}} = \frac{1}{N-1} \sum_{i=1}^{N} (\bar{x}_i - \bar{x})^2 \qquad (N-1 \text{ being the degrees of freedom}) \tag{A.4}$$

where:

$$\bar{x} = \frac{1}{nN} \sum_{i=1}^{N} \sum_{j=1}^{n} x_{ij} \qquad \text{(the mean of all data).} \tag{A.5}$$

Now, using a standard result from 'analysis of variance', we can also express σ^2_{EM} in terms of σ^2_{INT} and σ^2_{SST} (e.g. Scheffe 1959, p. 226):

$$\sigma^2_{\mathrm{EM}} = \sigma^2_{\mathrm{SST}} + \frac{1}{n} \sigma^2_{\mathrm{INT}}. \tag{A.6}$$

This essentially states that the variance of the ensemble means is a *biased* estimate of the variance due to SST forcing. This is because with finite n each ensemble mean ($\bar{x}_i$) still contains an element of internal variability (each ensemble mean is only an *estimate* of μ_i, not equal to μ_i), so that σ^2_{EM} overestimates σ^2_{SST}.

Thus, from Eqs. (A.6), (A.4) and (A.2), the variance due to SST forcing may be estimated as:

$$\hat{\sigma}^2_{\mathrm{SST}} = \hat{\sigma}^2_{\mathrm{EM}} - \frac{1}{n} \hat{\sigma}^2_{\mathrm{INT}} = \frac{1}{N-1} \sum_{i=1}^{N} (\bar{x}_i - \bar{x})^2 - \frac{1}{Nn(n-1)} \sum_{i=1}^{N} \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2 \tag{A.7}$$

[see note (i)].

This then leads to an estimate of the total variance:

$$\hat{\sigma}^2_{\mathrm{TOT}} = \hat{\sigma}^2_{\mathrm{INT}} + \hat{\sigma}^2_{\mathrm{SST}} \tag{A.8}$$

[see note (ii)],

and finally, the ratios of components of variance:

$$\frac{\hat{\sigma}^2_{\mathrm{SST}}}{\hat{\sigma}^2_{\mathrm{TOT}}} \quad \text{and} \quad \frac{\hat{\sigma}^2_{\mathrm{INT}}}{\hat{\sigma}^2_{\mathrm{TOT}}}.$$

Notes:

(i) Because of the subtracted term in Eq. (A.7), the distribution of σ^2_{SST} includes a few negative values, which are sometimes produced by chance. However, this is only likely when n and N are small and when $\sigma^2_{\mathrm{SST}}/\sigma^2_{\mathrm{TOT}}$ is small. Such negative values are best reset to zero, but in any case do not occur in the results presented here.

(ii) It is perhaps tempting to estimate σ^2_{TOT} as: $\frac{1}{Nn-1} \sum_{i=1}^{N} \sum_{j=1}^{n} (x_{ij} - \bar{x})^2$.

However, this is a *biased* estimate, because it fails to account for the impact that the make-up of the data has on the number of degrees of freedom, i.e. that the data has *two* components of variability. The total variance is correctly estimated only by the *sum* of its two components (Eq. (A.8)).

Where equation (A.2) is equation (1) in our manuscript
(A.4) is equation (2) in our manuscript
external variance/SST forced variance (A.7) is equation (3) in our manuscript.
Below is from Kang and Shukla (2006), *Dynamic Seasonal Prediction and Predictability of the Monsoon, in: The Asian Monsoon, edited by Wang, B pp. 585–612, Praxis, Springer, Berlin.*

Book Chapter by Kang and Shukla (2006)

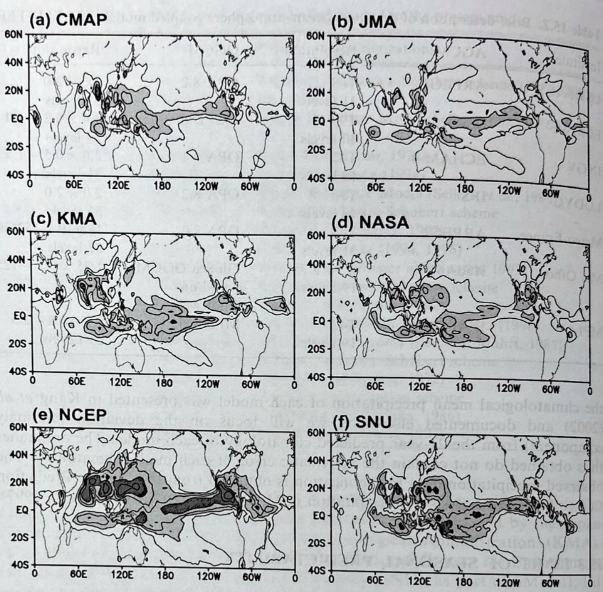

Figure 15.1. Variances of summer mean precipitation anomalies for the 21-year period (1979–1999). (a) CMAP observation precipitation, (b) JMA, (c) KMA, (d) NASA, (e) NCEP, and (f) SNU prediction models. The variance of each model is computed using all the ensemble members of the 21-year predictions. The contour interval is 1, 3, 6, 12, 24, and 48 $\mathrm{mm^2\,day^{-2}}$ and light and dark shadings indicate a variance of more than 3 and 12 $\mathrm{mm^2\,day^{-2}}$, respectively.

regions, particularly over the Asian monsoon region. The difference among the model variances is partly related to the difference in the mean climatology and to the different combinations of model physics. But, it is difficult to identify the model physics responsible for generating such large differences.

The total variance (σ^2_{TOT}) is divided into the external (σ^2_{SST}) and internal variances (σ^2_{INR}; Rowell, 1998). The ensemble mean is considered to be the external component of the prediction forced by the SST forcing, and the deviation from the ensemble mean is the stochastic internal component of the prediction. The

internal variance can then be expressed as:

$$\sigma^2_{INR} = \frac{1}{N(n-1)} \sum_{i=1}^{N} \sum_{j=1}^{n} (x_{ij} - \overline{x_i})^2 \qquad (15.1)$$

where x is the precipitation, i indicates the individual year, $N = 21$, j is the ensemble member, and n is 6 to 10 for different models. $\bar{x}_i$ is the ensemble mean. The external variance is obtained by the mean square of the deviation of each year's ensemble mean from the climatological mean and with a consideration of bias correction, as in Rowell (1998):

$$\sigma^2_{SST} = \sigma^2_{EN} - \frac{1}{n}\sigma^2_{INR} \quad \text{and} \quad \sigma^2_{EN} = \frac{1}{N-1} \sum_{i=1}^{N} (\overline{x_i} - \overline{\overline{x}})^2 \qquad (15.2)$$

where $\overline{\overline{x}}$ is the climatological mean and $\overline{\overline{x}} = 1/(Nn) \sum_{i=1}^{N} \sum_{j=1}^{n} x_{ij}$. It should be noted that the sum of external and internal variances expressed above is equal to the total variance.

Figure 15.2(a–e) show the external variances of various models, and Figure 15.2(f–j) the internal variances. The signal-to-noise ratio, the ratio of the external part to the internal part of the corresponding model, is shown in Figure 15.2(k–o). All the models produce large external variances over the tropical oceans that are much larger than the internal variance of the same model, particularly the ENSO region. This result indicates that tropical rainfall is less controlled by atmospheric internal processes and is thus less predictable for a given SST condition. In the extratropics, on the other hand, the internal variances are bigger than the external variances of the same model (Figures 15.2(k–o)), and therefore the extratropical atmosphere is more controlled by non-linear stochastic processes and less predictable

Comment 2: The article tries to discuss and analyze the paradox, but the purpose of using Nino3.4 to predict precipitation remains unclear. What is the intention behind comparing it with dynamic models? Is it to demonstrate whether the actual or potential forecast skill of dynamic models is higher or lower, reasonable or unreasonable? The objective is not clearly stated. Moreover, can using Nino3.4 to predict precipitation effectively achieve these goals? Would the forecast skill be reliable? Was the forecast skill mentioned in the article derived from training or test data? Similarly, were other modes affecting precipitation in the Indian region, such as IOD, considered?

Reply: The idea is to asses prediction skill of not only predictants (i.e. ISMR, PACR), but also the fidelity in simulating global predictors (e.g. ENSO) and their teleconnections. Figure 9 shows multiple correlations involving major global predictors (Niño3.4, IOD, PDO, AMO) and sub-seasonal components.

Comment: Your response does not address my question. Such a simple linear regression approach is unreliable and insufficient to explain any core issues discussed in this paper.

Reply: We would like to emphasize that *linear regression* is a well-established and reliable statistical tool, particularly when the relationship is found to be statistically significant. Some of the questions raised earlier (in black text above), such as *"Was the forecast skill mentioned in the article derived from training or test data?"*, are not relevant in this context and would not typically arise from a domain expert. Our analysis is based on coupled model re-forecast data; therefore, the concepts of "training" and "test" datasets do not apply here.

Comment 3: Rowell (1995) never defined signal variance and noise variance using ANOVA. While they did mention ANOVA, it was only used for statistical testing. The authors should revisit Rowell (1995) to better understand the content. ANOVA has exactly defintion in statistics, which should be followed to avoide unnecessary confusion.

Reply: Please look into page no 699 of Rowell et al. (1995).
https://rmets.onlinelibrary.wiley.com/doi/epdf/10.1002/qj.49712152311

,which mention "The approach we use to estimate the components of variance closely follows an 'analysis of variance' methodology ..."

Comment: I could not find the answers provided by the authors in Rowell (page 699) as below. It should be noted that ANOVA has a rigorous statistical definition. The authors, however, only performed variance partitioning, not ANOVA.

RAINFALL VARIABILITY OVER NORTH AFRICA 699

The next stage of research will be to explore the physical mechanisms which link the SST patterns to seasonal rainfall variability. Circulation changes over north Africa will be examined in a later publication, and some global-scale circulation patterns associated with Sahelian rainfall anomalies are presented by Ward *et al.* (1994).

Given that SST patterns are often predictable at least a few months in advance, this offers hope for the production of skilful forecasts of seasonal JAS rainfall anomalies averaged over the Sahel, Soudan and Guinea Coast. Indeed, such forecasts have now been issued by the UK Meteorological Office for the Sahel region since 1986, and for the Soudan and Guinea Coast regions since 1992, on an experimental basis (see Ward *et al.* (1993) for details). In order that such forecasts achieve maximum utility, further research is required on the variations of rainfall–SST relationships *within* the large regions used here and *within* the July to September season.

ACKNOWLEDGEMENTS

Reply: This comment appears to be misleading. The reviewer should refer to the entire page 699 of *Rowell et al.* (1995) rather than quoting an isolated paragraph, which risks misrepresenting the context. To clarify, we have now included pages 699 and 700 from *Rowell et al.* (1995) in our response to Comment 1. As clearly indicated in the highlighted section (marked with a red rectangle), the method indeed follows the "analysis of variance" (ANOVA) approach. Therefore, the argument questioning whether it is ANOVA-based is not meaningful.

Comment 4: I do not understand the meaning of the statement: "The use of the orthogonality assumption is a methodological simplification to partition variance across time scales; it does not imply the absence of physical co-variability." Do physical and mathematical co-variability have different interpretations? In my opinion, if two quantities are physically related, they cannot be assumed to be orthogonal in mathematics. Additionally, I do not comprehend the authors' claim that "sub-seasonal components are the building blocks of the seasonal mean." Following this logic, all time scales would be sources of error, since hourly components are the building blocks of the daily mean, and daily components are the building blocks of the weekly mean, and so on.

Reply: The argument why we are using assumption of orthogonality and not the actual one, lies on the fact that it is challenging (if not impossible) in a non-linear system to separate individual components.

Sub-seasonal components of the monsoon particularly have clear preferred band. Some of the band are more vigorous in terms of their spatial scale, strength than the others. In terms of their contribution to the mean and variability/predictability also varies. While MISOs have very large spatial structure and strong sub-seasonal variability, their contribution to year-to-year monsoon rainfall variability is minimum (weak negative correlation). So, clearly, we are not talking here about hourly/daily events but some known and prominent sub-seasonal variability/bands, which shape the seasonal monsoon rainfall of a year. Here are literatures, cited in support of our arguments (Saha et al., 2019; Borah et al., 2020). Some important papers in the similar lines but not cited here are.

Comment: It seems no basis to argue the "challenging to separate" as a justification for such an assumption. This is the most critical weakness of the study: on the one hand, it attempts to examine the effect of A on B using linear statistical analysis, while on the other hand, it assumes that A and B are orthogonal, implying that their covariance (or correlation coefficient) is zero.

I am drawing this inference based on the authors' own argument. You may choose to ignore or omit other scales of the atmospheric process, but I cannot overlook them. Isn't that ?

Reply: We thank for your comment. The assumption of orthogonality is a widely used and mathematically consistent methodological simplification in the analysis of complex, multiscale geophysical systems. It does not imply that the underlying physical processes are truly independent, but rather provides a tractable framework to *partition total variance* among distinct temporal (or spatial) scales. In nonlinear climate systems, exact separation of variability across scales is not possible because physical processes are dynamically coupled and often nonlinearly interacting. However, for the purpose of statistical decomposition and diagnostic analysis, an orthogonal representation allows us to quantify the *relative contributions* of different time-scale components (e.g., sub-seasonal, interannual, decadal) to the total variance, without double-counting shared variability.

This approach is conceptually similar to other well-established methods that rely on orthogonality, such as Empirical Orthogonal Function (EOF) analysis, spectral decomposition, and ANOVA, where the basis functions or components are constructed to be orthogonal in the statistical sense, even though the corresponding physical processes may interact. In this context, orthogonality is a mathematical convenience, not a physical claim of independence.

Thanks for your comment regarding “sub-seasonal components are the building blocks of the seasonal mean.”. We will add text here to make thing clearer to the reader.

Comment 5: So I have to feel sorry to decline this work again. The topic is interesting that is the reason why I agreed with reviewing it. Unfortunately I do not learn more from this work. To my understanding, the paradox should be from the "defintion" of potential predictability. The ratio of signal to noise may not well represent the potential predictability. If authors wish to work this problem, I suggest them to seek other measures to quantify the potential predictabilty.

Reply: We wish, if you could have read the full manuscript. The main content of the manuscript is the following:
i) Perfect model framework is used to estimate potential predictability of seasonal anomaly, which often shows paradoxical behaviour. ‘Analysis of variance’ framework is used for calculating ‘signal’ and ‘noise’ components using 52-ensemble member re-forecasts.

ii) Here we argue that ‘perfect model framework’ is not adequate, as the error growth is not from only initial condition errors but also from other sources, like physics, numerical scheme etc. We demonstrated that sub-seasonal component, which is part of the physics, adds error (biased contribution) in the seasonal forecast anomaly (i.e. Figure 7). However, ‘perfect model framework’ assumes, ensemble spread solely attributed to initial condition error. Consequently, true limit of predictability is not known. So, here our argument matches with your point of view that the method of estimating PPL based on perfect model framework is inadequate. We have already mentioned it in lines 337-344, in the last para of section 3.3

Comment : The authors appear to lack a clear understanding of the PPL issue. PPL is fundamentally a product of the “perfect model” framework. Once model errors are taken into account, it ceases to be a PPL problem. Therefore, the very premise of this study is conceptually inconsistent.

iii) Finally we propose a method for estimating PPL, which is free from paradox (section 3.4). Therefore, we believe the rationale provided for rejection does not fully capture the merits of the manuscrip

Reply: We respectfully disagree with the reviewer's interpretation. The PPL is indeed defined within the *perfect-model* framework; however, the objective of this study is to evaluate its practical limitations. We do not attempt to redefine PPL, but rather to show that the conventional signal-to-noise–based estimation becomes biased when ensemble spread is influenced not only by initial condition errors but also by internal model processes such as physics and sub-seasonal variability. In this sense, our work extends, not contradicts, the original concept by identifying how real-world model imperfections distort the theoretical upper bound of predictability. This perspective is consistent with earlier studies (e.g., Kumar & Hoerling, 2000; Scaife et al., 2014; Weisheimer et al., 2018) that recognized limitations in the perfect-model assumption and gave plausible reasons. Therefore, the premise of our study is conceptually coherent and offers a constructive refinement to the understanding of PPL.