



**HAL**  
open science

# Improving subseasonal precipitation forecasts through a statistical–dynamical approach : application to the southwest tropical Pacific

Damien Specq, Lauriane Batté

## ► To cite this version:

Damien Specq, Lauriane Batté. Improving subseasonal precipitation forecasts through a statistical–dynamical approach : application to the southwest tropical Pacific. *Climate Dynamics*, 2020, 55 (7-8), pp.1913-1927. 10.1007/s00382-020-05355-7 . meteo-03352976

**HAL Id: meteo-03352976**

**<https://meteofrance.hal.science/meteo-03352976>**

Submitted on 23 Sep 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.

1 **Improving subseasonal precipitation forecasts through a**  
2 **statistical-dynamical approach : application to the southwest**  
3 **tropical Pacific**

4 **Damien Specq · Lauriane Batté**

5  
6 Received: date / Accepted: date

7 **Abstract** Subseasonal forecasts are based on coupled general circulation models that of-  
8 ten have a good representation of large-scale climate drivers affecting rainfall. Yet, they  
9 have more difficulty in providing accurate precipitation forecasts. This study proposes a  
10 statistical-dynamical post-processing scheme based on a bayesian framework to improve  
11 the quality of subseasonal forecasts of weekly precipitation. The method takes advantage of  
12 dynamically-forecast precipitation (calibration) and large-scale climate features (bridging)  
13 to enhance forecast skill through a statistical model. It is applied to the austral summer pre-  
14 cipitation reforecasts in the southwest tropical Pacific, using the Météo-France and ECMWF  
15 reforecasts in the Subseasonal-to-seasonal (S2S) database. The large-scale predictors used  
16 for bridging are climate indices related to El Niño Southern Oscillation and the Madden-  
17 Julian Oscillation, that are the major sources of predictability in the area. Skill is assessed  
18 with a Mean Square Skill Score for deterministic forecasts, while probabilistic forecasts of  
19 heavy rainfall spells are evaluated in terms of discrimination (ROC skill score) and reliabil-  
20 ity. This bayesian method leads to a significant improvement of all metrics used to assess  
21 probabilistic forecasts at all lead times (from week 1 to week 4). In the case of the Météo-  
22 France S2S system, it also leads to strong error reduction. Further investigation shows that  
23 the calibration part of the method, using forecast precipitation as a predictor, is necessary  
24 to achieve any improvement. The bridging part, and particularly the ENSO-related informa-  
25 tion, also provides additional discrimination skill, while the MJO-related information is not  
26 really useful beyond week 2 over the region of interest.

27 **Keywords** Subseasonal prediction · Bayesian statistical post-processing · Calibration ·  
28 Bridging · El Niño Southern Oscillation · Madden-Julian Oscillation

29 **1 Introduction**

30 Subseasonal forecasts (from two weeks to two months) have met growing interest in the  
31 last few years, following the launch of the WWRP/WCRP Subseasonal-to-seasonal (S2S)

---

D. Specq · L. Batté  
Centre National de Recherches Météorologiques  
Université de Toulouse, Météo-France, CNRS  
42, avenue Gaspard Coriolis 31000 Toulouse, FRANCE  
E-mail: damien.specq@meteo.fr

prediction project (Vitart et al., 2017). Although it is usually considered a challenging time scale due to the chaotic nature of the atmosphere (Vitart, 2004), recent modeling progress combined with a better understanding of S2S sources of predictability have led to significant improvements in forecast skill (Vitart, 2014).

These sources of predictability include atmospheric boundary conditions varying slowly at subseasonal time scales, such as sea surface temperatures, soil moisture, snow cover and sea ice extent (Robertson and Vitart, 2019). They also include atmospheric waves and oscillations, the most important being the Madden-Julian Oscillation (MJO, Zhang, 2013). Finally, stratospheric processes, such the Quasi-Biennial Oscillation (QBO, Lim et al., 2019) and Sudden Stratospheric Warmings (SSW, Karpechko et al., 2018) are also involved. Under certain conditions, called "windows of opportunity", these sources of predictability may enhance prediction skill, not only locally but also remotely through atmospheric teleconnections, e.g active MJO conditions enhancing North Atlantic Oscillation (NAO) predictability (Vitart, 2014).

Consequently, a large number of studies show that S2S systems skillfully predict some meteorological variables at weeks 3 and 4, for instance temperature (e.g Liang and Lin, 2018; Tian et al., 2017; Mastrangelo and Malguzzi, 2019; Wang and Robertson, 2019; Pegion et al., 2019), the MJO (Vitart, 2017; Kim et al., 2018; Marshall and Hendon, 2019) and the NAO (Vitart, 2014). On the other hand, although it may extend beyond weather time scales, the ability to predict precipitation is often lower than that of prognostic variables such as temperature. This has been illustrated for a wide range of geographical areas, e.g North America (Vigaud et al., 2017a), the contiguous United States (Tian et al., 2017), monsoonal regions (Vigaud et al., 2017b), East Africa (Vigaud et al., 2018), East Asia (Liang and Lin, 2018), South America (Coelho et al., 2018), Australia (Hudson et al., 2011; Marshall et al., 2011; Marshall and Hendon, 2015), the southwest tropical Pacific (Specq et al., 2020, in press), and more generally at a global scale (de Andrade et al., 2018; Mastrangelo and Malguzzi, 2019). Yet, indications of significant skill can be found beyond week 2 when considering specific regions, such as the equatorial Pacific where precipitation is directly constrained by the state of El Niño Southern Oscillation (ENSO, de Andrade et al., 2018), or when considering specific rainfall indicators, such as the monsoon onset date (Bombardi et al., 2017) and the monthly accumulated precipitation (Tian et al., 2017).

Since S2S systems are generally able to forecast some large-scale climate features satisfactorily beyond short-range time scales (e.g Vitart, 2017), and since these climate features are known to impact precipitation patterns in wide areas of the globe, it can be assumed there is room for improvement of subseasonal precipitation forecasts through statistical post-processing. Indeed, as pointed out by Schepen et al. (2014) for the seasonal time scale, large-scale climate features in GCMs (General Circulation Models) can be partly disconnected to rainfall, because rainfall is influenced by processes that need to be parameterized as they occur at a much smaller spatial scale than the GCMs resolution. Similarly, S2S rainfall ensemble forecasts generally exhibit biases related to systematic errors and are poorly reliable (in terms of probabilities) beyond week 2 (e.g Vigaud et al., 2017a).

On the other hand, there is a growing demand of subseasonal forecasts among a wide variety of users, such as the energy, agriculture, water management, finance and insurance sectors (Hudson et al., 2011; White et al., 2017). These users obviously require skillful, reliable and bias-free forecasts. As a result, statistical correction and post-processing of S2S forecasts is a crucial matter, in particular for a parameterized variable like rainfall. Two general and complementary approaches can be developed for this post-processing: calibration and bridging (Schepen et al., 2014). A statistical-dynamical post-processing scheme is called calibration when the predictor for the observed variable (i.e precipitation) is the raw

81 variable itself as forecast by the dynamical forecasting system. It is referred to as bridging  
82 when the predictors are large-scale climate features, that are also forecast by the dynamical  
83 system. The link between predictors and the predicted variable is established through a  
84 statistical model that may have the same mathematical formulation for both calibration and  
85 bridging.

86 Calibration and bridging have been jointly applied in a large number of studies related to  
87 seasonal forecasting of temperature or precipitation (e.g. Schepen et al., 2012, 2014, 2016;  
88 Strazzo et al., 2019), and recent studies have started carrying out this work for S2S time  
89 scales. In terms of calibration-oriented approaches for precipitation, Li et al. (2019) have  
90 proposed a bias correction methodology for S2S precipitation at the scale of hydrological  
91 catchments, while Vigaud et al. (2019) have proposed a spatial correction methodology for  
92 multimodel subseasonal precipitation forecasts using local Laplacian eigenfunctions. Like-  
93 wise, Doss-Gollin et al. (2018) have tested several Model Output Statistics (MOS) methods  
94 for S2S precipitation prediction over central South America that also lead to better cali-  
95 bration. On the other hand, similar to the bridging approach, statistical-only subseasonal  
96 forecasts based on large-scale climate features have been developed using machine learning  
97 (Cohen et al., 2019, for temperature and precipitation) or empirical models (Johnson et al.  
98 2014 for temperature, Baggett et al. 2018 for tornadoes and hail).

99 In this study, we introduce a simple bayesian statistical-dynamical scheme dealing with  
100 both calibration and bridging in the same framework. This scheme is applied to precipita-  
101 tion forecast in the southwest tropical Pacific (SWTP) with the Météo-France and ECMWF  
102 S2S systems. It uses ENSO and MJO-related predictors for bridging, along with the fore-  
103 cast precipitation at grid point level for calibration. The SWTP domain ranges from 110°E  
104 to 200°E in longitude, and 30°S to 0° in latitude. Rainfall in this area is a crucial matter  
105 with the regular occurrence of tropical cyclones and heavy rainfall spells (McGree et al.,  
106 2014) but also droughts (McGree et al., 2016), affecting Pacific island territories as well as  
107 Australia. The focus is put on the austral summer season (December, January and February)  
108 as it corresponds to the peak season for MJO activity in the Southern Hemisphere (Zhang,  
109 2005), to the wet season in southwest Pacific island territories and to the monsoon season  
110 in Northern Australia (Marshall and Hendon, 2015). It has been widely demonstrated that  
111 precipitation patterns in the area are strongly influenced both by ENSO, as a slowly-varying  
112 process for the subseasonal time scales, and by the MJO (e.g. de Andrade et al., 2018). Since  
113 these phenomena are recognized to be well represented in subseasonal forecasting systems,  
114 using them as predictors in the statistical-dynamical scheme should provide additional in-  
115 formation and improve precipitation forecast skill compared to the raw precipitation output  
116 from the numerical models.

117 This article is structured as follows. Section 2 is dedicated to the description of the S2S  
118 reforecasts and the rest of the data used for this study. Section 3 goes into details of the  
119 statistical-dynamical approach implemented here and introduces the framework for forecast  
120 verification and comparison. Section 4 presents the main improvements obtained with the  
121 statistical-dynamical scheme in terms of forecast skill and reliability compared to the raw  
122 S2S forecasts, while Section 5 discusses the relative importance of each large-scale predictor  
123 used for bridging. Finally, Section 6 summarizes and discusses the main results of the article.

**Table 1** Reforecast attributes for the Météo-France and ECMWF systems

Attributes	ECMWF	MF
Time range	Day 1-46	Day 1-61
Atmospheric resolution	T639/319 L91	T255L91
Reforecast	On the fly	Fixed
Reforecast period	Past 20 yrs (1996-2015)	1993-2014
Reforecast frequency	Two per week	Four per month
Ensemble number	11	15
Coupling	Ocean and sea ice	Ocean and sea ice

**Table 2** Correspondence between the Météo-France and the ECMWF reference start dates

MF	12-01	12-08	12-15	12-22	01-01	01-08	01-15	01-22	02-01	02-08	02-15	02-22
ECMWF	12-01	12-08	12-15	12-22	12-31	01-07	01-14	01-21	02-01	02-08	02-15	02-22

## 124 2 Data

### 125 2.1 S2S reforecasts

126 The ensemble reforecasts from Météo-France (MF) and the European Centre for Medium-  
 127 range Weather Forecasts (ECMWF) were extracted from the S2S database (Vitart et al.,  
 128 2017). These two systems are both based on coupled GCMs and their main features are  
 129 summarized in Table 1. Their output data was extracted on a common  $1.5^\circ$  grid on the  
 130 SWTP domain ( $110^\circ\text{E}$ - $200^\circ\text{E}$ ;  $30^\circ\text{S}$ - $0^\circ$ ). The  $1.5^\circ$  resolution corresponds to the common  
 131 initial archiving resolution of the S2S database and has already been adopted in other S2S  
 132 studies (e.g Vigaud et al., 2017a).

133 In order to make a proper comparison between these two systems, it is essential to  
 134 choose a common reforecast period for verification and implementation of the statistical-  
 135 dynamical approach. The MF system uses a fixed reforecast period (1993-2014) while the  
 136 ECMWF system updates its reforecast for every real-time forecast by running the system at  
 137 the same calendar date over the past 20 years. We therefore chose to consider the 19-year  
 138 reforecast period 1996-2014, using the 2016 version of the ECMWF reforecasts.

139 In addition, we restrained our selection to the start dates in the DJF season. However,  
 140 these start dates also differ between the two systems (e.g MF starts on January 8 while  
 141 ECMWF starts on January 7) and ECMWF has actually more start dates (two per week) than  
 142 MF does (four per month). As we chose to keep all DJF start dates from the MF reforecast,  
 143 we only retained the closest corresponding start dates in the ECMWF reforecast in order  
 144 to compare the two systems' performance with the same sample size. This correspondence  
 145 is specified in Table 2. This leads to 216 start dates (18 years x 3 months x 4 startdates  
 146 per month) under consideration for each S2S system. Both systems are evaluated separately  
 147 on weekly windows after discarding the first four days since they belong to short-range  
 148 forecasting. If  $d$  is the day of the start date, the weekly verification calendar includes week  
 149 1 ( $[d + 5, d + 11]$ ) to week 4 ( $[d + 26, d + 32]$ ).

## 150 2.2 Precipitation data

151 S2S precipitation data for MF and ECMWF is extracted on the  $1.5^\circ$  grid over the domain  
152  $110^\circ\text{E}$ - $200^\circ\text{E}$ ;  $30^\circ\text{S}$ - $0^\circ$  ( $60 \times 21$  grid points) and averaged on weekly windows according to  
153 the procedure described in Section 2.1. It is verified against a reference precipitation dataset  
154 called Multi-Source Weighted-Ensemble Precipitation (MSWEP) version 1.2 (Beck et al.,  
155 2017). This dataset is available for the 1979-2015 period and covers the whole globe on a  
156  $0.25^\circ$  grid. It combines various precipitation sources: satellite data, World Meteorological  
157 Organization Global Telecommunication System (WMO GTS) rain gauges and reanalysis.  
158 MSWEP data was interpolated on the  $1.5^\circ$  reforecast grid through conservative remapping,  
159 and temporally averaged along the same weekly windows as the reforecasts.

## 160 2.3 Large-scale predictors

161 In this study, we considered the two most widely recognized sources of subseasonal pre-  
162 dictability for precipitation in the SWTP, ENSO and the MJO. We decided to characterize  
163 the ENSO state at a given day by the sea surface temperature anomalies, spatially-averaged  
164 over the standard Niño 3.4 box ( $170^\circ\text{W}$ - $120^\circ\text{W}$ ;  $5^\circ\text{S}$ - $5^\circ\text{N}$ ), and temporally-averaged over  
165 the previous 90 days. This index will be noted  $N34$  hereafter. Other ENSO-related indices  
166 were tested but showed little difference for the purpose of our statistical-dynamical scheme.  
167 As for the MJO, it is represented by the two indices  $RMM1$  and  $RMM2$  defined by Wheeler  
168 and Hendon (2004) and corresponding to the first two principal components of the mul-  
169 tivariate principal component analysis carried out on the meridionally-averaged fields of  
170 Outgoing Longwave Radiation (OLR) and zonal winds at 850 (U850) and 200 hPa (U200)  
171 in the  $15^\circ\text{S}$ - $15^\circ\text{N}$  equatorial band.

172 The determination of the indices in the S2S reforecasts first requires to compute anoma-  
173 lies of a given field (SST for  $N34$ , OLR, U850 and U200 for  $RMM1$  and  $RMM2$ ). The  
174 anomalies are calculated for every daily lead time by leave-one-year-out cross-validation:  
175 for a given start date in a given year and at a given lead, the anomaly is the variable in the re-  
176 forecast minus the averaged variable at this start date and lead in the reforecasts of all other  
177 years. Moreover, the large-scale predictors are also computed in the ERA-Interim reanalysis  
178 (Dee et al., 2011) to train the statistical model with real-world data. In order to be consistent  
179 between reforecast and reference data, the anomalies in ERA-Interim are computed with  
180 the same method by subtracting the average variable for the same calendar day in all other  
181 years of the 1996-2014 period. Finally, the computation of the indices in the S2S refore-  
182 casts often requires to subtract previous values averaged over a period that is longer than the  
183 length of the runs (90 days for  $N34$ , 120 days for MJO indices). Then, for the  $n$ -th day of  
184 the run, we subtract the average value composed of the  $n - 1$  previous days of the run and  
185 the  $90 - (n - 1)$  (respectively  $120 - (n - 1)$ ) days prior to the start dates in ERA-Interim.

186 Using this procedure, the MJO indices in the S2S reforecasts are computed following  
187 the recommendations of Gottschalck et al. (2010) and the ECMWF technical note available  
188 on the S2S website (ECMWF, 2017). As for the daily  $N34$  index at day  $d$ , it is obtained by  
189 averaging the spatially-averaged SST anomalies over the 90 days from day  $d - 89$  to day  
190  $d$ . Finally, although these indices are initially computed as daily values, they are averaged  
191 over the weekly windows of the S2S reforecast calendar (described in Section 2.1) for the  
192 implementation of the statistical-dynamical scheme.

### 3 Methods

#### 3.1 Statistical-dynamical prediction

##### 3.1.1 General framework

The bayesian approach proposed in this study is similar to that proposed by Coelho et al. (2004) that was also applied and completed by Luo et al. (2007), denoted L07 hereafter. Since this approach was initially designed for seasonal forecasting, we have adapted it to subseasonal forecasting. Let's consider that we intend to forecast the distribution of a predictand variable  $y$  conditional upon a set of  $n$  predictors  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ , noted  $p(y|\mathbf{x})$ . According to Bayes' theorem,

$$p(y|\mathbf{x}) = \frac{p(\mathbf{x}|y)p(y)}{p(\mathbf{x})} . \quad (1)$$

$p(y)$  is the prior distribution of the predictand  $y$  before any information is known from a specific forecast.  $p(\mathbf{x}|y)$  is called the likelihood: it expresses how  $\mathbf{x}$  is distributed when the predictand  $y$  is known.  $p(y|\mathbf{x})$  is the updated distribution of  $y$  when information about  $\mathbf{x}$  is available.

L07 have shown that an explicit formulation of  $p(y|\mathbf{x})$  is possible in a gaussian framework under the following assumptions:

1. The prior distribution  $p(y)$  is normal:  $p(y) \sim \mathcal{N}(\mu_p, \sigma_p^2)$
2. For each predictor  $x_i$ ,  $p(x_i|y)$  is a normal distribution determined through a linear regression:

$$p(x_i|y) \sim \mathcal{N}(a_i + b_i y, \sigma_i^2) , \quad (2)$$

where  $a_i$  is the intercept,  $b_i$  is the slope and  $\sigma_i^2$  is the variance of the residuals.

3. The residuals in the linear regressions from Equation (2) are independent from each other.

Then, according to L07,  $p(y|\mathbf{x})$  also follows a normal distribution, noted  $p(y|x) \sim \mathcal{N}(\mu_t, \sigma_t^2)$ , and the parameters of this conditional distribution are expressed as follows:

$$\frac{1}{\sigma_t^2} = \frac{1}{\sigma_p^2} + \sum_{i=1}^n \frac{b_i^2}{\sigma_i^2} \quad (3)$$

$$\frac{\mu_t}{\sigma_t^2} = \frac{\mu_p}{\sigma_p^2} + \sum_{i=1}^n \frac{b_i^2}{\sigma_i^2} \left( \frac{x_i - a_i}{b_i} \right) \quad (4)$$

##### 3.1.2 Scheme implementation

In this study, the predictors are the weekly-averaged forecast large-scale indices  $N34$ ,  $RMM1$ ,  $RMM2$ , and the weekly-averaged forecast rainfall  $r_f$ . The predictand variable is the weekly-averaged observed rainfall  $r_o$ . Assumption n°3 in the previous Section 3.1.1 suggests that any dependency between the predictors should be removed. For this purpose, since it has been widely shown that cross-timescale interactions between ENSO and the MJO might exist (e.g Doss-Gollin et al., 2018), it should be relevant to use the principal components of

223 ( $N34, RMM1, RMM2$ ) instead of the original indices. However, on a practical point of view,  
 224 this approach proves to be less relevant in terms of skill improvement (see Supplementary  
 225 material Figure S1, as compared to Figure 2). Then, it will not be further applied and the  
 226 original large-scale indices are kept as predictors.

227 Furthermore, as in L07, we need to overcome the fact that the rain rates  $r_o$  and  $r_f$  are  
 228 non-gaussian. This issue is tackled in a similar manner: the weekly rain rates  $r_o$  and  $r_f$   
 229 are converted into normally distributed variables of mean 0 and variance 1, noted  $\hat{r}_o$  and  
 230  $\hat{r}_f$  hereafter. This transformation is made up of two steps using the equal-quantile method:  
 231 application of the climatological Cumulative Distribution Function (CDF) of rain rates, fol-  
 232 lowed by the inverse CDF of the normal distribution  $\mathcal{N}(0, 1)$ .

233 Such a transformation is always mathematically feasible and leads to a normally-distributed  
 234 variable most of the time. One limitation appears in dry areas where the CDF for a 0 mm rain  
 235 rate is far greater than the 0% quantile, but it only applies to a small fraction of the SWTP do-  
 236 main (continental Australia, see Section 4.2). In such cases, the transformed variable cannot  
 237 be fully gaussian as it does not cover the full range of the normal distribution. For instance,  
 238 if 0 mm represents the lower 20% of the climatology, the transformation will necessarily  
 239 lead to values greater than the 20th percentile of the  $\mathcal{N}(0, 1)$  distribution (i.e.  $\sim -0.84$ ).

240 The climatological CDF of rain rates is determined separately for reforecast and obser-  
 241 vations with a leave-one-year-out cross-validation approach similar to Vigaud et al. (2017a).  
 242 In the observations, for a given target week, the same calendar weeks in all other 18 years  
 243 are considered, along with the week before and the week after. This leads to a climatologi-  
 244 cal sample of  $18 \times 3 = 54$  values for the target week, from which an empirical CDF is built.  
 245 Likewise, the S2S reforecast climatology for a given year, start date and lead time is con-  
 246 structed by taking the reforecasts for the same start date and lead time considering the 18  
 247 remaining years and all ensemble members. This leads to a climatological sample of  $18 \times n$   
 248 values to establish the empirical CDF, where  $n$  the ensemble size (15 members for MF and  
 249 11 for ECMWF).

250 The bayesian scheme described in Section 3.1.1 is implemented with  $y = \hat{r}_o$  and  $\mathbf{x} =$   
 251  $(\hat{r}_f, N34, RMM1, RMM2)$ . For the sake of robustness, we take the ensemble mean of the  
 252 forecast predictors, including  $\hat{r}_f$ , and do not consider the spread of the predictor around  
 253 the ensemble mean. According to the normalizing transformation from  $r_o$  to  $\hat{r}_o$ , the prior  
 254 distribution of  $\hat{r}_o$  is assumed to be  $\mathcal{N}(\mu_p = 0, \sigma_p^2 = 1)$ . Equations (3) and (4) are applied  
 255 to deduce the conditional distribution  $p(\hat{r}_o | \hat{r}_f, N34, RMM1, RMM2)$ . If needed, the resulting  
 256 mean  $\mu_i$  can be back-transformed to obtain a precipitation forecast in mm, by using the CDF  
 257 of  $\mathcal{N}(0, 1)$  followed by the inverse climatological CDF of observations.

### 258 3.1.3 Uncertainty in the predictors' forecasts

259 A major difference between the approach in L07 and ours is the inclusion of large-scale  
 260 climate features as predictors (bridging), while they only used the forecast precipitation as  
 261 a predictor for the observed precipitation. As a result, we have included a two-term decom-  
 262 position of the uncertainty  $\sigma_i^2$  in the relationship (defined in Equation (2)) between each  
 263 predictor  $x_i$  and the predictand  $y = \hat{r}_o$ . There are actually two possible values for predictor  
 264  $x_i$ : its real-world value in the reference data, noted  $x_{io}$ , and its forecast value, noted  $x_{if}$ . The  
 265 benefits of the predictor might be limited by its intrinsic predictive ability, but also by the  
 266 fact that S2S systems do not forecast it with sufficient quality.

267 The intrinsic predictive ability is established in the reference data with the linear regres-  
 268 sion:



$$p(x_{io}|y) \sim \mathcal{N}(a_{io} + b_{io}y, \sigma_{io}^2). \quad (5)$$

269 The ability of the S2S systems to forecast the predictor  $x_i$  is assessed in a second linear  
270 regression:

$$p(x_{if}|x_{io}) \sim \mathcal{N}(a_{if} + b_{if}x_{io}, \sigma_{if}^2). \quad (6)$$

271 The composition of the normal distribution from Equation (5) followed by the normal  
272 distribution from Equation (6) leads to the resulting expressions of the parameters  $a_i$ ,  $b_i$  and  
273  $\sigma_i^2$  from Equation (2):

$$a_i = a_{if} + b_{if}a_{io} \quad (7)$$

$$b_i = b_{if}b_{io} \quad (8)$$

$$\sigma_i^2 = b_{if}^2\sigma_{io}^2 + \sigma_{if}^2 \quad (9)$$

274  $\sigma_i^2$  is the sum of the uncertainty in the prediction of  $x_i$  by the S2S systems (i.e the error),  
275 noted  $\sigma_{if}^2$ , and an uncertainty term which is the intrinsic uncertainty of the predictor  $\sigma_{io}^2$ ,  
276 multiplied by  $b_{if}$ .  $b_{if}$  is proportional to the correlation score of the forecast predictor  $x_{if}$   
277 with the observed predictor  $x_{io}$ . When the predictor  $x_i$  is rainfall itself,  $\sigma_{io}^2 = 0$  and  $b_{io} = 1$   
278 so the only source of uncertainty is the forecast error  $\sigma_{if}^2$ . Both linear regressions from  
279 Equations (5) and (6) are fitted for each lead time using leave-one-year-out cross-validation.

### 280 3.2 Assessing predictor importance

281 Section 5 discusses the respective role of the each predictor in the predictive ability of the  
282 statistical-dynamical model. The importance of each predictor is assessed using likelihood-  
283 ratio tests (Buse, 1982) in a forward selection mode. Our predictive model is described by  
284 Equation (4). Because the prior probability distribution of  $\hat{r}_o$  has mean  $\mu_p = 0$ , the predictive  
285 model can be re-written:

$$\hat{r}_o = \sum_{i=1}^4 \sigma_i^2 \frac{b_i^2}{\sigma_i^2} \left( \frac{x_i - a_i}{b_i} \right), \quad (10)$$

286 This model consists in four predictive terms, the added value of which needs to be  
287 estimated. Similar to a stepwise forward predictor selection for a classical linear regression  
288 (Wilks, 2006), we start from a model where  $\hat{r}_o$  is a constant and compare this model with  
289 all the other models for which we consider each predictive term  $\sigma_i^2 \frac{b_i^2}{\sigma_i^2} \left( \frac{x_i - a_i}{b_i} \right)$  separately.  
290 The most important of the predictive terms is the one which returns the smallest p-value in  
291 the likelihood-ratio test when comparing with the constant model. In the next step of the  
292 procedure, we start from the new model including the most important predictive term, and  
293 we determine which of the remaining terms is the second most predictive, using the same  
294 p-value criterion. The procedure ends when all predictive terms have been included and it  
295 gives the order of importance of each predictive term. Using a p-value stop criterion, set  
296 at 0.05 in this study, also enables to eliminate predictive terms that do not bring additional  
297 information.

298 This procedure is carried out at each grid point with two sets of predictors:

1. Large-scale predictors *N34*, *RMM1* and *RMM2* only, in the reanalysis data, considering all lead times together
2. All predictors (including  $\hat{r}_f$ ), as forecast by S2S systems, at each lead time separately

The first option aims at identifying where each predictor influences precipitation in real-world data. The second enables to identify, at each lead time, which predictor actually brings information to the statistical-dynamical approach.

### 3.3 Verification

We use a verification framework with a deterministic and a probabilistic score, computed at every lead time (week 1 to week 4). This framework partly re-uses and adapts the verification procedure of subseasonal weekly accumulated precipitation predictions proposed by (Coelho et al., 2018) and called "all season hindcast verification". Deterministic verification is carried out with the Mean Square Skill Score (MSSS), which is defined as

$$MSSS = 1 - \frac{MSE}{MSE_{clim}}, \quad (11)$$

where  $MSE$  is the mean square error of the reforecasts, and  $MSE_{clim}$  is the mean square error of a climatological forecast that is constructed from the observations by cross-validation similarly to the procedure detailed in Section 3.1.2. When computing the MSSS, a simple cross-validated bias correction is applied to the raw forecasts so that the Mean Square Error does not include any systematic error that would lower the scores without being related to the predictive ability of the systems.

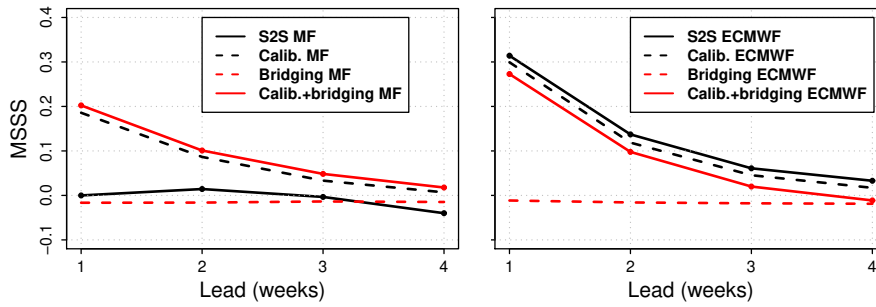
We also consider probabilistic forecasts of a binary event corresponding to weekly precipitation above the climatological upper quintile, because our main objective is to assess and improve the ability of S2S systems to detect spells of heavy rainfall. The forecast probability of the upper quintile in the raw S2S forecasts is evaluated by taking the fraction of ensemble members for which the forecast value exceeds the upper quintile threshold. For the output of the statistical-dynamical scheme, the forecast probability corresponds to the probability that a variable following a normal distribution with the parameters in Equations (3) and (4) exceeds the upper quintile of the  $\mathcal{N}(0, 1)$  normal distribution (i.e.  $\sim 0.84$ ).

The discrimination of the probabilistic forecasts (i.e. whether the forecasts correctly make the difference between occurrence and non-occurrence of the binary event) is assessed with the area under the Relative Operating Characteristic (ROC) curve  $A$ .  $A$  is rescaled as the ROC skill score (ROCSS), i.e.

$$ROCSS = 2A - 1, \quad (12)$$

so that values are comprised between 0 and 1 for forecasts that are more skillful than climatology. Finally, we also assess the reliability of the probabilistic forecasts using reliability diagrams and the reliability component of the Brier Score (Murphy, 1973).

Moreover, the verification scores MSSS and ROCSS will either be computed as a unique score at the scale of the whole SWTP domain, for the sake of brevity, or at grid point level. When the ROCSS is shown on maps, the value at a given grid point is actually computed by pooling the forecast/observation pairs for a neighborhood of nine grid points ( $3 \times 3$ ) around the central grid point. This enables to reduce spatial noise and to increase robustness of a comparison between the ROCSS of two forecasts.



**Fig. 1** Mean Square Skill Score at the scale of the SWTP basin for the subseasonal reforecasts with various statistical-dynamical post-processing: raw S2S (solid black line), calibrated (dashed black line), bridged (dashed red line) and calibrated + bridged (solid red line). Left: MF S2S system. Right: ECMWF S2S system. Note that a cross-validated bias correction has been performed to the raw S2S systems before computing the MSSS.

338 Such a comparison between forecast  $A$  and forecast  $B$  is made by taking the difference  
 339  $ROCSS_B - ROCSS_A$ . For this difference to be considered significant, we set two simultane-  
 340 ous conditions:

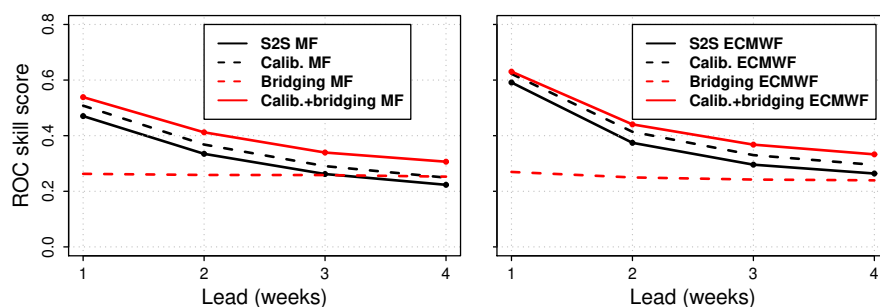
- 341 1. The greater of  $ROCSS_A$  and  $ROCSS_B$  must be significant at the 95% level according to a  
 342 Mann-Whitney U test (Wilks 2006)
- 343 2. The greater of  $ROCSS_A$  and  $ROCSS_B$  must be significantly greater than the other one, at  
 344 the 95% level, according to a one-sided DeLong test (DeLong et al., 1988)

## 345 4 Results

### 346 4.1 Comparison over the whole domain

347 Four types of subseasonal reforecasts of precipitation are assessed in this section: the raw  
 348 S2S reforecasts, the calibrated reforecasts (forecast precipitation  $\hat{r}_f$  is the only predictor),  
 349 the bridging reforecasts (large-scale predictors  $N34$ ,  $RMM1$  and  $RMM2$  without forecast  
 350 precipitation) and the calibration + bridging reforecasts (all predictors  $\hat{r}_f$ ,  $N34$ ,  $RMM1$  and  
 351  $RMM2$ ). The aim is to have an overview of the impact of each statistical-dynamical approach  
 352 on reforecast skill. For the sake of brevity, results are shown with skill scores computed by  
 353 pooling all the grid points in the SWTP domain (see Section 3.3).

354 Figure 1 illustrates such a comparison for the Mean Square Skill Score, once the mean  
 355 bias has been removed from the raw S2S reforecasts (solid black line). The first notable  
 356 result from Figure 1 is that the bridging approach alone leads to the same amount of errors  
 357 as a climatological forecast ( $MSSS = 0$ ), for both the MF and ECMWF systems. A second  
 358 notable result is that the raw MF system has apparently such large residual errors that it  
 359 does not really perform better than climatology, even after the simple bias correction. On  
 360 the contrary, this is not the case for the raw ECMWF system. A third important point is  
 361 that all approaches involving calibration exhibit good performances and less errors than  
 362 climatology up to week 4. This leads to a drastic improvement for the MF system, but  
 363 not for the ECMWF system. Indeed, the ECMWF system makes actually less error with  
 364 the simple bias correction than with any statistical-dynamical approach, a result which is  
 365 presumably due to the already good performance of the raw ECMWF reforecasts in terms



**Fig. 2** ROC skill score for the upper quintile of precipitation at the scale of the SWTP basin for the sub-seasonal reforecasts with various statistical-dynamical post-processing: raw S2S (solid black line), calibrated (dashed black line), bridged (dashed red line) and calibrated + bridged (solid red line). Left: MF S2S system. Right: ECMWF S2S system. The differences between the solid black line, the dashed black line and the solid red line are always significant at the 95% according to a one-sided DeLong test.

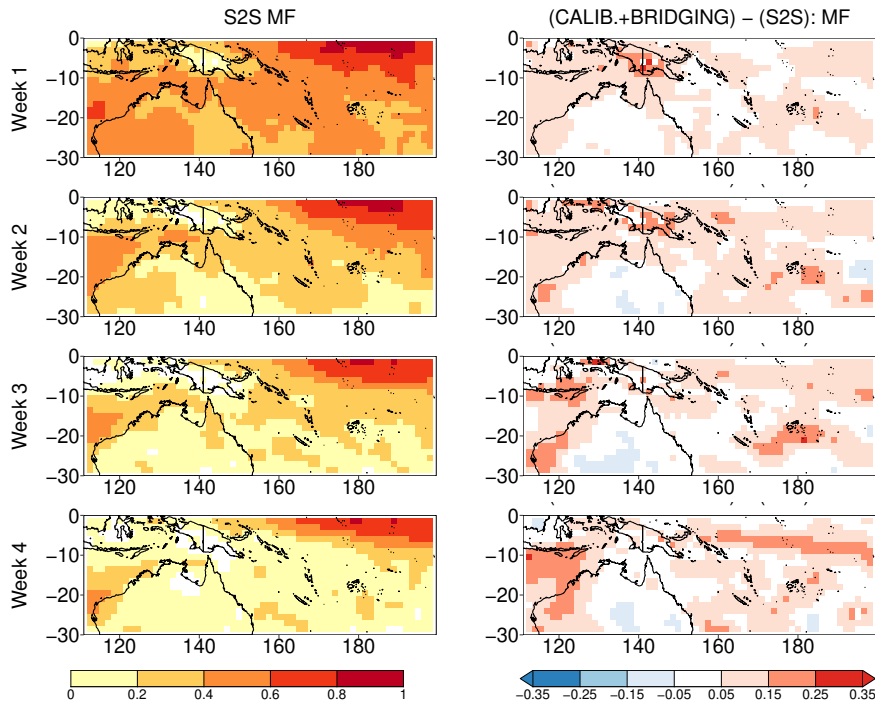
366 of MSSS. Finally, for both systems, there does not seem to be any added value of calibration  
 367 + bridging compared to calibration alone, suggesting that bridging does not bring any error  
 368 correction.

369 Figure 2 shows the same comparison for the ROC skill score of the upper quintile of  
 370 precipitation. For this score, results are similar for both S2S systems: the performance of the  
 371 calibrated reforecasts is better than that of the raw reforecasts, but the calibrated reforecasts  
 372 are also improved by the calibration + bridging approach, at least for the later lead times  
 373 (weeks 3 and 4). However, bridging alone leads to an almost constant skill score (no varia-  
 374 tion with lead) that is less than the scores obtained by any other approach. We find that this  
 375 constant score actually illustrates a baseline discrimination ability related to the  $N34$  pre-  
 376 dictor. Indeed, this predictor remains almost constant, and very close to the initial reference  
 377 value, in the first four weeks of the S2S runs (not shown).

378 The results from Figures 1 and 2 suggest that calibration is useful to error correction  
 379 to some extent, when the raw forecasts exhibit large errors, and it is also useful to improve  
 380 the detection of heavy rainfall events. In the meantime, bridging brings additional value for  
 381 discrimination, provided it is combined with calibration. These results are also confirmed  
 382 on four other S2S systems (see Supplementary material, Figures S2 and S3). Yet, it must  
 383 be acknowledged that the proposed approach is not able to make up for the skill difference  
 384 between the input dynamical models, i.e ECMWF remains better than Météo-France even  
 385 after statistical post-processing.

#### 386 4.2 Comparison of discrimination ability at grid point level

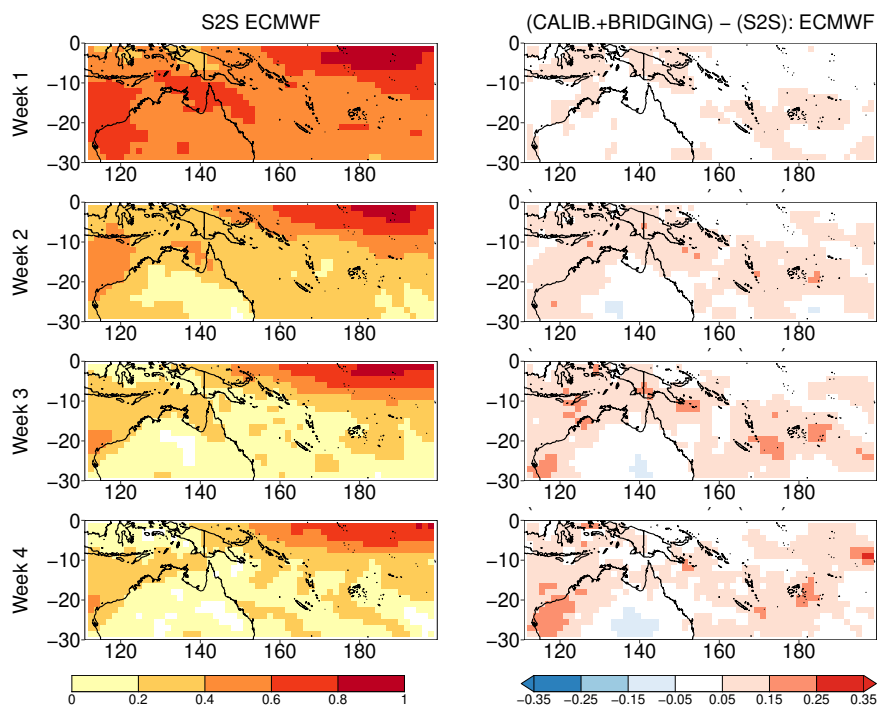
387 Figures 3 and 4 represent the difference in terms of ROC skill score between the calibration  
 388 + bridging reforecasts and the raw S2S reforecasts, using the MF and ECMWF systems  
 389 respectively. The differences are shown alongside the initial skill of the raw S2S system,  
 390 and the ROC skill scores are computed by pooling the grid points over a  $3 \times 3$  grid point  
 391 neighborhood as explained in Section 3.3. In agreement with Figure 2, the calibration +  
 392 bridging approach leads to a significant improvement of forecast skill in a large fraction of  
 393 the domain.



**Fig. 3** Left: ROC skill score for the upper quintile of weekly precipitation in the MF S2S reforecast. Right: Difference in ROC skill score between the calibration + bridging statistical-dynamical scheme and the raw MF S2S reforecast. The ROC skill score are computed by pooling values over a  $3 \times 3$  neighborhood. Grid points in white correspond to grid points where the difference is not significant according to the two criteria detailed in Section 3.3.

394 The benefits of using the statistical-dynamical approach appear to increase with lead  
 395 time. This is not surprising because the raw reforecasts have more detection failures and are  
 396 more noisy at later lead times. Consequently, there is more room for improvement through  
 397 calibration, that corrects distribution errors, and bridging, that dampens the importance of  
 398 spurious high-frequency variability relative to more relevant lower frequency variability.  
 399 The main regions of improvement are the Maritime Continent, the southern oceanic part of  
 400 the domain (including New Caledonia at  $166^\circ\text{E}$ - $21^\circ\text{S}$ , Vanuatu at  $168^\circ\text{E}$ - $17^\circ\text{S}$  and Fiji at  
 401  $178^\circ\text{E}$ - $18^\circ\text{S}$ ) and the western Australian coast.

402 One notable exception is a part of continental Australia for which there is no improve-  
 403 ment or even a decrease in skill, depending on the lead time. Further investigation (not  
 404 shown here) found this is related to the dryness of the area, for which the weekly pre-  
 405 cipitation distribution is highly positively skewed. When transforming precipitation into a  
 406 gaussian variable, small differences between precipitation values in mm may lead to larger  
 407 differences in the transformed precipitation. These differences in precipitation are not neces-  
 408 sarily related to important differences in large-scale conditions. They lead to non-significant  
 409 or inconsistent relationships between transformed precipitation and large-scale predictors,  
 410 such as  $N34$ , when fitting the linear regression in Equation (5), hence the degraded forecast  
 411 skill. This is a limitation of our methodology that only affects a restricted part of the chosen  
 412 domain.



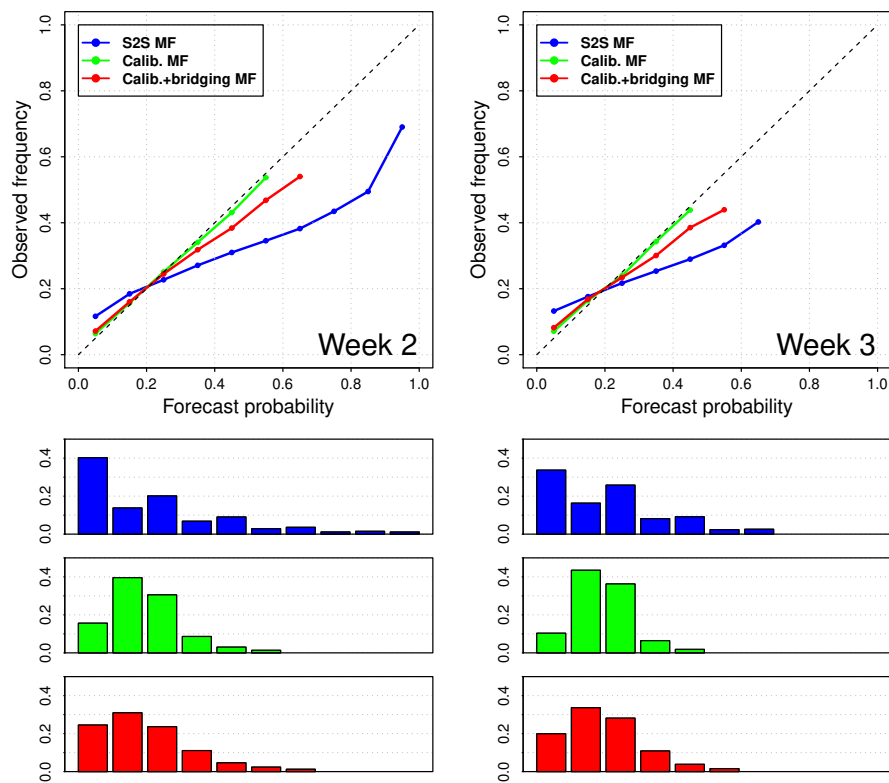
**Fig. 4** Left: ROC skill score for the upper quintile of precipitation in the ECMWF S2S reforecast. Right: Difference in ROC skill score between the calibration + bridging statistical-dynamical scheme and the raw ECMWF S2S reforecast.

#### 413 4.3 Reliability of probabilistic reforecasts

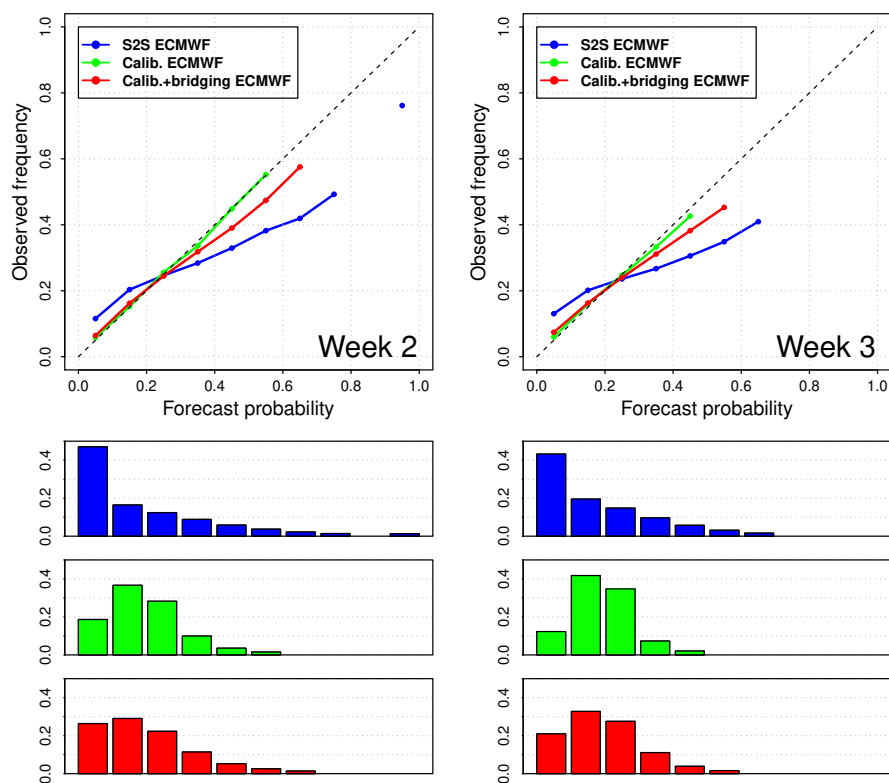
414 The overall reliability of the raw S2S reforecasts, the calibration and the calibration + bridging  
 415 reforecasts is evaluated in Figures 5 and 6, at the scale of the whole domain. These diagrams  
 416 are completed with the reliability and the resolution components of the Brier Score  
 417 in Table 3. The reliability component is negatively oriented (the smaller the more reliable)  
 418 while the resolution component is positively oriented. The two statistical-dynamical approaches  
 419 clearly outperform the raw S2S reforecasts for both reliability and resolution. The  
 420 result concerning resolution was foreseeable from Figures 2, 3 and 4, given that resolution  
 421 and discrimination (as assessed by the ROC skill score) are "two sides of the same coin"  
 422 (Bröcker, 2015), but the reliability improvement remained to be demonstrated. Moreover,  
 423 these improvements are notable for all lead times. The comparison between the calibration  
 424 and calibration + bridging diagrams would at first sight suggest that calibration alone, without  
 425 bridging, should be preferred because it leads to more reliable forecasts. However, this  
 426 would happen at the expense of resolution which is better when bridging is added, as could  
 427 have been inferred from Figure 2. Thanks to the information from large-scale predictors, the  
 428 calibration + bridging scheme is more skillful at detecting heavy precipitating events but it  
 429 also tends to be more overconfident.

**Table 3** Reliability and resolution components ( $\times 10^2$ ) of the Brier Score (whole domain) for the upper quintile of weekly precipitation at weeks 1 to 4. Reliability is negatively oriented and resolution is positively oriented.

		Week 1		Week 2		Week 3		Week 4	
		Rel.	Res.	Rel.	Res.	Rel.	Res.	Rel.	Res.
MF	Raw	1.41	2.13	0.95	1.07	0.79	0.74	0.81	0.60
	Calib.	0.020	2.34	0.0089	1.27	0.020	0.87	0.037	0.70
	Calib.+bridging	0.076	2.63	0.098	1.58	0.14	1.12	0.16	0.92
ECMWF	Raw	0.61	3.44	0.81	1.39	0.89	0.93	0.87	0.77
	Calib.	0.014	3.66	0.005	1.62	0.010	1.07	0.015	0.90
	Calib.+bridging	0.046	3.70	0.092	1.77	0.11	1.25	0.15	1.02



**Fig. 5** Comparison of the reliability diagrams (whole domain) for the upper quintile of weekly precipitation, between the raw MF S2S reforecast (blue), the calibration reforecast (green) and the calibration + bridging reforecast (red), at week 2 (left) and week 3 (right). The forecast frequency in each probability bin is represented below the bar plots for each of reforecasts, with the same color code. Bins with less than 1% of the total number of forecasts are not plotted.



**Fig. 6** Same as Figure 5 for the ECMWF S2S system. Bins with less than 1% of the total number of forecasts are not plotted.

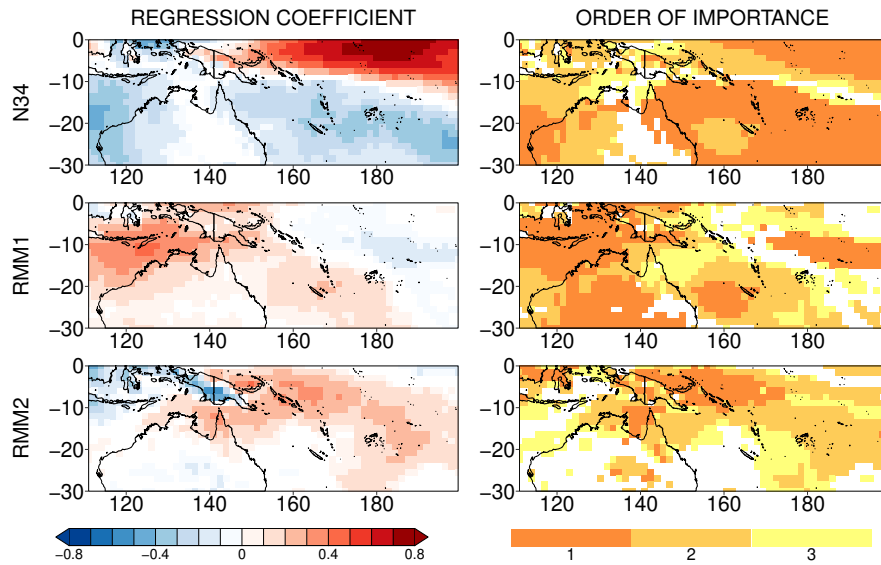
## 430 5 Role of the large-scale predictors

### 431 5.1 Theoretical relative added value

432 Following the methodology described in Section 3.2, the theoretical importance of each  
 433 predictor  $N34$ ,  $RMM1$  and  $RMM2$  in the reference datasets is assessed with a stepwise  
 434 forward selection scheme. For each of these predictors, the first column in Figure 7 shows the  
 435 coefficient of the linear regression in Equation (5) at each grid point, if it is significant at the  
 436 95% level. This is an estimation of the absolute role that each predictor plays individually.  
 437 The second column in Figure 7 shows the order in which the predictive term in Equation  
 438 (10), associated to a given predictor, is selected. This is an estimation of the relative role  
 439 of each predictor in the bridging part of the approach. Grid points in white correspond to  
 440 grid points where the predictive term is considered not to bring any additional information  
 441 according to the p-value threshold of 0.05 mentioned in Section 3.2.

442 The key result is that the  $N34$  predictor related to ENSO is the first large-scale source of  
 443 information over most of the SWTP, mainly in the southern oceanic part and in the north-  
 444 eastern equatorial region. If not ENSO, the most important predictor is quite often the first  
 445 MJO index  $RMM1$ , that is the main source of information in the northwestern part of the  
 446 domain but is also a significant source of information in a region that extends southeastward





**Fig. 7** Left: linear regression coefficient of each predictor against observed rainfall  $\hat{r}_o$ . Grid points are in white if the coefficient is not significant at the 95% level.

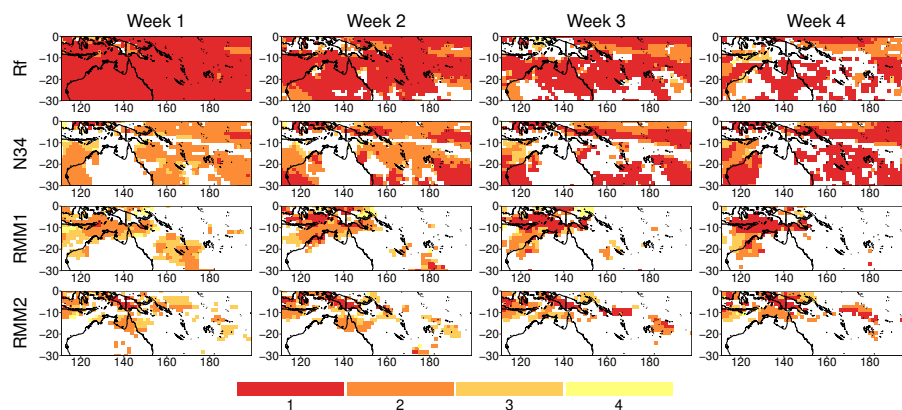
Right: Order of selection for each predictor in a stepwise forward selection scheme when considering large-scale predictors only in reference data. Grid points in white correspond to grid points where adding the predictor leads to a p-value greater than 0.05 in the likelihood-ratio test.

447 from the Maritime Continent to the subtropics and encompasses Pacific island territories  
 448 such as Vanuatu, New Caledonia and Fiji. As for *RMM2*, it is often a secondary source of  
 449 information bearing some added value in the equatorial and eastern part of the domain. The  
 450 patterns of significant added value for *N34*, *RMM1* and *RMM2* are similar to the regression  
 451 patterns of rainfall against these indices, as illustrated in the first column of Figure 7 and  
 452 also in several other studies (e.g de Andrade et al., 2018, Figure 2).

## 453 5.2 Added value in the statistical-dynamical scheme

454 On account of the decreasing quality of rainfall and MJO forecasts with lead time, it is  
 455 expected that the importance of the large-scale predictors as illustrated in Section 5.1, but  
 456 also the importance of the calibration predictor  $\hat{r}_f$ , might change as we get further away  
 457 from initialization. This aspect is investigated in Figure 8 with the order of importance of  
 458 the four predictors at each lead time in the statistical-dynamical approach, when applied to  
 459 the MF system. Results are similar for the ECMWF system (not shown). Because forecast  
 460 precipitation was expected to be the  $n^{\circ}1$  predictor most of the time, we shifted the color  
 461 code of Figure 7 (second column) from ranks 1-3 to 2-4, and added an extra color for the  
 462 most important predictor.

463 At weeks 1 and 2, the forecast precipitation (i.e calibration predictor) is indeed the first  
 464 source of information in the statistical-dynamical approach almost over the whole domain.  
 465 In the meantime, the *N34* predictor often comes second after forecast precipitation, at the  
 466 locations where it is identified as a useful predictor according to Figure 7. The main result  
 467 is that the prominence of forecast precipitation at these locations shrinks with lead time,



**Fig. 8** Order of selection for each predictor in a stepwise forward selection scheme, applied at each lead time for the calibration + bridging scheme applied to the MF system. Forecast rainfall  $\hat{r}_f$  is also included in the selection.

468 while ENSO gradually becomes the most relevant predictor. As a result, at week 4,  $N34$  has  
 469 become the  $n^{\circ}1$  predictor for the majority of grid points. This indicates that, as we go into  
 470 the subseasonal range, the statistical relationship between  $N34$  and rainfall actually better  
 471 captures the ENSO-related predictable signal than the coupled dynamics of the numerical  
 472 model.

473 The notable exceptions where  $N34$  is challenged by MJO indices are locations where  
 474 the  $RMM1$  predictor or, to a lesser extent, the  $RMM2$  predictor, are theoretically good pre-  
 475 dictors according to Figure 7. This corresponds mostly to the region between the Maritime  
 476 Continent and northern Australia, where ENSO does not account for much skill. In this area,  
 477  $RMM1$  is the second most important predictor at weeks 1 and 2 (after precipitation), and it  
 478 becomes the most important at weeks 3 and 4. As for  $RMM2$ , it plays a similar role over  
 479 the land grid points of New Guinea. Figure 7 suggested that these two indices could also  
 480 be sources of information for wider areas :  $RMM1$  can theoretically help predict rainfall  
 481 in a southeastward-extending region ranging from New Guinea to New Caledonia, and the  
 482 same can be said of  $RMM2$  along the South Pacific Convergence Zone (SPCZ) track. Yet,  
 483 their added value in these regions at the subseasonal time scales appears to be very lim-  
 484 ited or non-existent. Admittedly,  $RMM1$  comes as second or third predictor in a large zone  
 485 around New Caledonia at week 1, but this influence shrinks with lead time and ceases to  
 486 be at weeks 3 and 4. A similar effect can be noted for  $RMM2$  over Cape York peninsula in  
 487 northern Australia.

## 488 6 Conclusion

489 We have developed and applied a statistical-dynamical scheme to post-process subseasonal  
 490 forecasts of precipitation in a bayesian framework. Our framework encompasses a calibra-  
 491 tion approach that uses the forecast precipitation as a predictor in a statistical model. It  
 492 also includes bridging aspects under the assumption that large-scale climate features, ENSO  
 493 and the MJO, are better represented by S2S systems than precipitation itself. Because these  
 494 large-scale climate features are known to have a strong and direct impact on rainfall on the  
 495 SWTP domain, we chose this region as a relevant test bed for our scheme.

496 In this region, the statistical-dynamical approach definitely proves valuable to improve  
497 the probabilistic reforecast quality, especially in terms of discrimination for the occurrence  
498 vs non-occurrence of a binary event, e.g the upper quintile of weekly precipitation. How-  
499 ever, when it comes to deterministic forecasting, the benefits of the approach are more con-  
500 trasted : depending on the S2S system, it can perform better or worse than a simple bias  
501 correction. Therefore, the proposed statistical-dynamical scheme should preferably be ap-  
502 plied in a probabilistic forecasting context. These results are based on the Météo-France  
503 and ECMWF systems, but were confirmed with four other S2S systems (see Supplementary  
504 material, Figures S2 and S3), for which the qualitative benefits are close to those obtained  
505 with Météo-France.

506 The calibration part of the method is the most crucial step as it accounts alone for error  
507 reduction, provides reliable probabilistic forecasts, and reasonably improves their discrim-  
508 ination. Yet, adding information from the large-scale predictors is the source of an extra  
509 discrimination ability for heavy rainfall events. As a result, although calibration leads to  
510 good results on its own, the best of the statistical-dynamical approaches proposed in this  
511 study for probabilistic forecasting remains the calibration + bridging approach.

512 The benefits of using the large-scale predictors actually vary across space and lead times.  
513 Unsurprisingly, we have seen that a climate index provides valuable information where the  
514 corresponding phenomenon impacts rainfall. To this respect, this article confirms the known  
515 impacts of ENSO and the MJO at some locations, e.g the MJO over the Maritime Continent.  
516 It also highlights locations that are significantly impacted by the MJO while they do not lie  
517 directly within the MJO envelope track, for instance the zone extending southeastward from  
518 New Guinea to New Caledonia.

519 In these locations where MJO-related theoretical predictability is more limited, the  
520 degradation of the MJO forecast quality with lead time is such that the MJO predictors  
521 do not bring a real predictive improvement of the statistical-dynamical model beyond week  
522 2. The reason why bridging is still profitable at weeks 3 and 4 in these regions is related  
523 to the ENSO predictor, that exhibits little variability throughout the forecast, but does not  
524 correspond to subseasonal climate variability.

525 This result suggests that, in the case of SWTP precipitation forecasts, it is difficult to tap  
526 into the predictability of large-scale subseasonal variability (the MJO) to obtain improve-  
527 ments in aggregated scores through bridging after week 2. However, it does not exclude  
528 the existence of forecasts of opportunity for which particular precipitation events might  
529 be predictable three weeks in advance or more (e.g Doss-Gollin et al., 2018; Lin et al.,  
530 2019), thanks to subseasonal signals like the MJO. Moreover, it also indicates that other  
531 slowly-varying sources of seasonal predictability, analogous to ENSO, could be used to im-  
532 prove subseasonal forecasts. These assumptions could now be tested with similar statistical-  
533 dynamical approaches by considering other locations, predictors and predicted variables.  
534 For example, the Boreal Summer Intraseasonal Oscillation Index could be tested as a sub-  
535 seasonal predictor for precipitation in the Asian summer monsoon region (Lee et al., 2013),  
536 while dynamical weather types could be tested to improve subseasonal prediction of tem-  
537 peratures in Europe (Cassou et al., 2005; Ardilouze et al., 2017).

538 Finally, some findings in this study point that, beyond statistical post-processing tech-  
539 niques, the improvement of numerical prediction systems is still necessary for the advance-  
540 ment of S2S prediction. This is illustrated by the fact our method does not compensate for the  
541 initial difference in skill between the Météo-France and ECMWF systems, while it would if  
542 post-processing alone were sufficient to reach the theoretical skill limit. Nonetheless, post-  
543 processing proves valuable to diagnose the strengths and weaknesses of S2S systems. For  
544 instance, by revealing that a dynamically-forecast large-scale predictor (ENSO) can be more

545 informative than dynamically-forecast precipitation itself, it shows that higher skill could  
546 actually be obtained if prediction systems directly took advantage of their well-predicted  
547 signals. This advocates for a joint development of the S2S systems and their statistical post-  
548 processing approaches.

549 **Acknowledgements** The S2S reforecasts were accessed through the ECMWF MARS data portal (<http://apps.ecmwf.int/datasets/data/s2s/>).  
550 The MSWEP precipitation data is developed by H. Beck (Beck et al., 2017) at Princeton University and is  
551 available at <http://www.gloh2o.org/>. The EOFs used for computation of the RMM indices are provided by the  
552 Australian Bureau of Meteorology at <http://poama.bom.gov.au/project/maproom/RMM/>.

553 All calculations were performed using the open source language 'R'. The functions in the package  
554 's2dverification' (Manubens et al., 2018) were used for data manipulation. All calculations related to the  
555 ROC curve were made with the 'pROC' package (Robin et al., 2011). Empirical CDF determination was  
556 made using the 'EnvStats' package (Millard, 2013). Likelihood-ratio tests were performed with the 'lmtree'  
557 package.

558 We thank the two anonymous reviewers whose comments helped improve the former version of this  
559 article.

## 560 References

- 561 de Andrade FM, Coelho CAS, Cavalcanti IFA (2018) Global precipitation hindcast qual-  
562 ity assessment of the Subseasonal to Seasonal (S2S) prediction project models. *Climate*  
563 *Dynamics* DOI 10.1007/s00382-018-4457-z
- 564 Ardilouze C, Batté L, Déqué M (2017) Subseasonal-to-seasonal (S2s) forecasts with  
565 CNRM-CM: a case study on the July 2015 West-European heat wave. *Advances in Sci-*  
566 *ence and Research* 14:115–121, DOI 10.5194/asr-14-115-2017
- 567 Baggett CF, Nardi KM, Childs SJ, Zito SN, Barnes EA, Maloney ED (2018) Skill-  
568 ful Subseasonal Forecasts of Weekly Tornado and Hail Activity Using the Madden-  
569 Julian Oscillation. *Journal of Geophysical Research: Atmospheres* 123(22), DOI  
570 10.1029/2018JD029059
- 571 Beck HE, van Dijk AIJM, Levizzani V, Schellekens J, Miralles DG, Martens B, de Roo  
572 A (2017) MSWEP: 3-hourly 0.25° global gridded precipitation (1979-2015) by merging  
573 gauge, satellite, and reanalysis data. *Hydrology and Earth System Sciences* 21(1):589–  
574 615, DOI 10.5194/hess-21-589-2017
- 575 Bombardi RJ, Pegion KV, Kinter JL, Cash BA, Adams JM (2017) Sub-seasonal Predictabil-  
576 ity of the Onset and Demise of the Rainy Season over Monsoonal Regions. *Frontiers in*  
577 *Earth Science* 5, DOI 10.3389/feart.2017.00014
- 578 Bröcker J (2015) Resolution and discrimination—two sides of the same coin: Resolution and  
579 Discrimination. *Quarterly Journal of the Royal Meteorological Society* 141(689):1277–  
580 1282, DOI 10.1002/qj.2434
- 581 Buse A (1982) The Likelihood Ratio, Wald, and Lagrange Multiplier Tests: An Expository  
582 Note. *The American Statistician* 36(3):153–157, DOI 10.2307/2683166
- 583 Cassou C, Terray L, Phillips AS (2005) Tropical Atlantic Influence on European Heat  
584 Waves. *Journal of Climate* 18(15):2805–2811, DOI 10.1175/JCLI3506.1
- 585 Coelho CA, Pezzulli S, Balmaseda M, Doblás-Reyes F, Stephenson D (2004) Forecast Cal-  
586 ibration and Combination: A Simple Bayesian Approach for ENSO. *Journal of Climate*  
587 17(7):1504–1515
- 588 Coelho CA, Firpo MA, de Andrade FM (2018) A verification framework for South Amer-  
589 ican sub-seasonal precipitation predictions. *Meteorologische Zeitschrift* 27(6):503–520,  
590 DOI 10.1127/metz/2018/0898

- 591 Cohen J, Coumou D, Hwang J, Mackey L, Orenstein P, Totz S, Tziperman E (2019) S2s  
592 reboot: An argument for greater inclusion of machine learning in subseasonal to seasonal  
593 forecasts. *Wiley Interdisciplinary Reviews: Climate Change* 10(2), DOI 10.1002/wcc.567
- 594 Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, et al. (2011) The  
595 ERA-Interim reanalysis: configuration and performance of the data assimilation sys-  
596 tem. *Quarterly Journal of the Royal Meteorological Society* 137(656):553–597, DOI  
597 10.1002/qj.828
- 598 DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the Areas under Two or  
599 More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach.  
600 *Biometrics* 44(3):837–845, DOI 10.2307/2531595
- 601 Doss-Gollin J, Muñoz AG, Mason SJ, Pastén M (2018) Heavy Rainfall in Paraguay dur-  
602 ing the 2015/16 Austral Summer: Causes and Subseasonal-to-Seasonal Predictive Skill.  
603 *Journal of Climate* 31(17):6669–6685, DOI 10.1175/JCLI-D-17-0805.1
- 604 ECMWF (2017) MJO Index Computation at ECMWF. Tech. rep., ECMWF, URL  
605 [ftp://acquisition.ecmwf.int/RMMS/RMMS\\_computation.docx](ftp://acquisition.ecmwf.int/RMMS/RMMS_computation.docx)
- 606 Gottschalck J, Wheeler M, Weickmann K, Vitart F, Savage N, Lin H, et al. (2010) A Frame-  
607 work for Assessing Operational Madden–Julian Oscillation Forecasts: A CLIVAR MJO  
608 Working Group Project. *Bulletin of the American Meteorological Society* 91(9):1247–  
609 1258, DOI 10.1175/2010BAMS2816.1
- 610 Hudson D, Alves O, Hendon HH, Marshall AG (2011) Bridging the gap between weather  
611 and seasonal forecasting: intraseasonal forecasting for Australia. *Quarterly Journal of the*  
612 *Royal Meteorological Society* 137(656):673–689, DOI 10.1002/qj.769
- 613 Johnson NC, Collins DC, Feldstein SB, L’Heureux ML, Riddle EE (2014) Skillful Winter-  
614 time North American Temperature Forecasts out to 4 Weeks Based on the State of ENSO  
615 and the MJO\*. *Weather and Forecasting* 29(1):23–38, DOI 10.1175/WAF-D-13-00102.1
- 616 Karpechko AY, Charlton-Perez A, Balmaseda M, Tyrrell N, Vitart F (2018) Predicting Sud-  
617 den Stratospheric Warming 2018 and Its Climate Impacts With a Multimodel Ensemble.  
618 *Geophysical Research Letters* 45(24):13,538–13,546, DOI 10.1029/2018GL081091
- 619 Kim H, Vitart F, Waliser DE (2018) Prediction of the Madden–Julian Oscillation: A Review.  
620 *Journal of Climate* 31(23):9425–9443, DOI 10.1175/JCLI-D-18-0210.1
- 621 Lee JY, Wang B, Wheeler MC, Fu X, Waliser DE, Kang IS (2013) Real-time multivariate  
622 indices for the boreal summer intraseasonal oscillation over the Asian summer monsoon  
623 region. *Climate Dynamics* 40(1-2):493–509, DOI 10.1007/s00382-012-1544-4
- 624 Li W, Chen J, Li L, Chen H, Liu B, Xu CY, Li X (2019) Evaluation and Bias Correction of  
625 S2s Precipitation for Hydrological Extremes. *Journal of Hydrometeorology* 20(9):1887–  
626 1906, DOI 10.1175/JHM-D-19-0042.1
- 627 Liang P, Lin H (2018) Sub-seasonal prediction over East Asia during boreal summer us-  
628 ing the ECCO monthly forecasting system. *Climate Dynamics* 50(3-4):1007–1022, DOI  
629 10.1007/s00382-017-3658-1
- 630 Lim Y, Son SW, Marshall AG, Hendon HH, Seo KH (2019) Influence of the QBO on  
631 MJO prediction skill in the subseasonal-to-seasonal prediction models. *Climate Dynam-*  
632 *ics* 53(3-4):1681–1695, DOI 10.1007/s00382-019-04719-y
- 633 Lin H, Mo R, Vitart F, Stan C (2019) Eastern Canada Flooding 2017 and its Subseasonal  
634 Predictions. *Atmosphere-Ocean* 57(3):195–207, DOI 10.1080/07055900.2018.1547679
- 635 Luo L, Wood EF, Pan M (2007) Bayesian merging of multiple climate  
636 model forecasts for seasonal hydrological predictions. *Journal of Geophys-*  
637 *ical Research: Atmospheres* 112(D10), DOI 10.1029/2006JD007655, URL  
638 <http://doi.wiley.com/10.1029/2006JD007655>

- 639 Manubens N, Caron LP, Hunter A, Bellprat O, Exarchou E, Fučkar NS, et al. (2018) An R  
640 package for climate forecast verification. *Environmental Modelling & Software* 103:29–  
641 42, DOI 10.1016/j.envsoft.2018.01.018
- 642 Marshall AG, Hendon HH (2015) Subseasonal prediction of Australian summer  
643 monsoon anomalies. *Geophysical Research Letters* 42(24):10913–10919, DOI  
644 10.1002/2015GL067086
- 645 Marshall AG, Hendon HH (2019) Multi-week prediction of the Madden–Julian oscilla-  
646 tion with ACCESS-S1. *Climate Dynamics* 52(5-6):2513–2528, DOI 10.1007/s00382-  
647 018-4272-6
- 648 Marshall AG, Hudson D, Wheeler MC, Hendon HH, Alves O (2011) Assessing the simula-  
649 tion and prediction of rainfall associated with the MJO in the POAMA seasonal forecast  
650 system. *Climate Dynamics* 37(11-12):2129–2141, DOI 10.1007/s00382-010-0948-2
- 651 Mastrangelo D, Malguzzi P (2019) Verification of Two Years of CNR-ISAC Subseasonal  
652 Forecasts. *Weather and Forecasting* 34(2):331–344, DOI 10.1175/WAF-D-18-0091.1
- 653 McGree S, Whan K, Jones D, Alexander LV, Imielska A, Diamond H, et al. (2014) An  
654 updated assessment of trends and variability in total and extreme rainfall in the western  
655 Pacific. *International Journal of Climatology* 34(8):2775–2791, DOI 10.1002/joc.3874
- 656 McGree S, Schreider S, Kuleshov Y (2016) Trends and Variability in Droughts in the  
657 Pacific Islands and Northeast Australia. *Journal of Climate* 29(23):8377–8397, DOI  
658 10.1175/JCLI-D-16-0332.1
- 659 Millard SP (2013) *EnvStats*. Springer New York, DOI 10.1007/978-1-4614-8456-1
- 660 Murphy AH (1973) A New Vector Partition of the Probability Score. *Journal of Applied*  
661 *Meteorology* 12(4):595–600
- 662 Pegion K, Kirtman BP, Becker E, Collins DC, LaJoie E, Burgman R, et al. (2019) The Sub-  
663 seasonal Experiment (SubX): A multi-model subseasonal prediction experiment. *Bulletin*  
664 *of the American Meteorological Society* DOI 10.1175/BAMS-D-18-0270.1
- 665 Robertson AW, Vitart F (2019) Sub-seasonal to seasonal prediction: the gap between  
666 weather and climate forecasting. Elsevier
- 667 Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M (2011) pROC:  
668 an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioin-*  
669 *formatics* 12(1):77, DOI 10.1186/1471-2105-12-77
- 670 Schepen A, Wang QJ, Robertson DE (2012) Combining the strengths of statistical and dy-  
671 namical modeling approaches for forecasting Australian seasonal rainfall. *Journal of Geo-*  
672 *physical Research: Atmospheres* 117(D20), DOI 10.1029/2012JD018011
- 673 Schepen A, Wang QJ, Robertson DE (2014) Seasonal Forecasts of Australian Rainfall  
674 through Calibration and Bridging of Coupled GCM Outputs. *Monthly Weather Review*  
675 142(5):1758–1770, DOI 10.1175/MWR-D-13-00248.1
- 676 Schepen A, Wang QJ, Everingham Y (2016) Calibration, Bridging, and Merging to Im-  
677 prove GCM Seasonal Temperature Forecasts in Australia. *Monthly Weather Review*  
678 144(6):2421–2441, DOI 10.1175/MWR-D-15-0384.1
- 679 Specq D, Batté L, Déqué M, Ardilouze C (2020, in press) Multimodel forecasting of pre-  
680 cipitation at subseasonal timescales over the southwest tropical Pacific. *Earth and Space*  
681 *Science* DOI 10.1029/2019EA001003
- 682 Strazzo S, Collins DC, Schepen A, Wang QJ, Becker E, Jia L (2019) Application of a Hybrid  
683 Statistical–Dynamical System to Seasonal Prediction of North American Temperature  
684 and Precipitation. *Monthly Weather Review* 147(2):607–625, DOI 10.1175/MWR-D-18-  
685 0156.1
- 686 Tian D, Wood EF, Yuan X (2017) CFSv2-based sub-seasonal precipitation and temperature  
687 forecast skill over the contiguous United States. *Hydrology and Earth System Sciences*

- 21:1477–1490, DOI 10.5194/hess-21-1477-2017
- Vigaud N, Robertson AW, Tippett MK (2017a) Multimodel Ensembling of Subseasonal Precipitation Forecasts over North America. *Monthly Weather Review* 145(10):3913–3928, DOI 10.1175/MWR-D-17-0092.1
- Vigaud N, Robertson AW, Tippett MK, Acharya N (2017b) Subseasonal Predictability of Boreal Summer Monsoon Rainfall from Ensemble Forecasts. *Frontiers in Environmental Science* 5:67, DOI 10.3389/fenvs.2017.00067
- Vigaud N, Tippett MK, Robertson AW (2018) Probabilistic Skill of Subseasonal Precipitation Forecasts for the East Africa–West Asia Sector during September–May. *Weather and Forecasting* 33(6):1513–1532, DOI 10.1175/WAF-D-18-0074.1
- Vigaud N, Tippett MK, Yuan J, Robertson AW, Acharya N (2019) Spatial correction of multi-model ensemble subseasonal precipitation forecasts over North America using local Laplacian eigenfunctions. *Monthly Weather Review* DOI 10.1175/MWR-D-19-0134.1
- Vitart F (2004) Monthly Forecasting at ECMWF. *Monthly Weather Review* 132(12):2761–2779, DOI 10.1175/MWR2826.1
- Vitart F (2014) Evolution of ECMWF sub-seasonal forecast skill scores. *Quarterly Journal of the Royal Meteorological Society* 140(683):1889–1899, DOI 10.1002/qj.2256
- Vitart F (2017) Madden-Julian Oscillation prediction and teleconnections in the S2s database. *Quarterly Journal of the Royal Meteorological Society* 143(706):2210–2220, DOI 10.1002/qj.3079
- Vitart F, Ardilouze C, Bonet A, Brookshaw A, Chen M, Codorean C, et al. (2017) The Subseasonal to Seasonal (S2S) Prediction Project Database. *Bulletin of the American Meteorological Society* 98(1):163–173, DOI 10.1175/BAMS-D-16-0017.1
- Wang L, Robertson AW (2019) Week 3–4 predictability over the United States assessed from two operational ensemble prediction systems. *Climate Dynamics* 52(9-10):5861–5875, DOI 10.1007/s00382-018-4484-9
- Wheeler MC, Hendon HH (2004) An All-Season Real-Time Multivariate MJO Index: Development of an Index for Monitoring and Prediction. *Monthly Weather Review* 132(8):1917–1932, DOI 10.1175/1520-0493(2004)132;1917:AARMMI;2.0.CO;2
- White CJ, Carlsen H, Robertson AW, Klein RJ, Lazo JK, Kumar A, et al. (2017) Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteorological Applications* 24(3):315–325, DOI 10.1002/met.1654
- Wilks DS (2006) *Statistical methods in the atmospheric sciences*, 2nd edn. Academic Press, Boston, MA
- Zhang C (2005) Madden-Julian Oscillation. *Reviews of Geophysics* 43(2), DOI 10.1029/2004RG000158
- Zhang C (2013) Madden-Julian Oscillation: Bridging Weather and Climate. *Bulletin of the American Meteorological Society* 94(12):1849–1870, DOI 10.1175/BAMS-D-12-00026.1

#### Conflict of interest

The authors declare that they have no conflict of interest.