



**HAL**  
open science

# Multimodel Forecasting of Precipitation at Subseasonal Timescales Over the Southwest Tropical Pacific

Damien Specq, Lauriane Batté, Michel Déqué, Constantin Ardilouze

► **To cite this version:**

Damien Specq, Lauriane Batté, Michel Déqué, Constantin Ardilouze. Multimodel Forecasting of Precipitation at Subseasonal Timescales Over the Southwest Tropical Pacific. *Earth and Space Science*, 2020, 7 (9), 10.1029/2019EA001003 . meteo-03352969

**HAL Id: meteo-03352969**

**<https://meteofrance.hal.science/meteo-03352969>**

Submitted on 18 Nov 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

# Earth and Space Science



## RESEARCH ARTICLE

10.1029/2019EA001003

# Multimodel Forecasting of Precipitation at Subseasonal Timescales Over the Southwest Tropical Pacific

### Key Points:

- Subseasonal prediction of SW tropical Pacific intense precipitation events in the S2S models can be improved through a multimodel ensemble
- The increased ensemble size of the multimodel is a key driver of the increase in forecast skill
- The strong role played by ENSO and the MJO in subseasonal predictability is similar between individual models and the multimodel

### Correspondence to:

D. Specq,  
damien.specq@meteo.fr

### Citation:

Specq, D., Batté, L., Déqué, M., & Ardilouze, C. (2020). Multimodel forecasting of precipitation at subseasonal timescales over the southwest tropical Pacific. *Earth and Space Science*, 7, e2019EA001003. <https://doi.org/10.1029/2019EA001003>

Received 10 DEC 2019

Accepted 23 APR 2020

Accepted article online 6 MAY 2020

Damien Specq<sup>1,2</sup> , Lauriane Batté<sup>1</sup> , Michel Déqué<sup>1</sup>, and Constantin Ardilouze<sup>1</sup> 

<sup>1</sup>CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France, <sup>2</sup>Direction de la recherche, École des Ponts, Paris, France

**Abstract** Multimodel ensemble (MME) reforecasts of rainfall at subseasonal time scales in the southwest tropical Pacific are constructed using six models (BoM, CMA, ECCO, ECMWF, Météo-France, and UKMO) from the Subseasonal-to-Seasonal (S2S) database by member pooling. These reforecasts are verified at each grid point of the 110°E to 200°E; 30°S to 0° domain for the 1996–2013 DJF period. The evaluation is based on correlation and on the ROC skill score of the upper quintile of precipitation for both weekly targets and Weeks 3–4 outlook. Confirming previous results at the seasonal time scales, the MME reaches the highest skill and also improves the reliability of probabilistic forecasts. However, an equivalent ensemble size comparison between the MME and the individual models shows that the better performance of the MME compared to the best individual models is significantly related to the larger ensemble size of the MME. Forecast skill is then explained in light of potential sources of predictability by evaluating the performance of the models depending on the initial ENSO and MJO state. While the role of ENSO on predictability is quite consistent with its related rainfall anomalies, the role of the MJO is more ambiguous and strongly depends on the location: An initialization in active MJO conditions does not necessarily imply better forecasts. This influence of ENSO and the MJO on predictability does not change when switching from individual models to the MME.

## 1. Introduction

While short-range numerical weather forecasting (up to 4 days ahead) provides deterministic information about the chronological day-to-day evolution of precipitation, seasonal forecasting (from 3 to 6 months) is only able to provide climate information about larger time spans, for example, the next month or the next season, as it mostly relies on slowly varying oceanic boundary conditions that constrain the possible evolution of the climate system. Subseasonal-to-seasonal (S2S) prediction aims at connecting short-range and seasonal predictions by considering lead times from 1 week up to 2 months.

As pointed out by Vitart (2004), subseasonal forecasting is usually considered a challenging forecasting time scale. However, due to modeling improvements and advances in the understanding of key phenomena, subseasonal time scales have encountered growing interest in the last years and are no longer regarded as a predictability desert. Indeed, an important number of sources of S2S predictability are now well identified and understood, including the Madden-Julian oscillation (MJO Zhang, 2013), sea surface temperatures, and their control by El Niño Southern Oscillation (ENSO), land surface conditions (soil moisture, snow cover), sea ice extent, and stratospheric processes.

However, owing to the chaotic nature of the climate system, both subseasonal and seasonal forecasts bear uncertainties that grow larger as lead time increases, due to uncertainties in the initial and boundary conditions, as well as modeling uncertainties (Slingo & Palmer, 2011). In order to provide useful information to end users, these uncertainties need to be taken into account, for example, by framing the calibrated forecasts in a probabilistic way (Palmer, 2002). For these probabilistic forecasts to be accurate and reliable, large ensembles, representing a range of possible evolutions of the system given the uncertainties, are needed (Richardson, 2000). In addition, modeling uncertainties can be taken into account by combining ensemble forecasts from several models and forming a “superensemble” (Krishnamurti, 1999), also called multimodel ensemble (MME).

The S2S (WMO/WWRP/WCRP) prediction project was launched in 2013 (Vitart et al., 2017). It includes a database of forecasts and reforecasts (hindcasts) made with subseasonal forecasting systems from several

©2020. The Authors.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

global prediction centers (11 currently). The S2S database offers the opportunity to carry out an intercomparison of predictions and to build MMEs. The MME approach has already been carried out at other time scales, and particularly in seasonal forecasting. At this time scale, it has often been highlighted that the MME approach bears some added value compared to using an individual model. For instance, in the DEMETER seasonal reforecasts (Palmer et al., 2004), Hagedorn et al. (2005) have shown that the MME performs better than a single model when considering the three components of a probabilistic evaluation (consistency, reliability, resolution).

Does the multimodel ensemble approach improve subseasonal forecasts as much as seasonal forecasts? This question has been addressed in few studies so far. Among the studies involving multimodel combination of subseasonal forecasts, Ferrone et al. (2017) have used a weighted multimodel approach with two S2S models for probabilistic prediction of monthly 2-m temperature anomalies over the globe, but do not compare it to the use of individual models alone. Karpechko et al. (2018) have shown that an MME constructed from nine individual S2S models could satisfactorily forecast the surface climate anomalies following the early 2018 Sudden Stratospheric Warming event over a period of 1 month, taking advantage of the different strengths and weaknesses of the individual models. In the case of precipitation, Vigaud et al. (2017a) have developed multimodel subseasonal forecasts of weekly precipitation terciles over North America using three different S2S models, by calibrating the probabilistic forecasts of the models separately using extended logistic regression before averaging the probabilities. They have then applied their methodology to other regions: monsoon areas (Vigaud et al., 2017b) and East Africa (Vigaud et al., 2018). More recently, Pegion et al. (2019) have evaluated multimodel forecast skill of global temperature, precipitation, and climate indices in the SubX subseasonal forecasting project, for which the multimodel construction is carried out by lagged average ensemble. These studies suggest that, as could have been expected from the results at the seasonal time scales, the MMEs tend to improve forecast skill compared to individual models, be it evaluated in terms of Ranked Probability Skill Score (Vigaud et al., 2017a) or anomaly correlation (Pegion et al., 2019), and this holds true for all lead times. Moreover, as illustrated by Vigaud et al. (2017a), using an MME increases the reliability of probabilistic forecasts.

However, it is unclear whether the MMEs perform better in the S2S time scales because they take into account different modeling choices, or because of their larger ensemble size. On the one hand, some of the aforementioned MMEs from the S2S database (Ferrone et al., 2017; Vigaud et al., 2017a) were constructed using elaborate approaches that lead to a substantial increase in forecast skill but make it difficult to disentangle the role of these two effects. On the other hand, Karpechko et al. (2018) have constructed an MME for the needs of a case study and do not carry out a systematic evaluation of forecast skill over a large sample of reforecasts. The present study therefore intends to address the importance of the ensemble size when comparing MMEs to individual models. For this reason, it uses an MME constructed with a simple member pooling approach (i.e., combining together the members of the individual models to build a new ensemble before computing the probabilities), which makes it easier to adjust ensemble size and carry out fair ensemble size comparisons. In order to have a substantial pool of members for this approach, it also includes a large number of prediction systems from the S2S database (six).

In this framework, the potentialities of subseasonal MME forecasting are explored with a focus on precipitation in the southwest tropical Pacific. The domain retained for this study ranges from 110°E to 200°E in longitude and 30°S to 0° in latitude. It includes the southern part of the Maritime Continent, the tropical and subtropical part of Australia, and a wide number of Pacific island territories such as Solomon Islands, Vanuatu, New Caledonia, Fiji, Samoa, and Tonga. Rainfall prediction in this area is a paramount matter, as the various territories mentioned above are subject to heavy precipitating events and tropical cyclones, but are also regularly exposed to a risk of drought. S2S prediction of rainfall in the region has already been partially explored, in particular in Australia (Hudson et al., 2011; Marshall et al., 2014), where there is a strong interest in the prediction of the onset and demise of the austral summer North Australian monsoon (Bombardi et al., 2017; Marshall & Hendon, 2015). In the oceanic part of the domain, rainfall patterns are dominated by the location of the South Pacific Convergence Zone (SPCZ Vincent, 1994). Therefore, the variability in the SPCZ location, described in details by Matthews (2012), has a huge impact on rainfall for the Pacific island territories. For instance, New Caledonia (166°E; 21°S) is located south of the mean position of the SPCZ but is impacted by its frequent southward shifts during austral summer, which bring a significant part of its annual rainfall input (Lefort, 2005). This variability in the SPCZ location is influenced by the interaction between the SPCZ and other large-scale drivers: Rossby waves (Matthews, 2012) and the

**Table 1**  
*Reforecast Attributes for the 6 Systems From the S2S Database*

| Attributes             | BoM            | CMA             | ECCC          | ECMWF                   | MF             | UKMO            |
|------------------------|----------------|-----------------|---------------|-------------------------|----------------|-----------------|
| Time range             | Day 1–62       | Day 1–60        | Day 1–32      | Day 1–46                | Day 1–61       | Day 1–60        |
| Atmospheric resolution | T47L17         | T106L40         | 0.45x0.45 L40 | T639/319 L91            | T255L91        | N216L85         |
| Reforecast             | Fix            | Fix             | On the fly    | On the fly              | Fix            | On the fly      |
| Reforecast period      | 1981–2013      | 1994–2014       | 1995–2014     | Past 20 yrs (1996–2015) | 1993–2014      | 1993–2015       |
| Reforecast frequency   | Four per month | Daily           | Weekly        | Two per week            | Four per month | Four per month  |
| Reforecast size        | 33             | 4               | 4             | 11                      | 15             | 7               |
| Coupling               | Ocean          | Ocean + sea ice | No coupling   | Ocean                   | Ocean          | Ocean + sea ice |

MJO (Matthews et al., 1996; Matthews & Li, 2005) at the subseasonal time scales, ENSO at the interannual time scale (Salinger et al., 2014; Vincent et al., 2011), and climate change trends (Griffiths et al., 2003) at interdecadal time scales.

On account of the well-known large-scale drivers mentioned above, rainfall is expected to exhibit subseasonal predictability at least in some parts of the chosen domain. Indeed, at subseasonal time scales, ENSO is a slowly varying process that constrains the precipitation distribution and should therefore provide some background predictability. On top of the signal related to ENSO, the MJO might provide information about the evolution of rainfall patterns at shorter time scales, and there is hope that the numerical models are able to reproduce these patterns consistently. de Andrade et al. (2018) have already investigated the role of ENSO and the MJO on subseasonal predictability of precipitation at a global scale using correlation. By isolating the ENSO and MJO components of rainfall with linear regression, they indeed prove that these two phenomena bring subseasonal predictability, at least in the tropical Indian and Pacific basins. This study will also extend their approach for the case of the southwest tropical Pacific by (1) using multimodel ensemble prediction and (2) focusing on the probabilistic prediction of exceeding a heavy rainfall threshold. This article presents an evaluation of the skill of an S2S multimodel reforecast for rainfall in the austral summer season (December, January, February). The austral summer season was chosen as it corresponds to the peak period of MJO activity in the Southern Hemisphere (Zhang, 2005), to the wet season in southwest Pacific island territories and to the monsoonal season in northern Australia (Marshall & Hendon, 2015).

This paper is organized as follows. Section 2 introduces the data and methods used for this study. Section 3 focuses on the evaluation of multimodel performance by comparing it with the performance of the individual models, while section 4 investigates the relationship between forecast skill, ENSO and the MJO. Finally, a discussion of the results is presented in section 5.

## 2. Data and Methodology

### 2.1. S2S Reforecasts

The S2S database (Vitart et al., 2017) gathers the outputs from the subseasonal forecasting systems of 11 different centers. Daily precipitation has been extracted from the reforecasts of six different systems for the months of December, January, and February. Given that all atmospheric models have different native horizontal resolution, the data are extracted on a common 1.5° grid on the southwest tropical Pacific domain (110°E to 200°E; 30°S to 0°). The selected prediction systems are the Bureau of Meteorology (BoM, Australia), the China Meteorological Administration (CMA), the Environment and Climate Change Canada (ECCC), the European Center for Medium-range Weather Forecast (ECMWF), the Météo-France (MF), and the United Kingdom Met Office (UKMO) systems. The features of each S2S system are summarized in Table 1.

The choice of these six systems was motivated by the dates of the reforecast periods. Selecting these systems enables us to have a common 18-year period for our analysis, therefore less sensitive to possible multiannual variability of skill. Four start dates were considered for each month in the DJF season on the common 18-year period 1996–2013. There is a total of  $4 \times 3 \text{ months} \times 18 \text{ years} = 216$  start dates, indicated in Table 2. The time range of the reforecasts had to be restricted to the shortest one, from Day 1 to Day 32. In this study, the lead times considered correspond to weekly periods: Week 1 (Days 5 to 11), Week 2 (Days 12 to 18), Week 3

**Table 2**  
*Correspondence Between the Météo-France (MF) Reference Start Dates and the Start Dates for the Five Other Systems*

| MF    | BoM   | CMA   | ECCC  | ECMWF | UKMO  |
|-------|-------|-------|-------|-------|-------|
| 01-01 | 01-01 | 01-01 | 01-04 | 12-31 | 01-01 |
| 01-08 | 01-06 | 01-08 | 01-11 | 01-07 | 01-09 |
| 01-15 | 01-16 | 01-15 | 01-18 | 01-14 | 01-17 |
| 01-22 | 01-21 | 01-22 | 01-25 | 01-21 | 01-25 |
| 02-01 | 02-01 | 02-01 | 02-01 | 02-01 | 02-01 |
| 02-08 | 02-06 | 02-08 | 02-08 | 02-08 | 02-09 |
| 02-15 | 02-16 | 02-15 | 02-15 | 02-15 | 02-17 |
| 02-22 | 02-21 | 02-22 | 02-22 | 02-22 | 02-25 |
| 12-01 | 12-01 | 12-01 | 11-30 | 12-01 | 12-01 |
| 12-08 | 12-06 | 12-08 | 12-07 | 12-08 | 12-09 |
| 12-15 | 12-16 | 12-15 | 12-14 | 12-15 | 12-17 |
| 12-22 | 12-21 | 12-22 | 12-21 | 12-22 | 12-25 |

(Days 19 to 25), and Week 4 (Days 26 to 32). The first four days were discarded as belonging to short-range forecasting.

## 2.2. Verification Methodology

Verification is carried out against the Multi-Source Weighted-Ensemble Precipitation (MSWEP) version 1.2 (Beck et al., 2017) global historic daily precipitation data set, which covers the period 1979–2015 on a global 0.25° grid. This data set combines several sources: satellite data, World Meteorological Organisation Global Telecommunication System (WMO GTS) network rain gauges and reanalysis. This data set was interpolated on the common 1.5° model grid using conservative remapping. The reforecast performance is evaluated by computing verification scores separately at each of the 60 × 21 grid points. However, in order to summarize the information about the systems' performance at the scale of the whole domain, scores will also be represented as the average of all grid point scores.

Following Murphy (1993), the quality of reforecasts can be evaluated on different aspects. In this study, the focus is put on linear association, reliability, and discrimination. These three aspects were chosen because they are elementary bricks of forecast quality, contrary to accuracy or bias that include several of these elementary aspects. Although developed independently, this verification framework shares common features with the one proposed by Coelho et al. (2018) over South America and corresponds to what they call “all season hindcast verification.”

Linear association is assessed through the correlation of the weekly average of rainfall between the model ensemble mean and the MSWEP data. Such grid point correlations are calculated at each lead time with a set of  $N$  values corresponding to the  $N$  start dates. Significance testing of correlation is performed as follows. For a given correlation  $r$ , under the no-skill null hypothesis that the forecast and the observations are independent, the sample statistic

$$t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}} \quad (1)$$

follows a Student distribution with  $N - 2$  degrees of freedom. However, owing to seasonal and interannual variability, rainfall values are not necessarily independent from one day to another so it cannot be ascertained that the real number of degrees of freedom is  $N - 2$ . An equivalent number of independent observations  $N_{eq}$  can be computed following the method proposed by Guemas et al. (2014). The correlation is considered as significant at the 95% level if the  $p$  value of the sample statistic is less than 0.05.

In this study, reliability and discrimination apply to the probabilistic reforecasts of a binary event, the occurrence of a week with accumulated precipitation in the upper quintile (i.e., exceeding the 80th percentile) of climatology. The climatology of observations is built in the same manner as in Vigaud et al. (2017a). For observations at a given grid point, a climatology is constructed for each year, start date, and lead time

(Week 1–4) by considering the MSWEP weekly accumulated rainfall in a 3-week window, composed of the week of interest and 1 week on either side, in the other 17 years of the 1996–2013 period. The actual year of observation is removed in agreement with the leave-one-year-out cross-validation approach. The resulting climatological sample contains  $3 \times 17 = 51$  values. The observed weeks with accumulated precipitation exceeding the 80th percentile of this climatology are assigned with the value 1, and 0 otherwise. The climatology of the reforecasts is also built with the leave-one-out approach. For each year, start date, and target week, the climatology is made up of the weekly accumulated rainfall for the same start date and target week, in the other 17 years and in all members. For instance, this makes a climatological sample of  $17 \times 15 = 255$  values for the 15-member Météo-France reforecast. Reforecasts are then binarized in the same way as observations, according to their own climatology. The probabilistic reforecast of the binary event is the fraction of members where the event occurs. Additional verification is carried out for the Weeks 3–4 outlook (i.e., Days 19 to 32), which is another common verification target at the S2S time scales (e.g., Hudson et al., 2011; Vigaud et al., 2017a; Zhu et al., 2014). In this case, the correlation is computed on the average rainfall over the whole fortnight, and the event of interest is the upper quintile of the fortnight accumulated precipitation. For the latter score, the climatology of observations is constructed with 6-week windows including the target fortnight and one fortnight on either side.

Reliability is the correspondence between the forecast probability of an event and the probability that the event actually occurs in observations for a given forecast probability. As pointed out by Weisheimer and Palmer (2014) for the case of seasonal forecasts, reliability is the most important aspect to determine how good a forecast is, given that forecast-based decisions can only be made with reliable forecasts. In this article, reliability is examined with reliability diagrams. However, reliability is not an intrinsic property of the probabilistic forecasts as it can be improved through calibration (not applied here). Then, we also assess discrimination of the binary event by the reforecasts using the Receiver Operating Characteristic (ROC) skill score. Note that if  $A$  is the area under the ROC curve, the ROC skill score is defined as

$$ROCSS = 2A - 1, \quad (2)$$

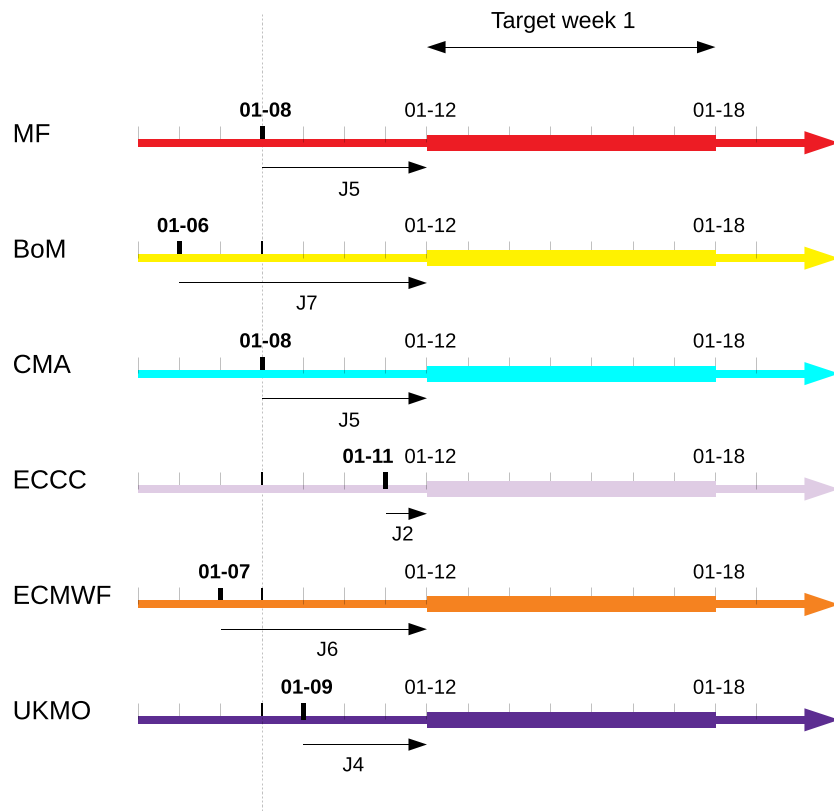
so that the no-skill value for  $ROCSS$  is 0 (and not 0.5 as it is for the area under the ROC curve). Statistical significance of the area under the ROC curve is determined with a Mann-Whitney  $U$  test as in Wilks (2006). Given the above considerations about reliability and discrimination, the ROC skill score is not sensitive to a lack of reliability and can be used for the verification of the upper quintile of weekly precipitation without any further calibration.

### 2.3. Multimodel Ensemble Construction

The MME construction and intercomparison is carried out using the calendar of the Météo-France reforecast. The four start dates considered are the ones indicated in the first column of Table 2: the 1st, the 8th, the 15th, and the 22nd of each month. However, in the S2S database, there is no unique standard for reforecasts and start dates differ from one system to another. Here we chose to overcome this drawback by selecting, for each system, the start dates that were closest to the corresponding Météo-France start date. The lines of Table 2 show the correspondence between start dates.

The intraseasonal lead times considered are also based on the Météo-France start dates. For example, Week 1 for January 8 start date corresponds to Météo-France Days 5 to 11, that is, January 13 to 19. For the BoM prediction system, the equivalent start date is January 6 so that the January 13 to 19 Week-1 period no longer corresponds to Days 5 to 11 but to Days 7 to 13. In other words, the predicted values for a given weekly lead time might correspond to slightly different daily lead times depending on the system. Such a process is illustrated with an example in Figure 1. Note that in most cases, start dates are evenly distributed between earlier and later dates than the Météo-France calendar. Moreover, the maximum offset between start dates is three days. Therefore, it may be assumed that this choice will not favour a particular system. It could be argued that the UKMO system might be at an advantage for the earlier lead times as it starts with a systematic delay (+1, +2, or +3 days) compared to the reference start dates (see Table 2). However, additional verification carried out with the UKMO reforecast calendar did not exhibit any significant difference in our results (not shown).

In this study, the expression “balanced multimodel” will be used to refer to an ensemble reforecast that is constructed by pooling members from the six individual prediction systems, with the additional requirement that there must be the same number of members coming from each system (e.g., four members from BoM



**Figure 1.** Example of multimodel reforecast construction for January 8 start date and target Week 1.

and four members from CMA). In the following sections, we will consider balanced multimodels with 24, 18, 12, and six members (with, respectively, four, three, two, and one member per system). In order to evaluate the uncertainty of the scores obtained with the multimodel ensemble, 1,000 different sets of members are considered when randomly selecting the members from each system. The score is calculated separately for the 1,000 sets and the overall multimodel score corresponds to the median, while the 90% confidence interval is bounded by the 5th and 95th percentiles. When calculating a correlation, the multimodel is built with the members' rainfall output values and the correlation is based on the MME mean. When calculating a ROC skill score, the multimodel is built with the members' binary values and the corresponding multimodel probability of occurrence is deduced from the multimodel fraction of members where the event occurs.

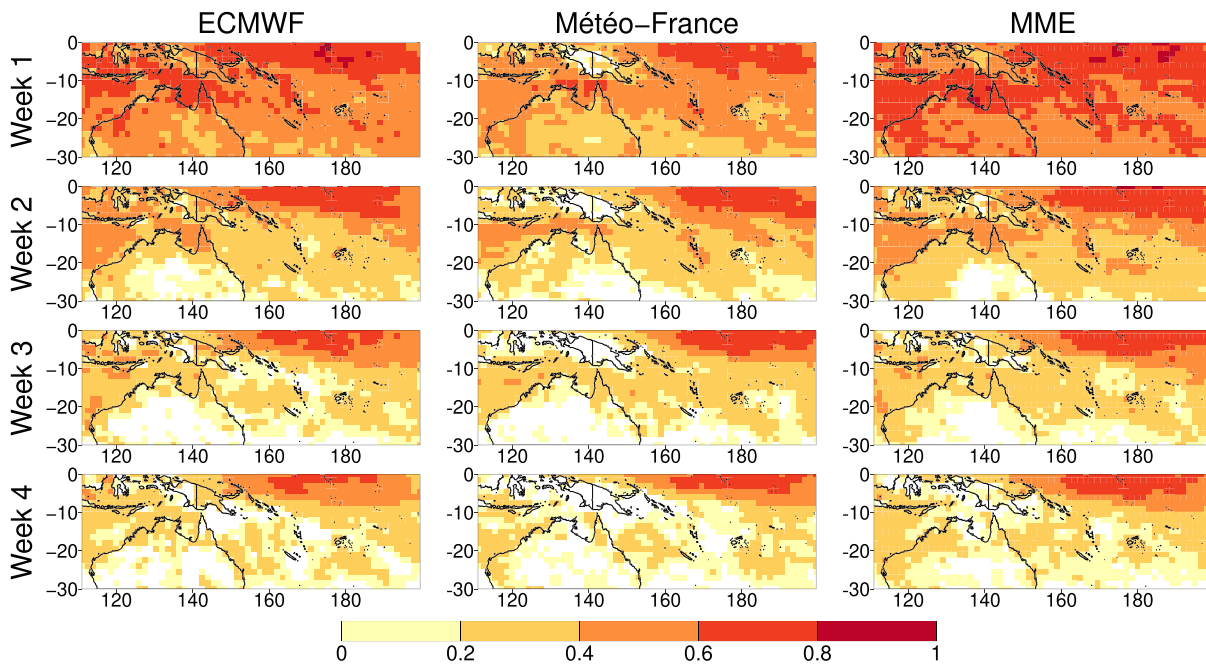
#### 2.4. State of ENSO and the MJO

Reforecast skill conditioned on the state of ENSO and the MJO will be computed in section 4. For this purpose, records of MJO and ENSO indices on the 1996–2013 reforecast period are needed. The ENSO index considered is the CPC Oceanic Niño Index (ONI), which corresponds to the 3-month running mean of ERSST v5 SST anomalies in the Niño 3.4 region. There is one value for each 3-month period. The value associated with a given start date is the value for the DJF period including the start date. A decomposition in three discrete phases is defined with thresholds of  $-0.5^{\circ}\text{C}$  and  $+0.5^{\circ}\text{C}$ . The start date is considered to be in La Niña phase if the ONI index is less than  $-0.5^{\circ}\text{C}$ , in El Niño phase if it is greater than  $+0.5^{\circ}\text{C}$ , and in neutral phase otherwise.

The MJO indices are the daily RMM indices (Wheeler & Hendon, 2004) at the start date, provided by the Australian Bureau of Meteorology. The start date will be considered in active or inactive MJO depending on whether the MJO amplitude (i.e., the norm of the  $(RMM1, RMM2)$  vector) is greater than or less than 1.

### 3. Model Intercomparison and Multimodel Forecast Skill

The performance of the 24-member balanced multimodel is compared to that of the individual models for the DJF start dates of the common period 1996–2013. Figure 2 shows the correlation in the southwest tropical



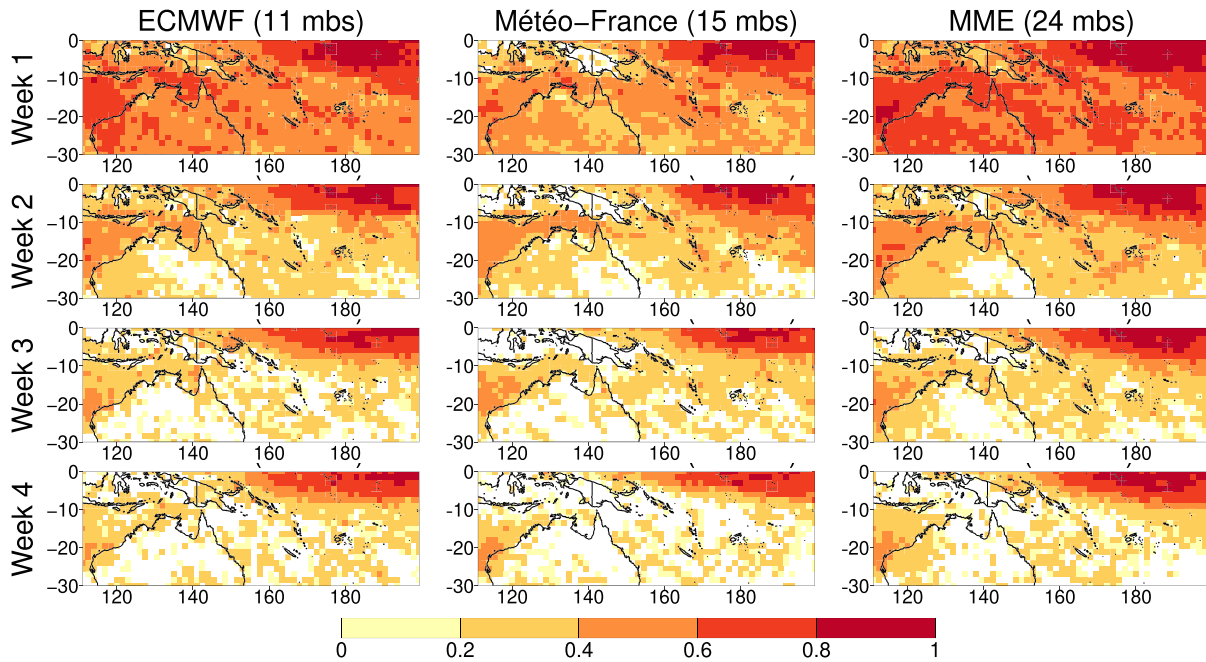
**Figure 2.** Ensemble mean correlation of the weekly average of rainfall between the MSWEP dataset and the ECMWF model (first column), the Météo-France model (second column), and the 24-member MME (third column), for the 12 start dates in 1996–2013 DJF. The different rows correspond to different lead times from 1 to 4 weeks. Grid points with correlation not exceeding the 95% significance threshold are in white (see section 2.2 for full details).

Pacific at each weekly lead time for the 24-member multimodel, the 15-member Météo-France system, and the 11-member ECMWF system. Only two individual models are represented in this figure for the sake of space and clarity. These two models were chosen because (1) they have a reasonably high ensemble size compared to the MME and (2) the reference calendar comes from the Météo-France reforecasts. Figure 3 illustrates a similar comparison with the ROC skill score for the prediction of the upper quintile of weekly accumulated rainfall. The multimodel exhibits better reforecast skill for almost every grid point and every lead time, although the contrast reduces as lead time increases, as confirmed by Figures 4 and 5. The MME also reduces the extent of the areas where skill score values are not significant. Similar spatial patterns of reforecast skill can be noted between models, for correlation and ROC skill score alike. Predictability is higher and remains significant at larger lead times in the northeastern part of the area, which corresponds to the equatorial part where the SST and the subsequent circulation patterns are most constrained by the state of ENSO. This result is very consistent with that of Wheeler et al. (2017). On the contrary, in the eastern part of Australia and in the southeastern part of the domain, there is no more significant skill as early as Week 2. At Week 3, most of the Australian continent exhibits insignificant score values, and so does a large fraction of Maritime Continent islands. At Week 4, significant skill is mostly concentrated in the central equatorial Pacific region and in the small portion of the Indian Ocean west of Australia that is included in the domain.

To complete this evaluation, Figure 4 shows an intercomparison of the correlation and ROC skill score averaged over the whole domain. The comparison is established between all individual models and the 24-member multimodel. For a fair comparison between independent models, the scores were all computed with the same ensemble size, that is, four members per model. For each model with more than four members, the median of 1,000 random subsets of four members is shown. Similarly, the multimodel score is represented by the median of 1,000 multimodel subsets. For the sake of clarity, the 95% confidence intervals around the median values are not shown but would otherwise be very small and nonoverlapping when considering the spatial mean of the scores over the whole domain. From Figure 4, it clearly appears that with such ensemble sizes, the multimodel is better than any model at all lead times and for both scores. As for individual models, the ECMWF and UKMO models are the best for both scores up to at least Week 3.

A common argument to account for the better skill of a multimodel compared with a single model is that the model delivering the best prediction is not the same every time, such that when aggregating over a large

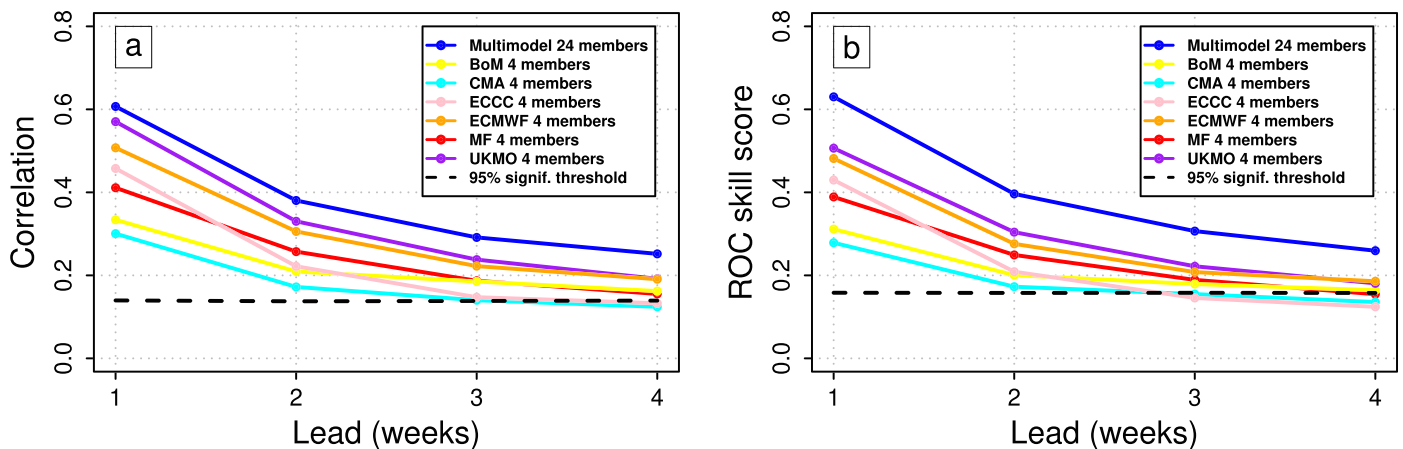




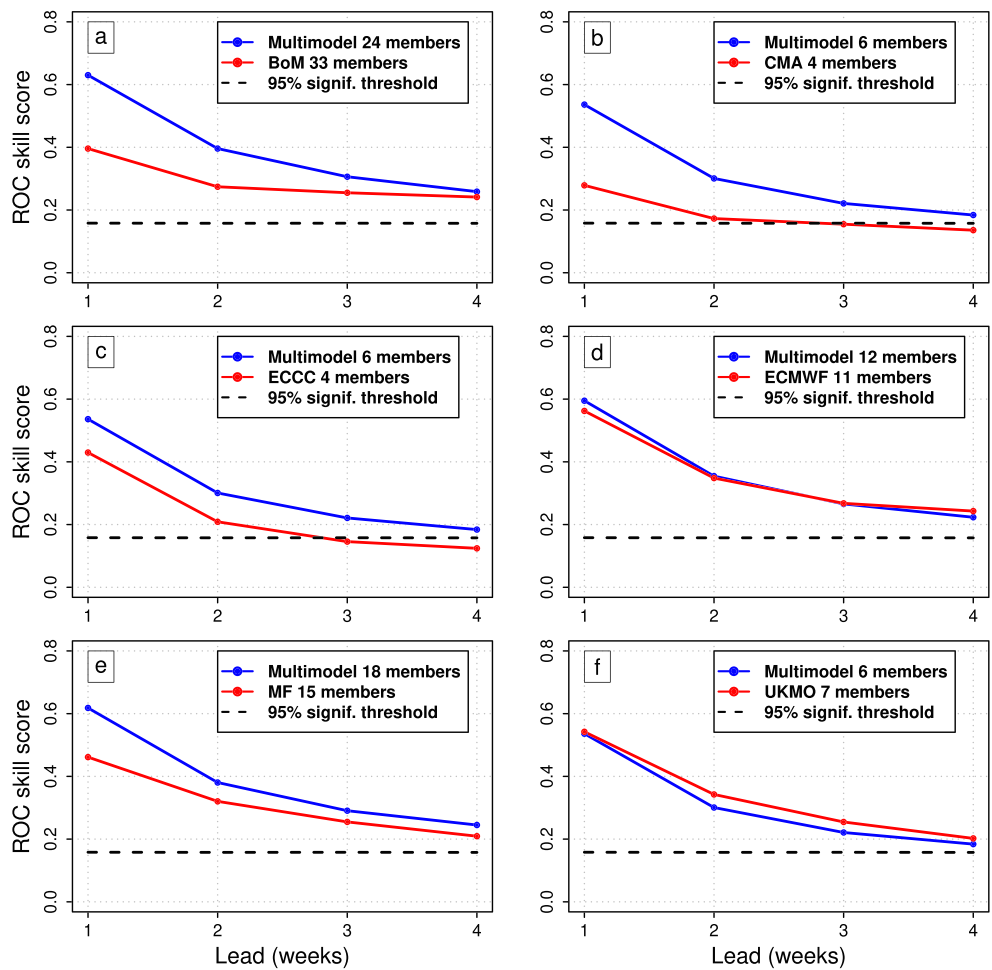
**Figure 3.** Same as Figure 2 for the ROC skill score of the probability that the accumulated weekly rainfall is in the upper quintile of climatology.

number of cases, the multimodel should win (diversity effect). However, as already pointed out for the seasonal time scale in Hagedorn et al. (2005), the higher scores of the multimodel in Figure 4 may also arise from the difference in ensemble size (24 vs. four members). Indeed, a larger ensemble size increases skill and reduces uncertainty (ensemble size effect). The contributions of the diversity effect and the ensemble size effect to the better performance of the multimodel are hard to separate. Yet, it remains possible to make fairer comparisons between the multimodel and individual models in terms of ensemble size. Figure 5 shows the spatially averaged ROC skill score between the multimodel and individual models with their full ensemble size. The ensemble size of the balanced multimodel is chosen as close as possible to the individual model ensemble size.

Figure 5 shows that, even in the context of a fair comparison, the multimodel is more skillful in terms of ROC skill score than the BoM, CMA, ECCC, and Météo-France models, while it is as skillful as the ECMWF



**Figure 4.** Spatial mean of (a) correlation for weekly average of rainfall and (b) ROC skill score of the probability that the accumulated weekly rainfall is in the upper quintile, as a function of reforecast lead time, in the 24-member multimodel and the four-member individual models. The dashed black line represents the 95% significance threshold for (a) correlation and (b) ROCSS, averaged over the whole domain. The scores are computed as the median of the scores obtained with 1,000 random draws of members (see section 2.3 for full details).



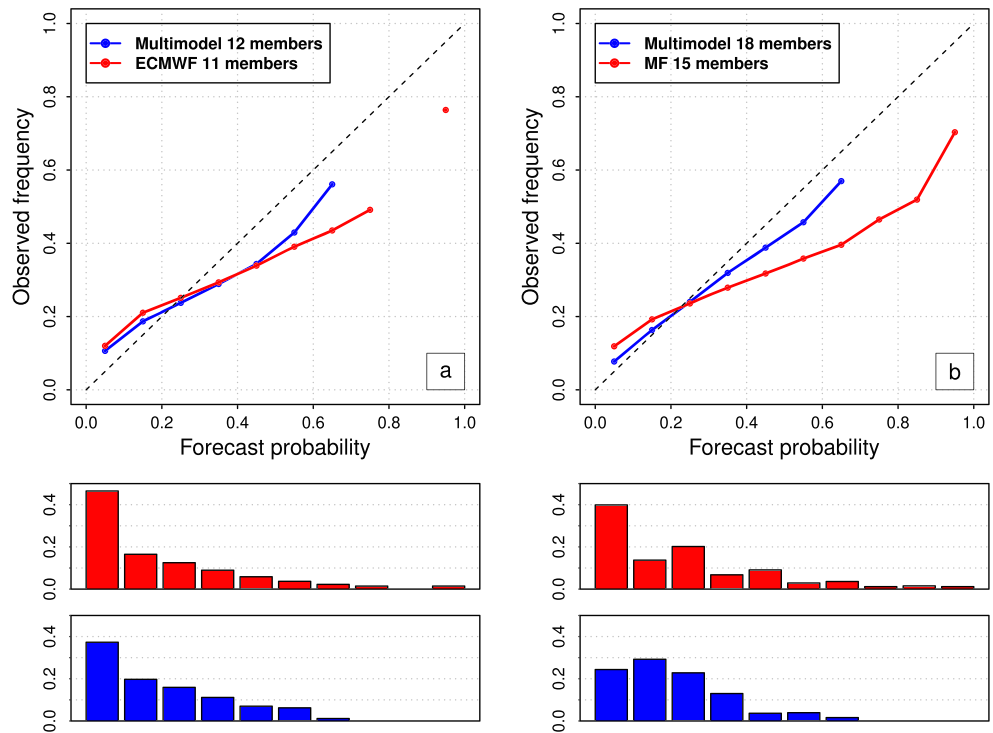
**Figure 5.** Comparison between multimodel and (a) BoM, (b) CMA, (c) ECCC, (d) ECMWF, (e) Météo-France, and (f) UKMO spatially averaged ROC skill score that weekly precipitation is in the upper quintile. The dashed black line represents the spatially averaged 95% significance threshold for ROC skill score. Note that the confidence interval for the MME is so close to the MME median that it has not been represented for the sake of clarity.

model and slightly less skillful than the UKMO model. The same conclusions can be drawn when looking at the correlation (not shown).

The benefits of the MME appear clearly when considering the reliability of forecasts of the upper quintile of weekly rainfall. Figure 6 shows the reliability diagrams of the MME at lead time Week 2, compared to that of two individual models, the ECMWF and Météo-France models. Reforecasts for every point of the domain are pooled together so as to provide robust estimates. The MME is more reliable than the individual models, even though the ensemble sizes are similar. The difference is particularly sharp at high probability bins, for which the MME points are closer to the first diagonal. The individual models are overconfident when predicting high or low probabilities due to lack of spread, even when the predicted outcome is erroneous. This flaw is partly corrected by combining information from different models, as shown by the associated histograms.

#### 4. Modulation of Skill by ENSO and the MJO

In this section, the influence of ENSO and the MJO, assumed to be the most important sources of predictability, is explored by evaluating the performance of the reforecasts conditioned by the state of ENSO and the MJO in the initial conditions. As in Zhu et al. (2014), the evaluation is shown for the fortnight Weeks 3–4 since these are the lead times for which skill exhibited the biggest dependence to ENSO and the MJO (not

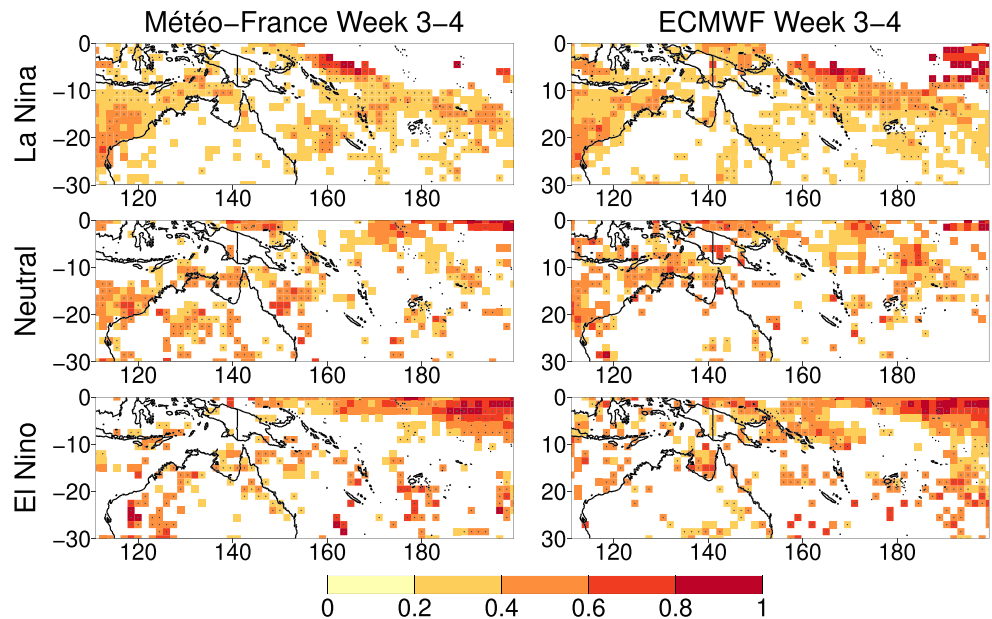


**Figure 6.** Comparison between multimodel (blue) and (a) ECMWF and (b) Météo-France (red) Week 2 reliability diagram for the upper quintile of weekly precipitation. The forecast frequency in each probability bin is represented below on the bar plots for each of the two reforecasts with the same color code. Bins with less than 1% of the total number of forecasts are not plotted.

shown). Indeed, skill at earlier lead times (Weeks 1 and 2) mainly comes from initial atmospheric conditions (White et al., 2017), so the slowly varying processes will play a less important role than at later lead times. The same conclusions could be drawn from the analysis of Week 3 and Week 4 separately.

#### 4.1. Influence of ENSO

Figure 7 represents the map of the ROC skill score of the upper quintile of accumulated precipitation in Weeks 3–4 for the ECMWF and Météo-France models, when separating the 216 start dates of our reforecast between the three ENSO phases as explained in section 2.4. Grid points for which there is increased skill in the considered ENSO phase compared to the other two phases are indicated with dots. For increased skill to be considered, two requirements must be met: The ROC skill score in the phase must be greater than the maximum of the ROC skill score in the other two phases, and it must also be intrinsically significant at the 95% level according to the Mann-Whitney  $U$  test. The significance level computed from this test depends on two elements: (1) the verification sample size (i.e., the number of start dates) within the given phase and (2) the number of times the event occurs in the sample observations. Figure 7 shows that there is added value for reforecasts of both systems launched in La Niña conditions, except in the northeastern part (the western tip of the equatorial Niño 3.4 box) and the subtropical part of the domain. Higher skill can especially be noted along the SPCZ track and on the western coast of Australia. On the contrary, reforecasts in El Niño conditions show less skill over the domain, except in the northeastern part, complementary to La Niña phases. It must be highlighted that the regions of higher skill for a particular phase correspond to the regions where rainfall is increased during this phase. Indeed, according to Figure 8, the regions of increased ROC skill score correspond to the regions where it is more likely to observe rainfall in the upper quintile. Finally, neutral ENSO conditions do not show a consistent region of increased skill. These findings are quite similar between the different individual models, as illustrated by the comparison between the ECMWF and Météo-France models in Figure 7.

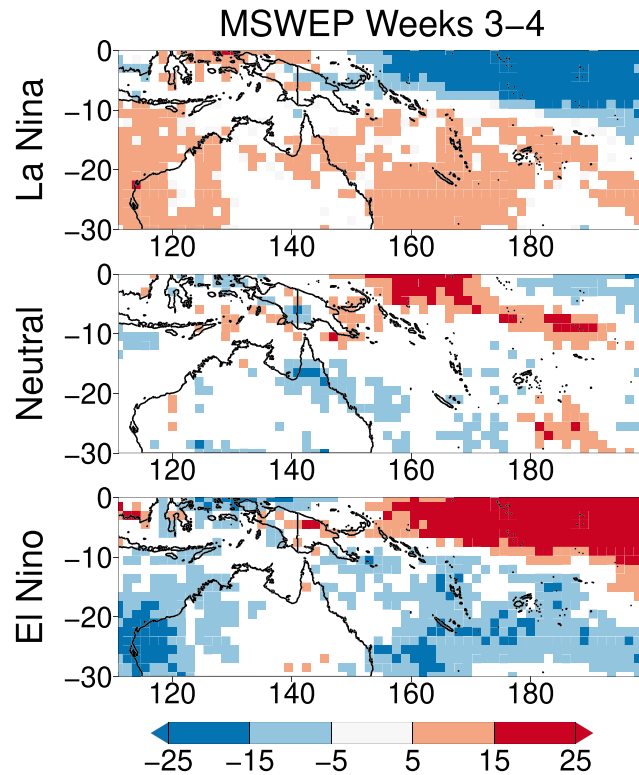


**Figure 7.** ROC skill score for the upper quintile of 2-week accumulated precipitation at lead time Weeks 3–4 for ECMWF (first column) and Météo-France (second column), conditioned on the ENSO phase in the initial conditions. La Niña includes 104 start dates, neutral ENSO includes 52 start dates and El Niño includes 60 start dates. Grid points with ROC skill score not exceeding the 95% significance threshold are in white. Dots indicate grid points for which the skill score is significant and greater than the ones in the other two ENSO phases.

#### 4.2. Influence of the MJO

Similar to Figure 7, Figure 9 represents the reforecast evaluation conditioned on the MJO activity in the initial conditions, as in Marshall et al. (2011, Figure 8) and Liang and Lin (2018, Figure 15). The 216 start dates were split depending on whether the MJO was active or inactive at the start date, as explained in section 2.4. Considering the influence of these initial conditions on predictability at Weeks 3–4 makes sense under the assumption that the S2S models are reasonably able to represent the propagation of the MJO. Indeed, Vitart (2017) has demonstrated that the S2S models had skill to predict the MJO between 2 and 4 weeks, and he shows the models represented in our Figure 9, ECMWF and Météo-France, are among the best for MJO bivariate correlation at Weeks 3–4. Moreover, it can be verified that Weeks 3 and 4 are indeed the lead times for which the ROC skill score is the most strongly dependent to MJO initial conditions (not shown).

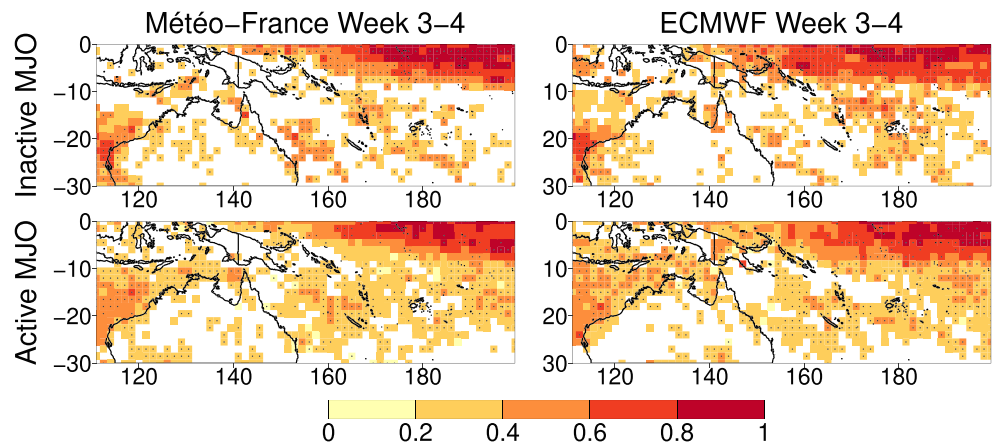
Dots in Figure 9 indicate the grid points for which there is increased skill in the start dates in the considered MJO state (e.g., active MJO) compared to the complementary set of start dates (e.g., inactive MJO). Similar to the verification for the ENSO phases (section 4.1), the patterns of increased skill are still very close between the different models in terms of location and spatial extent. However, at the scale of the whole domain, no initial MJO configuration leads to higher skill than the other one. It is true that reforecasts initialized in active MJO conditions exhibit significant skill at Weeks 3–4 in a larger number of grid points than those initialized in inactive MJO. Yet, the vast majority (about 75%) of the grid points where scores are significant for both initial conditions (which in total represent one third of the points), the inactive MJO reforecast skill is higher than the active MJO reforecast skill. The region where this effect can be observed most easily is the northeastern part of the domain, which again corresponds to the equatorial area and the edge of the Niño 3.4 region. In the subequatorial part of the domain (south of 10°S), regions for which skill is higher in active MJO conditions appear alternating with other regions for which skill is higher in inactive MJO conditions. These findings tend to moderate the conclusions that were drawn by Marshall et al. (2011). If at a global scale, skill over the western Pacific appears to be higher in active MJO initial conditions (Marshall et al., 2011, Figure 8), at a more regional level, this is likely less the case. For instance, our Figure 9 shows that New Caledonia (166°E; 21°S) lies in an area where skill is higher in active initial MJO while Vanuatu (168°E; 17°S) and Fiji (178°E; 18°S) lie in areas where skill is higher when the MJO is inactive.



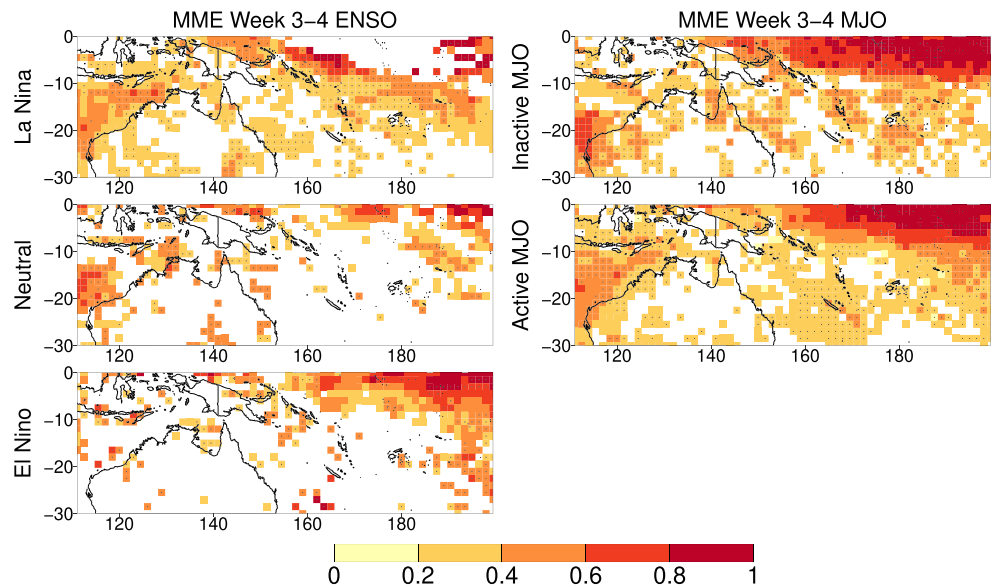
**Figure 8.** Week 3–4 frequency anomalies (%) of the upper quintile of precipitation in each ENSO phase in the MSWEP data set. Significance of the anomalies at the 90% level is evaluated with 1,000 random draws of the same number of start dates as the given ENSO phase. Grid points with insignificant anomaly are indicated in white.

**4.3. Does the Influence of ENSO and the MJO on Predictability Change with the MME Approach?**

As pointed out in the previous paragraphs with the example of ECMWF and Météo-France, the conditional reforecast skill patterns do not vary a lot between the different models. Therefore, the MME technique is not expected to bring any change regarding this matter. For instance, it is not expected to increase the sensitivity of reforecasts to the initial MJO activity. This assumption is tested in Figure 10, which represents the same verification as Figures 7 and 9 applied to the MME. It appears that the MME does indeed exhibit the same



**Figure 9.** ROC skill score for the upper quintile of 2-week accumulated precipitation at lead time weeks 3-4 for ECMWF (first column) and Météo-France (second column), conditioned on the state of the MJO in the initial conditions. Inactive MJO includes 78 start dates and active MJO includes 138 start dates. Grid points with ROC skill score not exceeding the 95% significance threshold are in white. Dots indicate grid points for which the skill score is significant and greater than the one in the other initial MJO state.



**Figure 10.** Same as Figure 7 (first column) and Figure 9 (second column) for the MME.

behavior as the individual models in terms of ENSO or MJO conditional skill. For instance, in Figure 10, the patterns of higher skill in the subequatorial band for active MJO are similar to those in Figure 9. As a result, the influence of ENSO and the MJO on predictability is the same in the MME as it is in individual models.

## 5. Discussion and Conclusion

In this study, we built a MME combining six different models from the S2S database to assess rainfall prediction over the southwest tropical Pacific at subseasonal time scales. The focus was put on the ability of S2S models to distinguish weeks with more intense rainfall than the others, in this case weeks with precipitation in the upper quintile. Such information can be valuable for potential users, for instance, to predict the end of the dry season in Pacific island territories (e.g., New Caledonia, Alexandre Peltier, personal communication).

It has been regularly pointed out that the MME technique performed better than single models in seasonal forecasting (e.g., Hagedorn et al., 2005), and a few studies tend to show that it is also the case in subseasonal forecasting (Pegion et al., 2019; Vigaud et al., 2017a). This is confirmed by our results for three aspects of forecast quality in Murphy's framework (Murphy, 1993): association, reliability, and discrimination. However, the overall better performances of the MME over single models could be due to two effects: the cancellation or compensation of individual model errors by increasing model diversity (diversity effect), and the larger ensemble size that increases skill, whether the members come from the same model or not (ensemble size effect). We have been able to investigate this question with our MME, which uses a large number of prediction systems from the S2S database. When carrying out a fair ensemble size comparison, it appears that this MME does not outperform the best individual models in ROC skill score and correlation. Therefore, the ensemble size effect is undoubtedly involved for a significant part. In other words, running one of the best models with a larger ensemble size could possibly lead to similar or higher scores than the 24-member MME.

Besides, this study illustrates subseasonal predictive skill of precipitation in S2S forecasting systems, in a region where this variable exhibits high predictability compared to the rest of the globe. In this respect, it confirms the results of previous studies that have investigated subseasonal forecast skill of precipitation at a global scale (Li & Robertson, 2015; Wheeler et al., 2017). It shows that the limit of predictive skill of the upper quintile of weekly rainfall in the southwest tropical Pacific ranges from Week 1 in continental Australia up to Week 4 (at least) in the northeastern equatorial part of the domain. We also show that the spatial patterns and the evolution of forecast skill with lead times are consistent from one model to another.

As a consequence, there is no huge difference in sensitivity to the sources of subseasonal predictability (ENSO and the MJO) between the individual models and the MME.

Furthermore, zooming in on this particular region bears some added value when considering the role of these two sources of predictability. We show that the initial state of ENSO and the MJO both condition predictability. The role played by ENSO is quite straightforward, with predictability for the upper quintile of precipitation increased at the grid points where this event becomes more frequent in a given ENSO state. On the contrary, the role of the MJO, for the specific case of the southwest tropical Pacific, is more ambiguous than what could be expected from previous studies (e.g., de Andrade et al., 2018; Marshall et al., 2011). Indeed, an active MJO in the initial conditions is not necessarily associated with higher skill. First, it depends on the location within the domain. Second, the extent of the areas for which predictability is higher in inactive MJO is greater than the extent of the areas for which predictability is higher in active MJO. MJO-related forecasts of opportunity are not the same for the whole domain, so they need to be identified in relation with the precise location under study.

One hypothesis explaining the contrast between this result and previous works lies in the fact that the southwest tropical Pacific is located in the track of the MJO convective envelope. In the regions considered in previous studies (e.g., Jones et al., 2011; Vignaud et al., 2017a), the link between the MJO and precipitation mostly relies on teleconnections. Under the assumption that the initial MJO activity also reflects the MJO activity in the model at Weeks 3–4 (Vitart, 2017), an active MJO implies specific rainfall anomaly patterns in the southwest tropical Pacific, and these patterns might be misplaced by the model. We conjecture this is the reason why, at some locations of our domain, higher skill is observed at large lead times when the MJO is initially inactive.

This work, by confirming and extending the results of previous studies, advocates for the use of the multimodel ensemble approach in subseasonal prediction studies when data are already available. However, the benefits of the MME relative to the best individual models should be clarified. Investigating this matter in further detail would require (1) to run larger ensembles for some individual S2S models and (2) to build a more consistent multimodel ensemble by defining common reforecast start dates and periods in the S2S database, similarly to what has been done in several past seasonal forecasting initiatives (e.g. DEMETER) (3) to carry out additional verification diagnostics for other areas and predictands. Finally, aside from the MME approach, the role of ENSO and the MJO presented in this paper calls for attempts to improve the prediction of their impacts in the models through both dynamical modeling and statistical-dynamical approaches.

#### Acknowledgments

The S2S reforecasts were accessed through the ECMWF MARS data portal (<https://apps.ecmwf.int/datasets/data/s2s/>). The MSWEP precipitation data is developed by H. Beck (Beck et al., 2017) at Princeton University and is available at <https://www.gloh2o.org/>. MJO RMM indices from the Australian Bureau of Meteorology were retrieved from <https://www.bom.gov.au/climate/mjo/> and the values of the CPC Oceanic Niño Index were obtained from [https://origin.cpc.ncep.noaa.gov/products/analysis\\_monitoring/ensostuff/ONI\\_v5.php](https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php). All calculations were performed using the open source language R. The functions in the package *s2dverification* (Manubens et al., 2018) were used for data manipulation. Area under the ROC curve calculations were performed using the *roc* function from the *pROC* package (Robin et al., 2011). We thank the four anonymous reviewers and the editor, whose comments helped to substantially improve the former versions of this work.

#### References

- Beck, H. E., van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., & de Roo, A. (2017). MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. *Hydrology and Earth System Sciences*, 21(1), 589–615. <https://doi.org/10.5194/hess-21-589-2017>
- Bombardi, R. J., Pegion, K. V., Kinter, J. L., Cash, B. A., & Adams, J. M. (2017). Sub-seasonal predictability of the onset and demise of the rainy season over monsoonal regions. *Frontiers in Earth Science*, 5, 1–17. <https://doi.org/10.3389/feart.2017.00014>
- Coelho, C. A. S., Firpo, M. A. F., & de Andrade, F. M. (2018). A verification framework for South American sub-seasonal precipitation predictions. *Meteorologische Zeitschrift*, 27(6), 503–520. <https://doi.org/10.1127/metz/2018/0898>
- de Andrade, F. M., Coelho, C. A. S., & Cavalcanti, I. F. A. (2018). Global precipitation hindcast quality assessment of the Subseasonal to Seasonal (S2S) prediction project models. *Climate Dynamics*, 52, 5451–5475. <https://doi.org/10.1007/s00382-018-4457-z>
- Ferrone, A., Mastrangelo, D., & Malguzzi, P. (2017). Multimodel probabilistic prediction of 2m-temperature anomalies on the monthly timescale. *Advances in Science and Research*, 14, 123–129. <https://doi.org/10.5194/asr-14-123-2017>
- Griffiths, G. M., Salinger, M. J., & Leleu, I. (2003). Trends in extreme daily rainfall across the South Pacific and relationship to the South Pacific Convergence Zone. *International Journal of Climatology*, 23(8), 847–869. <https://doi.org/10.1002/joc.923>
- Guemas, V., Auger, L., & Doblas-Reyes, F. J. (2014). Hypothesis testing for autocorrelated short climate time series. *Journal of Applied Meteorology and Climatology*, 53(3), 637–651. <https://doi.org/10.1175/JAMC-D-13-064.1>
- Hagedorn, R., Doblas-Reyes, F. J., & Palmer, T. N. (2005). The rationale behind the success of multi-model ensembles in seasonal forecasting - I. Basic concept. *Tellus A*, 57(3), 219–233. <https://doi.org/10.1111/j.1600-0870.2005.00103.x>
- Hudson, D., Alves, O., Hendon, H. H., & Marshall, A. G. (2011). Bridging the gap between weather and seasonal forecasting: Intraseasonal forecasting for Australia. *Quarterly Journal of the Royal Meteorological Society*, 137(656), 673–689. <https://doi.org/10.1002/qj.769>
- Jones, C., Gottschalk, J., Carvalho, L. M. V., & Higgins, W. (2011). Influence of the Madden-Julian oscillation on forecasts of extreme precipitation in the contiguous United States. *Monthly Weather Review*, 139(2), 332–350. <https://doi.org/10.1175/2010MWR3512.1>
- Karpechko, A. Y., Charlton Perez, A., Balmaseda, M., Tyrrell, N., & Vitart, F. (2018). Predicting sudden stratospheric warming 2018 and its climate impacts with a multimodel ensemble. *Geophysical Research Letters*, 45, 13,538–13,546. <https://doi.org/10.1029/2018GL081091>
- Krishnamurti, T. N. (1999). Improved weather and seasonal climate forecasts from multimodel superensemble. *Science*, 285, 1548–1550. <https://doi.org/10.1126/science.285.5433.1548>
- Lefort, T. (2005). Starting up medium-range forecasting for New Caledonia in the South-West Pacific Ocean—A not so boring tropical climate. ECMWF Newsletter No. 102 - Winter 2004/05, 2–7.

- Li, S., & Robertson, A. W. (2015). Evaluation of submonthly precipitation forecast skill from global ensemble prediction systems. *Monthly Weather Review*, *143*(7), 2871–2889. <https://doi.org/10.1175/MWR-D-14-00277.1>
- Liang, P., & Lin, H. (2018). Sub-seasonal prediction over East Asia during boreal summer using the ECCO monthly forecasting system. *Climate Dynamics*, *50*(3-4), 1007–1022. <https://doi.org/10.1007/s00382-017-3658-1>
- Manubens, N., Caron, L.-P., Hunter, A., Bellprat, O., Exarchou, E., & Fukar, N. S. (2018). An R package for climate forecast verification. *Environmental Modelling & Software*, *103*, 29–42. <https://doi.org/10.1016/j.envsoft.2018.01.018>
- Marshall, A. G., & Hendon, H. H. (2015). Subseasonal prediction of Australian summer monsoon anomalies. *Geophysical Research Letters*, *42*, 10,913–10,919. <https://doi.org/10.1002/2015GL067086>
- Marshall, A. G., Hudson, D., Hendon, H. H., Pook, M. J., Alves, O., & Wheeler, M. C. (2014). Simulation and prediction of blocking in the Australian region and its influence on intra-seasonal rainfall in POAMA-2. *Climate Dynamics*, *42*(11-12), 3271–3288. <https://doi.org/10.1007/s00382-013-1974-7>
- Marshall, A. G., Hudson, D., Wheeler, M. C., Hendon, H. H., & Alves, O. (2011). Assessing the simulation and prediction of rainfall associated with the MJO in the POAMA seasonal forecast system. *Climate Dynamics*, *37*(11-12), 2129–2141. <https://doi.org/10.1007/s00382-010-0948-2>
- Matthews, A. J. (2012). A multiscale framework for the origin and variability of the South Pacific Convergence Zone. *Quarterly Journal of the Royal Meteorological Society*, *138*(666), 1165–1178. <https://doi.org/10.1002/qj.1870>
- Matthews, A. J., Hoskins, B. J., Slingo, J. M., & Blackburn, M. (1996). Development of convection along the SPCZ within a Madden-Julian oscillation. *Quarterly Journal of the Royal Meteorological Society*, *122*(531), 669–688. <https://doi.org/10.1002/qj.49712253106>
- Matthews, A. J., & Li, H. Y. Y. (2005). Modulation of station rainfall over the western Pacific by the Madden-Julian oscillation. *Geophysical Research Letters*, *32*, L14827. <https://doi.org/10.1029/2005GL023595>
- Murphy, A. H. (1993). What is a good forecast? An essay on the nature of goodness in weather forecasting. *Weather and Forecasting*, *8*(2), 281–293. [https://doi.org/10.1175/1520-0434\(1993\)008<0281:WIAGFA>2.0.CO;2](https://doi.org/10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2)
- Palmer, T. N. (2002). The economic value of ensemble forecasts as a tool for risk assessment: From days to decades. *Quarterly Journal of the Royal Meteorological Society*, *128*(581), 747–774. <https://doi.org/10.1256/0035900021643593>
- Palmer, T. N., Alessandri, A., Andersen, U., Cantelaube, P., Davey, M., & Délecluse, P. (2004). Development of a European Multimodel Ensemble System for Seasonal-to-Interannual Prediction (DEMETER). *Bulletin of the American Meteorological Society*, *85*(6), 853–872. <https://doi.org/10.1175/BAMS-85-6-853>
- Pegion, K., Kirtman, B. P., Becker, E., Collins, D. C., LaJoie, E., & Burgman, R. (2019). The Subseasonal Experiment (SubX): A multi-model subseasonal prediction experiment. *Bulletin of the American Meteorological Society*, *100*(10), 2043–2060. <https://doi.org/10.1175/BAMS-D-18-0270.1>
- Richardson, D. S. (2000). Skill and relative economic value of the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, *56*(3), 649–667. <https://doi.org/10.1002/qj.49712656313>
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12*(1), 77. <https://doi.org/10.1186/1471-2105-12-77>
- Salinger, M. J., McGree, S., Beucher, F., Power, S. B., & Delage, F. (2014). A new index for variations in the position of the South Pacific Convergence Zone 1910/11-2011/2012. *Climate Dynamics*, *43*(3-4), 881–892. <https://doi.org/10.1007/s00382-013-2035-y>
- Slingo, J., & Palmer, T. N. (2011). Uncertainty in weather and climate prediction. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *369*(1956), 4751–4767. <https://doi.org/10.1098/rsta.2011.0161>
- Vigaud, N., Robertson, A. W., & Tippett, M. K. (2017a). Multimodel ensembling of subseasonal precipitation forecasts over North America. *Monthly Weather Review*, *145*(10), 3913–3928. <https://doi.org/10.1175/MWR-D-17-0092.1>
- Vigaud, N., Robertson, A. W., Tippett, M. K., & Acharya, N. (2017b). Subseasonal predictability of boreal summer monsoon rainfall from ensemble forecasts. *Frontiers in Environmental Science*, *5*, 67. <https://doi.org/10.3389/fenvs.2017.00067>
- Vigaud, N., Tippett, M. K., & Robertson, A. W. (2018). Probabilistic skill of subseasonal precipitation forecasts for the East Africa West Asia sector during September May. *Weather and Forecasting*, *33*(6), 1513–1532. <https://doi.org/10.1175/WAF-D-18-0074.1>
- Vincent, D. G. (1994). The South Pacific Convergence Zone (SPCZ): A review. *Monthly Weather Review*, *122*(9), 1949–1970. [https://doi.org/10.1175/1520-0493\(1994\)122<1949:TSPCZA>2.0.CO;2](https://doi.org/10.1175/1520-0493(1994)122<1949:TSPCZA>2.0.CO;2)
- Vincent, E. M., Lengaigne, M., Menkes, C. E., Jourdain, N. C., Marchesio, P., & Madec, G. (2011). Interannual variability of the South Pacific Convergence Zone and implications for tropical cyclone genesis. *Climate Dynamics*, *36*(9-10), 1881–1896. <https://doi.org/10.1007/s00382-009-0716-3>
- Vitart, F. (2004). Monthly Forecasting at ECMWF. *Monthly Weather Review*, *132*(12), 2761–2779. <https://doi.org/10.1175/MWR2826.1>
- Vitart, F. (2017). Madden-Julian Oscillation prediction and teleconnections in the S2S database. *Quarterly Journal of the Royal Meteorological Society*, *143*(706), 2210–2220. <https://doi.org/10.1002/qj.3079>
- Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., & Codorean, C. (2017). The Subseasonal to Seasonal (S2S) Prediction Project Database. *Bulletin of the American Meteorological Society*, *98*(1), 163–173. <https://doi.org/10.1175/BAMS-D-16-0017.1>
- Weisheimer, A., & Palmer, T. N. (2014). On the reliability of seasonal climate forecasts. *Journal of The Royal Society Interface*, *11*(96), 1–10. <https://doi.org/10.1098/rsif.2013.1162>
- Wheeler, M. C., & Hendon, H. H. (2004). An All-Season Real-Time Multivariate MJO Index: Development of an index for monitoring and prediction. *Monthly Weather Review*, *132*(8), 1917–1932. [https://doi.org/10.1175/1520-0493\(2004\)132<1917:AARMMI>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2)
- Wheeler, M. C., Zhu, H., Sobel, A. H., Hudson, D., & Vitart, F. (2017). Seamless precipitation prediction skill comparison between two global models: Seamless precipitation prediction skill in two global models. *Quarterly Journal of the Royal Meteorological Society*, *143*(702), 374–383. <https://doi.org/10.1002/qj.2928>
- White, C. J., Carlsen, H., Robertson, A. W., Klein, R. J. T., Lazo, J. K., & Kumar, A. (2017). Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteorological Applications*, *24*(3), 315–325. <https://doi.org/10.1002/met.1654>
- Wilks, D. S. (2006). *Statistical methods in the atmospheric sciences* (2nd ed.). Boston, MA: Academic Press. 9780127519661.
- Zhang, C. (2005). Madden-Julian oscillation. *Reviews of Geophysics*, *43*, 1–36. <https://doi.org/10.1029/2004RG000158>
- Zhang, C. (2013). Madden-Julian oscillation: Bridging weather and climate. *Bulletin of the American Meteorological Society*, *94*(12), 1849–1870. <https://doi.org/10.1175/BAMS-D-12-00026.1>
- Zhu, H., Wheeler, M. C., Sobel, A. H., & Hudson, D. (2014). Seamless precipitation prediction skill in the tropics and extratropics from a global model. *Monthly Weather Review*, *142*(4), 1556–1569. <https://doi.org/10.1175/MWR-D-13-00222.1>