Final Draft
of the original manuscript


First published online by Elsevier: 20.05.2020

https://dx.doi.org/10.1016/j.atmosres.2020.105053
PERFORMANCE EVALUATION OF ETA/HADGEM2-ES AND ETA/MIROC5 PRECIPITATION SIMULATIONS OVER BRAZIL

André Almagro¹*, Paulo Tarso S. Oliveira¹, Rafael Rosolem², Stefan Hagemann³ & Carlos A. Nobre⁴.

¹Federal University of Mato Grosso do Sul, CxP 549, Campo Grande, MS, 79070-900, Brazil.

²University of Bristol, Faculty of Engineering, University Walk, Clifton BS8 1TR, United Kingdom.

³Helmholtz-Zentrum Geesthacht, Institute of Coastal Research, Max-Planck Str. 1, 21502 Geesthacht, Germany.

⁴University of São Paulo, Institute of Advanced Studies, São Paulo, SP, 05508-970, Brazil.

E-mail adress: andre.almagro@gmail.com (A. Almagro)

paulotarsoms@gmail.com (P. T. S. Oliveira)

rafael.rosolem@bristol.ac.uk (R. Rosolem)

stefan.hagemann@hzg.de (S. Hagemann)

cnobre.res@gmail.com (C. A. Nobre)

*corresponding author: André Almagro. PhD candidate in Water Resources at Federal University of Mato Grosso do Sul, Campo Grande, MS, Brazil. Phone: +55 67 981696219.
Abstract

Climate change effects can have significant impacts worldwide. Extreme events can modify water availability and agricultural production, making climate change planning an essential task. The National Institute for Space Research (INPE in Portuguese) in Brazil has made a large dataset of regional climate model outputs (simulations and projections) available, which opens up many possibilities of carrying out high-resolution climate change studies. However, there is still no performance evaluation of the model-derived rainfall output against high-resolution ground-based observation data considering the Brazilian biomes. This paper attempts to fill this gap and evaluates the simulated precipitation throughout Brazil. We used gridded observed precipitation data and historical climate simulations from the Model for Interdisciplinary Research on Climate, version 5 (MIROC5) and from the Hadley Center Global Environment Model, version 2 (HadGEM2-ES), which were downscaled by the Eta RCM (Regional Climate Model). For the overlapping period (1980-2005), there is good agreement (PBIAS up to 10%) of downscaled annual simulations for the Amazon and Cerrado biomes and large biases (reaching 40%) in the Pampa biome, compared to the observations. Our results showed that HadGEM2-ES is capable of representing long-term mean monthly precipitation for large areas well, such as the Amazon and Cerrado. Furthermore, the Eta RCM has considerably improved the driving GCM MIROC5 simulations. In conclusion, we recommend using the HadGEM2-ES simulations for the Amazon, Eta/HadGEM2-ES for the Atlantic Forest, Cerrado, and Pampa, and Eta/MIROC5 for the Caatinga and Pantanal. Our study provides an overview of two downscaled simulation datasets in Brazil that may help verify the models’ suitability for further climate change assessments.

Keywords: Climate change, general circulation model, rainfall, regional climate model.
1. INTRODUCTION

Effects of climate change (e.g., warming of the atmosphere, extreme weather events contributing to a lack or excess of water) have significant socio-economic and environmental impacts worldwide (Hansen and Cramer, 2015). In Brazil, more severe and frequent hydrometeorological events (e.g., droughts, floods, landslides) are expected. Such extreme events can significantly alter water availability and agricultural production (PBMC, 2013). To ensure appropriate climate change planning and mitigation policies, accurate projections of climate changes are an essential but yet challenging task for the scientific research community due to the large number of climate-sensitivity factors that need to be considered (McNutt, 2013).

Future climate projections are usually produced by Global Climate Models (GCMs) or Earth System Models (ESM), which resolve the physics and dynamics of the Earth System as a whole (IPCC, 2013). GCMs/ESMs are the most advanced scientific tools for simulating responses of global climate regarding to variations of greenhouse gas concentration and to support climate change studies (Mello et al., 2015). The community-wide use of GCMs/ESMs is widely recognized in the fifth phase of the Coupled Model Intercomparison Project (CMIP5). The CMIP5 initiative comprises new sets of climate model experiments, which are coordinated by the World Climate Research Programme and include more than 50 complex GCMs/ESMs (Knutti and Sedláček, 2012; Taylor et al., 2012). Climate data are produced under different Representative Concentration Pathway scenarios (RCP 8.5, 6, 4.5, 2.6 Wm⁻²) of the 5th Assessment Report (AR5) from the IPCC (IPCC, 2014). These pre-determined climatic conditions have been widely used to understand the climate and project climate changes on Earth (Zhao et al., 2013).

To provide a large volume of climatic data at global coverage, GCM historical simulations and future projections are usually produced at relatively coarser resolution
(grid sizes in the order of 100 km-200 km) to assess climate change impacts. To evaluate the potential impacts of climate change on regional scales at finer spatial resolution (grid sizes in the order of 20 km), simulations and projections from GCMs are usually downscaled by employing Regional Climate Models (RCMs) (Giorgi, 1990; Maraun et al., 2017). The National Institute for Space Research (INPE) developed four sets of downscaled products based on the Eta RCM for Brazil, parts of South America and adjacent oceans, forced with both RCP 8.5 and RCP 4.5 scenarios obtained from the AR5 taken from global simulations and projections from two GCMs/ESMs, namely HadGEM2-ES and MIROC5 GCMs, respectively (Chou et al., 2014a). Choosing GCMs/ESMs was based on a satisfactory performance in resolving precipitation and atmospheric circulation over South America, and also importantly due to ease accessibility of the data on public domain (Brazil, 2016; Flato et al., 2013). These downscaling simulations were performed in support of strategic climate change studies and the Brazilian Third National Communication to the United Nations Framework Convention on Climate Change (Brasil, 2016). Before the CMIP6 new data sets becomes fully available, those data based on AR5 scenarios are the latest and most advanced products in terms of spatial resolution available for climate change studies in South America. As a result, these regional-scale scenarios have been adopted for the Brazilian National Adaptation Plan for Climate Change.

For over 30 years, the RCMs have been satisfying the need of high-spatial resolution climatologies for climate change impacts assessments, overcoming the inability of the GCMs to deal with it. During this time, we had phases of development, maturing, and exploration of contradictions and limitations (Tapiador et al., 2019). Due the large publicly availability, the comparison between simulations of the RCMs against their driving-GCM turned into a common and necessary procedure to assess and evaluate
the added values of the dynamical downscaling employed. For the South America, some of the state-of-the-art of added value from RCMs simulations were performed using CORDEX experiments from the CMIP5 datasets. Llopart et al. (2019) assessed the added value of a pair of multi-model ensembles of RCMs and GCMs. They found, in terms of precipitation, an added value of the RCMs in the coastal portion of the South Atlantic Convergence Zone (SACZ) and the Intertropical Convergence Zone (ITCZ), showing the clear improvement of the finer resolution in resolving convective schemes, such as the convergence zones. In their study, Falco et al. (2018) found that all the multi-model ensembles analyzed could reproduce the main features of seasonal climatology, while the individual analyses presented more biases, showing the needed for individual assessing of models. Moreover, they concluded that added value of RCMs simulations over historical periods were not general, being found only in certain combinations of model-region. The most noticeable degrading of simulations was found on winter climatology. Solman & Blázquez (2019) evaluated the ability of RCMs and their corresponding driving-GCM in reproducing the precipitation spatial distribution and behavior in a multi-temporal scale over South America. According to their results, spatial patterns and seasonal means are close related with large-scale atmospheric circulation, such as SACZ and ITCZ, and for some regions large biases (underestimating or overestimating) in rainfall amount were found for a group of RCMs. On common point was found for all above mentioned studies: this kind of analysis is the basis for further work using climatic projections. The ability of GCM/RCM in simulate historical spatial distribution and patterns of climatologies tell us a lot about their projections for the future.

As mentioned above, to increase the degree of confidence in these model projections, their simulations need to be evaluated compared to historical observations. However, a proper investigation of Eta/HadGEM-ES and MIROC5 has not yet been fully
addressed. A study conducted by Chou et al. (2014b) is the only one that has previously evaluated these RCM products (Eta/HadGEM2-ES and Eta/MIROC5). Their analyses use long-term monthly and seasonal mean fields of temperature and precipitation from 1961 to 1990 against a relatively coarse resolution CRU TS 3.1 global gridded dataset (Mitchell and Jones, 2005) focusing mainly on summer and winter seasons. However, the results were obtained by averaging the simulations over oversized areas with heterogeneous hydroclimatic characteristics and multiple precipitation regimes, limiting the representation of specific local climatic characteristics. When averaging precipitation over oversized areas or regions, we can mask some regional characteristics and local features, and the use of large areas to analyze precipitation seems to be viable only when using spatial distribution. As we can see in supplementary Figure S1, the averaged observed precipitation (1980-2005) over North and Central-West of Brazil (administrative regions) does not reproduce the precipitation regime of any biome within this area (Amazon, Cerrado, and Pantanal). The use of the smaller and more coherent areas for averaging precipitation was done before. For the Europe territory, Christensen & Christensen (2007) and Dosio & Paruolo (2011) used eight sub-areas taking in account topography and climate features. Gregory et al. (1991) adopted nine spatially coherent precipitation regions for analyzing area-averaged statistics of precipitation in Great Britain. In a global scale, some of world biomes were adopted by Huxman et al. (2004) for an average rain-use efficiency in aboveground net primary production using global precipitation data. Thus, we claim that the most appropriate is to use hydrometeorological divisions, such as biomes or hydrographic regions. Therefore, to improve the RCM evaluation across Brazil, a more suitable division should consider areas with similar ecoclimatic dynamics and characteristics. In this context, the Brazilian biomes appear as a viable alternative. These biomes were defined by the Ministry of Environment in Brazil
and are large ecosystems with relatively similar and uniform climate, vegetation and biodiversity, relative to the overall extent of the country (Brown and Maurer, 1989; Coutinho, 2016).

The objective of this study is to evaluate the Eta/HadGEM2-ES and Eta/MIROC5 performance to represent long-term monthly and seasonal mean precipitation over key Brazilian biomes. We compared the simulated precipitation of the driving GCMs (HadGEM2-ES and MIROC5) and their respective downscaled datasets (Eta/HadGEM2-ES and Eta/MIROC5) against a high-resolution gridded dataset derived from observational networks in Brazil. In addition, we investigated the possible origin of biases in the simulated precipitation.

2. MATERIAL AND METHODS

2.1. Brazil: Study area and biomes

Given its large spatial extent (8,511,000 km²) including a large range of elevation (sea level to 2900 m altitude) and heterogeneous patterns of precipitation seasonality, Brazil’s diverse vegetation types are classified into six main biomes (see Figure 1). We followed previous definitions (Li et al., 2006; Murray-Tortarolo et al., 2017; Myneni et al., 2007) that consider 100 mm per month as a threshold for defining a month within the dry season to verify if the simulations are capable of estimating the onset, duration, and termination of both rainy and dry seasons. This value is a global precipitation threshold that considers the amount to begin runoff (Zhang et al., 2004), to maintain the vegetation growth (Li et al., 2006; Murray-Tortarolo et al., 2017), and sensitivity analysis (Murray-Tortarolo et al., 2017).
Figure 1. Brazilian biomes map and their respective mean monthly precipitation shown as bar plots. Monthly means were calculated for the 1980-2013 period from daily dataset originally developed by Xavier et al. (2016). The red lines represent a threshold used as a criterion to identify the dry months (< 100 mm).

The Amazon is the largest tropical biome in the world, consisting of a densely vegetated rainforest with the highest annual mean precipitation (average annual precipitation of 2.3 m and greater than 4 m/year in some portions of western Amazon) and a short dry season (Sombroek, 2001). The Cerrado biome mainly consists of woodlands and savanna and is crucial for water and food supplies, for maintaining ecological services and for economic activities in Brazil. Moreover, it is considered as one of the most important biomes in Brazil related to food-energy-water security (Oliveira et al., 2014). The Pantanal is one of the largest flooded areas in the world and it is the most intact biome in Brazil (Junk et al., 2006). It has very defined dry and rainy seasons, and the flood cycles – caused not by an excess of precipitation but due to drainage deficiency – are the most important ecological phenomenon (Ribas and Schoereder, 2007). The Caatinga biome is characterized by a semi-arid region in the Northeast of Brazil. It comprises mostly secondary vegetation (herbaceous and arboreous) and presents the lowest values of annual precipitation with a severe dry season – about 70%
of annual precipitation occurs in February-April period (Menezes et al., 2012; Pinheiro et al., 2013). The Atlantic Forest is characterized by rainforest cover in the coastal area and the semi-deciduous forest in the continental area with very defined wet and dry seasons (Morellato and Haddad, 2000). Finally, the Pampa biome is located in the South of Brazil and has no defined dry season. Natural grasslands are predominant, with tree formations and sparse shrub, and it is referred to as “Campos” (Lupatini et al., 2013; Roesch et al., 2009).

2.2. Data acquisition and processing

We used rainfall data from the Eta Regional Climate Model, available from INPE (http://projeta.cptec.inpe.br/). These products were generated using climate forcing data derived from the r1i1p1 ensemble member of two GCMs/ESMs: the British HadGEM2-ES and the Japanese MIROC5 (both original GCM/ESM products available at https://esgf-data.dkrz.de/search/esgf-dkrz/). The MIROC5 is an atmosphere-ocean general circulation model (AOGCM), with a 1.4°x1.4° spatial resolution for the atmospheric parcel, that brought several improvements on Intertropical Convergence Zone (ITCZ) and El Niño Southern Oscillation (ENSO) simulations (Watanabe et al., 2010). The HadGEM2-ES is also a coupled AOGCM, with atmospheric resolution of 1.875°x1.25°, that brings some components generated interactively by the model, instead of being assigned as boundary conditions (Jones et al., 2011).

The baseline period is defined from 1961 through 2005. The INPE dataset was produced at approximately 20 km spatial grid resolution for South America and with two different Representative Concentration Pathways (RCP4.5 and RCP8.5), respectively. The Eta RCM downscaling procedure is described in more detail in Chou et al. (2014a, 2014b).
To evaluate the simulated monthly precipitation from both datasets, we compared model-derived precipitation against a gridded-interpolated product derived from observations (0.25° by 0.25° spatial resolution), developed by Xavier et al. (2016), and available at http://careyking.com/data-downloads/). This product used observed precipitation data derived from approximately 4,000 rain gauges from the Brazilian Water Agency (ANA), the National Institute of Meteorology (INMET), and the Water and Electric Energy Department of São Paulo state (DAEE/SP) from 1980-2013. Furthermore, this reference dataset has been extensively applied in many fields of study, such as evaluation of remote sensing products (Melo et al., 2015; Paredes-trejo et al., 2018, 2017; Paredes-trejo and Barbosa, 2017), vegetation response to rainfall variability (Souza et al., 2016), impacts of climatic extremes (Melo et al., 2016), and climate change assessments (Almagro et al., 2017).

We applied an interpolation method on the global climate model outputs and on the reference gridded observations to a common spatial resolution of the Eta RCM spatial resolution, which is the focus of this study. We applied a first-order conservative remapping method (Jones, 1999) using the Climate Data Operators (CDO) to preserve the main characteristics of each dataset and to ensure that any area average – area of a pixel or an area of a biome – would be similar to the original dataset, allowing more detailed comparisons when considering the different resolutions and introducing the less possible kind of error associated to the remapping. As we show in the supplementary material (Figure S2), the highest mean error added to the dataset due the first-order conservative remap is about 0.01 mm, a value lower than the uncertainty associated to the most of common ground rain gauges observations and to our reference dataset (Villarini et al., 2008; Xavier et al., 2016). The same procedure was widely done in previous studies (Diaconescu et al., 2015; Nadeem et al., 2019; Wu et al., 2020).
2.3. Metrics to evaluate the simulated precipitation

To evaluate the performance and quality of simulated precipitation against the observed data product, we used the following statistical metrics: Percentage Bias (PBIAS), Root Mean Squared Error (RMSE), Correlation Coefficient (CC) and Coefficient of Variation (CV) (Equations 1, 2, 3 and 4).

\[ BIAS = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{P_{sim,i} - P_{obs,i}}{P_{obs,i}} \right) \]

\[ CC = \frac{n \left( \sum_{i=1}^{n} P_{sim,i} P_{obs,i} \right) - \left( \sum_{i=1}^{n} P_{sim,i} \right) \left( \sum_{i=1}^{n} P_{obs,i} \right)}{\sqrt{\left[ n \sum_{i=1}^{n} P_{sim,i}^2 - \left( \sum_{i=1}^{n} P_{sim,i} \right)^2 \right] \left[ n \sum_{i=1}^{n} P_{obs,i}^2 - \left( \sum_{i=1}^{n} P_{obs,i} \right)^2 \right]}} \]

\[ RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (P_{obs,i} - P_{sim,i})^2} \]

\[ CV = \frac{\sigma}{P} \]

where \( P \) is the long-term mean precipitation from observations “obs” and simulations “sim”; \( \sigma \) is the standard deviation of the annual precipitation; and \( n \) is the number of points for each biome.

We also ranked the performance of GCMs and RCMs to simulate some proprieties of precipitation, such as the rainy and dry periods, seasonal and annual precipitation for each biome. To do so, we followed the methodology proposed by Deidda et al. (2013), which is commonly applied on GCM-RCM comparison (see Mascaro et al., 2018). This methodology calculates a single dimensionless error metric, the \( \epsilon_j \), for each precipitation property combining multiple variables that characterizes that property (Table 1). Then, models are ranked by the value of \( \epsilon_j \) (lower rank means better performance), calculated by Equation 5.

<table>
<thead>
<tr>
<th>Property</th>
<th>Variable, ( k )</th>
<th>Error, ( E_{k,j} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Properties and the variables (k) considered to calculate the errors (\( E_{k,j} \)) for each property.
| Dry-season                      | Number of dry months, DM | $|DM_{obs} - DM_j|$ |
|-------------------------------|--------------------------|----------------|
| Seasonal cycle                | Seasonal root mean square error, RMSE | $|RMSE_{obs} - RMSE_j|$ |
|                               | Seasonal correlation coefficient, CC | $1 - CC_j$ |
| Annual cycle                  | Mean annual P, P          | $|\bar{P}_{obs} - \bar{P}_j|$ |
|                               | Coefficient of variation of annual P, CV | $|CV_{obs} - CV_j|$ |

where $\varepsilon_j$ is the dimensionless error for the $j$ ($j=1,\ldots,N$) simulation; $E_{k,j}$ is the error between observed and simulated values of the variable $k$ ($k=1,\ldots,S$), and dividing it by the sum of the errors of all models, we obtain a dimensionless contribution for the error of variable $k$. Then, summing and taking the square root of the error parcel for all variables $S$, we reach the $\varepsilon_j$ for the rank.

### 2.4. Regional and spatial analysis

Using the baseline period (1980-2005) from simulations and observations, we calculated the long-term precipitation averages at monthly, seasonal (December, January and February – DJF; March, April and May – MAM; June, July and August – JJA; and September, October and November – SON), and annual scales for each grid point for all datasets across Brazilian biomes (Figure 1). Then, we performed two distinct analyses with regards to regional averages and spatial distributions and patterns using the metrics described in Subsection 2.3.

We computed the PBIAS, CC and CV separately for each biome, considering grid points within the biome to represent biome-average quantity (regional analysis) and the PBIAS and CV on a grid point scale for the whole of Brazil (spatial analysis). The
first analysis identifies a general behavior for each biome while the second one enables an investigation of the biases at the same time, as well as spatial patterns and their possible causes. To assess the reliability of the models to represent the wet and dry months, we computed the biome-specific long-term mean for each month of the year.

3. RESULTS AND DISCUSSION

3.1. Spatial patterns on annual and seasonal precipitation

In this section, we present the results of spatial distribution of the biases among observations (Figure 2a) and models’ simulations/projections (Figure 2b to 2e) and observations for the entire Brazilian territory. The analysis identifies the spatial patterns of precipitation, and consequently the biases (Figure 3) in relation to the observed annual means.
Figure 2. Spatial distribution of the annual precipitation (P) over the Brazilian biomes for a) Observed dataset; b) HadGEM2-ES dataset; c) Eta/HadGEM2-ES dataset; d) MIROC5 dataset; and e) Eta/MIROC5 dataset.
The HadGEM2-ES presented an overall positive bias throughout Brazil with negative biases observed in some areas in the northern region. The downscaling process barely corrected the biases on annual precipitation simulations, but a negative bias was generated in the Caatinga biome portion, where the lowest values of annual precipitation occur and a small change in rainfall corresponds to a relatively large percentual change. At the same time, in the coastal area of the Northeast region, in the Atlantic Forest biome, a negative bias was spread to the Eta simulations. Even with the 20 km resolution, the Eta RCM is not capable of capturing the rainfall system neither improve the representation of the phenomenon. This is also true with respect to sea-breeze induced rainfall along the Amazonian coastal zone. Due to the coarse resolution, we cannot expect that the GCMs/ESMs capture sea-breeze induced rainfall. However, we expect that Eta RCM would at least improve the simulations, but it did not.

For the MIROC5 simulations, extremely high values of PBIAS (which reached up to 200%) were found in the midwestern and northeastern areas of Brazil. However, the Eta/MIROC5 downscaled simulations reduced the biases while remaining just few grid points with positive values in the coast of southeastern region. These results show the inability of MIROC5 to simulate mean annual precipitation over a large area of Brazil. At the same time, the results demonstrate the great improvement of Eta/MIROC5 in relation to its original coarse-scale GCM product.
Figure 3. Spatial distribution of the PBIAS on annual precipitation (P) and on the annual coefficient of variation (CV) over the Brazilian biomes. a) and e) represents the PBIAS and CV on HadGEM2-ES simulations; b) and f) represents the PBIAS and CV on Eta/HadGEM2-ES simulations; c) and g) represents the PBIAS and CV on MIROC5 simulations; and d) and h) represents the PBIAS and CV on Eta/MIROC5 simulations.
Figure 4 shows the bias on the long-term mean precipitation for each season (DJF, MAM, JJA and SON) calculated for all evaluated model products against observations (see also supplementary Figure S2). As we can note in the DJF and MAM seasons, there is an overestimation of all simulations for the extreme north of the Amazon, which can be related to the inability of the models to capture the sea-breeze influence on the rainfall systems in this part of Brazil. The Amazonian coastal regions of Amapá and Pará states are strongly influenced by sea-breeze and are affected by it up to 300 days a year. The negative bias observed in this region could be related to the wrong simulation of this phenomenon, which can be expected by a GCM (due to its coarse resolution). However, it was expected that this local feature would be captured and reproduced by the Eta model. For other locations and biomes, the simulations significantly improved the use of Eta RCM, especially for the MIROC5, showing the good suitability of Eta RCM downscaling process on these seasons over Brazil. During the JJA season, high-pressure systems (< 1,013 hPa), or anticyclonic, dominate the subtropical region (see Figure S5), making the formation of clouds more difficult and blocking the occurrence of rainfall. The amount of precipitation in much of Brazil is very low, except for the northern part of the Amazon and southern part of the Pampa. In general, both GCMs represented these features of the dry season well and we noted more improvement of the Eta RCM in the MIROC5 data. Despite this, due to the very low rainfall amounts observed in the JJA months, any minimal over/underestimation generates an expressive relative bias, as we can see in Figure S6. For the SON season, the simulations presented low biases in absolute terms and higher biases in relative terms, in the same way as JJA. The downscaling process did not improve the simulations of HadGEM2-ES, maintaining the spatial behavior of the biases. For the MIROC5 simulations, there was a great improvement in
the Amazon biome and a change in the signal of the biases in the Caatinga biome. Once again, this is due to low precipitation totals in this biome.

Figure 4. Absolute biases (BIAS) in each season (DJF, MAM, JJA and SON) simulated precipitation in Brazilian biomes. a) to d) represent the BIAS for the HadGEM2-ES for all seasons; e) to h) represent the BIAS for the Eta/HadGEM2-ES for all seasons; i) to l) represent the BIAS for the MIROC5 for all seasons; and m) to p) represent the BIAS for the Eta/MIROC5 for all seasons. Shades of blue indicate a positive BIAS while shades of red indicate a negative BIAS.

3.2. Long-term means and annual variability at the biomes

In this section, we present and discuss the results of mean monthly and annual precipitation, and the annual variability during the 1980-2005 period for each Brazilian
biome. The long-term mean monthly precipitation analysis identifies how well the models can simulate precipitation patterns and the dry/rainy seasons. Figure 5 shows the comparison of the downscaled and driver-family models’ simulations against the observations. We used an error bar (standard deviation) of observations to create a range of acceptable values for the simulations. Eta/HadGEM2-ES generally underestimates the mean monthly rainfall in the rainy season and represents it well in the dry season. On the other hand, HadGEM2-ES overestimates the means for rainy and underestimates means for dry seasons, but most of these under/overestimations are inside the range of acceptable values. For larger areas such as the Atlantic Forest, the Amazon and Cerrado, the GCM means are closer to observations than RCM simulations in the rainy season (DJF to MAM). In the Amazon, Cerrado and Pantanal biomes, Eta/HadGEM2-ES has more negative precipitation biases during the rainy season, which reach almost -50%. Considering the error bar of observations, the Eta RCM simulated the rainy and dry seasons over the biomes well, except for the Pantanal (one month longer) and Pampa (no defined dry season). Throughout the Amazon, we verified very large error bars due its large area and spatial variability of rainfall regimes. Taking this into account, downscaled models were capable of simulating the short dry season (three months), showing close values to the observations and a significant drop from the rainy season rain. In the Pampa biome, where there is no defined dry season, the HadGEM2-ES model was capable of capturing this characteristic but, at the same time, produced considerable errors in the DJF and SON months. These large errors found in the Pampa are related to the coarse resolution of the GCMs, which makes Pampa’s area relatively small for 100-200 km simulations. In the Caatinga, most of the GCM and RCM simulations were considered acceptable, except for HadGEM2-ES in February and Eta/HadGEM2-ES in January.
Moreover, the dry season and distribution of rainfall over the year were well simulated by all models.

The mean monthly rainfall simulated by Eta/MIROC5 is overestimated in the rainy season and underestimated in the dry season in the Atlantic Forest and Caatinga biomes, while in the Cerrado and Pantanal, the opposite pattern is observed. In relation to the original MIROC5 product, the downscaled Eta/MIROC5 version mostly improved the values, thereby showing a clear added value to the downscaling process. The Amazon means are underestimated in all months of the year, and there is a noticeable improvement of the values simulated by Eta/HadGEM2-ES in the DJF and JJA months. The poorest simulation was performed for the Pampa biome, where the model cannot represent the absence of a dry season. Moreover, simulations for this biome has the largest difference to the observations with just two months of simulations inside the error bar of the observations. In general, Eta/MIROC5 can capture the rainy and dry season except for the Pampa biome. Finally, the Eta/MIROC5 is generally drier than the driving GCM in the wet season and wetter in the dry season.
Figure 5. Long-term mean monthly rainfall for the 1980-2005 period for observations (black dashed line), HadGEM2-ES (dark blue line), Eta/HadGEM2-ES (light blue line), MIROC5 (dark red line), and Eta/MIROC5 (light red line) simulations in the a) Amazon, b) Atlantic Forest, c) Cerrado, d) Caatinga, e) Pampa, and f) Pantanal. The red reference line represents a threshold used as a criterion to identify the dry season (< 100 mm).

In terms of seasonality, for most biomes, the Eta/HadGEM2-ES is drier than HadGEM2-ES in the wet season, but the wetter behavior in the dry season is less obvious than the Eta/MIROC5 one. In general, the downscaling process applied by the Eta RCM improved the long-term mean monthly values, but those on the MIROC5 were more notable than for HadGEM2-ES. The simulations of MIROC5 were originally not as good as the one of the simulations of HadGEM2-ES and this led to a more notable improvement of the downscaling process for the first model. At the same time, our analysis of the
annual cycle clearly showed that the downscaled simulations are more suitable for the biomes where there are large amounts and well defined rainy and dry seasons.

For the mean annual rainfall during the period 1980-2005, Eta/HadGEM2-ES performed well for the Atlantic Forest (+2%), Caatinga (-6%) and Pampa (-3%), while Eta/MIROC5 shows lower biases over the Amazon (-7%), Cerrado (+6%) and Pantanal (-14%) (Figure 6a). We found a poor performance of Eta/HadGEM2-ES for the mean annual rainfall in the Cerrado with negative biases up to -26%, and Eta/MIROC5 in the Pampa biome with underestimates up to -39%. No model was capable of capturing the annual variability (Figure 6b) for the Pantanal and just the HadGEM2-ES captured it for the Caatinga biome. These two biomes presented the highest values of observed CV once they presented the lowest observed amounts of annual precipitation.

Figure 6. Annual characteristics of precipitation over the Brazilian biomes for the 1980-2005 period. The mean annual precipitation (a) is presented in absolute values to differ the magnitudes between the biomes. The annual variability (b) is calculated by the division of annual standard deviation by the annual mean precipitation.

3.3. Long-term mean seasonal precipitation

Figure 7 shows the percent bias (PBIAS) of HadGEM2-ES, Eta/HadGEM2-ES, MIROC5 and Eta/MIROC5 in terms of amount in seasonal precipitation simulations. In general, Eta/HadGEM2-ES underestimates the rainfall in DJF and MAM in the Brazilian biomes, while the JJA and SON rainfalls are overestimated in the Amazon,
Atlantic Forest and Pampa. In the Cerrado and Pantanal, the largest biases occurred in the JJA season (dry period), which were increased by the downscaling process (Eta RCM) (Figure 7c and 7f). All seasons in the Caatinga were underestimated by the HadGEM2-ES, in terms of the amount in the season. On the other hand, Eta/MIROC5 underestimates rainfall throughout all the seasons in the Amazon and Pampa. Overestimates were simulated for MAM, JJA and SON in the Cerrado, DJF and MAM in the Caatinga, JJA and SON in the Pantanal and for all seasons in the Atlantic Forest, with lower overestimates in the dry season. The Eta/MIROC5 shows some improvements of the rainfall biases compared with its driving GCM, especially in the rainy seasons (DJF and MAM) for all biomes. These improvements were more distinct than for Eta/HadGEM2-ES, where improvements were restricted to some season/biome combinations.
Figure 7. PBIAS on the long-term seasonal precipitation for the 1980-2005 period simulated by models against observations in the a) Amazon, b) Atlantic Forest, c) Cerrado, d) Caatinga, e) Pampa, and f) Pantanal.

Correlation coefficients (Equation 3) are presented in Figure 8 for the GCM and RCM simulations. The results for the Eta/GCM corroborate those presented by Chou et al. (2014b), who found spatial correlations above 0.50 for all simulations in their regional analysis for the whole of Brazil. These better values of seasonal means compared to mean monthly values are expected, once that spatial errors are being reduced when averaging (Pierce et al., 2009). In general, Eta/HadGEM2-ES simulates the mean seasonal precipitation better than the Eta/MIROC5 in the Atlantic Forest, Caatinga, Cerrado and Pampa. We highlight the smaller correlation of Eta/HadGEM2-ES simulations in the Pantanal during the JJA (dry season) and minimal correlation (up to 0.07) found between Eta/MIROC5 and observations in Pampa during the MAM and SON and Pantanal’s JJA. On the other hand, both models show good results (up to 0.95) in simulating the seasonal cycle of precipitation for the Amazon, Caatinga, and Cerrado biomes. Once again, we noted better improvements of Eta RCM for MIROC5 than HadGEM2-ES and the best performance of GCM simulations for large areas such as the Amazon and Cerrado.
Figure 8. Taylor diagrams of seasonal mean precipitation over a) Amazon, b) Atlantic Forest, c) Cerrado, d) Caatinga, e) Pampa, and f) Pantanal for simulations and observations. Means of the observed seasonal precipitation are marked as a black star. The azimuth and the radial distance from the origin of the plot represents the correlation coefficient and the standard deviation (mm) of simulated data in relation to the observed value, respectively.
3.4. Investigating the origin of the biases

We consider that precipitation biases in the Eta/HadGEM2-ES and Eta/MIROC5 simulations for Brazilian biomes may have three possible reasons: a) they are produced by the GCM and not corrected by Eta RCM; b) they are produced by the Eta RCM and are absent in the CGM; or c) they are related to uncertainty in the observations. Figure 9 provides some insight concerning these reasons.

Figure 9. Mean bias error for precipitation simulations of a) HadGEM2-ES, b) Eta/HadGEM2-ES, c) MIROC5, and d) Eta/MIROC5 for the 1980-2005 period.

Figures 9a and 9b show that simulations of HadGEM2-ES and Eta/HadGEM2-ES have a positive bias in the western part of the Amazon and the southern part of the Atlantic Forest and negative biases in the northeastern area of the
Atlantic Forest and Caatinga biomes. In turn, Fig. 9c and 9d indicate negative biases in
the northern part of the Amazon, the southern region of the Atlantic Forest and all
throughout the Pampa biome, and strong positive biases in the Cerrado and the central
part of the Atlantic Forest. These biases are likely related to inherent driving GCM biases
that can be carried out from GCM to RCM via the lateral boundary conditions (Ehret et
al., 2012; Xu and Yang, 2015). The mean bias error estimated for the Pantanal is positive
for the GCM simulations and negative for the RCM simulations. This kind of error, and
the lower biases seen in Figure 3 of Section 3.1 for HadGEM2-ES instead of
Eta/HadGEM2-ES, could be explained by the downscaling process, but more in-depth
analysis in the data generation in the downscaling process and also in the physical
processes involved leading to deforestation need to be made. According to Chou et al.
(2014b), the Eta RCM is especially suited for regions with steep topography (particularly
because of the Eta vertical coordinate). Some of the surface physical processes of the
Pantanal biome may not be simulated accurately by the Eta RCM, leading to errors in the
precipitation outputs. We noted that all simulated data show a high negative mean bias
error in the extreme north of Brazil. A logical reason is the failure of the GCMs to capture
local features such as the sea breeze induced rainfall. The sea breeze circulation typifies
a mesoscale atmospheric system from coastal areas. It is a specific local wind system
(from sea to land) due to thermal differences between land and sea surfaces, which leads
to low level pressure anomalies. In the tropics, the mesoscale diurnal processes, such as
the sea breeze, are particularly important and may occur in 3 out of every 4 days (Ahrens,
2010; National Research Council, 1992). As shown in Kousky (1980), the Amazonian
coastal area is highly influenced by the sea breeze, with formation and propagation of the
line of convective activity inland. Thereof, we can relate the negative bias found in this
region to the inability of the GCMs to capture this mesoscale system, resulting in lower
amounts simulated than the observed ones. Moreover, using Eta RCM did not resolve this local feature. A logical reason is the uncertainty of the observations. The rain gauge stations in Brazil are not equally distributed over the biomes and the northern part of Brazil has the lowest density of stations (Xavier et al., 2016). Moreover, interpolation methods for generating the observational grid have uncertainties that can impact our results and must be considered. The inability of the GCMs to capture the local sea breeze influenced rainfall along with the observational gap in the northern part of the Amazon, resulted in a strong negative bias.

Based on the biases of the models to simulate temporal and spatial precipitation patterns (Equation 5), we calculated their suitability for the Brazilian biomes, which is graphically represented in a heatmap (Figure 10). The heatmap divides the errors into classes and each class has a color, ranging from blue (best) to yellow (worst). In general, models were capable of simulating the phase and the amplitude of the rainy and dry seasons for the large and well-defined season biomes. We highlight the excellent performance of models to simulate the dry season in the Amazon, Atlantic Forest, and Cerrado biomes. The Caatinga and Pantanal presented significant improvements on this property for the Eta/MIROC5. For the seasonal cycle, there was an overall good performance of models. MIROC5 presented the highest errors (worst performance) for Cerrado and Caatinga biomes, but the downscaling process improved their ability to simulate seasonal precipitation. Related to the annual precipitation (represented by the mean annual precipitation and annual variability), as well as for seasonal precipitation, the GCM MIROC5 was not capable of capturing the main characteristics of the Cerrado and Caatinga biomes. For all properties, simulations were improved by using Eta RCM in HadGEM2-ES data for the Pampa biome. For the same biome, MIROC5’s simulations did not improve by using Eta RCM. In a general view of
simulated precipitation over Brazil, the Eta RCM improved the results of HadGEM2-ES for many biomes, except for the Amazon and Caatinga, where the original GCM is more suitable than the downscaled data. Related to the MIROC5 family, the GCM simulations were improved in the Amazon, Cerrado, Caatinga and Pantanal by the Eta RCM. For the Atlantic Forest and Pampa, the Eta RCM worsened the simulations.

As we showed above, it is not a rule that the downscaling procedure will provide more suitable values of precipitation. The driving GCM HadGEM2-ES proved to be more suitable than Eta/HadGEM2-ES for large biomes such as the Amazon and Caatinga. At the same time, Eta/MIROC5 significantly improved the monthly means for almost all biomes and, consequently, the annual totals. For the Pampa biome, only the HadGEM2-ES family was capable of simulating the precipitation on acceptable levels. These results must be considered when the projections of these models are used. Moreover, our results supports previous studies that aimed in the same RCM-GCM evaluation. Liang et al. (2008) observed very high spatial correlation between RCM minus GCM differences in precipitation and temperature between present and future climates, indicating that a major portion of the biases found on simulations (either for RCM and GCM) are systematically propagated into their future projections. In addition,
they concluded that, even the uncertainty of future climate projections is sensitive to present climate simulation biases, there is no linear relationship between simulation and projections biases, depending on regions and models. We can infer that a model that better reproduce the present climate lead us to be more confident in the physical and dynamical processes considered and represented by this model under boundary conditions applied, such as the historical GHG concentration. Considering that only the boundary conditions and scenarios are changed for project future climate, we can also expect a good representation of the climate for a given scenario by the model. And even with the advances in model developments and computational power, biases are still occurring (and sometimes increasing) and the identification of their causes is an actual need for assessing future impacts (Addor and Seibert, 2014). These highlights the importance of a more accurate assessment of the origin and incidence of models’ biases, using adequate regions for the evaluation. As showed in Teutschbein & Seibert (2012), there is always a best bias correction method for a group of regions – achieving the best mean statistical results – but it is not always the best for all regions.

As is well known, GCMs and RCMs suffer from substantial biases, especially regarding precipitation (Flato et al., 2013; Kotlarski et al., 2014), and climate model precipitation usually needs to be bias corrected before these data are used for impact assessments. The most accurate chose of regions of assessment implies in the a more accurate chose of bias correction method to be applied in future projections, enhancing climate change impact studies.

4. CONCLUSIONS

We evaluated the performance of the downscaled precipitation data from higher-resolution RCM simulations driven by two coarser-resolution GCM products (Eta/HadGEM2-ES and Eta/MIROC5). Both products have been used in the Brazilian
Third National Communication to the UN Framework on Climate Change. We statistically analyzed long-term means of the simulated precipitation compared to high-resolution observation-based gridded products in order to better understand the reliability of these simulations. Our analysis was conducted for the six main Brazilian biomes in order to consider areas with rather homogeneous (eco)climatological patterns and we evaluated the precipitation simulations in terms of monthly and seasonal means, thereby considering the separation into rainy and dry seasons. To the best of our knowledge, it is the most appropriate evaluation of high-resolution climate change datasets of precipitation for large areas in Brazil.

For the long-term mean monthly analysis, HadGEM2-ES and Eta/HadGEM2-ES simulated rainy and dry seasons very well in the Amazon, Atlantic Forest and Cerrado biomes. This result expresses the potential reliability of the GCM to simulate mean fields of precipitation in large areas. The GCMs require lower time and computational effort to process long-term data for large areas than RCMs and in this case, HadGEM2-ES presents itself as a viable alternative for larger Brazilian biomes. In turn, Eta/MIROC5 showed great improvements when compared to its driving-GCM MIROC5. In most cases, for all biomes, the downscaling brought the simulated means close to the observational means. In the Pampa biome, no model was able to represent the mean monthly precipitation well. However, in some cases, the biases embedded in the model simulations interfered in the identification and duration of rainy/dry seasons. The long-term mean seasonal analysis showed that the Eta RCM modifies the range of precipitation, with less reliability of models to simulate means in the dry season (JJA and SON). According to our heatmap, we recommend the following model for each biome: HadGEM2-ES for the Amazon, Eta/HadGEM2-ES for the Atlantic Forest, the Cerrado, and the Pampa, and Eta/MIROC5 for the Caatinga and the Pantanal.
The development of regional climate models for Brazil increases the country’s ability to better understand the impacts of climate change. However, these data must be used with caution, as RCM simulations have systematic errors. Our results show that Eta/HadGEM2-ES and Eta/MIROC5 data for Brazil have various biases, which can be originated from the driving GCMs, introduced by the downscaling RCM, and related to uncertainties in the observational data. As these models project rainfall data for the future as well, it is expected that these biases are also present in these projections and if these data are not corrected, any hydrological application will be compromised. When corrected, the climate change simulations and projections become a valuable tool for increasing the resilience and decreasing the environmental, social, and economic vulnerability.

Acknowledgments

This study was supported by grants from the Ministry of Science, Technology, Innovation and Communication – MCTIC and National Council for Scientific and Technological Development – CNPq [grants numbers 441289/2017-7 and 306830/2017-5]. This study was also financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 and CAPES Print. Rafael Rosolem kindly acknowledges support from the Natural Environment Research Council (NERC) under the project “Brazilian Experimental datasets for MUlti-Scale interactions in the critical zone under Extreme Drought” (BEMUSED) [grant number NE/R004897/1]. Moreover, the authors are grateful to Dr. Chou Sin Chan of the National Institute of Space Research (INPE) for providing Eta HadGEM2-ES and MIROC5 data.
5. REFERENCES


Deidda, R., Marrocui, M., Caroletti, G., Pusceddu, G., Langousis, A., Lucarini, V.,


Junk, W.J., Cunha, C.N., Wantzen, K.M., Petermann, P., Strüßmann, C., Marques,


Murray-Tortarolo, G., Jaramillo, V.J., Maass, M., Friedlingstein, P., Sitch, S., 2017. The decreasing range between dry- and wet-season precipitation over land and its effect...
on vegetation primary productivity. PLoS One 12, 1–11.

https://doi.org/10.1073/pnas.0611338104

https://doi.org/10.1007/s41748-019-00116-x


https://doi.org/10.3390/w9060377


https://doi.org/10.1016/j.jaridenv.2016.12.009


https://doi.org/10.1073/pnas.0900094106


Watanabe, M., Suzuki, T., O’ishi, R., Komuro, Y., Watanabe, S., Emori, S., Takemura,
T., Chikira, M., Ogura, T., Sekiguchi, M., Takata, K., Yamazaki, D., Yokohata, T.,
Nozawa, T., Hasumi, H., Tatebe, H., Komio, M., 2010. Improved Climate
Simulation by MIROC5: Mean States, Variability, and Climate Sensitivity. J.
Clim. 23, 6312–6335. https://doi.org/10.1175/2010JCLI3679.1

Wu, M., Nikulin, G., Kjellström, E., Beluši, D., Lindstedt, D., 2020. The impact of
regional climate model formulation and resolution on simulated precipitation in

Xavier, A.C., King, C.W., Scanlon, B.R., 2016. Daily gridded meteorological variables
https://doi.org/10.1002/joc.4518

Xu, Z., Yang, Z., 2015. A new dynamical downscaling approach with GCM bias
https://doi.org/10.1002/2014JD022958.Received

vegetation characteristics derived from Moderate Resolution Imaging
Spectroradiometer (MODIS) leaf area index and normalized difference vegetation

https://doi.org/10.3724/SP.J.1248.2013.137