University of Reading

# *Configuration and hindcast quality assessment of a brazilian global sub-seasonal prediction system*

Article

Accepted Version

Guimarães, B. S., Coelho, C. A. S., Woolnough, S., J. ORCID: https://orcid.org/0000-0003-0500-8514, Kubota, P. Y., Bastarz, C. F., Figueroa, S. N., Bonatti, J. P. and Souza, D. C. (2020) Configuration and hindcast quality assessment of a brazilian global sub-seasonal prediction system. Quarterly Journal of the Royal Meteorological Society, 146 (728). pp. 1067-1084. ISSN 0035-9009 doi: 10.1002/qj.3725 Available at https://centaur.reading.ac.uk/88559/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1002/qj.3725

Publisher: Wiley

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

University of Reading

www.reading.ac.uk/centaur

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

# CONFIGURATION AND HINDCAST QUALITY ASSESSMENT OF A BRAZILIAN GLOBAL SUB-SEASONAL PREDICTION SYSTEM

Bruno S. Guimarães[1,2], Caio A. S. Coelho[1], Steve J. Woolnough[2], Paulo Y. Kubota[1], Carlos F. Bastarz[1], Silvio N. Figueroa[1], José P. Bonatti[1] and Dayana C. de Souza[1]

1- Center for Weather Forecast and Climate Studies,
National Institute for Space Research, Cachoeira Paulista, SP, Brazil;

2- National Centre for Atmospheric Science, Department of Meteorology,
University of Reading, UK.

## ABSTRACT

This paper presents the Center for Weather Forecast and Climate Studies (CPTEC) developments for configuring a global sub-seasonal prediction system and assessing its ability in producing retrospective predictions (hindcasts) for meteorological conditions of the following 4 weeks. Six Brazilian Global Atmospheric Model version 1.2 (BAM-1.2) configurations were tested in terms of vertical resolution, deep convection and boundary layer parameterizations, as well as soil moisture initialization. The aim was to identify the configuration with best performance when predicting weekly accumulate precipitation, weekly mean 2-meter temperature (T2M) and the Madden and Julian Oscillation (MJO) daily evolution. Hindcasts assessment was performed for 12 extended austral summers (November to March - 1999/2000 to 2010/2011) with two start dates for each month for the six configurations and two ensemble approaches. The first approach, referred to as Multiple Configurations Ensemble (MCEN), was formed of one ensemble member from each of the six configurations. The second, referred to as Initial Condition Ensemble (ICEN), was composed of six ensemble members produced with the chosen configuration as the best using an Empirical Orthogonal Function (EOF) perturbation methodology. The chosen configuration presented high correlation and low root mean squared error (RMSE) for precipitation and T2M anomaly predictions at the first week and these indices degraded as lead time increased, maintaining moderate performance up to week 4 over the tropical Pacific and northern South America. For MJO predictions, this configuration crossed the 0.5 bivariate correlation threshold in 18 days. The ensemble approaches improved the

correlation and RMSE of precipitation and T2M anomalies. ICEN improved precipitation and T2M predictions performance over eastern South America at week 3 and over northern South America at week 4. Improvements were also noticed for MJO predictions. The time to cross the above mentioned threshold increased to 21 days for MCEN and to 20 days for ICEN.

**Keywords**: MJO, Intraseasonal Variability, Forecast Verification.

## 1. INTRODUCTION

Forecasting for the time scale between two weeks and two months is known as sub-seasonal prediction (Vitart et al., 2017). This type of forecast is a major challenge because the predictability contribution from the atmospheric initial conditions is reduced compared to shorter (weather) timescales, and the predictability from slowly varying boundary conditions is small for 1-2 week averages, typically the focus of sub-seasonal prediction, compared to seasonal timescales (Kumar et al., 2011; Lin et al., 2016). The main source of predictability for sub-seasonal forecasting is the Madden - Julian Oscillation (MJO) (Zhang, 2013). However, General Circulation Models (GCMs) still show limitations in simulating this oscillation (Green et al., 2017; Wang et al., 2018) even with important improvements achieved in recent years (Saha et al., 2014; Vitart, 2014). As a consequence of these limitations, the predictive ability of GCMs in the sub-seasonal scale is lower than in the weather and seasonal scales (Zhu et al., 2014). For example, de Andrade et al. (2019) showed the limited predictive ability of GCM for sub-seasonal precipitation predictions for lead times beyond 15 days. In general, the GCMs show modest performance in specific areas such as the equatorial

regions of the Atlantic and Pacific oceans and over a few regions in South America at this lead time.

In spite of these results, a tendency towards improvements in GCMs for sub-seasonal predictions is seen and several meteorological centres currently operationally produce this type of forecasts (Vitart, 2004; Hudson et al., 2011; Mastrangelo et al., 2012; Liu et al., 2017; Weber and Mass, 2017; Liang and Lin, 2018). The Center for Weather Forecast and Climate Studies [Centro de Previsão de Tempo e Estudos Climáticos (CPTEC)], which plays a leading role in South America with respect to weather and seasonal forecasts, is now following the trend of these meteorological centres and started to develop a sub-seasonal prediction system. This is motivated by the fact that, as in seasonal forecasting, South America is located in a privileged region for sub-seasonal prediction, with GCMs showing better predictive ability in this region when compared to other continental regions (Li and Robson, 2015; de Andrade et al., 2019).

The identified evolution in sub-seasonal predictions is mainly due to improvements in the representation of the MJO in GCMs. For example, The European Centre for Medium-Range Weather Forecasts (ECMWF) showed a mean gain of one day in MJO prediction performance per year (Vitart, 2014). This indicates that in addition to improvements in predictive ability for a phenomenon that manifests in the tropical region, there is also associated improvement in the extratropics due to teleconnections generated by the MJO (Vitart, 2017).

These findings are documented, in large part, thanks to the effort generated by Sub-seasonal to Seasonal (S2S) Prediction Project. This project was launched jointly by the

World Weather Research Program (WWRP) and the World Climate Research Program (WCRP) of the World Meteorological Organization (WMO) and aims to improve forecast skill and understanding on the sub-seasonal to seasonal time scales and also to promote its uptake by operational centres and by the applications community. Currently, the S2S Prediction Project stores and disseminates near-real-time forecasts and hindcasts of eleven operational and research centres for research purposes (Vitart et al., 2017).

The Brazilian Global Atmospheric model [BAM (Figueroa et al., 2016)] is the current CPTEC global atmospheric model for weather forecasting. The performance of this model for sub-seasonal predictions has not been documented yet. Therefore, this study presents the first outcomes of this model for sub-seasonal predictions and aims to determine which model configuration presents the best performance for this time scale. Special attention is given to characteristics such as vertical resolution, deep convection and boundary layer parameterizations and as well as initialization of the soil moisture because they have an important influence on the MJO and sub-seasonal predictions. A similar approach was taken by Green et al (2017) in order to identify a model configuration with best performance when producing MJO predictions. Green et al (2017) evaluated the MJO predictive ability in multiphysics and multimodel global ensembles, by performing two sets of hindcasts in order to test the impact of using the Grell–Freitas (2014) versus the revised simplified Arakawa–Schubert (Han and Pan, 2011) deep convection parameterization. They revealed that the Grell–Freitas (2014) convection parameterization showed better MJO prediction performance than the revised simplified Arakawa–Schubert scheme.

The paper is organized as follows. Section 2 presents the model description, datasets used for model initialization and hindcast quality assessment, the definition of the experiments, ensemble approaches and the metrics used for evaluation. The retrospective performance of the produced precipitation, 2-meter temperature (T2M) and MJO predictions with different BAM configuration experiments, including two ensemble approaches, is shown in section 3. The final section is intended for the conclusion.

## 2. MODEL DESCRIPTION, DATASETS AND EXPERIMENTAL CONFIGURATIONS, EVALUATION METRICS AND ENSEMBLE APPROACHES

### 2.1. Model description

The model version used in this study is the current operational CPTEC global spectral atmospheric model developed for numerical weather forecasting, which is known as BAM version 1.2 (BAM-1.2). This model version has different options for dynamical and physics parameterizations. The Eulerian advection scheme option with a two-time-level Semi-Lagrangian scheme for moisture transport and microphysics prognostic variables is used in this study. The physical processes of this recent operational version are similar to the previous version (BAM-1.0) and are described in Figueroa et al. (2016), which are: microphysics from Morrison et al. (2009), the IBIS-CPTEC surface model (Kubota, 2012), the long-wave radiation scheme developed by Chou et al. (2001) (CLIRAD-LW), the short-wave radiation scheme developed by Chou and Suarez (1999) (CLIRAD-SW), the latter modified by Tarasova and Fomin (2000), the modified Mellor-Yamada diffusion scheme for the planetary boundary layer (PBL), which is

based on Mellor-Yamada (1982) and is referred to as dry-PBL, and the modified Grell-Dévényi deep convection scheme, which is based on Grell-Dévényi (2002). The two new BAM-1.2 components are Bretherton-Park moist diffusion scheme (Bretherton and Park, 2009) for the PBL, which is referred to as moist-PBL, and the revised version of the simplified Arakawa-Shubert deep convection scheme (Han and Pan. 2011), which were recently implemented. Following Yu et al. (2006) aerosol optical depth in the first 2 km of the atmosphere is specified as 0.22 over the continents and as 0.14 over the oceans. The horizontal resolution used in this study is triangular truncation at 126 waves (TQ126, corresponding to a grid of approximately 1.0° in latitude and longitude) and two vertical resolutions are examined: 42 (L42) and 64 (L64) sigma vertical levels.

One of the objectives of this study is to investigate the performance of the two PBL and deep convection schemes mentioned above for sub-seasonal predictions. The main difference of the newly implemented moist-PBL Bretherton-Park scheme compared to the dry-PBL modified Mellor-Yamada diffusion scheme is the use of moist-conserved variables and an explicit entrainment closure for convective layers. Regarding the convection schemes, the revised simplified Arakawa-Shubert and the previously implemented modified Grell-Dévényi deep convection parameterization schemes were both derived from Grell (1993), in which the cloud spectrum of the original Arakawa-Schubert (1974) scheme is reduced to a single cloud using a single mass flux closure. The main differences between these convection schemes implemented in BAM-1.2 are the fractional entrainment rate and convection trigger formulations [see Han and Pan (2011) and Figueroa et al., (2016) for additional information].

### 2.2. Datasets and experimental configurations

Sub-seasonal hindcasts were performed over the period defined as the extended austral summer (from November to March) over the 1999/2000-2010/2011 period. Two hindcasts for two selected start dates were produced for each month of a given year. Starts dates vary from one month to the next and are presented in Table 1. Each hindcast was run for the following 35 days after the start date (35 days of lead time). For the production of these hindcasts, BAM-1.2 was not coupled with an ocean model. Instead, the total Sea Surface Temperature (SST) field (not the anomaly) of each start date was kept constant during the 35 days of integration (persisted SST). It is worth highlighting that coupled ocean-atmosphere processes are recognized as being important on these timescales (Reichler and Roads, 2005; Chen et al., 2010, Kumar et al., 2011; Shelly et al., 2014), but a number of centres contributing to the S2S database [e.g., Japan Meteorological Agency (JMA) and Environment and Climate Change Canada (ECCC)] produce operational sub-seasonal forecasts using un-coupled systems (Vitart et al., 2017). The CPTEC couple ocean-atmosphere model version which uses BAM-1.2 as atmospheric component is under development. The sub-seasonal hindcast quality assessment of this coupled model version will be reported in future work.

ERA-Interim reanalyses (Dee et al., 2011) produced by ECMWF were used in two ways. Firstly, the reanalyses were used as atmospheric initial conditions for the hindcasts produced with BAM-1.2. The variables required for initialization are zonal and meridional wind, specific humidity, virtual temperature and ozone in 35 vertical levels between 1000 hPa and 50 hPa, surface pressure and SST. The horizontal resolution chosen for initialization was 1.5° x 1.5° degrees in latitude and longitude, which was interpolated to the model spectral resolution (TQ126L42, ~100km).

Secondly, ERA-Interim data were used as reference to assess the quality of the produced hindcasts. The variables selected for this assessment are T2M and zonal and meridional winds at 850 hPa and 200 hPa.

To assess precipitation hindcasts quality, daily data from the Global Precipitation Climatology Project (GPCP) were used (Huffman, 2001). GPCP is a product derived from observed rainfall data and precipitation estimates by geostationary and polar-orbiting satellites. The spatial resolution of GPCP is 1° x 1° degrees in latitude and longitude. Additionally, estimates of Outgoing Longwave Radiation (OLR) from National Oceanic and Atmospheric Administration (NOAA), with a spatial resolution of 2.5° x 2.5° degrees in latitude and longitude, were used for assessing the model ability to represent the MJO in conjunction with zonal wind at 850 and 200 hPa from ERA-Interim. This OLR estimation is generated through interpolations in time of polar-orbiting satellite data (for additional information, see Liebmann and Smith, 1996).

Six BAM-1.2 configurations for sub-seasonal prediction have been defined for evaluation. Characteristics such as vertical resolution, convection and boundary layer parameterizations were evaluated as well as the impact of soil moisture initialization. Single member hindcasts over the 1999/2000 – 2010/2011 extended austral summer period were produced for each configuration. Five of the configurations were defined by combining two convection schemes, the revised simplified Arakawa Shubert and the modified Grell-Dévényi, and two vertical diffusion schemes for the PBL, the moist-PBL Bretherton-Park scheme and the dry-PBL modified Mellor-Yamada, and two vertical resolutions, 42 and 64 sigma levels. These physical processes and vertical resolutions of

the model were selected because they have an important influence on the predictive ability of the MJO (Vitart, 2014; Boyle et al., 2015; Wang and Chen, 2017) and consequently in the sub-seasonal precipitation and T2M predictions. It is important to highlight that other aspects such as horizontal resolution, radiation and microphysics parameterizations are also important for the good representation of the MJO (Zhang, 2005; Vitart, 2014; Wang et al., 2018). However, such characteristics were not evaluated in the present work.

The sixth configuration evaluates the impact of soil moisture. This characteristic is a source of predictability for the sub-seasonal timescale and has a positive impact on GCM predictive ability, especially in longer lead times such as when predicting weeks 3 and 4 (Koster et al., 2010). In this part of the study, the mean soil moisture from the previous month of the start date of each hindcast from the Global Land Data Assimilation System (GLDAS) version 2 product (Rui and Beaudoing, 2017) was used to initialize the soil moisture rather than using the monthly climatological soil moisture estimate in order to assess whether a more realistic soil moisture condition has an impact on the predictive ability of BAM-1.2. The monthly climatological soil moisture data estimates used in this study were obtained from the balance analyses of Willmott et al. (1985). Both GLDAS and climatological soil moisture data estimates were interpolated to the model Gaussian grid and converted to soil moisture fraction for hindcasts initialization.

The six examined configurations are defined in Table 2 and are summarized below:

- 42ABC: BAM-1.2 with 42 vertical levels, revised simplified Arakawa-Schubert deep convection parameterization, moist Bretherton-Park boundary layer parameterization, and climatological soil moisture initialization;
- 64ABC: BAM-1.2 with 64 vertical levels, revised simplified Arakawa-Schubert deep convection parameterization, moist Bretherton-Park boundary layer parameterization, and climatological soil moisture initialization;
- 42ABG: BAM-1.2 with 42 vertical levels, revised simplified Arakawa-Schubert deep convection parameterization, moist Bretherton-Park boundary layer parameterization, and soil moisture initialized through the GLDAS version 2 product;
- 42GBC: BAM-1.2 with 42 vertical levels, modified Grell-Dévényi deep convection parameterization, moist Bretherton-Park boundary layer parameterization, and climatological soil moisture initialization;
- 64GBC: BAM-1.2 with 64 vertical levels, modified Grell-Dévényi deep convection parameterization, moist Bretherton-Park boundary layer parameterization, and climatological soil moisture initialization;
- 42AMC: BAM-1.2 with 42 vertical levels, revised simplified Arakawa-Schubert deep convection parameterization, dry modified Mellor-Yamada boundary layer parameterization, and climatological soil moisture initialization.

### 2.3. Evaluation metrics and ensemble approaches

We assessed the ability of the six BAM-1.2 configurations to predict precipitation, T2M and the MJO. For precipitation and T2M, the deterministic assessment consists of

computing the Pearson correlation and Root-Mean-Square Error (RMSE) between the prediction and observed anomalies. Each metric was calculated for each grid point and for four lead times: days 1-7 (week-1), 8-14 (week-2), 15-21 (week-3) and 22-28 (week-4). The results were evaluated in the form of weekly averages because the model is expected to have a greater ability to predict weekly anomalies than daily values when producing sub-seasonal predictions (Vitart, 2014).

The performance of MJO prediction was evaluated using the Real-time Multivariate MJO indices (RMMs) (Wheeler and Hendon, 2004). Reference RMMs were calculated using the meridional wind at 850 and 200 hPa from the ERA-Interim reanalyses and satellite observed OLR. RMMs for hindcasts were calculated as proposed by Rashid et al. (2011). The metrics used for the MJO prediction quality assessment were bivariate correlation and RMSE (Lin et al., 2008) with lead times in days.

In addition to the single member deterministic prediction assessment, we evaluated the ability of the Multiple Configurations Ensemble (MCEN) mean prediction formed by the six here investigated BAM-1.2 configurations with each configuration representing one ensemble member. This was compared to an Initial Condition Ensemble (ICEN) produced with an Empirical Orthogonal Function (EOF) perturbation methodology (Mendonça and Bonatti, 2009), using the configuration that showed the best performance among the six evaluated configurations for producing six ensemble members. The EOF-based perturbation methodology is in operation at CPTEC for extended range forecasts up to 15 days. The methodology produces optimally perturbed analyses by applying the EOFs to $n$ time series formed by the differences between a

model run initialized with a control initial condition and *n* model runs initialized with randomly perturbed initial conditions. The initial random perturbations are drawn from a Gaussian distribution with zero mean and standard deviation comparable to the model short length forecast error [e.g., 3 ms−1 for the horizontal wind field components, 0.6 K for the air temperature field, 1 hPa for the surface pressure field and a vertical standard deviation profile for the specific humidity derived from the ECMWF background error covariance matrix (Derber and Boutier, 1999)]. The EOF analysis is performed over the Northern and Southern Hemispheres, over the tropical domain and also regionally over southern and northern South America. The EOF perturbations are the ones associated with the fast growth coefficients. To be used as optimal perturbations, these fast growth modes are rescaled in order to have a standard deviation of the same order of magnitude as the initial perturbations. Finally, the optimal perturbation is added and subtracted to/from the control analysis and an ensemble of 2*n* initial perturbed states is produced. A more detailed revision of the EOF-based perturbation methodology used at CPTEC can be found in Cunningham et al. (2015).

The above mentioned deterministic metrics for precipitation and T2M anomaly hindcasts, as well as for the hindcast MJO indices, were calculated for the ensemble mean of the two equal size (six members) ensemble (MCEN and ICEN) to assess and compare the value of utilizing multiple sub-seasonal predictions using two approaches.

In order to have an assessment of the differences in the obtained scores for the investigated model configurations and the two ensemble mean approaches, 95% confidence intervals were computed for the mean correlation and RMSE (for

precipitation and T2M anomalies), globally averaged between 60° N and 60° S, and for the bivariate correlation and RMSE (for the MJO) using a bootstrap resampling procedure with replacement with 1000 samples.

## 3. HINDCAST QUALITY ASSESSMENT

### 3.1. Deterministic evaluation of the six investigated BAM-1.2 configurations

Figure 1 shows the correlation between predicted and observed (GPCP) precipitation anomalies for the six BAM-1.2 configurations (first 6 rows) and four lead times (four columns representing week-1, week-2, week-3 and week-4). The 10 hindcasts per extended austral summer (5 months times 2 start dates) over 12 austral summers produce a sample with a total of 120 hindcasts. Applying a two-side Student's t test, the correlation value of 0.2 is statistically significant, different from zero at the 5% level.

For the six configurations, in general, the correlation is high during the first week in most regions and drops rapidly as lead time increases. This fall is more pronounced between the first and second week for all six configurations as the forecasts extend beyond the deterministic limit of predictability for many scales and we are considering only a single member for this initial analysis. It is noted that all configurations show greater correlation over the North Hemisphere than the South Hemisphere during week-1 and week-2. This is because GCMs are more likely to predict winter baroclinic weather systems and associated fronts (Zhu et al., 2014). As of the third week, BAM-1.2 correlation values are smaller than 0.2 in practically all extratropical regions. This illustrates that the predictive ability of BAM-1.2 over mid-latitudes beyond 15 days is limited for single member hindcasts. For weeks 3 and 4, significant correlation values

292 are only seen in the Tropical Pacific Ocean, over a few areas in northern South America
293 and in the equatorial Atlantic Ocean. The high correlation values in the first two lead
294 times, especially at week-1, are associated to the predictability provided by the initial
295 conditions, and the high correlation values observed in the last two lead times over the
296 equatorial Pacific Ocean are mainly associated to the predictability provided by the El
297 Nino-Southern Oscillation (ENSO) and the MJO (Li and Robertson, 2015; de Andrade
298 et al., 2019). All six configurations show negligible correlation values near the
299 Subtropical Highs and desert regions from week-1. These characteristics are also
300 noticed in other GCMs configured for sub-seasonal predictions (Zhu et al, 2014; Li and
301 Robertson, 2015; Wheeler et al., 2017; de Andrade et al., 2019) and are associated with
302 the low capacity of GCMs to simulate small precipitation rates in these regions.

303 The spatial correlation pattern is similar for the four weeks of each of the six BAM-1.2
304 configurations. However, this pattern for the configurations with revised simplified
305 Arakawa-Schubert (deep convection) and moist Bretherton-Park (boundary layer)
306 parameterizations shows slightly larger values in the first two weeks than for the other
307 configurations (Figure 1, first two columns of configurations 42ABC, 64ABC and
308 42ABG vs. first two columns of configurations 42GBC, 64GBC and 42AMC). In the
309 week-3 and week-4 (last two columns of Figure 1), correlation levels are similar in
310 terms of both spatial pattern and intensity for the six configurations. Increasing the
311 vertical resolution shows very little change in the precipitation correlation levels at any
312 lead time. For example, the 42 vertical level configuration, 42ABC (first row of Figure
313 1), and the 64 vertical level configuration, 64ABC (second row of Figure 1), have
314 nearly identical correlation values for most regions. The same is noticed for

315 configurations 42GBC (fourth row of Figure 1) and 64GBC (fifth row of Figure 1).
316 Initialization of the soil moisture also shows no increase in the correlation values for all
317 four investigated weeks. Hindcasts initialised with climatological soil moisture
318 (42ABC, first row of Figure 1) have the same correlation levels as 42ABG (third row of
319 Figure 1) hindcasts, which were initialized with GLDAS soil moisture.

320 Figure 2 shows the precipitation anomaly RMSE spatial features for the six BAM-1.2
321 configurations. Highest RMSE values are found over the Intertropical Convergence
322 Zone (ITCZ), Indian Ocean, Maritime Continent, South Pacific Convergence Zone
323 (SPCZ) and South American Convergence Zone (SACZ), which are regions of strong
324 sub-seasonal variability (Liu et al., 2014). The errors grow as lead time increases. As for
325 the correlation assessment, the errors grow more between week-1 and week-2 than from
326 weeks 2 to 3. Again, configurations with revised simplified Arakawa-Shubert and moist
327 Bretherton-Park parameterizations have the fewest errors and do not differ greatly from
328 each other (rows 42ABC, 64ABC and 42ABG in Figure 2). Configurations with
329 modified Grell-Dévényi parameterizations (rows 42GBC and 64GBC in Figure 2) also
330 do not differ much from each other and have larger errors when compared to
331 configurations with revised simplified Arakawa-Shubert parameterizations. That is, the
332 revised simplified Arakawa-Shubert parameterization seems to be better than the
333 modified Grell-Dévényi parameterization and increase of the vertical resolution and soil
334 moisture initialization do not reduce the errors of the hindcasts in any lead time.

335 To better note the differences between the six configurations, the mean global
336 correlation between 60°N and 60°S was calculated as a function of lead time (Figure
337 3a). Vertical bars represent bootstrap 95% confidence intervals. The six configurations

show a near-exponential drop in correlation as a function of lead time. The configurations with the revised simplified Arakawa-Shubert and moist Bretherton-Park parameterizations (black, orange and blue lines) have the largest correlation values when compared to the other configurations in the week-1 and week-2. Important improvements are noticed when comparing the configurations with revised simplified Arakawa-Shubert and modified Grell-Dévényi deep convection parameterizations at the first two lead times For example, 42ABC (black line) has a global mean correlation equals to 0.45 at week-1 and drops to 0.18 at week-2, whereas 42GBC (yellow line) has a global mean correlation values equals to 0.40 at week-1 and drops to 0.14 at week-2. The 95% confidence intervals for the 42ABC (black vertical bars on top of solid black line) do not overlap the 95% confidence intervals for the 42GBC (yellow vertical bars on top of solid yellow line), illustrating the superiority of 42ABC over 42GBC. However, the six configurations show similar correlation levels at week-4. As noted in the previous figures, the increase of vertical resolution does not result in an increase in the correlation values. This feature is noticed when we compare the 42ABC (black line) and 64ABC (orange line) configurations or the 42GBC (yellow line) and 64GBC (green line) configurations, which show practically the same behaviour, with the differences between configurations smaller than the 95% confidence intervals (vertical bars). This is also noticed for soil moisture initialization, where the 42ABG configuration (blue line) shows similar correlation levels to the 42ABC configuration (black line).

The global RMSE mean between 60°N and 60°S (Figure 3b) further emphasizes the differences between the revised simplified Arakawa-Shubert and modified Grell-Dévényi deep convection parameterizations. The modified Grell-Dévényi

parameterization (green and yellow lines) produces larger errors than the revised simplified Arakawa-Shubert parameterization (other lines) at all lead times. The differences between the errors of these two parameterizations are much larger than the 95% confidence intervals (vertical bars) shown in Figure 3b, illustrating the superiority of the revised simplified Arakawa-Shubert over the modified Grell-Dévényi parameterization. The configurations 42ABC (black line) and 64ABC (orange line) show very similar values at the four lead times, with overlapping 95% confidence intervals. This is also noticed with the 42GBC (yellow line) and 64GBC (green line) configurations. These results suggest that increasing the vertical resolution does not decrease the RMSE. The initialization of soil moisture also does not contribute to the reduction of error (black vs. blue lines). An interesting aspect is that the configuration with dry modified Mellor-Yamada boundary layer parameterization (red line) has the smallest error in the last two lead times.

Figure 4 shows the correlation between predicted and reanalyses (ERA-Interim) T2M anomalies for the six configurations and four lead times. The six BAM-1.2 configurations show better prediction performance for T2M anomalies than precipitation anomalies (see Figures 1 and 4). The correlation values decrease with lead time. The highest correlation values are seen over cloud free oceanic regions (e.g., 42ABC row in Figure 4). However, significant sub-seasonal correlation values exist over a large portion of the global land domain. Over extratropical continental regions, strong correlation values are observed in restricted regions at week 3 and 4, for example, over the southeast of the United States and some regions over Asia. Over

tropical regions, correlation values are low in regions with high convective activity (e.g., over the Maritime Continent).

The spatial correlation pattern is similar for the four weeks of the six BAM-1.2 configurations. The difference in performance between configurations with revised simplified Arakawa-Shubert and modified Grell-Dévényi deep convection parameterizations is not observed for T2M (e.g., row 42ABC vs. row 42GBC in Figure 4). Configurations with these two parameterizations have the same performance level for the four weeks of lead time. There are differences when comparing the two boundary layer parameterizations. The configuration with the dry modified Mellor-Yamada parameterization (row 42AMC in Figure 4) shows reduced performance than the other five configurations at all lead times, which were configured with the moist Bretherton-Park boundary layer parameterization. Increasing vertical resolution from 42 (rows 42ABC and 42GBC in Figure 4) to 64 (rows 64ABC and 64GBC in Figure 4) levels seems to slightly reduced prediction performance of extratropical T2M anomalies in the first two weeks of lead times. Predictions with the initialisation of soil moisture (row 42ABG in Figure 4) rather than the climatology soil moisture (row 42ABC in Figure 4), show a slight improvement in correlation in the continental regions (Australia, South and North Americas and Africa) at week 2 and 3 lead time.

Figure 5 shows the T2M anomaly RMSE spatial features for the six BAM-1.2 configurations. In all configurations, RMSE values grow with the lead time and are generally lower over oceanic regions than over continental regions for all 4 lead times. The highest RMSE values are noticed over Northern Hemisphere regions where there are interactions between mid-latitude baroclinic system, and tropical convective

406 anomalies, which are usually associated with the MJO and circulation teleconnections
407 through Rossby waves (Stan et al., 2017; Hu et al., 2019). Over northern Asia, high
408 RMSE values are also noticed. The RMSE values are lower over the Southern
409 Hemisphere because there are fewer continental regions than over the Northern
410 Hemisphere, and baroclinic instability is lower at this time of the year in the Southern
411 Hemisphere. The latter makes the interaction between the convective anomalies over
412 tropical regions and circulation over mid-latitudes regions less pronounced. As a result,
413 the sub-seasonal variability over the Southern Hemisphere extratropical regions is also
414 reduced during the austral summer.

415 Configurations with the revised simplified Arakawa-Shubert and modified Grell-
416 Dévényi deep convection parameterizations present similar RMSE patterns (Figure 5).
417 Some differences are found in specific regions. For example, the 42GBC configuration
418 shows slightly lower RMSE values over southern South America than the 42ABC
419 configuration at week 3. The opposite is noticed over the Iberian Peninsula. Similar
420 features are noticed for the increase in vertical resolution. Concerning the initialization
421 of soil moisture, subtle differences are noticed between 42ABC and 42ABG, with slight
422 improvements over continental regions such as Australia, South America, southern
423 Africa and North America with initialized soil moisture (42ABG). Large differences are
424 found when comparing configurations with moist Bretherton-Park and dry modified
425 Mellor-Yamada boundary layer parameterizations. For example, the RMSE values are
426 lower in high latitude regions over North America and Asia for the 42AMC
427 configuration when compared to the 42ABC configuration. The opposite is over tropical
428 and medium latitudes regions.

Figure 6 shows the global mean T2M anomaly correlation (Figure 6a) and RMSE (Figure 6b) averaged between 60°N and 60°S as a function of lead time with 95% confidence intervals (vertical bars). The six configurations show a similar drop (rise) in correlation (RMSE) as a function of lead time. The increase of vertical resolution from 42 to 64 levels, change of deep convection scheme and soil moisture initialization do not influence the levels of correlation and error values for T2M anomalies predictions for the global perspective. This feature is noticed by the proximity or overlap of correlation and RMSE lines of most investigated configurations shown in Figure 6, with overlapping 95% confidence intervals. Differences in performance levels are noticed when comparing moist Bretherton-Park and dry modified Mellor-Yamada boundary layer parameterizations. The dry modified Mellor-Yamada parameterization (red line) produces smaller correlation values and larger errors than the other five configurations at all lead times, with the differences between these configurations and the others larger than the 95% confidence intervals (vertical bars) illustrating the superiority of the other configurations.

Figures 7a and 7b show MJO bivariate correlation and bivariate RMSE of all six BAM-1.2 configurations, respectively. Vertical bars represent bootstrap 95% confidence intervals. The MJO predictive ability is determined when the bivariate correlation is lower than 0.5 and the bivariate RMSE grows to $\sqrt{2}$. The lead times for these two thresholds to be reached are usually found to be close (Rashid et al., 2011). The bivariate correlation decreases with the increase in lead time and crosses the threshold of 0.5 in 18-19 days for all configurations, except for the 42AMC configuration (red line), which uses the dry modified Mellor-Yamada boundary layer parameterization and

has a much reduced performance, with the bivariate correlation reaching the 0.5 threshold in 12 days. The bivariate RMSEs (Figure 7b) increase with lead time and each configuration reaches the bivariate RMSE value of $\sqrt{2}$ at approximately the same lead time as the bivariate correlation. The 42AMC configuration crosses the threshold value of $\sqrt{2}$ in 11 days, whereas the other five configurations cross the threshold value of $\sqrt{2}$ in around 18 to 19 days. The overlap of the 95% confidence intervals (vertical bars) for most configurations (expect 42AMC) illustrates their similarity in MJO predictive ability.

### 3.2.Deterministic assessment of two investigated ensemble approaches

With the precipitation anomalies, T2M anomalies and MJO hindcast evaluation of the six configurations shown in the previous section, a preferred BAM-1.2 configuration was determined for defining an ensemble sub-seasonal forecasting system for CPTEC. The increase of the vertical resolution from 42 levels to 64 levels did not result in increase in predictive ability, therefore a vertical resolution of 42 levels was selected. The moist Bretherton-Park boundary layer parameterization was selected because it contributed to a better performance than the dry modified Mellor-Yamada boundary layer parameterization, especially for T2M anomalies and MJO predictions. The revised simplified Arakawa-Shubert and modified Grell-Dévényi deep convection parameterizations showed similar ability for T2M anomalies and MJO prediction with a slight advantage to the modified Grell-Dévényi parameterization for MJO prediction. On the other hand, the revised simplified Arakawa-Shubert parameterization showed a large advantage for sub-seasonal precipitation, with higher correlation and smaller errors values than the modified Grell-Dévényi parameterization. Based on this

assessment, the revised simplified Arakawa-Shubert deep convection parameterization was chosen. Soil moisture initialization instead of the climatology led to subtle improvements in T2M anomalies predictions in specific regions (e.g., Australia). However, these improvements were lower than expected and given limitations in the availability of accurate real-time soil moisture data, the use of climatological soil moisture was selected for the BAM-1.2 system. Therefore, the chosen BAM-1.2 version for ensemble sub-seasonal forecasting was the 42ABC configuration.

The possible physical reasons why the selected configuration (42ABC) performed better than the other investigated configurations, particularly in terms of the tested boundary layer and deep convections parameterizations are as follows. The use of the moist Bretherton-Park boundary layer parameterization resulted in better MJO and T2M predictions performance than the use of the dry modified Mellor-Yamada boundary layer parameterization. This is because the moist Bretherton-Park has several advantages compared to the dry modified Mellor-Yamada parameterization. The main contribution of the moist Bretherton-Park parameterization is to improve the representation of the stable night boundary layer, where the predominant physical process in flat areas such as the oceans is the surface radiative cooling. The evolution of the stable nocturnal boundary layer depends on the radiative cooling rate, and therefore the presence of clouds is essential for reducing radiative loss. At sunrise, the state of the stable boundary layer will be important for the evolution of vertical instability and the mixing boundary layer. Therefore, the higher or lower the energy released during the evolution of the stable boundary layer this energy surplus or deficit will contribute to the formation of shallow and deep clouds, and consequently impacts the daytime

498 temperature and precipitation cycle. The energy scales produced by these processes
499 directly or indirectly impact the atmospheric conditions on the sub-seasonal time scale.
500 As for the comparative performance of sub-seasonal precipitation predictions, important
501 differences were noted when changing the deep convection parameterizations. The
502 revised simplified Arakawa-Shubert parameterization showed better performance than
503 the modified Grell-Dévényi parameterization. This is likely due to the revision made by
504 Han and Pan (2011) in the simplified Arakawa-Shubert parameterization to suppress
505 unrealistic grid point storms due to remaining instability in the atmospheric column.We
506 next further assess BAM-1.2 sub-seasonal hindcast quality through a deterministic
507 evaluation of ensemble mean predictions. Two ensemble types are evaluated and
508 presented here. The first ensemble consists of one ensemble member from each of the
509 six configurations previously presented, which was denominated Multiple
510 Configurations Ensemble (MCEN). The second ensemble was denominated Initial
511 Condition Ensemble (ICEN) and is composed of six members produced with the chosen
512 42ABC configuration consisting of a control member and five perturbed members
513 produced with an EOF method (Mendonça and Bonatti, 2009; Cunningham et al.,
514 2015).

515 The assessment of the ensemble mean of the two ensemble types (MCEN and ICEN)
516 revealed important increase in global mean performance at four lead times for
517 precipitation anomalies predictions when compared to the single member assessment of
518 the six investigated BAM-1.2 configurations (Figure 3), with the increase in
519 performance larger than the 95% confidence intervals (vertical bars). The two ensemble
520 mean approaches show similar correlation levels (dashed lines in Figure 3a) and

521 overlapping 95% confidence intervals (vertical grey and black bars on top of dashed
522 lines). This shows that BAM-1.2 performance increases when more (six) members are
523 used to form an ensemble with the 42ABC configuration or when using the six
524 configurations as an ensemble. The predictive ability of GCMs increases as the number
525 of members increases because the ensemble mean acts as a filter for decreasing the
526 uncertainties of the initial conditions used to run the model (Cheung, 2001). This is
527 noticed over several regions in the four investigated lead times (e.g., Figure 1). For
528 example, over extratropical regions at week 2, over eastern South America at week 3
529 and over northern South America at week 4. Precipitation anomaly hindcasts also show
530 lower RMSE values for both ICEN and MCEN at all four lead times (dashed lines in
531 Figure 3b), with the reduction of error much larger than the 95% confidence intervals
532 (vertical bars). Improvements are noticed primarily over the ITCZ, Indian Ocean,
533 Maritime Continent, SPCZ and SACZ regions (last two rows in Figure 2). This suggests
534 that the ensemble mean helps BAM-1.2 to better represent the sub-seasonal variability
535 in these regions. The same feature is noticed for the T2M hindcasts ensemble means.
536 The two ensemble means show improved T2M anomalies performance when compared
537 to the single member performance with increased correlation values and decreased error
538 (see last two rows in Figure 4 and 5 and dashed lines in Figures 6). Improvements in
539 MJO forecast performance are also noticed when using the two ensemble approaches.
540 The prediction ability limits are around lead times 18 and 19 days for the single member
541 configurations, except for the 42AMC which is much reduced (solid lines in Figures 7).
542 For the MCEN this limit increases to 21 days (dashed grey line in Figures 7a-b) and to
543 20 days (dashed black line in Figures 7a-b) for the ICEN. However, these improvements

are less prominent than those identified for precipitation and T2M, because the 95% confidence intervals of the two ensemble approaches largely encompass the 95% confidence intervals of the single members of the individual investigated configurations.

## 4. CONCLUSIONS

Vertical resolution and physical parameterizations (deep convection and boundary layer) were changed in BAM-1.2 to form five model configurations to determine the model configuration with greater performance for sub-seasonal predictions. These components were selected because these parameters have an important impact on GCMs simulated MJO (Zhang, 2005; Wang and Chen, 2017; Wang et al., 2018). Given the soil moisture initialization potential to increase the sub-seasonal predictions performance (Koster et al., 2010; Koster et al., 2011; Guo et al., 2012), a further configuration initialized with monthly soil moisture from the previous month, rather than the climatological mean soil moisture, was formed to investigate the impact of soil moisture initialization on BAM-1.2 predictive ability. The configuration with the best result was selected to form an initial condition ensemble (ICEN) with six members, one control member and five perturbed members produced using an EOF-based methodology. The six configurations, individually evaluated in the first part of this work, were also used to form another ensemble (multiple configurations ensemble-MCEN) to compare the improvements inherent in the ensemble mean between ICEN and MCEN.

All six BAM-1.2 configurations produced high precipitation and T2M anomalies correlation levels for the first week and decreased correlation levels for weeks 2-4. For weeks 3-4, moderate precipitation anomaly correlation levels were restricted to the

Equatorial Pacific Ocean region. This feature was also noticed in other models (e.g., Li and Robertson, 2015; de Andrade et al., 2019). Precipitation anomaly RMSE increased with lead time and the highest RMSE values were found over regions with strong sub-seasonal variability, for example, over the ITCZ, Indian Ocean, Maritime Continent, SPCZ and SACZ (Liu et al., 2013). For T2M, this characteristic was identified over the Northern Hemisphere where interaction between anomalous convection and mid-latitudes circulation anomalies are noticed (Hu et al., 2019). The six BAM-1.2 configurations showed better prediction performance for T2M anomalies than for precipitation anomalies.

The increase of the vertical resolution from 42 levels to 64 levels did not result in an increase in predictive ability. Comparing 42ABC with 64ABC (revised simplified Arakawa-Shubert deep convection configurations with 42 and 64 levels, respectively) and 42GBC with 64GBC (modified Grell-Dévényi deep convection configurations with 42 and 64 levels, respectively), it was noticed that the correlation and RMSE showed nearly identical levels for all lead times for precipitation anomalies, T2M anomalies and MJO predictions. These results may sound contradictory since other studies have shown that the increase in vertical resolution contributes to improvements in predictive ability in the sub-seasonal timescale (Zhang, 2005; Vitart, 2014). Other factors might be contributing to this finding. For example, the use of initial conditions with only 37 vertical levels, which had to be interpolated to cover 42 and 64 levels.

BAM-1.2 configurations with revised simplified Arakawa-Shubert deep convection parameterization showed better performance than BAM-1.2 configurations with modified Grell-Dévényi deep convection parameterization for sub-seasonal

precipitation prediction, with the largest correlation levels found in the first two weeks and the smallest RMSE in the four lead times. However, these two parameterizations showed practically the same performance for T2M anomalies and MJO, with the commonly used performance thresholds reached at 18/19 days of lead time. The fact that the BAM-1.2 model presents increased ability for sub-seasonal precipitation forecast with the revised simplified Arakawa-Shubert deep convection parameterization compared to modified Grell-Dévényi and very similar ability for T2M and MJO is intriguing. Han and Pan (2011) provided a revision of the simplified Arakawa-Shubert deep convection in the National Centers for Environmental Prediction's (NCEP) global forecast system. This revision aimed to suppress unrealistic gridpoint storms due to remaining instability in the atmospheric column. This might be a possible reason for the better BAM-1.2 performance in sub-seasonal precipitation anomaly forecasting with revised simplified Arakawa-Shubert deep convection parameterization.

The moist Bretherton-Park boundary layer parameterization produced better performance for precipitation anomalies, T2M anomalies and MJO predictions than the dry modified Mellor-Yamada boundary layer parameterization. The greatest differences were noticed for MJO predictions, where the bivariate correlation decreased more sharply as a function of lead time with the configuration using the dry modified Mellor-Yamada parameterization (42AMC) than for the other configurations. For this, the 42AMC has correlation below 0.5 and RMSE above $\sqrt{2}$ around the $11^{th}$ day of lead time. Large discrepancies were also noticed for T2M anomalies predictions when comparing 42AMC with the other configurations.

We did not find important impacts of soil moisture initialization when compared to climatological initialization on the predictive ability of precipitation anomalies in the four investigated lead times. Slight improvements were seen for T2M anomaly predictions in some continental regions. These improvements were smaller than those found in Koster et al. (2010), Koster et al. (2011), Guo et al. (2012) and Van den Hurk and et al. (2012), and might be related to differences in the investigated seasons, time window, experiments or/and even to low BAM-1.2 sensitivity to soil moisture initialization.

With the evaluation of the six configurations, it was possible to determine a configuration for use as CPTEC sub-seasonal ensemble system. For this, the determined configuration was the 42ABC. This configuration consists of a model version at TQ126 spatial resolution, 42 vertical sigma levels, revised simplified Arakawa-Shubert deep convection parameterization, moist Bretherton-Park boundary layer scheme, initialization with climatological soil moisture, CLIRAD-LW, CLIRAD-SW, Morrison microphysics and the IBIS-2.6-CPTEC surface model.

The deterministic evaluation of ensembles (MCEN and ICEN), through the computation of the ensemble means, presented considerable improvements when compared to the single (control) member evaluation (42ABC). For precipitation and T2M anomalies predictions, this improvement was noticed mainly in extratropical continental regions. For MJO predictions, the ensemble means extended in two days the MJO prediction ability limit (e.g., up to 20 days). It is interesting to note that the MCEN mean, formed from the six configurations, showed very similar level of improvements to ICEN when compared to the control member.

This work focused on determining a global CPTEC model configuration for sub-seasonal prediction through a deterministic assessment using a limited number of ensemble members (six). The results presented in this paper suggest that BAM-1.2 has competitive performance to other S2S models (Vitart et al., 2017; Lim et al., 2018; de Andrade et al. 2019). In a forthcoming paper, we plan to perform a probabilistic assessment of the defined configuration with an increased number of ensemble members and a more detailed comparison of BAM-1.2 with other S2S models. It is worth mentioning that although the extended austral summer is a fundamental season for Brazil (particularly in terms of precipitation) further work is needed in order to evaluate the performance of the Brazilian model during other seasons for identifying regions where best to trust the model for issuing operational sub-seasonal predictions.

**Acknowledgements**

We thank two anonymous reviewers for providing valuable comments and suggestions that contributed for improving the quality of this manuscript. The first author was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeicoamento de Pessoal de Nível Superior (CAPES) and University of Reading (ref GS18-179). CASC thanks CNPq, process 304586/2016-1, and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), process 2015/50687-8 (CLIMAX Project) for the support received. SJW was supported by the National Centre for Atmospheric Science ODA national capability programme ACREW (NE/R000034/1), which is supported by NERC and the GCRF. This research was partially supported by the Climate Science for Services Partnership Brazil project

657 (CSSP-Brazil) funded by the Newton Fund. DCS was supported by CNPq (process
658 167804/2018-9).

659

660 **References**

661 Arakawa A and Schubert WH. 1974. Interaction of a cumulus cloud ensemble

662 Boyle JS, Klein SA, Lucas DD, Ma HY, Tannahill J and Xie S. 2015. The parametric
663 sensitivity of CAM5's MJO. Journal of Geophysical Research: Atmospheres. **120**:1424-
664 44.

665 Bretherton CS and Park S. 2009. A new moist turbulence parameterization in the
666 Community Atmosphere Model. Journal of Climate. **22**:3422-48.

667 Chen M, Wang W and Kumar A. 2010. Prediction of monthly-mean temperature: The
668 roles of atmospheric and land initial conditions and sea surface temperature. Journal of
669 Climate, **23:**717-725.

670 Cheung KK. 2001. A review of ensemble forecasting techniques with a focus on
671 tropical cyclone forecasting. Meteorological Applications. **8**:315-32.

672 Chou MD and Suarez, MJ. 1999. A solar radiation parameterization (CLIRAD-SW) for
673 atmospheric studies. NASA/TM-1999-104606, M. J. Suarez, Ed., Series on Global
674 Modeling and Data Assimilation. **15**, 40 pp.

675 Chou MD, Suarez MJ, Liang XZ, Yan MM and Cote C. 2001. A thermal infrared
676 radiation parameterization for atmospheric studies. NASA/TM-2001-104606, Vol. 19.

677 Cunningham C, Bonatti JP and Ferreira M. 2015. Assessing improved CPTEC
678 probabilistic forecasts on medium-range timescale. Meteorological Applications. **22**:
679 378-384.

680 de Andrade, FM, Coelho, CA and Cavalcanti, IF. 2019. Global precipitation hindcast
681 quality assessment of the Subseasonal to Seasonal (S2S) prediction project models.
682 Climate Dynamics, **52**: 5451-5475.

683 Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U,
684 Balmaseda MA, Balsamo G, Bauer DP and Bechtold P. 2011. The ERA-Interim
685 reanalysis: Configuration and performance of the data assimilation system. Quarterly
686 Journal of the royal meteorological society. **137**:553-97.

687 Derber J and Bouttier F. 1999. A reformulation of the background error covariance in
688 the ECMWF global data assimilation system. Tellus A: Dynamic Meteorology and
689 Oceanography. **51**: 195-221.

690 Figueroa SN, Bonatti JP, Kubota PY, Grell GA, Morrison H, Barros SR, Fernandez JP,
691 Ramirez E, Siqueira L, Luzia G and Silva J. 2016. The Brazilian global atmospheric
692 model (BAM): performance for tropical rainfall forecasting and sensitivity to
693 convective scheme and horizontal resolution. Weather and Forecasting, **31**: 1547-1572.

694 Green BW, Sun S, Bleck R, Benjamin SG and Grell GA. 2017. Evaluation of MJO
695 predictive skill in multiphysics and multimodel global ensembles. Monthly Weather
696 Review. **145**: 2555-2574.

697 Grell GA and Freitas SR. 2014: A scale and aerosol aware stochastic convective
698 parameterization for weather and air quality modeling. Atmos. Chem. Phys., **14**: 5233–
699 5250.

700 Grell GA. 1993. Prognostic evaluation of assumptions used by cumulus
701 parameterizations. Mon. Wea. Rev. **121:** 764–787.

Grell, GA and Dévényi D. 2002. A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. Geophysical Research Letters, 29: 38-1.

Guo Z, Dirmeyer PA, DelSole T and Koster RD. 2012. Rebound in atmospheric predictability and the role of the land surface. Journal of Climate. **25**: 4744-9.

Han J and Pan HL. 2011. Revision of convection and vertical diffusion schemes in the NCEP global forecast system. Weather and Forecasting. **26**: 520-33.

Hu W, Liu P, Zhang Q and He B. 2019. Dominant patterns of winter-time intraseasonal surface air temperature over the CONUS in response to MJO convections. Climate Dynamics. 1-20.

Hudson D, Alves O, Hendon HH and Marshall AG. 2011. Bridging the gap between weather and seasonal forecasting: intraseasonal forecasting for Australia. Quarterly Journal of the Royal Meteorological Society, **137**: 673-689.

Huffman GJ, Adler RF, Morrissey MM, Bolvin DT, Curtis S, Joyce R, McGavock B and Susskind J. 2001. Global precipitation at one-degree daily resolution from multisatellite observations. Journal of hydrometeorology. **2**:36-50.

Koster RD, Mahanama SP, Yamada TJ, Balsamo G, Berg AA, Boisserie M, Dirmeyer PA, Doblas-Reyes FJ, Drewitt G, Gordon CT, Guo Z. 2010. Contribution of land surface initialization to subseasonal forecast skill: First results from a multi-model experiment. Geophysical Research Letters. **37**.

722 Koster RD, Mahanama SP, Yamada TJ, Balsamo G, Berg AA, Boisserie M, Dirmeyer
723 PA, Doblas-Reyes FJ, Drewitt G, Gordon CT and Guo Z. 2011. The second phase of the
724 global land–atmosphere coupling experiment: soil moisture contributions to subseasonal
725 forecast skill. Journal of Hydrometeorology. **12**: 805-22.

726 Kubota PY. 2012. Variability of storage energy in the soil-canopy system and its impact
727 on the definition of precipitation standard in South America (in Portuguese with abstract
728 in English). Ph. D. thesis, Instituto Nacional de Pesquisas Espaciais (INPE), São José
729 dos Campos, Brazil.

730 Kumar A, Chen M and Wang W. 2011. An analysis of prediction skill of monthly mean
731 climate variability. Climate dynamics, **37:** 1119-1131.

732 Li S and Robertson AW. 2015. Evaluation of submonthly precipitation forecast skill
733 from global ensemble prediction systems. Monthly Weather Review, 143: 2871-2889.

734 Liang P and Lin H. 2018. Sub-seasonal prediction over East Asia during boreal summer
735 using the ECCC monthly forecasting system. Climate dynamics, **50:** 1007-1022.

736 Liebmann B and Smith CA. 1996. Description of a complete (interpolated) outgoing
737 longwave radiation dataset. Bulletin of the American Meteorological Society. **77**:1275-
738 7.

739 Lim Y, Son SW and Kim D. 2018. MJO prediction skill of the subseasonal-to-seasonal
740 prediction models. Journal of Climate, **31(10)**: 4075-4094.

741 Lin H, Brunet G and Derome J. 2008. Forecast skill of the Madden–Julian oscillation in
742 two Canadian atmospheric models. Monthly Weather Review. **136**: 4130-49.

743 Lin H, Gagnon N, Beauregard S, Muncaster R, Markovic M, Denis B and Charron M.
744 2016. GEPS-based monthly prediction at the Canadian Meteorological Centre. Monthly
745 Weather Review, **144:** 4867-4883.

746 Liu X, Wu T, Yang S, Li T, Jie W, Zhang L, Wang Z, Liang X, Li Q, Cheng Y and Ren
747 H. 2017. MJO prediction using the sub-seasonal to seasonal forecast model of Beijing
748 Climate Center. Climate Dynamics, **48:** 3283-3307.

749 Liu X, Yang S, Li Q, Kumar A, Weaver S and Liu S. 2014. Subseasonal forecast skills
750 and biases of global summer monsoons in the NCEP Climate Forecast System version
751 2. Climate dynamics. **42**:1487-508.

752 Mastrangelo D, Malguzzi P, Rendina C, Drofa O and Buzzi A. 2012. First outcomes
753 from the CNR-ISAC monthly forecasting system. Advances in Science and Research,
754 **8:**77-82.

755 Mellor GL and Yamada T. 1982: Development of a turbulence closure model for
756 geophysical fluid problems. Rev. Geophys. Space Phys. **20**: 851–875

757 Mendonça AM and Bonatti J. 2009. Experiments with EOF-based perturbation methods
758 and their impact on the CPTEC/INPE ensemble prediction system. Monthly Weather
759 Review. **137**:1438-59.

760 Morrison H, Thompson G and Tatarskii V. 2009. Impact of cloud microphysics on the
761 development of trailing stratiform precipitation in a simulated squall line: Comparison
762 of one-and two-moment schemes. Monthly weather review. **137**: 991-1007.

763 Rashid HA, Hendon HH, Wheeler MC and Alves O. 2011. Prediction of the Madden–
764 Julian oscillation with the POAMA dynamical prediction system. Climate Dynamics.
765 **36**: 649-61.

766 Reichler T and Roads JO. 2005. Long-range predictability in the tropics. Part I: monthly
767 averages. Journal of climate, **18**: 619-633.

768 Rui H and Beaudoing H. 2017. Readme document for GLDAS Version 2 data products.
769 NASA Goddard Space Flight Center. **610**: 1-21.

770 Saha S, Moorthi S, Wu X, Wang J, Nadiga S, Tripp P, Behringer D, Hou YT, Chuang,
771 HY, Iredell M and Ek M. 2014. The NCEP climate forecast system version 2. Journal of
772 Climate. **27:** 2185-2208.

773 Shelly A, Xavier P, Copsey D, Johns T, Rodríguez JM, Milton S and Klingaman N.
774 2014. Coupled versus uncoupled hindcast simulations of the Madden-Julian Oscillation
775 in the Year of Tropical Convection. Geophysical Research Letters. **41**: 5670-7.

776 Stan C, Straus DM, Frederiksen JS, Lin H, Maloney ED and Schumacher C. 2017.
777 Review of tropical-extratropical teleconnections on intraseasonal time scales. Reviews
778 of Geophysics. **55**: 902-37.

779 Tarasova TA and Fomin BA. 2000. Solar radiation absorption due to water vapor:
780 Advanced broadband parameterizations. Journal of Applied Meteorology. **39**: 1947-
781 1951.

van den Hurk B, Doblas-Reyes F, Balsamo G, Koster RD, Seneviratne SI and Camargo H. 2012. Soil moisture effects on seasonal temperature and precipitation forecast scores in Europe. Climate Dynamics. **38**: 349-62.

Vitart F, Ardilouze C, Bonet A, Brookshaw A, Chen M, Codorean C, Déqué M, Ferranti L, Fucile E, Fuentes M and Hendon H. 2017. The subseasonal to seasonal (S2S) prediction project database. Bulletin of the American Meteorological Society. **98**: 163-173.

Vitart F. 2004. Monthly forecasting at ECMWF. Monthly Weather Review, **132**: 2761-2779.

Vitart F. 2014. Evolution of ECMWF sub-seasonal forecast skill scores. Quarterly Journal of the Royal Meteorological Society. **140**: 1889-1899.

Vitart F. 2017. Madden—Julian oscillation prediction and teleconnections in the S2S database. Quarterly Journal of the Royal Meteorological Society. **143**: 2210-2220.

Wang B and Chen G. 2017. A general theoretical framework for understanding essential dynamics of Madden–Julian oscillation. Climate Dynamics. **49**: 2309-28.

Wang B, Lee SS, Waliser DE, Zhang C, Sobel A, Maloney E, Li T, Jiang X and Ha KJ. 2018. Dynamics-oriented diagnostics for the Madden–Julian Oscillation. Journal of Climate. **31**: 3117-3135.

Weber, NJ and Mass, CF. 2017. Evaluating CFSv2 subseasonal forecast skill with an emphasis on tropical convection. Monthly Weather Review, **145**: 3795-3815.

802 Wheeler MC and Hendon HH. 2004. An all-season real-time multivariate MJO index:
803 Development of an index for monitoring and prediction. Monthly Weather Review. **132**:
804 1917-32.

805 Wheeler MC, Zhu H, Sobel AH, Hudson D and Vitart F. 2017. Seamless precipitation
806 prediction skill comparison between two global models. Quarterly Journal of the Royal
807 Meteorological Society, **143**: 374-383.

808 Willmott CJ, Rowe CM and Mintz Y. 1985. Climatology of the terrestrial seasonal
809 water cycle. International Journal of Climatology. **5**: 589-606.

810 with the large-scale environment, Part I. J.Atmos. Sci. **31:** 674–701.

811 Yu H, Kaufman YJ, Chin M, Feingold G, Remer LA, Anderson TL, Balkanski Y,
812 Bellouin N, Boucher O, Christopher S and DeCola P. 2006. A review of measurement-
813 based assessments of the aerosol direct radiative effect and forcing. Atmospheric
814 Chemistry and Physics. **3**: 613-66.

815 Zhang C. 2005. Madden-julian oscillation. Reviews of Geophysics. **43**.

816 Zhang C. 2013. Madden–Julian oscillation: Bridging weather and climate. Bulletin of
817 the American Meteorological Society. **94**: 1849-1870.

818 Zhu H, Wheeler MC, Sobel AH and Hudson D. 2014. Seamless precipitation prediction
819 skill in the tropics and extratropics from a global model. Monthly Weather Review, **142**:
820 1556-1569.

821

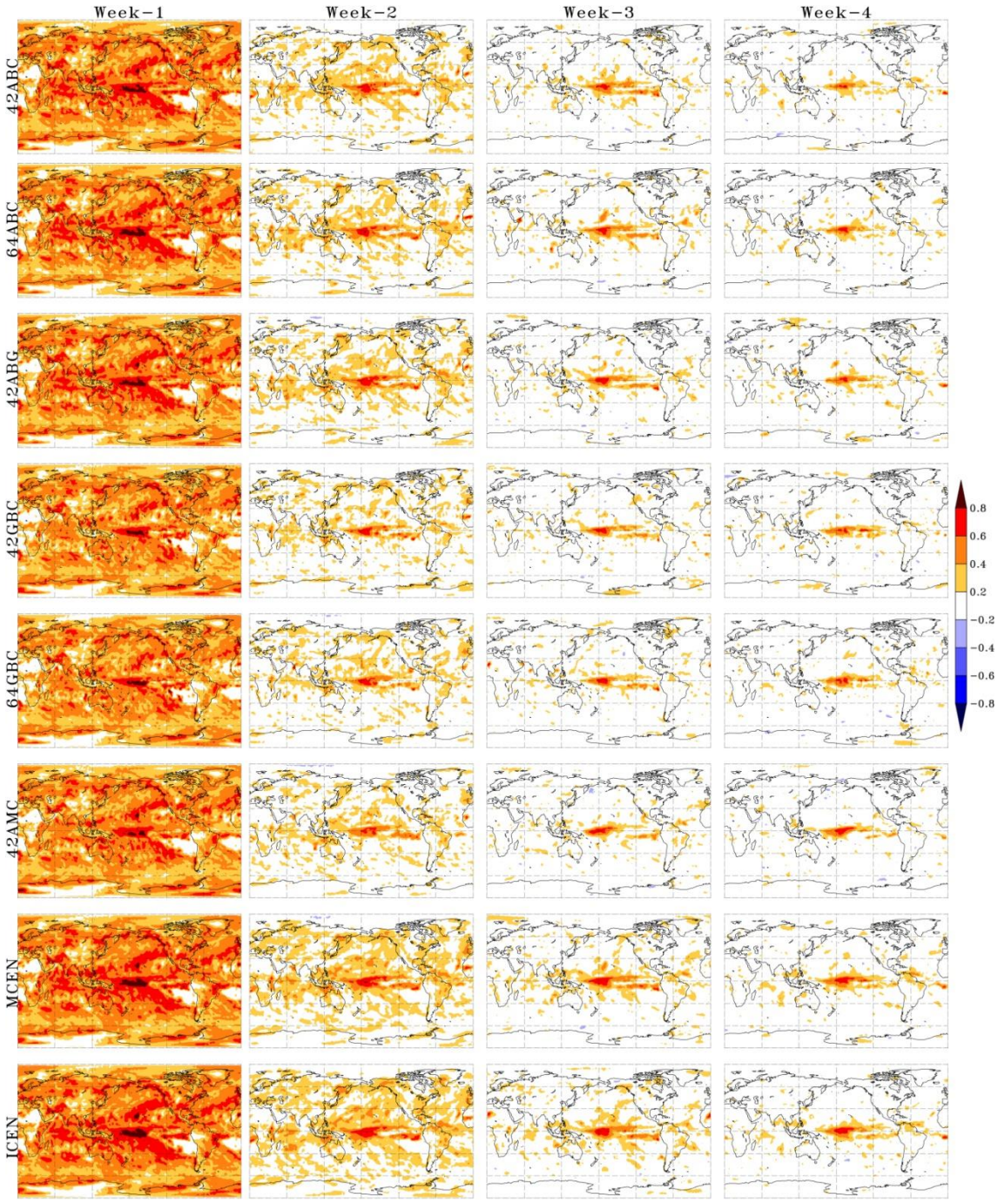


Figure 1: Correlation between the predicted and observed (GPCP) precipitation anomalies for the six BAM-1.2 configurations (42ABC, 64ABC, 42ABG, 42GBC, 64GBC and 42AMC) and two ensemble means (MCEN and ICEN) (rows) for week-1, week-2, week-3 and week-4 (columns). The hindcasts were initialized within the

extended austral summer period (from November to March 1999/2000 – 2010/2011) on the dates shown in Table 1.

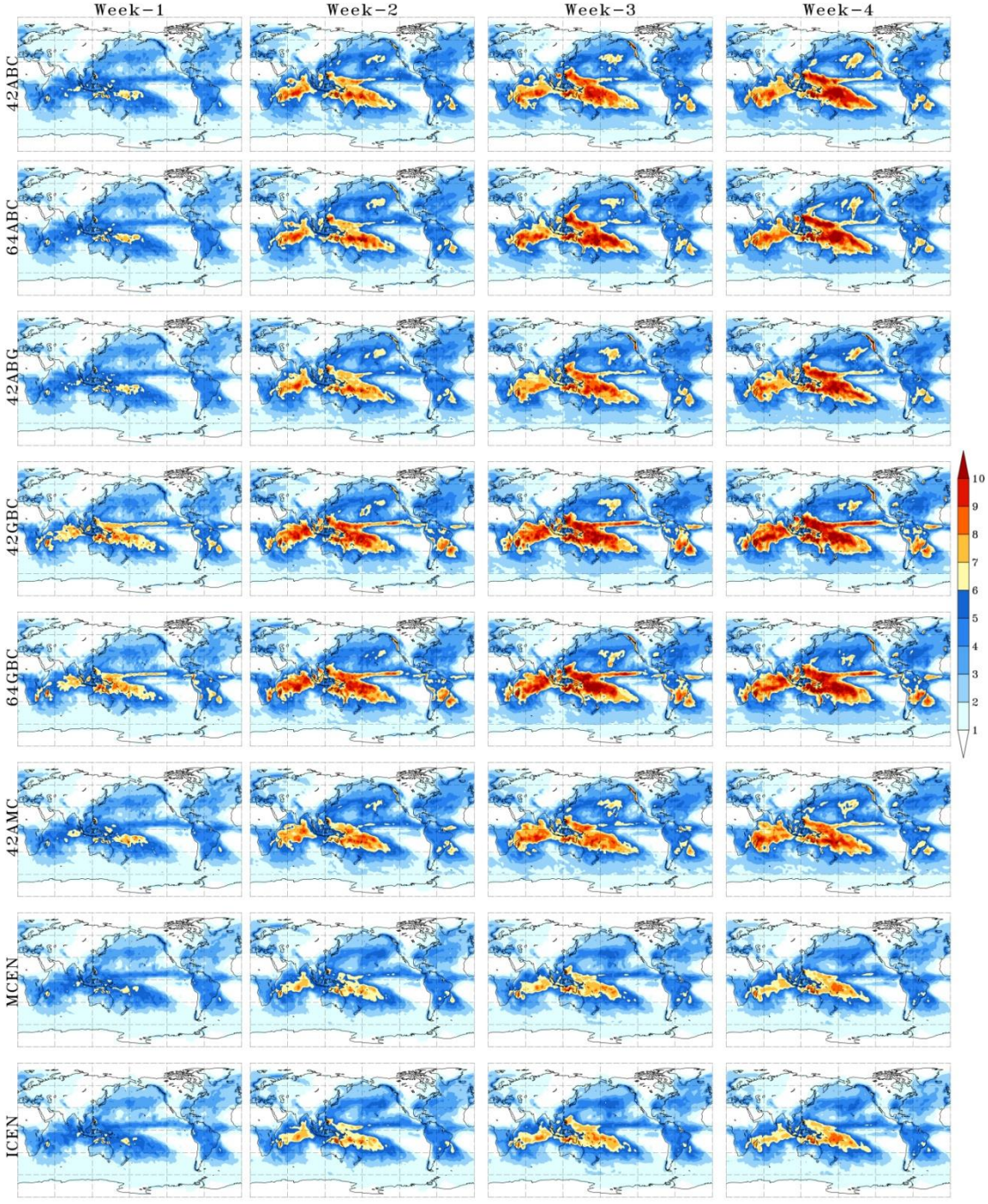


Figure 2: Six configurations (42ABC, 64ABC, 42ABG, 42GBC, 64GBC and 42AMC) and two ensemble means (MCEN and ICEN) (rows) RMSE precipitation anomaly (units are mm day-1) for week-1, week-2, week-3 and week-4 (columns). The hindcasts

were initialized within the extended austral summer period (from November to March 1999/2000 – 2010/2011) on the dates shown in Table 1.

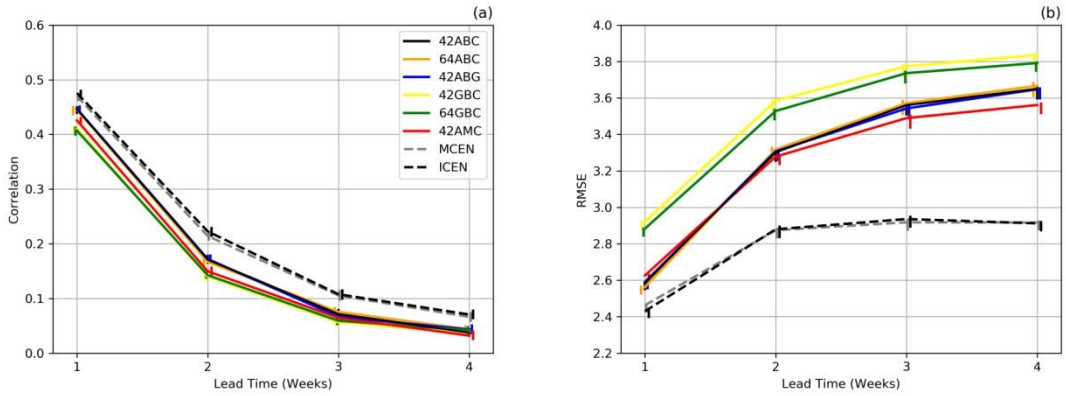


Figure 1: Global mean correlation between predicted and observed precipitation anomalies (a) and RMSE (b) for six BAM-1.2 configurations (42ABC, 64ABC, 42ABG, 42GBC, 64GBC and 42AMC) and two ensemble approaches (MCEN and ICEN) assessed against GPCP averaged over the latitudinal band 60°N-60°S for four lead times (weeks 1 to 4). The hindcasts were initialized within the extended austral summer period (from November to March 1999/2000 – 2010/2011) on the dates shown in Table 1. The vertical bars plotted around the four lead times represent 95% confidence intervals produced using a bootstrap resampling procedure with replacement with 1000 samples. These vertical bars are slightly displaced from the exact lead time location in the horizontal axis to facilitate visualization.

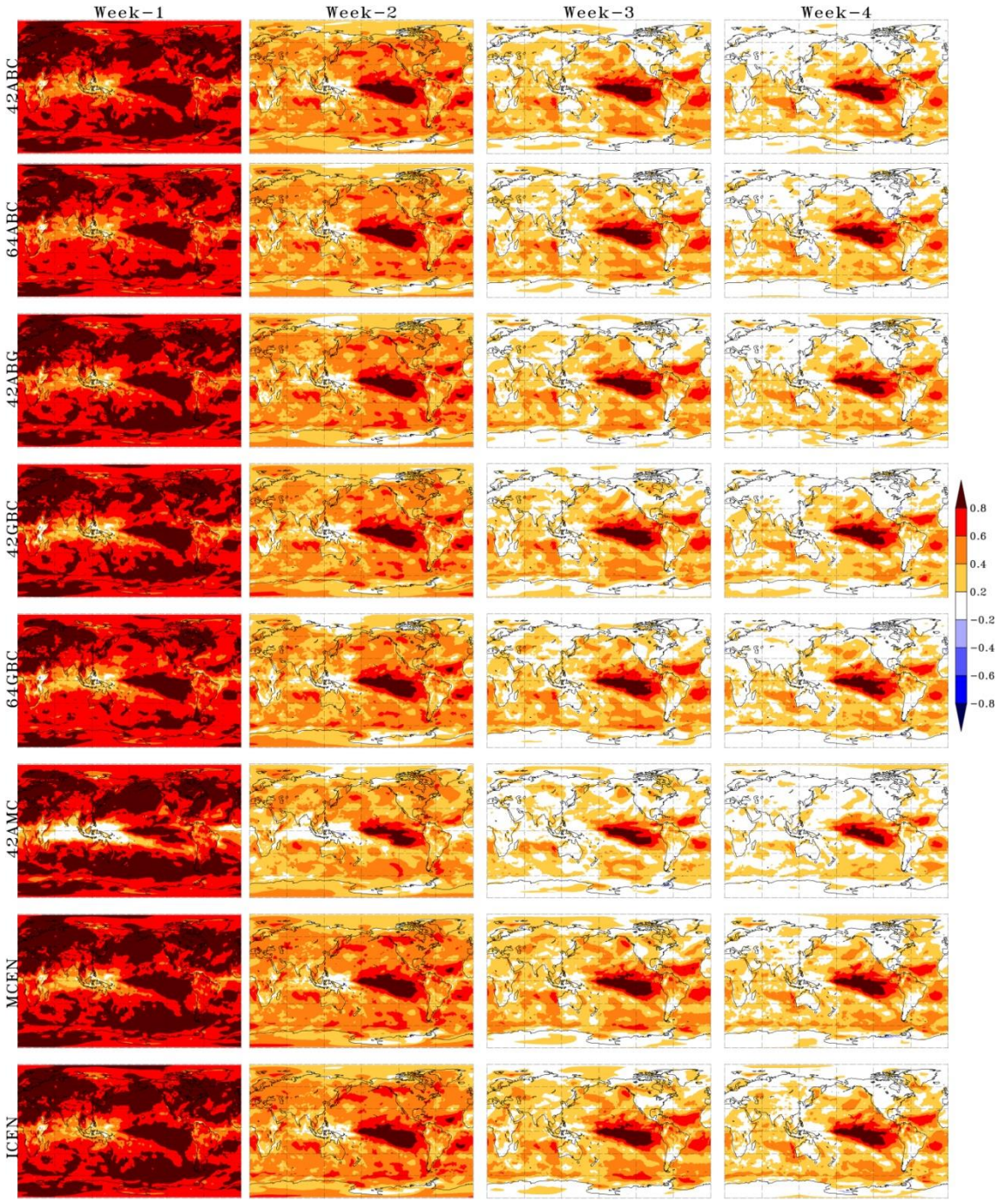


Figure 4: Same as Figure 1, except for T2M anomaly.

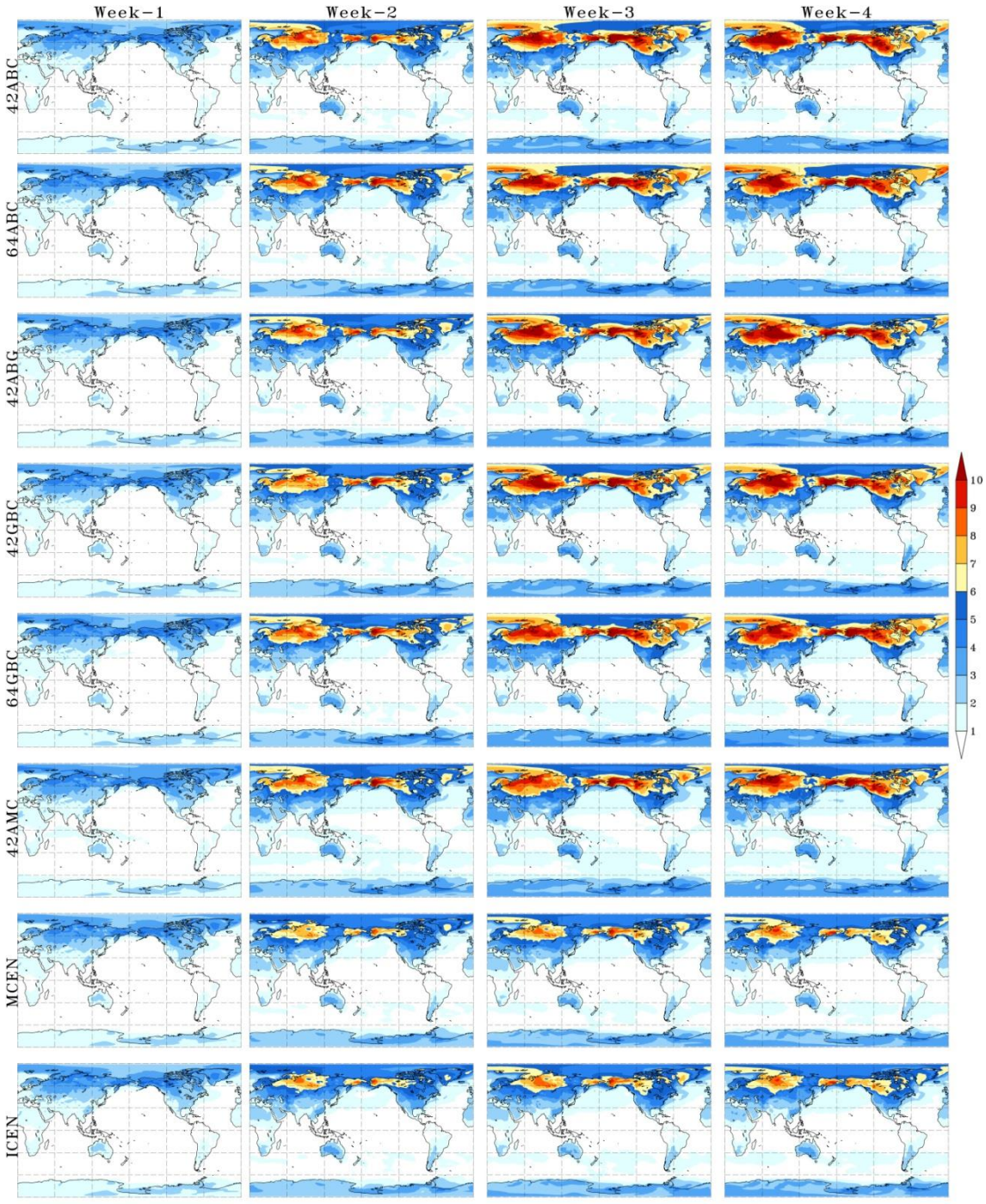


Figure 5: Same as Figure 2, except for T2M anomaly (units are °C).

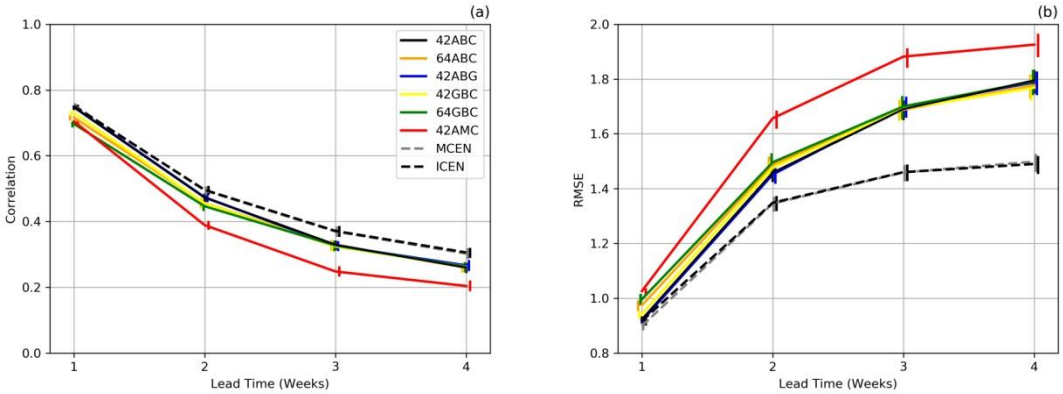


Figure 6: Same as Figure 3, except for T2M anomaly.

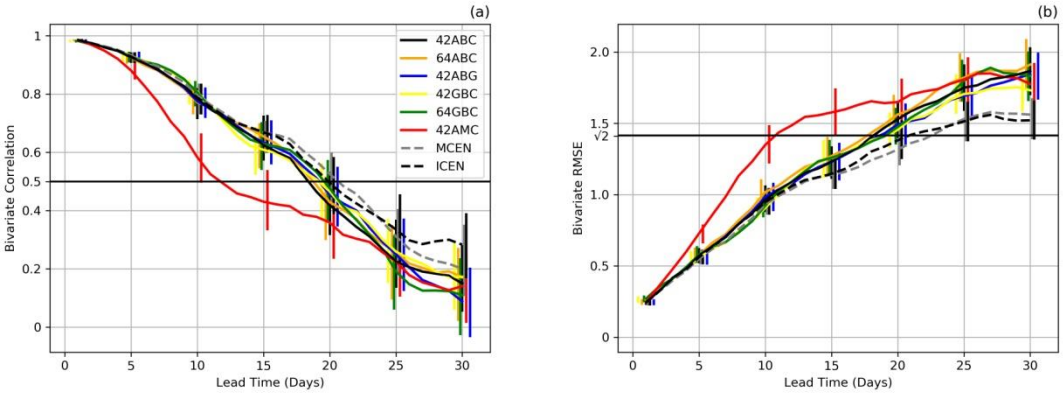


Figure 2: Bivariate correlation (a) and bivariate RMSE (b) for six BAM-1.2 configurations (42ABC, 64ABC, 42ABG, 42GBC, 64GBC and 42AMC) and the two ensemble approaches (MCEN and ICEN) as a function of forecast lead time (in days). The hindcasts were initialized within the extended austral summer period (from November to March 1999/2000 – 2010/2011) on the dates shown in Table 1. The vertical bars around lead times 1 to 30 days plotted every 5 days represent 95% confidence intervals produced using a bootstrap resampling procedure with replacement with 1000 samples. These vertical bars are slightly displaced from the exact lead time location in the horizontal axis to facilitate visualization. Note that two black vertical bars are plotted every 5 days, with the first of these bars corresponding to 42ABC and the second corresponding to ICEN.