



POTSDAM-INSTITUT FÜR
KLIMAFOLGENFORSCHUNG

Originally published as:

Boers, N., Bookhagen, B., Marengo, J., Marwan, N., Storch, J.-S. von, Kurths, J. (2015): Extreme rainfall of the South American monsoon system: A dataset comparison using complex networks. - *Journal of Climate*, 28, 3, 1031-1056

DOI: [10.1175/JCLI-D-14-00340.1](https://doi.org/10.1175/JCLI-D-14-00340.1)

Available at <http://journals.ametsoc.org>

© American Meteorological Society

Extreme Rainfall of the South American Monsoon System: A Dataset Comparison Using Complex Networks

NIKLAS BOERS,* BODO BOOKHAGEN,⁺ JOSÉ MARENGO,[#] NORBERT MARWAN,[@]
JIN-SONG VON STORCH,[&] AND JÜRGEN KURTHS**

* *Potsdam Institute for Climate Impact Research, Potsdam, and Department of Physics, Humboldt University Berlin, Berlin, Germany*

⁺ *Institute of Earth and Environmental Science, University of Potsdam, Potsdam, Germany*

[#] *Centro de Ciência do Sistema Terrestre, Instituto Nacional de Pesquisa Espacial, Cachoeira Paulista, São Paulo, Brazil*

[@] *Potsdam Institute for Climate Impact Research, Potsdam, Germany*

[&] *Max Planck Institute for Meteorology, Hamburg, Germany*

** *Potsdam Institute for Climate Impact Research, Potsdam, and Department of Physics, Humboldt University Berlin, Berlin, Germany, and Department of Control Theory, Nizhny Novgorod State University, Nizhny Novgorod, Russia*

(Manuscript received 5 May 2014, in final form 22 October 2014)

ABSTRACT

In this study, the authors compare six different rainfall datasets for South America with a focus on their representation of extreme rainfall during the monsoon season (December–February): the gauge-calibrated TRMM 3B42 V7 satellite product; the near-real-time TRMM 3B42 V7 RT, the GPCP 1° daily (1DD) V1.2 satellite–gauge combination product, the Interim ECMWF Re-Analysis (ERA-Interim) product; output of a high-spatial-resolution run of the ECHAM6 global circulation model; and output of the regional climate model Eta. For the latter three, this study can be understood as a model evaluation. In addition to statistical values of local rainfall distributions, the authors focus on the spatial characteristics of extreme rainfall covariability. Since traditional approaches based on principal component analysis are not applicable in the context of extreme events, they apply and further develop methods based on complex network theory. This way, the authors uncover substantial differences in extreme rainfall patterns between the different datasets: (i) The three model-derived datasets yield very different results than the satellite–gauge combinations regarding the main climatological propagation pathways of extreme events as well as the main convergence zones of the monsoon system. (ii) Large discrepancies are found for the development of mesoscale convective systems in southeastern South America. (iii) Both TRMM datasets and ECHAM6 indicate a linkage of extreme rainfall events between the central Amazon basin and the eastern slopes of the central Andes, but this pattern is not reproduced by the remaining datasets. The authors' study suggests that none of the three model-derived datasets adequately captures extreme rainfall patterns in South America.

1. Introduction

Extreme rainfall plays a crucial role for the hydrological cycle in large parts of South America. In the subtropics, events above the 90th percentile account for more than 50% of total rainfall during the core monsoon season from December to February [see Fig. 5 but also Boers et al. (2014a)]. Considerable contributions are made by mesoscale convective systems (MCS), in particular in northern

Argentina, Paraguay, and southern Brazil (Zipser et al. 2006; Salio et al. 2007). However, apart from their role for the overall water budget, understanding extreme rainfall events and their synchronization is important because of the associated natural hazards. For example, extreme rainfall frequently triggers landslides and flash floods, in particular along the eastern slopes of the Andes (Schuster et al. 2002; Coppus and Imeson 2002; O'Hare and Rivas 2005; Moreiras 2005; Boers et al. 2014a), as well as in urban areas in southeastern South America (Marengo et al. 2013a).

The far-reaching impacts of extreme rainfall events call for a better understanding of how their frequency, magnitude, and spatial covariability are represented by

Corresponding author address: Niklas Boers, Potsdam Institute for Climate Impact Research, Telegraphenberg A31, 14412 Potsdam, Brandenburg, Germany.
E-mail: boers@pik-potsdam.de

different datasets. Furthermore, in order to obtain a sound assessment of future extreme rainfall development, it is crucial to evaluate climate models with respect to their performance in reproducing observed spatiotemporal characteristics of extreme rainfall. The representation of extreme events in a dataset or model is usually only assessed by investigating local rainfall distributions and in particular the behavior of their respective tails. Because of their outstanding hydrological, but also societal relevance, a detailed investigation of the spatiotemporal interrelations of extreme events and how they are represented by different datasets and climate models is needed.

In this study, we compare the representation of extreme rainfall events between three gridded observational, as well as three gridded model-derived datasets for South America: the satellite–gauge combined Tropical Rainfall Measuring Mission (TRMM) 3B42 V7; the corresponding real-time product TRMM 3B42 V7 RT, which is not gauge adjusted; the gauge–satellite GPCP combined product; the model-derived Interim European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-Interim) data; and high-spatial-resolution data obtained from the global circulation model ECHAM6 and from the Eta regional climate model.

Several previous studies have analyzed rainfall variability in the South American monsoon system (SAMS), but only a few studies have compared different rainfall datasets in South America: Matsuyama et al. (2002) have analyzed the Climate Prediction Center Merged Analysis of Precipitation and more recently Negrón Juárez et al. (2009) have compared satellite and gauge products over the Amazon basin, while Silva et al. (2011) evaluated different reanalysis products over the entire continent. Carvalho et al. (2012) provide a comparison of station-based, satellite-derived, and reanalysis data with a focus on daily gridded precipitation (see also references therein). A specific analysis of the implementation of rainfall processes over tropical South America in CMIP5 was recently carried out by Yin et al. (2013). Typically, these studies use principal component analysis (PCA) in order to investigate the spatial characteristics of rainfall covariability in terms of empirical orthogonal functions (EOFs).

Complex networks (CN) have proven to provide a well-suited general framework for the analysis of climatic time series (Tsonis et al. 2007; Yamasaki et al. 2008; Donges et al. 2009; Malik et al. 2012; Boers et al. 2014b; Stolbova et al. 2014). For many problems in climate data analysis, CN can be regarded as alternative to more traditional techniques such as PCA (Boers et al. 2013). However, in the context of daily rainfall and specifically of extreme events, methods based on PCA

are not applicable because of non-Gaussian data distributions and in particular because of the binary-like structure of extreme-event time series. In this case, CN provide the only known framework to analyze corresponding spatial characteristics of extreme rainfall synchronicity (e.g., Boers et al. 2013). The basic idea behind our approach is that climatic mechanisms characteristically influence the way how extreme rainfall events at different locations synchronize and that these influences are encoded in the corresponding high-dimensional synchronization matrix. Complex network theory allows us to extract that information from this matrix: Spatial patterns exhibited by suitable CN measures have been related to and used to explain the underlying climatic mechanisms. The complex atmospheric mechanisms typically involve processes represented by many different climatic observables such as temperature, pressure, or wind. However, their influence on just a single variable, rainfall, can by means of complex networks be used to identify these mechanisms and their role for the modulation of extreme events (Boers et al. 2013).

In an earlier study, Boers et al. (2013) applied a CN approach to the high-spatiotemporal-resolution TRMM 3B42 V7 dataset in order to investigate the spatial structure of synchronicity of extreme rainfall events in the SAMS. This dataset has been concluded to provide a reasonably good representation of rainfall variability in different regions (Xue et al. 2013; Chen et al. 2013) but particularly in South America (Carvalho et al. 2012; Zulkafii et al. 2014). Our methodology is based on the idea that relevant and important features of a climatic system such as the SAMS influence the way extreme rainfall events synchronize at different locations. These synchronizations can be represented by a CN and the CN's internal structure is assumed to contain the influences of the relevant climate features. Boers et al. (2013) documented that CN identify the most relevant climatic features of the SAMS. These include the main convergence zones (the intertropical and South Atlantic convergence zones), the most important moisture transport routes, and areas of frequent development of MCS (Maddox 1980; Durkee et al. 2009; Durkee and Mote 2010). Here, we extend the methodology by two more CN measures and compare results for TRMM 3B42 V7 to corresponding results for the two other observational as well as for the three model-derived datasets.

2. Climatic setting

We only provide a brief overview of the key climatic features relevant for rainfall in South America. For detailed reviews on the SAMS, we recommend, for example, Vera et al. (2006) and Marengo et al. (2012).

Throughout this study, we will focus on the core monsoon season in South America from December through February (DJF): that is, the austral summer season. This season is characterized by enhanced low-level flow of moist air from the tropical Atlantic Ocean toward the South American continent because of the southward displacement of the intertropical convergence zone (ITCZ) as well as the differential heating between ocean and land (Zhou et al. 1998). After crossing the central Amazon basin (CAB), these winds are blocked by the eastern Andean cordillera and redirected southward. In course, the moisture content of these low-level winds remains high, since precipitation is counterbalanced by strong evapotranspiration over the rain forest (Eltahir and Bras 1994; Marengo 2006). At the eastern slopes of the Andes, the winds lead to high rainfall amounts because of substantial orographic lifting (Bookhagen and Strecker 2008). There is large variability in the direction of the subsequent flow toward the subtropics, with typical exit regions from central Argentina to eastern Brazil. Southward anomalies of the flow are associated with increased rainfall in southeastern South America (SESA), while eastward anomalies are associated with increased rainfall in the South Atlantic convergence zone (SACZ) (Carvalho et al. 2002; Liebmann et al. 2004; Carvalho et al. 2004). The alternation between these two regimes with rainfall either in SESA or in the SACZ and vice versa is commonly called the South American rainfall dipole and has been described as the most important pattern of rainfall variability in the SAMS (Paegle and Mo 2002; Carvalho et al. 2002; Vera et al. 2006; Marengo et al. 2012). If the southward component is particularly strong east of the Bolivian Andes, these winds are referred to as the South American low-level jet (SALLJ) (Marengo et al. 2004), which has been associated with the development of MCS in SESA (Salio et al. 2007).

3. Data

We employ six different daily rainfall datasets:

- 1) TRMM: The research-grade TRMM 3B42 V7 gauge-calibrated satellite product at 3-hourly temporal and $0.25^\circ \times 0.25^\circ$ spatial resolutions (Huffman et al. 2007) is available from 1998 to 2013.
- 2) TRMM RT: the (near) real-time satellite product TRMM 3B42 V7 RT at 3-hourly temporal and $0.25^\circ \times 0.25^\circ$ spatial resolutions (Huffman et al. 2007) is available from 2001 to 2013.
- 3) GPCP: The GPCP 1° daily (1DD) V1.2, a satellite–gauge combination at daily temporal and $1.0^\circ \times 1.0^\circ$ spatial resolutions (Huffman et al. 2001), is available from 1996 to 2013.

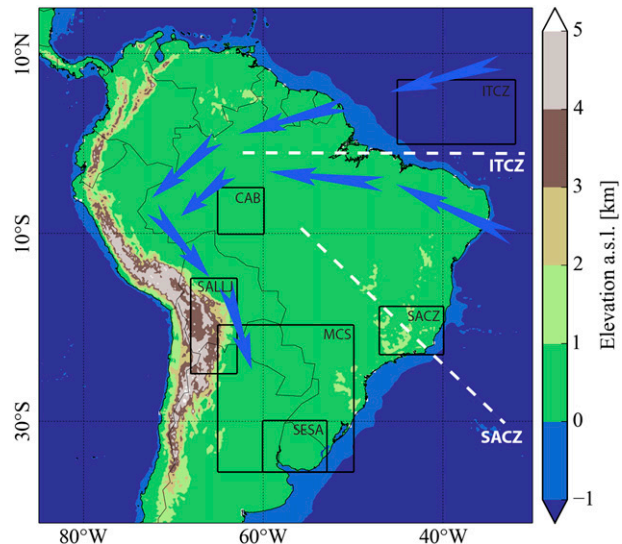


FIG. 1. Topography of South America and important atmospheric features of the SAMS. The black boxes delineate the following geographical regions: central Amazon basin, South American low-level jet, mesoscale convective systems, southeastern South America, South Atlantic convergence zone, and intertropical convergence zone. These regions are referred to in the text.

- 4) ERA-Interim: The ERA-Interim product with daily temporal and $0.75^\circ \times 0.75^\circ$ spatial resolutions is available from 1979 to 2013 (Dee et al. 2011).
- 5) ECHAM6: A standard AMIP simulation for the time period 1998 to 2008 (forced by the boundary conditions specified for CMIP5) is carried out with the ECHAM6 model (Stevens et al. 2013) at 6-hourly temporal resolution and horizontal resolution of T255 (~ 50 km) with 95 vertical levels. This simulation, which is carried out within the German consortium project STORM, is described in Hertwig et al. (2014).
- 6) Eta: Output from the regional climate model Eta (Mesinger et al. 2012) is provided by the Centro de Previsão de Tempo e Estudos Climáticos (CPTEC), driven by ERA-Interim. The original temporal resolution of this model run is 6 hourly, and the native horizontal resolution is $50 \text{ km} \times 50 \text{ km}$ with 38 vertical levels. The run covers the period from 1990 to 2008 (Solman et al. 2013; Marengo et al. 2013b).

For all datasets, we focused on daily values of the DJF seasons confined to the spatial domain from 85° to 30° W and from 40° to 5° S (see Fig. 1). TRMM and TRMM RT are first compared on their native $0.25^\circ \times 0.25^\circ$ grid for the time period 2001–12. The GPCP product is compared to TRMM for the period 1998–08 at its native resolution of $1.0^\circ \times 1.0^\circ$ after a bilinear interpolation of the TRMM data to this resolution. All other comparisons are carried out for the common time period from

1998 to 2008, with all datasets bilinearly interpolated to a common spatial resolution of $0.75^\circ \times 0.75^\circ$.

4. Methods

Our dataset comparison consists of two steps: First, we analyze and compare local rainfall distributions and several traditional statistical values with a focus on the tail behavior of the distributions. Second, we investigate the spatial synchronicity structure of strong, extreme, and most extreme rainfall events. For the time period 1998–2008, we define these events as the top 200, top 100, and top 50 events at each location, corresponding to events above the 80th, 90th, and 95th percentiles of the 11 DJF seasons, respectively. For strong events (above 80th percentile) we have $0.2 \times 11 \text{ yr} \times 92 \text{ days} = 202$ events, for extreme events (above 90th percentile) we have $0.1 \times 11 \text{ yr} \times 92 \text{ days} = 101$ events, and for the most extreme events (above the 95th percentile) we obtain $0.05 \times 11 \text{ yr} \times 92 \text{ days} = 51$ events. For the comparison between TRMM and TRMM RT, we have 12 DJF seasons (2001–12), which results in the top 216, 108, and 54 events at each grid cell. Locations with less than the respective number of events are discarded from the analysis. Previous studies have demonstrated that TRMM 3B42 V7 performs very well in reproducing the main spatial patterns of the SAMS (Boers et al. 2013) and found overall good agreement between TRMM and GPCP (Carvalho et al. 2012).

a. Event synchronization

To assess the similarity of event series at different locations, we compute the amount of synchronicity of events using the nonlinear measure event synchronization (ES; Quiroga et al. 2002; Malik et al. 2012; Boers et al. 2013). For two given event series e_i and e_j at grid cells i and j with l extreme events, consider two events e_i^μ and e_j^ν , with $e_j^\nu \leq e_i^\mu$ and $0 \leq \mu, \nu \leq l$. To decide whether these two events can be uniquely assigned to each other, we compute for $d_{ij}^{\mu,\nu} := e_i^\mu - e_j^\nu$ the dynamical delay,

$$\tau = \frac{\min\{d_{ii}^{\mu,\mu-1}, d_{ii}^{\mu,\mu+1}, d_{jj}^{\nu,\nu-1}, d_{jj}^{\nu,\nu+1}\}}{2}.$$

Furthermore, we introduce a filter by declaring a maximum delay $\tau_{\max} = 5$ days. We put $S_{ij}^{\mu,\nu} = 1$ if $|d_{ij}^{\mu,\nu}| \leq \tau$ and $|d_{ij}^{\mu,\nu}| \leq \tau_{\max}$ and $S_{ij}^{\mu,\nu} = 0$ otherwise. ES between e_i and e_j is then given as the normalized sum of

$$\text{ES}_{ij} := \frac{\sum_{\mu,\nu} S_{ij}^{\mu,\nu}}{l}.$$

A major advantage of this similarity measure is the introduction of the dynamical delay τ between events at different locations, which can vary from one time step to the next (within τ_{\max}). For the classical lead–lag analysis using, for example, Pearson’s correlation coefficient, this is not the case, since the prescribed time lags are assumed to be valid for the entire pair of time series. Because of the pronounced nonlinearities, but also because many different time scales are involved, we do not think that such temporal homogeneity is a justifiable assumption in the context of extreme rainfall.

All pairs of grid cells for which the value of ES is among the top 2% of all values will be represented by a CN link. This link density is chosen such that all links correspond to significant values of ES at a significance level of 5%. Significance is tested against a null model based on a uniformly random placement of events. A common representation of a CN is the adjacency matrix $A = (A_{ij})_{1 \leq i, j \leq N}$, for which $A_{ij} = 1$ if grid cells i and j are connected and $A_{ij} = 0$ otherwise. Here, N denotes the total number of grid cells. In the following, we will say that extreme events at two grid cells occur synchronously if these two grid cells are connected by a CN link.

b. Complex network measures

For each dataset, we will consider the following CN measures (Boers et al. 2013): First, the degree (DG) of a grid cell i is defined as the number of links attached to i ,

$$\text{DG}_i = \sum_{j=1}^N A_{ij}.$$

Thus, DG_i gives the number of grid cells where extreme events occur synchronously (within 5 days) with grid cell i . In this sense, DG assesses the importance of a grid cell for the distribution of extreme events to other locations.

Second, we are interested in the CN’s tendency to form clusters of mutually connected grid cells. This can be quantified as the probability of two CN neighbors of a grid cell to be connected themselves. Clustering (CL) is thus defined as

$$\text{CL}_i := \frac{\sum_{j < k} A_{ij} A_{jk} A_{ik}}{\sum_{j < k} A_{ij} A_{ik}}.$$

This CN measure estimates the spatial coherence of the occurrence of extreme events. High values have been observed in regions with frequent development of MCS (Boers et al. 2013).

Third, betweenness centrality (BC) at a grid cell is a nonlocal measure based on the notion of shortest paths

in a CN. A shortest path between two grid cells l and k is a shortest sequence of links that have to be passed to travel from l to k . BC of a node i is then defined as the number of shortest paths from l to k that pass through i [denoted by $\sigma_{kl}(i)$] divided by the total number of shortest paths from l to k (denoted by σ_{kl}),

$$BC_i := \frac{\sum_{l < k \neq i} \sigma_{kl}(i)}{\sum_{l < k \neq i} \sigma_{kl}}.$$

This measure estimates the importance of a grid cell for the long-ranged, directed propagation of extreme events. For TRMM, regions with high BC values have been found in the ITCZ, over the Amazon basin, and along the eastern slopes of the Andes in the vicinity of the SALLJ (Boers et al. 2013).

The spatial distribution of these CN measures is affected by the spatial embedding of the CN into a two-dimensional rectangular spatial domain. We correct for these effects following Rheinwalt et al. (2012) in the same way we did in Boers et al. (2013): The effect of the spatial embedding is quantified as the conditional probability of a network link between two nodes, given the geographical distance between these nodes. We therefore construct 200 surrogate networks from each original network that preserve the conditional link probability distribution given the corresponding geographical distances. These surrogate networks thus capture the topological properties that are only due to the spatial embedding. To correct for these spatial embedding effects, each network measure value of the original networks is therefore divided by the value of that measure that should be expected from the spatial embedding alone, as derived from the surrogate networks.

Finally, we are interested in how the six datasets represent local connectivities of certain regions that have been known to be of particular importance for the SAMS. The regional connectivity (RC) of a grid cell i to a region R is defined as the number of links that go from i to any grid cell within that region,

$$RC_i(R) = \sum_{j \in R} A_{ij}.$$

This measure can thus be used to assess where extreme events occur synchronously with events in a given region under consideration. We will focus on the CAB, SESA, and the SACZ (see Fig. 1).

5. Results and discussion

As noted above, several studies have concluded that TRMM 3B42 V7 provides reliable estimates of rainfall

variability (Xue et al. 2013; Chen et al. 2013; Carvalho et al. 2012; Zulkaffi et al. 2014). We will therefore consider the results for the TRMM dataset as a reference and discuss deviations from these results for the other five datasets.

a. Statistical values of rainfall distributions

1) MEAN DAILY RAINFALL

Mean daily rainfall rates (Fig. 2) show considerable differences between the spatial distribution exhibited by TRMM and the remaining five datasets. For TRMM, we observe high values at the ITCZ, over the Amazon basin, along the SACZ, near the Colombian Pacific coast, and at the eastern slopes of the Peruvian and Bolivian Andes.

The TRMM RT dataset shows a very similar spatial distribution but, at the eastern slopes of the Peruvian and Bolivian Andes, mean daily rainfall values are up to 5 mm day^{-1} higher. The GPCP dataset shows slightly lower values than TRMM in the ITCZ and at the eastern slopes of the Peruvian and Bolivian Andes but is otherwise in good agreement with TRMM.

In contrast, the model-derived datasets (ERA-Interim, ECHAM6, and Eta) deviate substantially from TRMM. ERA-Interim mean daily rainfall values are higher in the eastern Amazon basin but lower in the western part of the basin and along the SACZ. Furthermore, rainfall values at the eastern slopes of the Argentinean Andes are much higher than for TRMM.

The mean daily values of ECHAM6 are higher than TRMM in the western Amazon basin and along the eastern slopes of the Andes in northern Argentina and southern Bolivia but diminished at the climatological position of the SACZ. East of 40°W , over the subtropical Atlantic Ocean, there is a band of slightly higher values than for TRMM in this region.

Eta mean daily values are lower than for TRMM in the entire Amazon basin, along the SACZ, in most of SESA, in western Colombia, and in northeastern Brazil. On the other hand, values are higher than for TRMM at the eastern slopes of the Ecuadorian Andes.

2) THE 90TH AND 95TH PERCENTILES OF DAILY RAINFALL

For TRMM, scores at the 90th percentile (Fig. 3) follow a very similar spatial distribution as the mean values. For all five remaining datasets, the deviations from TRMM's spatial distribution at the 90th percentile scores are qualitatively similar to the deviations observed for mean daily rainfall, but with higher differences in absolute terms. TRMM RT overestimates 90th percentiles at the eastern Andean slopes of Bolivia and Peru, while GPCP underestimates 90th percentiles in

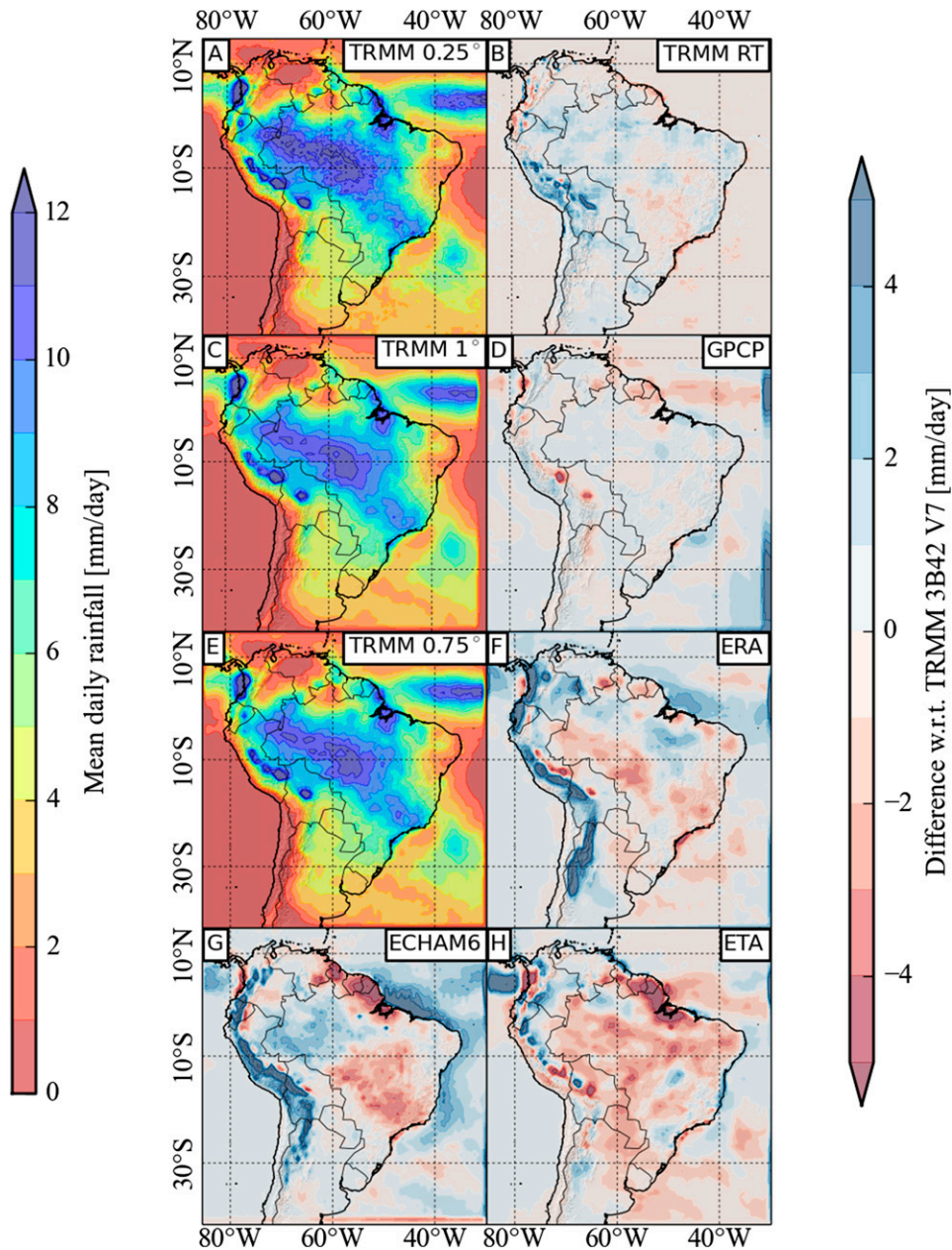


FIG. 2. Mean daily rainfall during DJF for TRMM 3B42 V7 at (a) 0.25°, (c) 1°, and (e) 0.75° spatial resolutions. Differences with respect to TRMM 3B42 V7 for (b) TRMM 3B42 V7 RT, (d) GPCP 1DD V1.2, (f) ERA-Interim, (g) ECHAM6, and (h) Eta.

this region. Apart from additional discrepancies between TRMM and GPCP over the ITCZ, the observational datasets (TRMM RT and GPCP) are still the ones with best agreement with TRMM.

Compared to TRMM, the model-derived datasets show substantial deviations: ERA-Interim exhibits lower values than TRMM in the ITCZ, the entire Amazon basin, and the SACZ. Meanwhile, we observe strongly overestimated rainfall scores at the eastern slopes and

high-elevation regions of the central and southern Andes. The ECHAM6 model performs well in reproducing TRMM's spatial distribution of extreme scores in the western and central Amazon basin, while values are too high along the ITCZ and too low in the eastern Amazon basin and along the continental part of the SACZ. Rainfall scores are strongly overestimated along the entire Andes mountain range. In contrast, the Eta model reproduces the values obtained for TRMM reasonably

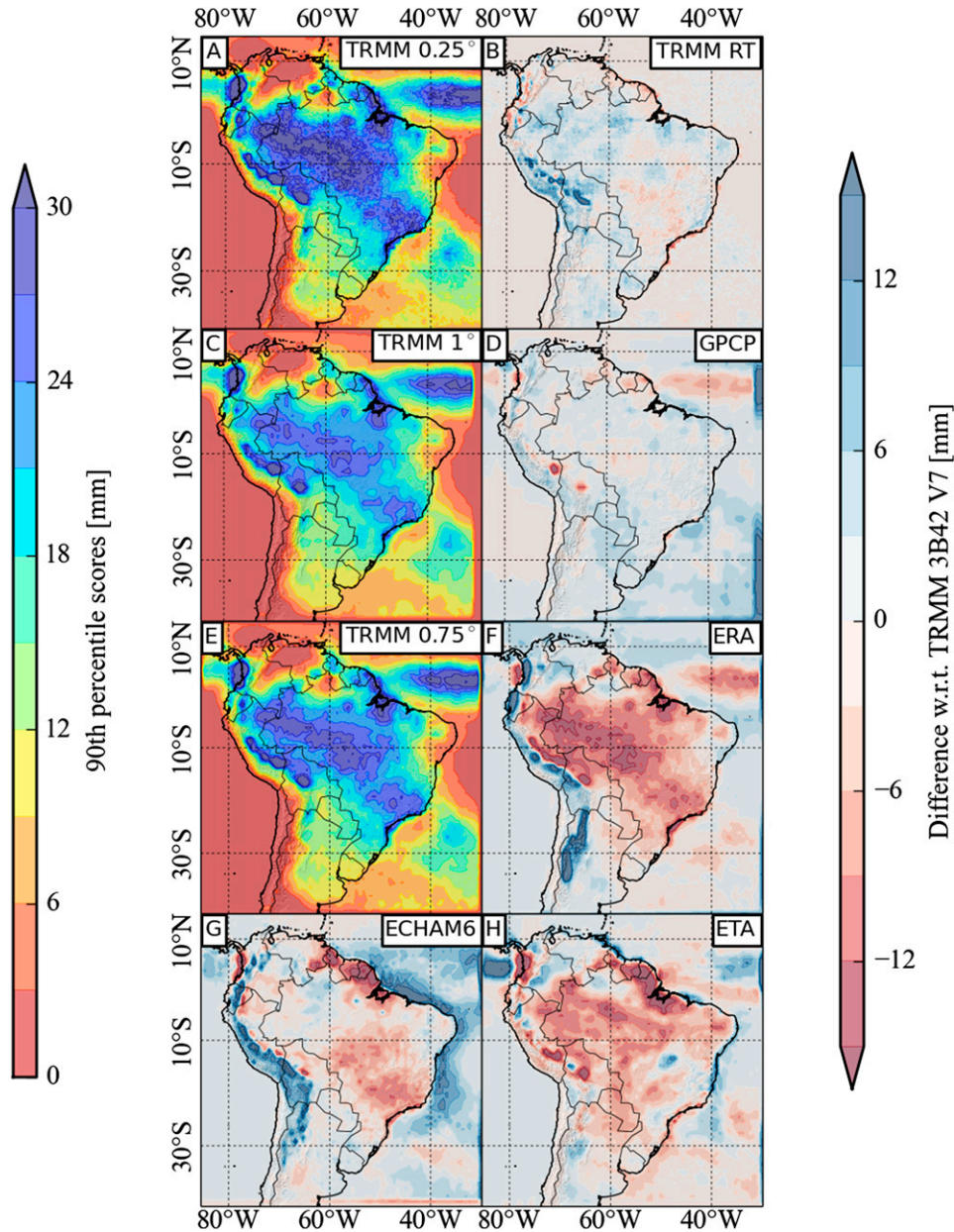


FIG. 3. Scores at the 90th percentile of daily rainfall during DJF for TRMM 3B42 V7 at (a) 0.25°, (c) 1°, and (e) 0.75° spatial resolutions. Differences with respect to TRMM 3B42 V7 for (b) TRMM 3B42 V7 RT, (d) GPCP 1DD V1.2, (f) ERA-Interim, (g) ECHAM6, and (h) Eta.

well in the SACZ but not in the Amazon basin, northeastern Brazil, or SESA, where values are too low.

For all six datasets, the spatial distribution (but not the absolute values) of 95th percentile scores (Fig. 4) resembles that of the 90th percentile, with the exception being locally higher scores in northeastern Argentina and southern Brazil for TRMM, TRMM RT, GPCP, and ECHAM6 but not for ERA-Interim and Eta.

3) FRACTIONS OF TOTAL SEASONAL RAINFALL

The fraction of total DJF rainfall accounted for by events above the 90th percentile (Fig. 5) shows a pronounced gradient from the tropics to the subtropics for all six datasets. For TRMM and TRMM RT, more than 70% of total DJF rainfall is caused by these events in large parts of South America south of 20°S, while they account for about 50% in the Amazon basin and in the

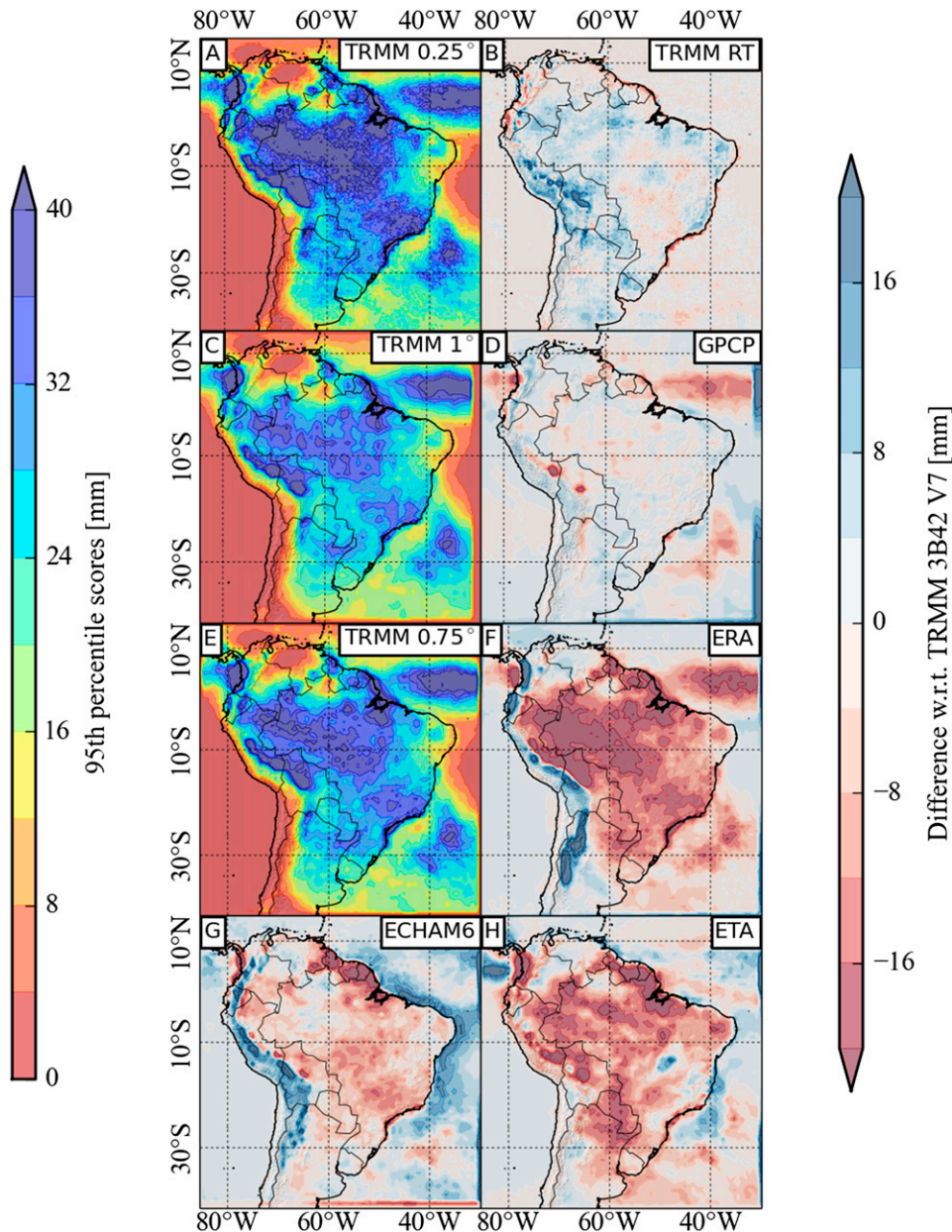


FIG. 4. Scores at the 95th percentile of daily rainfall during DJF for TRMM 3B42 V7 at (a) 0.25°, (c) 1°, and (e) 0.75° spatial resolutions. Differences with respect to TRMM 3B42 V7 for (b) TRMM 3B42 V7 RT, (d) GPCP 1DD V1.2, (f) ERA-Interim, (g) ECHAM6, and (h) Eta.

vicinity of the SACZ. For GPCP, we find only very small deviations from TRMM over the South American continent, while values are considerably lower over the Atlantic Ocean north of the equator and south of 20°. These fractions are smaller for the ERA-Interim dataset, with 40%–80% in southern South America and less than 30% in the Amazon basin and along the SACZ. For ECHAM6, fractions are closer to the fractions found for TRMM, with 30%–60% in the Amazon basin and at the SACZ and 60%–80% farther south. The Eta model

produces values between the fractions found for ERA-Interim and ECHAM6. The spatial pattern for all six datasets is remarkably similar. We note that fractions accounted for by events above the 95th percentile have a very similar spatial distribution as fractions for events above the 90th percentile, with, of course, reduced values (not shown).

The fact that the model-derived datasets (ERA-Interim, ECHAM6, and Eta) show lower values in SESA indicates that the models fail to reproduce the low-frequency and

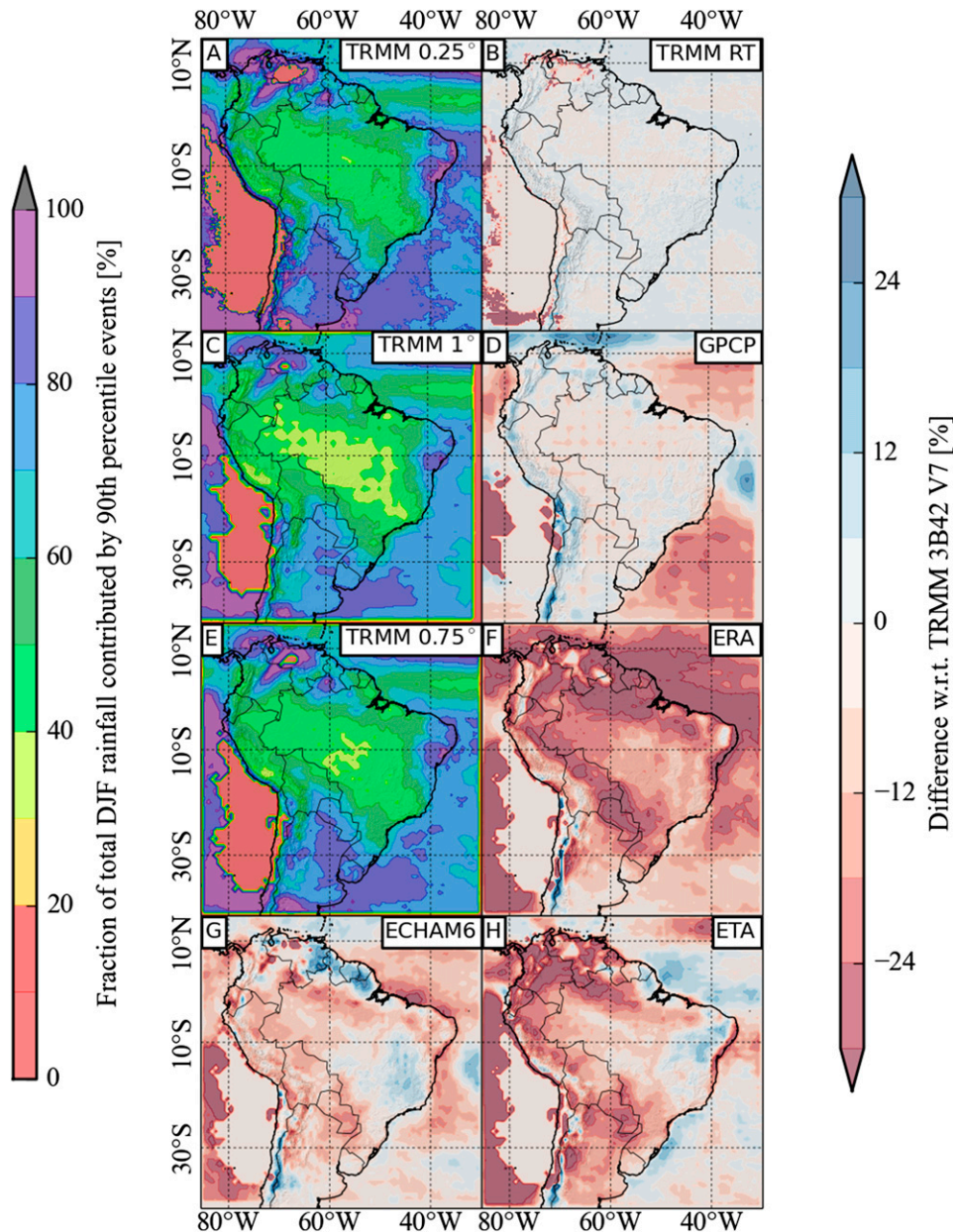


FIG. 5. Fraction of total DJF rainfall accounted for by events above the 90th percentile of daily rainfall for TRMM 3B42 V7 at (a) 0.25°, (c) 1°, and (e) 0.75° spatial resolutions. Differences with respect to TRMM 3B42 V7 for (b) TRMM 3B42 V7 RT, (d) GPCP 1DD V1.2, (f) ERA-Interim, (g) ECHAM6, and (h) Eta. Note that this fraction typically decreases for lower grid resolutions.

high-magnitude characteristics found for the observational datasets (TRMM, TRMM RT, and GPCP) in this region. Furthermore, the above-mentioned latitudinal gradient is less pronounced for the three model-derived datasets.

4) DIFFERENCE BETWEEN 90TH AND 50TH PERCENTILES

The difference between local scores at the 90th and at the 50th percentile (Fig. 6) yields a relatively simple

estimate of the tail behavior of the local rainfall distributions at each grid point. For TRMM, we observe high values in the ITCZ and the northern Brazilian coast, at the Colombian Pacific coast, and at the eastern slopes of the central Andes in Peru and Bolivia. Intermediate values can be found in the Amazon basin and the adjacent SACZ. Again, TRMM RT has higher values than TRMM at the eastern slopes of the central Andes but is otherwise in good agreement with TRMM. The GPCP

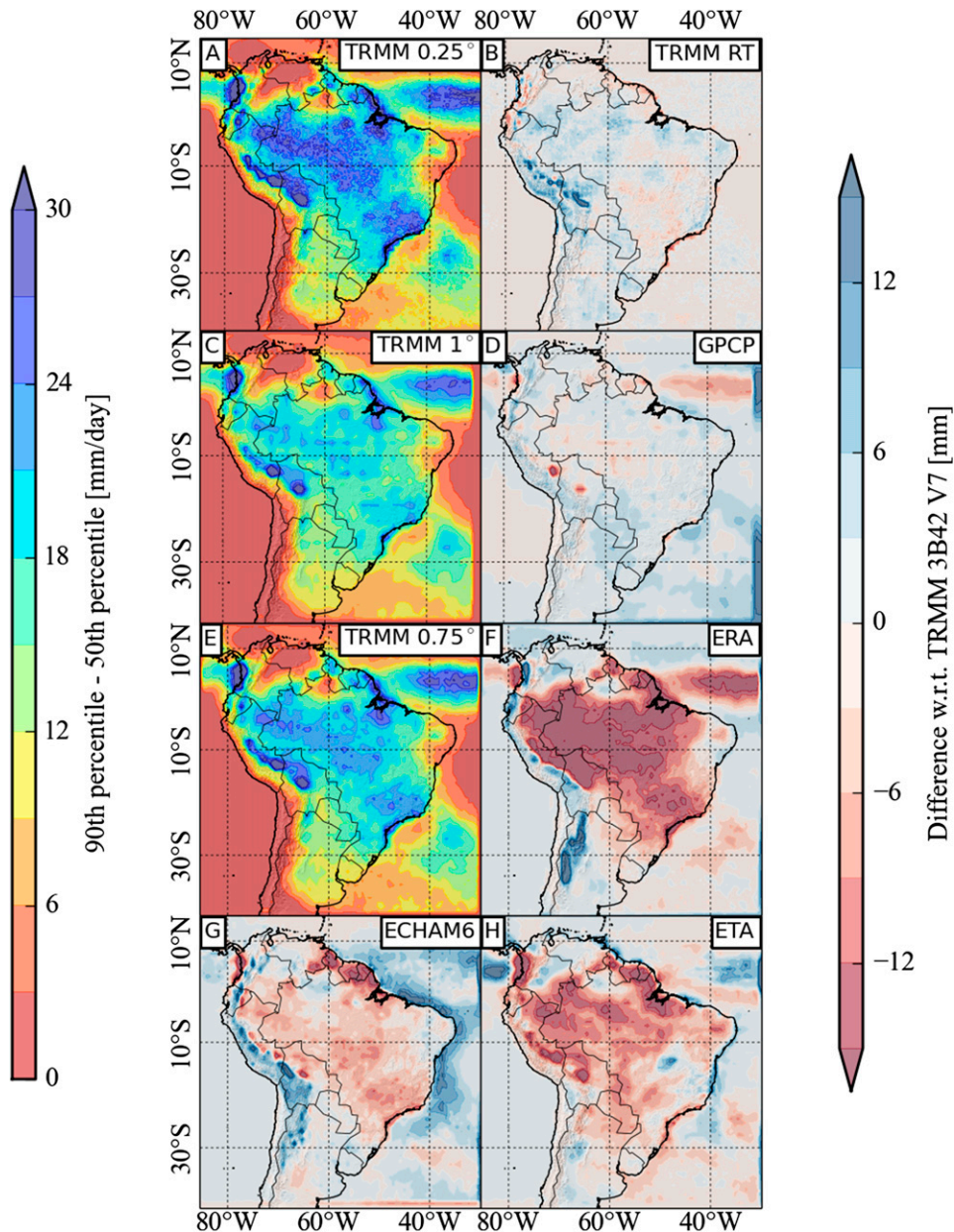


FIG. 6. Difference between scores at the 90th and 50th percentiles (90th percentile – 50th percentile) of daily rainfall during DJF for TRMM 3B42 V7 at (a) 0.25°, (c) 1°, and (e) 0.75° spatial resolutions. Differences with respect to TRMM 3B42 V7 for (b) TRMM 3B42 V7 RT, (d) GPCP 1DD V1.2, (f) ERA-Interim, (g) ECHAM6, and (h) Eta.

dataset exhibits lower values than TRMM in the ITCZ and at the eastern central Andean slopes but is otherwise also in good agreement.

The model-derived datasets perform rather poorly in reproducing the values found for TRMM. Most notably, they strongly underestimate the distributions' tails in the entire Amazon basin. For ERA-Interim, values are lower than for TRMM, with locally higher values only at the Colombian Pacific coast and the eastern slopes of the

Argentinean Andes. In the entire Amazon basin and the SACZ, values are strongly reduced as compared to TRMM (values $< 10 \text{ mm day}^{-1}$ instead of between 15 and 25 mm day^{-1}). For ECHAM6, values are higher than for ERA-Interim but still lower than for TRMM in the Amazon basin and the continental part of the SACZ. At the ITCZ and the adjacent coast but also at the eastern slopes of the Peruvian and Bolivian Andes, values are close to the values found for TRMM. However, values are too high at

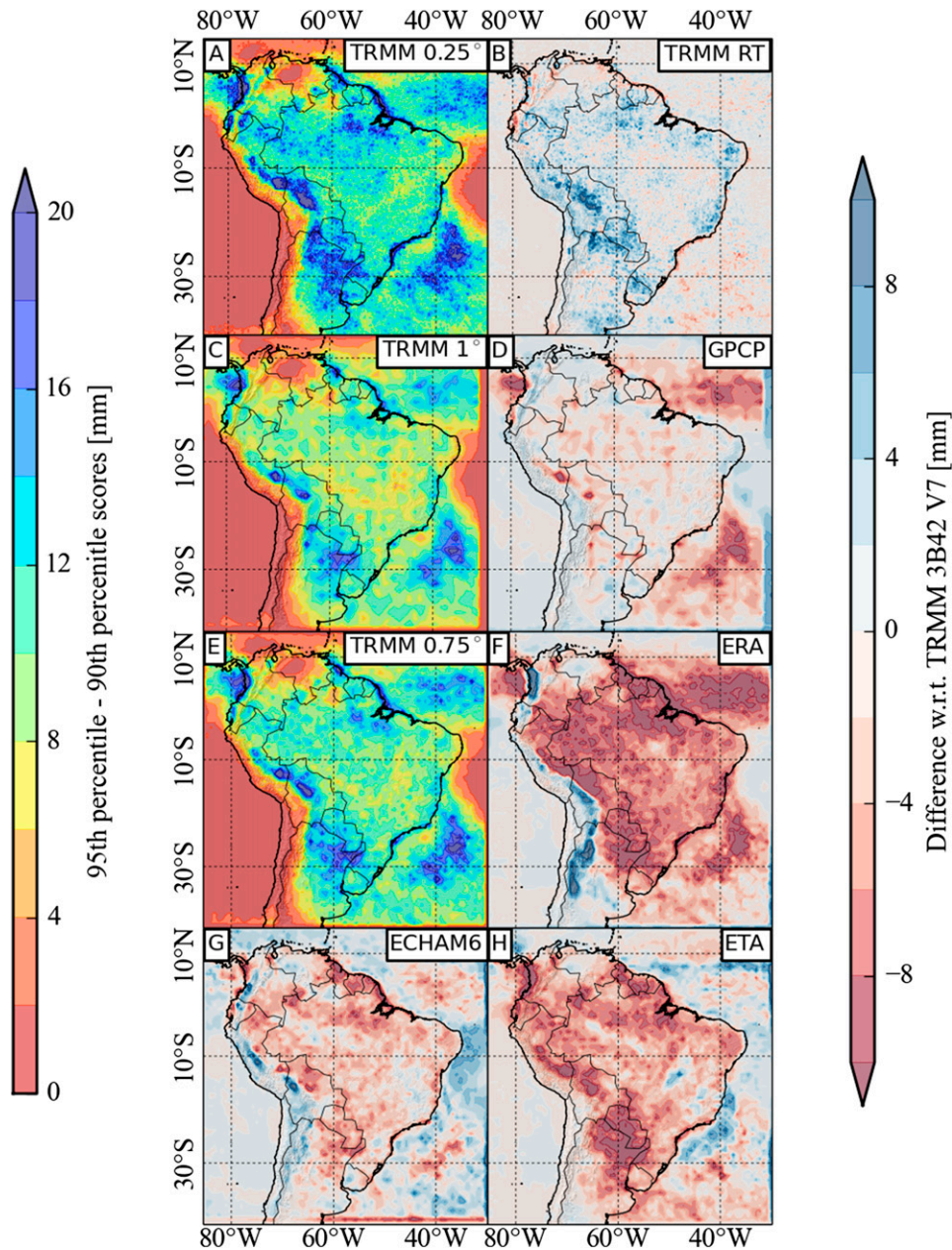


FIG. 7. Difference between scores at the 95th and 90th percentiles (95th percentile – 90th percentile) of daily rainfall during DJF for TRMM 3B42 V7 at (a) 0.25°, (c) 1°, and (e) 0.75° spatial resolutions. Differences with respect to TRMM 3B42 V7 for (b) TRMM 3B42 V7 RT, (d) GPCP 1DD V1.2, (f) ERA-Interim, (g) ECHAM6, and (h) Eta.

the northern Argentinean Andes. The Eta dataset exhibits lower values than TRMM in the Amazon basin but higher values in some eastern parts of the continental SACZ.

5) DIFFERENCE BETWEEN 95TH AND 90TH PERCENTILES

To assess the different datasets' behaviors for the most extreme events, we computed the difference between scores at the 95th and 90th percentiles (Fig. 7).

For the TRMM dataset, the difference between scores at the 95th and 90th percentiles exhibits high values in the ITCZ and the coastal areas around the equator, at the Colombian Pacific coast, and at the eastern slopes of the Peruvian and Bolivian Andes. Furthermore, high values can be observed in northern and northeastern Argentina and in the oceanic part of the SACZ. Meanwhile, the continental part of the SACZ does not show a particularly high difference between the 95th and 90th percentiles.

In large parts of the continent, TRMM RT shows higher values than TRMM, particularly in the Bolivian Andes, northern Argentina, Paraguay, and southern Brazil. GPCP is in fair agreement with TRMM, apart from underestimated values at the ITCZ and the eastern central Andes. However, rather large differences from TRMM can be observed over the ocean (in particular over the Atlantic Ocean south of 20°), where no rain gauge data are available.

For the ERA-Interim dataset, the differences between scores at the 95th and 90th percentiles exhibit lower values than for the TRMM dataset over the entire continent, except for the western slopes of the Colombian Andes and the eastern slopes of the eastern slopes of the Bolivian and Argentinean Andes, where values are locally high. The ECHAM6 model data show high values over the ITCZ, the Colombian Pacific coast, the eastern Andes from Peru to Argentina, and the subtropical Atlantic Ocean east of 40°W. For Eta, we observe high values over the ITCZ and the oceanic part of the SACZ comparable to TRMM. However, in the remainder of the continent values are lower than for TRMM: in particular, the high values in northwestern Argentina are not well reproduced by the Eta model.

High values for the observational datasets (TRMM, TRMM RT, and GPCP) in northeastern Argentina can be identified with the development of MCS in this region (Durkee et al. 2009), which are thus responsible for the most extreme rainfall events. None of the three model-derived datasets features a particularly heavy tail of the rainfall distribution in this area, suggesting that all of them fail to reproduce the most extreme rainfall events originating from the MCS.

The strongest orographic rainfall peaks at the eastern flanks of the Peruvian and Bolivian Andes (Bookhagen and Strecker 2008) are overestimated by TRMM RT, ERA-Interim, and ECHAM6 but underestimated by GPCP and Eta.

6) REGIONAL CHARACTERISTICS OF RAINFALL DISTRIBUTIONS

For the spatial boxes denoted by CAB, SALLJ, MCS, SESA, SACZ, and ITCZ in Fig. 1, we computed the 50th to 95th percentiles (in steps of 5) of the respective rainfall distributions (Fig. 8, left) and the difference of each dataset from the reference TRMM dataset (Fig. 8, right).

(i) CAB

In the CAB, the TRMM RT dataset shows small positive deviations for the highest percentiles, while we observe small positive deviations for GPCP in particular for lower percentiles. In contrast, with increasing percentiles,

ERA-Interim and Eta show increasing negative deviations. Underestimation of rainfall over the central Amazon by a collection of regional models (including Eta) was also observed by Solman et al. (2013). The ECHAM6 model, in contrast, overestimates daily rainfall scores up to the 80th percentile in this region but underestimates the 95th percentile score.

(ii) SALLJ

In the vicinity of the SALLJ east of the central Andes, TRMM RT shows considerable positive deviations only for the 90th and 95th percentiles and overall good agreement with TRMM for the remaining percentiles. GPCP scores are close to TRMM scores on the entire domain from the 50th to the 95th percentile. ERA-Interim and ECHAM6 show strong positive deviations from TRMM for the entire domain. However, deviations for ERA-Interim become small for high percentiles, while for ECHAM6 differences increase up to the 90th percentile. This overestimation of orographic rainfall, in particular at the eastern slopes of the Andes, is a well-known problem shared by most global and regional circulation models (Urrutia and Vuille 2009; Solman et al. 2013). In contrast, the Eta model produces values close to TRMM up to the 85th percentile; for higher percentiles, the scores are lower than for TRMM. The Eta model thus performs better than ERA-Interim and ECHAM6 over this mountainous region for the lower percentiles of the distribution, probably because of the step-like representation of mountains in its vertical coordinate system (the so-called eta coordinate) (Mesinger 1984; Mesinger and Black 1992; Pesquero et al. 2010). Still, the 90th and 95th percentiles are underestimated by Eta when compared to TRMM.

(iii) MCS

In the MCS region in subtropical South America, TRMM RT and GPCP are in good agreement with TRMM for all percentiles in this region. ERA-Interim, ECHAM6, and Eta exhibit higher scores than TRMM up to the 85th percentile and negative deviations for the 90th and 95th percentiles. However, ECHAM6 remains close to TRMM, while ERA-Interim and Eta deviate strongly for the 95th percentile. Problems to implement the particularly strong cyclogenetic activity in this region have been observed for a number of global and regional models (Solman et al. 2008; Chou et al. 2012; Solman et al. 2013). Our results here suggest that these deficiencies are mainly due to the underestimation of the most extreme events.

(iv) SESA

Because of considerable spatial overlap, it is little surprising that for SESA, we observe a similar behavior

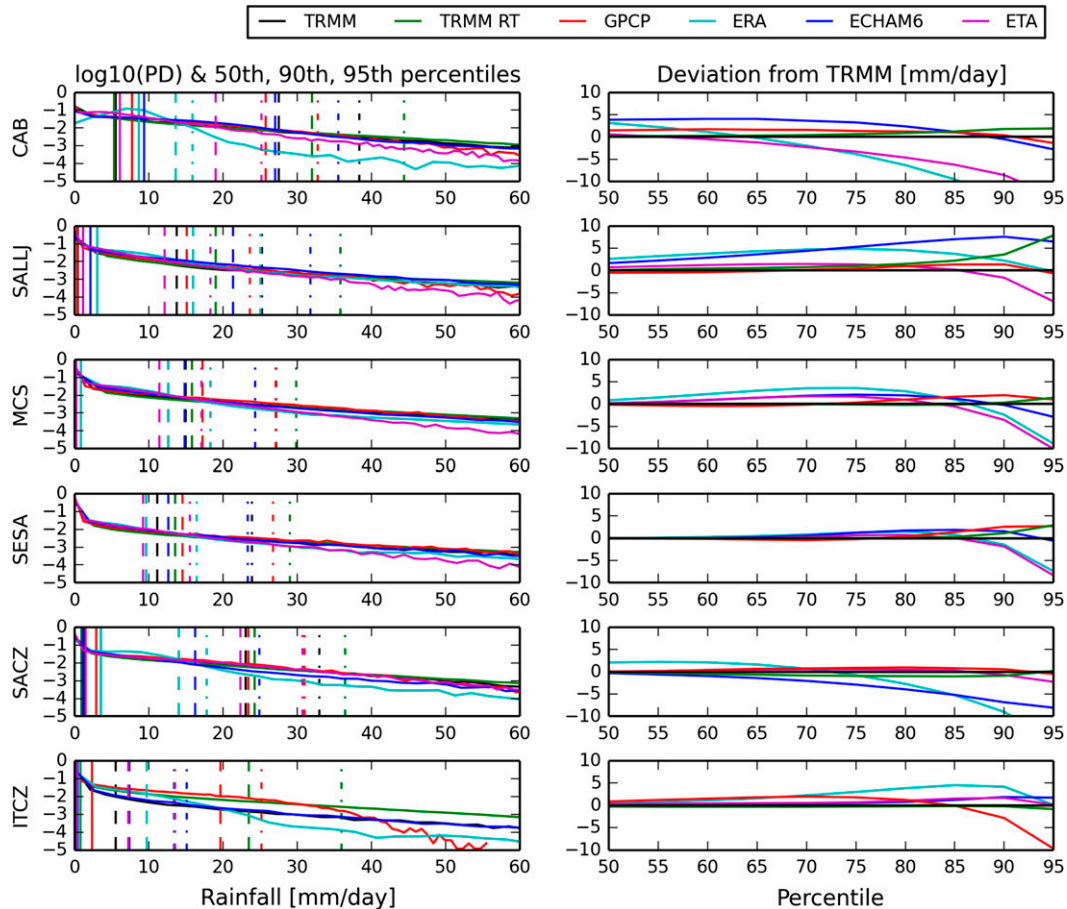


FIG. 8. (left) Comparison of logarithmic daily rainfall distributions for TRMM (black), TRMM RT (green), GPCP (red), ERA-Interim (cyan), ECHAM6 (blue), and Eta (magenta). Rainfall distributions are obtained for CAB, the SALLJ region, the MCS region, SESA, SACZ, and ITCZ (see Fig. 1 for definitions of these regions). Vertical lines denote median (solid), 90th percentile (dashed), and 95th percentile (dotted-dashed) scores of the respective distributions. (right) Percentile score differences with respect to TRMM 3B42 V7 for the five remaining datasets for the same regions.

as for the MCS region. However, discrepancies are smaller up to the 85th percentile and larger for the 90th and 95th percentiles than for the latter region.

(v) SACZ

In the SACZ region, values agree fairly well between TRMM, TRMM RT, GPCP, and Eta. The good performance of the Eta model in reproducing rainfall in this area was also observed by Solman et al. (2013). For higher percentiles, ERA-Interim and ECHAM6 show considerable negative deviations from TRMM, indicating that these two model-derived datasets fail to reproduce strong and extreme precipitation in the SACZ region.

(vi) ITCZ

Over the Atlantic ITCZ, TRMM RT has almost no deviations from TRMM. This is not surprising, since there

are no gauge stations that can be used for postprocessing adjustment. GPCP exhibits considerably lower 90th and 95th percentile scores than TRMM. ERA-Interim shows moderate positive deviations for low percentiles; however, they increase to about 5 mm at the 90th percentile. ECHAM6 and Eta are in good agreement with TRMM on the entire percentile domain.

The results described so far all concern statistical properties of local rainfall distributions. To investigate the spatial structure of synchronization between extreme events at different locations, we will now turn to the results of our CN approach.

b. Complex network measures

In the following, we will describe the spatial characteristics of extreme rainfall synchronicity, which are reflected by the network measures DG, BC, CL, and RC,

introduced in section 4b. All measures are computed for network measures derived for a maximum delay of $\tau_{\max} = \pm 5$ days. Corresponding results for $\tau_{\max} = \pm 7$ days are very similar (not shown), indicating that most extreme events synchronize on time scales shorter than 5 days. For all six datasets, these network measures are obtained by constructing networks as described in section 4 for rainfall events above the 80th, 90th, and 95th percentiles, respectively.

Instead of presenting differences from TRMM as for local statistical values in the last section, we will show absolute values for each dataset because, rather than specific local values, the overall spatial patterns are important for this kind of analysis. For the three network measures DG, BC, and CL, we will also compare spatial averages between the different datasets for the Amazon, SALLJ, MCS, and ITCZ areas (Fig. 15).

1) DEGREE

As explained in section 4b, we expect the DG to be high at locations that are particularly important for the direct distribution of extreme rainfall over the continent. Extreme events at these grid cells are strongly synchronized with extreme events at many other locations.

DG for 90th percentile events of the TRMM dataset (Fig. 9) exhibits a connected area of high values from northeastern Brazil westward across the Amazon basin and along the eastern slopes of the central Andes toward central Argentina, SESA, and the adjacent subtropical Atlantic Ocean. The SACZ is characterized by low values of DG. This spatial pattern indicates the main and well-known climatological moisture pathways along which extreme rainfall events synchronize (Vera et al. 2006; Marengo et al. 2012; Boers et al. 2013): These pathways lead from the mouth of the Amazon River across the Amazon basin and continue farther west to the Andes, where the mountain range blocks the low-level winds and channels them southward. Extreme events follow this wind channel (the SALLJ), cause abundant rainfall at the eastern slope of the Andes through orographic lifting, and continue farther toward northern Argentina and SESA.

For the TRMM RT dataset, we observe a very similar spatial distribution of DG as for the gauge-calibrated TRMM version. Also for GPCP, the spatial pattern is similar to the one found for TRMM. However, lower values along the eastern slopes of the Andes can be observed, which may be due to low numbers of measurement stations in these regions, as well as interpolations to a regular 1° grid.

In contrast, for ERA-Interim we get high values only north of the ITCZ over the Atlantic Ocean, in Uruguay and eastern Argentina, and over the adjacent subtropical

Atlantic Ocean. All other parts of the continent exhibit quite continuously low DG values and the pattern that was observed for TRMM is absent. ERA-Interim thus fails to reproduce the main synchronization pathway of extreme events exhibited by TRMM.

DG for the ECHAM6 model data is high in northeastern Brazil, over the Amazon basin, at the eastern slopes of the Bolivian Andes, in most of SESA, and over the adjacent subtropical Atlantic Ocean. However, values for ECHAM6 are lower compared to TRMM over the western Amazon basin and at the slopes of the Peruvian and northern Bolivian Andes, indicating that this model does not accurately represent the southwestward propagation of extreme rainfall from the Amazon.

The Eta model data yield high DG over the western Amazon basin, in northern Argentina, and over the subtropical Atlantic Ocean south of 30°S . No high values can be found along the eastern slopes of the Andes. Despite these substantial discrepancies, we argue that ECHAM6 and the Eta model reproduce the large-scale DG pattern to a reasonable extent.

2) BETWEENNESS CENTRALITY

For BC (Fig. 10), we expect a strong emphasis on long-ranged, directed synchronization pathways, because this measure is defined on the basis of shortest paths in the network. In contrast, DG only measures the connectivity to direct CN neighbors of a grid cell.

For the 90th percentile events of TRMM, BC exhibits high values over the ITCZ, from northeastern Brazil across the Amazon basin and along the eastern slopes of the Andes southward to central Argentina. In addition, relatively high values can be observed over the SACZ. This spatial distribution of BC has been found to be consistent with the interpretation of large-scale, directed moisture pathways (Boers et al. 2013). Apart from being less homogeneous, the spatial pattern is similar to the one for DG over eastern Brazil, the Amazon basin, along the Andes, and over SESA. On the other hand, BC is high in the vicinity of the ITCZ and, to weaker extent, in the vicinity of the SACZ, where DG is particularly low. This highlights the role of these convergence zones for large-scale propagation of extreme events over the entire continent, since their high connectivity only becomes apparent when taking into account network paths incorporating several other locations, as opposed to only local next neighbor connections.

The spatial pattern of BC for the TRMM RT dataset is in very good agreement with the results for TRMM, indicating that the large-scale propagation of extreme events is already captured well by the satellite-only product.

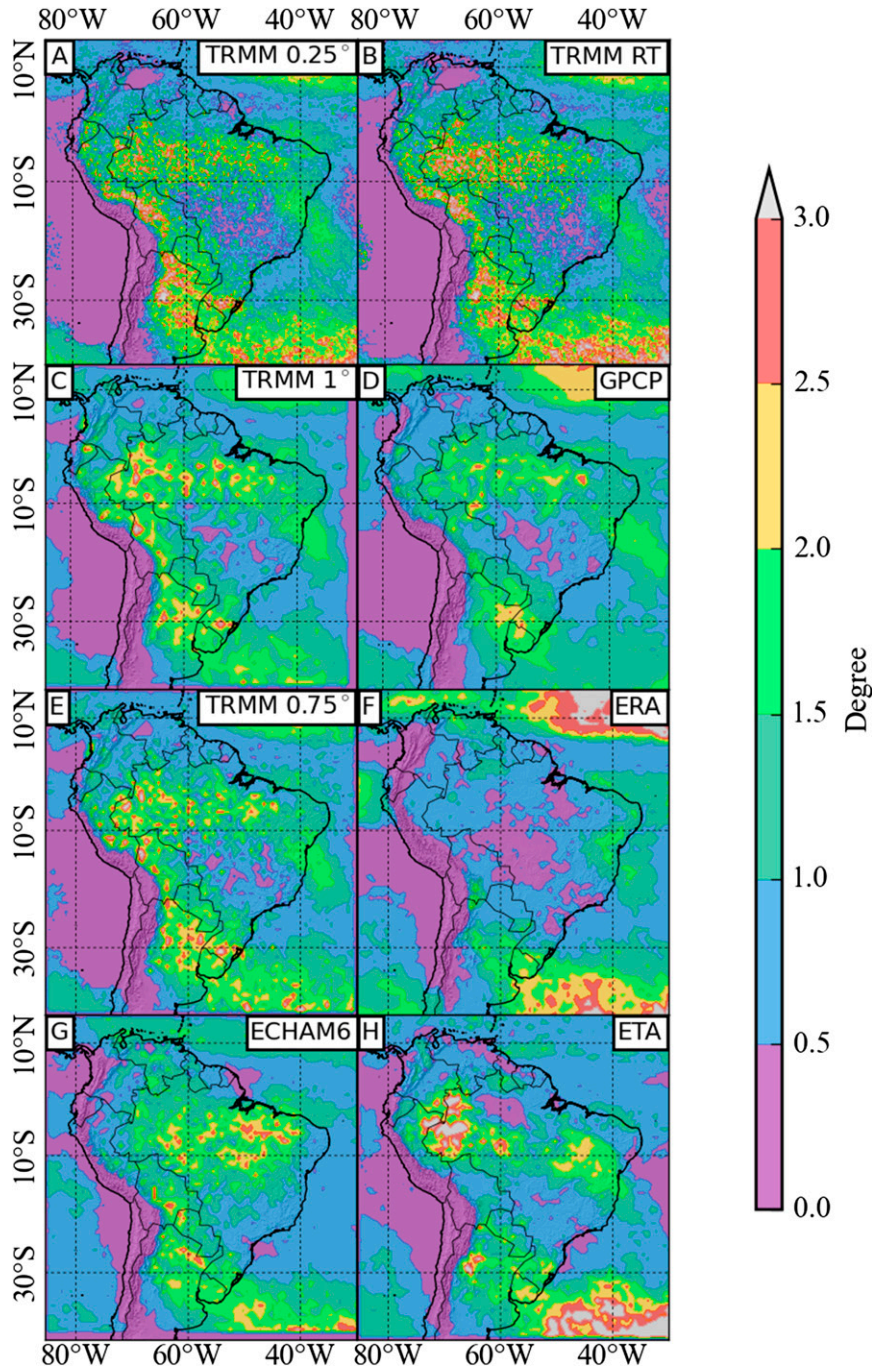


FIG. 9. Network measure degree calculated for the 90th percentile of all DJF seasons for TRMM3B42 V7 at (a) 0.25°, (c) 1°, and (e) 0.75° spatial resolutions and for (b) TRMM 3B42 V7 RT, (d) GPCP 1DD V1.2, (f) ERA-Interim, (g) ECHAM6, and (h) Eta.

Similarly, results for the GPCP dataset are in fair agreement with TRMM's spatial pattern, with only small discrepancies at the eastern slopes of the southern central Andes.

In strong contrast to DG, the spatial pattern of BC exhibited by ERA-Interim does quite accurately resemble

the spatial pattern found for TRMM, with high values marking the large-scale propagation pathway from the ITCZ toward the Andes and southward to northern Argentina (Fig. 10). Since BC is the more sophisticated measure for the long-range, directed propagation pathways, we conclude that ERA-Interim reproduces the

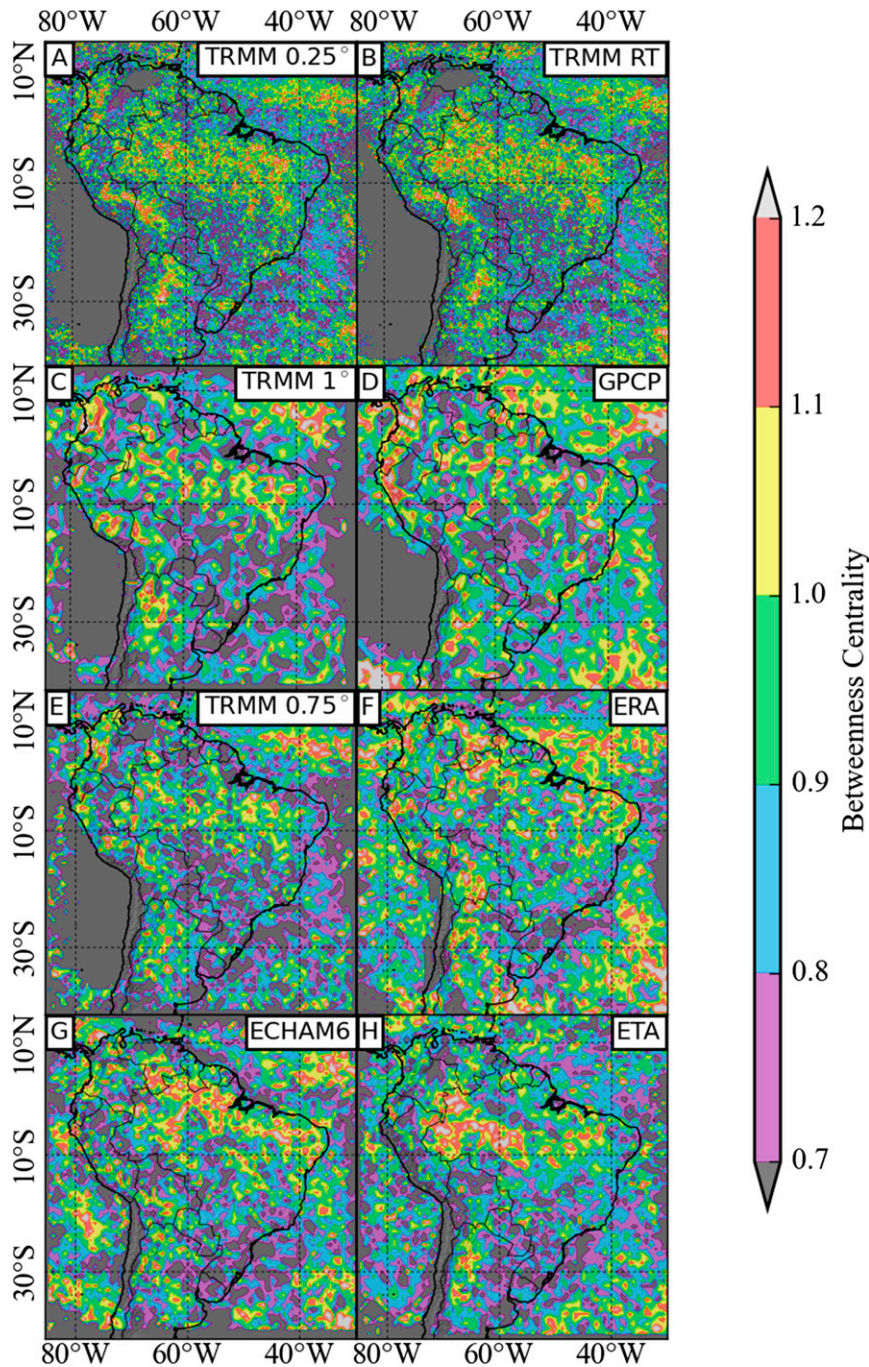


FIG. 10. Network measure betweenness centrality calculated for the 90th percentile of all DJF seasons for TRMM 3B42 V7 at (a) 0.25°, (c) 1°, and (e) 0.75° spatial resolutions and for (b) TRMM 3B42 V7 RT, (d) GPCP 1DD V1.2, (f) ERA-Interim, (g) ECHAM6, and (h) Eta.

climatological synchronization pathway, although it does not perform well in estimating the local connectivities along the pathway.

For the ECHAM6 model, there are high values of BC over the ITCZ as well but also over the entire coast of northeastern South America and the CAB. Furthermore,

high values can be observed in Bolivia and northern Argentina east of the Andes. Hence, this model accurately incorporates the large-scale propagation paths.

The Eta model yields high BC in the central and western Amazon basin. Values are also relatively high at the eastern slopes of the Southern Andes of Argentina.

However, over the ITCZ and in the vicinity of the SALLJ in Bolivia they are not as high as for the other five datasets. Hence, the Eta model does not perform well in reproducing the large-scale transport route from the ITCZ across the Amazon and southward along the Andes to the subtropics. In particular, problems with implementing the SALLJ's impact on extreme rainfall and corresponding orographic effects at the eastern slopes of the central Andes are apparent. The latter observation, which is corroborated by the results obtained for DG, is in agreement with the underestimation of 90th and 95th percentile scores by the Eta model at the eastern Andean slopes described in [section 5a](#).

3) CLUSTERING

Extreme rainfall in regions with high CL values ([Fig. 11](#) and right column of [Fig. 15](#)) can be expected to occur in a spatially coherent manner. In particular, this is the case for large thunderstorms, squall lines, and MCS.

For events above the 90th percentile ([Fig. 11](#) and right column of [Fig. 15](#)), the TRMM dataset exhibits high CL at the mouth of the Amazon river, in eastern Brazil, at the slopes of the northern Argentinean Andes, and in most of SESA including most of Paraguay. The Amazon basin and the SACZ are characterized by low CL. In ([Boers et al. 2013](#)), we attributed high CL values in northern Argentina, Paraguay, and southern Brazil to the frequent development of MCS in this region. Similarly, high values in northern Brazil close the Atlantic coast may correspond to the so-called Amazonian squall lines ([Cohen et al. 1995](#)). Furthermore, rainfall on the Altiplano and Puna de Atacama plateaus in northwestern Argentina and southwestern Bolivia has been found to typically occur in a way that either it rains on the entire plateau for several days in a row or it does not rain on the entire plateau ([Garreaud 2000](#)). This spatial coherence of rainfall events is expressed by high CL values over the plateau for the TRMM dataset.

While the relative spatial pattern of CL for TRMM RT looks similar to the one for TRMM, we observe substantially higher values for the satellite-only product in all regions and for all three event types.

Although the GPCP dataset yields too low CL values in the entire spatial domain, it has to be noted that the only region of relatively enhanced values is the subtropical region around Paraguay. Thus, GPCP does in this sense still discover this area of frequent MCS development, although values are much lower than for TRMM for all three percentiles.

ERA-Interim and ECHAM6 do not show any concise spatial pattern of CL, suggesting that these models have problems to reproduce large convective systems (e.g., MCS or large thunderstorms).

In contrast, the Eta model shows a clear spatial signature for CL, with relatively low values over the CAB, although they are still much higher than for TRMM. However, while some similarities to the pattern obtained for TRMM are apparent, values are not particularly high in Paraguay and northern Argentina.

4) REGIONAL CONNECTIVITY

Regional connectivity of a given region yields an estimate of where extreme events typically occur synchronously with events in that region.

(i) CAB

Events above the 90th percentile in the CAB ([Fig. 12](#)) are synchronized with events in the entire Amazon basin but also with events at the eastern slopes of the Peruvian and northern Bolivian Andes for TRMM. Furthermore, there are weak linkages to northern Argentina. TRMM RT exhibits very similar regional connectivity for the central Amazon. For both TRMM and TRMM RT, weak connections to the eastern Amazon basin can be interpreted as a consequence of the dominant easterly direction of low-level winds in this area, driving extreme rainfall events from the eastern part of the Amazon to the west. However, we also observe connection to the west until the eastern slopes of the central Andes in northern Bolivia. This shows that extreme events in the Amazon basin are climatologically synchronized with extreme events at the eastern slopes of the Andes. On longer spatial scales, we observe a linkage between extreme rainfall in the Amazon and a small area in northern Argentina at the foothills of the Andes. These results support the hypotheses that the Amazon plays a crucial role for distributing and propagating extreme rainfall events to large parts of the remaining continent, as it was suggested by high DG and BC values.

For GPCP, the CAB is only connected to the larger area of the Amazon basin and the slopes of the Peruvian Andes but not of the Bolivian Andes, possibly because of a lack of gauge stations in this area.

In comparison, for ERA-Interim the connectivity of the central Amazon is more localized, with almost no links pointing out of the region. In particular, the pattern of long-ranged connection between the Amazon basin and the eastern slopes of the Peruvian and Bolivian Andes is not reproduced by ERA-Interim.

In contrast, the ECHAM6 model produces strong connections to the eastern part of the Amazon basin and to the Peruvian and Bolivian Andes to the west in a similar way as observed for TRMM. However, the teleconnection from the Amazon basin to northern Argentina is not visible.

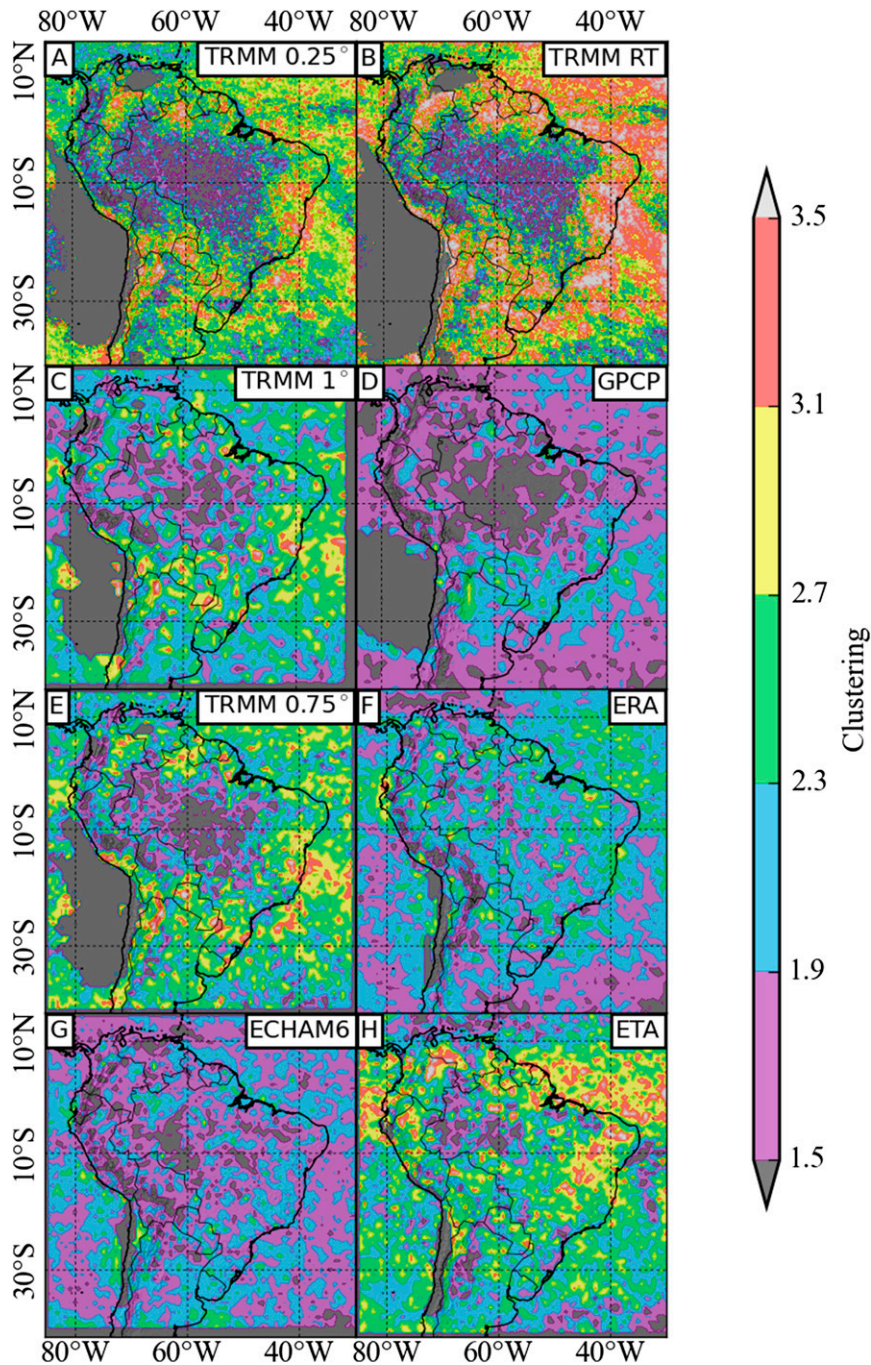


FIG. 11. Network measure clustering calculated for the 90th percentile of all DJF seasons for TRMM 3B42 V7 at (a) 0.25°, (c) 1°, and (e) 0.75° spatial resolutions and for (b) TRMM 3B42 V7 RT, (d) GPCP 1DD V1.2, (f) ERA-Interim, (g) ECHAM6, and (h) Eta.

For the Eta model, events in the central Amazon are less synchronized to the eastern parts of the basin as compared to TRMM but still present. No connectivity to the eastern slopes of the Andes and the adjacent lowlands can be observed. Thus, the Eta model reasonably resembles the connectivity of the central Amazon to the

eastern and northern parts of the basin, but it does not show any connectivity to the eastern slopes of the Peruvian and Bolivian Andes. This suggests that the impact of the low-level flow from the Amazon basin toward the Andes and relevant orographic lifting effects on extreme rainfall are not correctly implemented in this model.

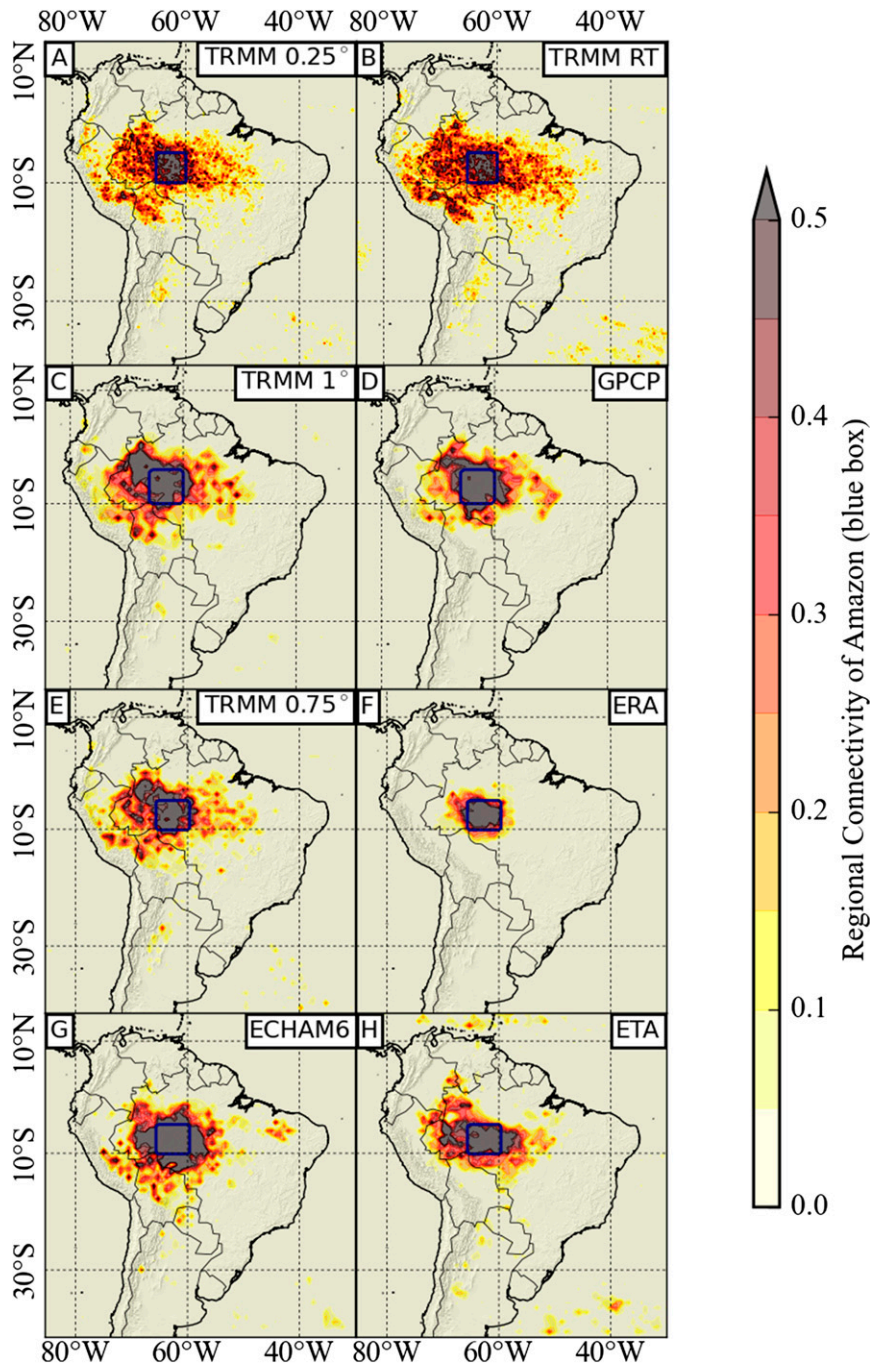


FIG. 12. Network measure regional connectivity for the CAB (blue box) calculated for the 90th percentile of all DJF seasons for TRMM 3B42 V7 at (a) 0.25°, (c) 1°, and (e) 0.75° spatial resolutions and for (b) TRMM 3B42 V7 RT, (d) GPCP 1DD V1.2, (f) ERA-Interim, (g) ECHAM6, and (h) Eta.

(ii) *SESA*

The regional connectivity of *SESA* (Fig. 13) is represented quite similarly by all six datasets. The main discrepancy is that only TRMM and TRMM RT show a connection to the

eastern slopes of the southern central Andes in northwestern Argentina and southern Bolivia. This signature can be assigned to a certain class of MCS, which form over *SESA* and then migrate toward the Bolivian Andes (Anabor et al. 2008). While the GPCP dataset seems to incorporate them

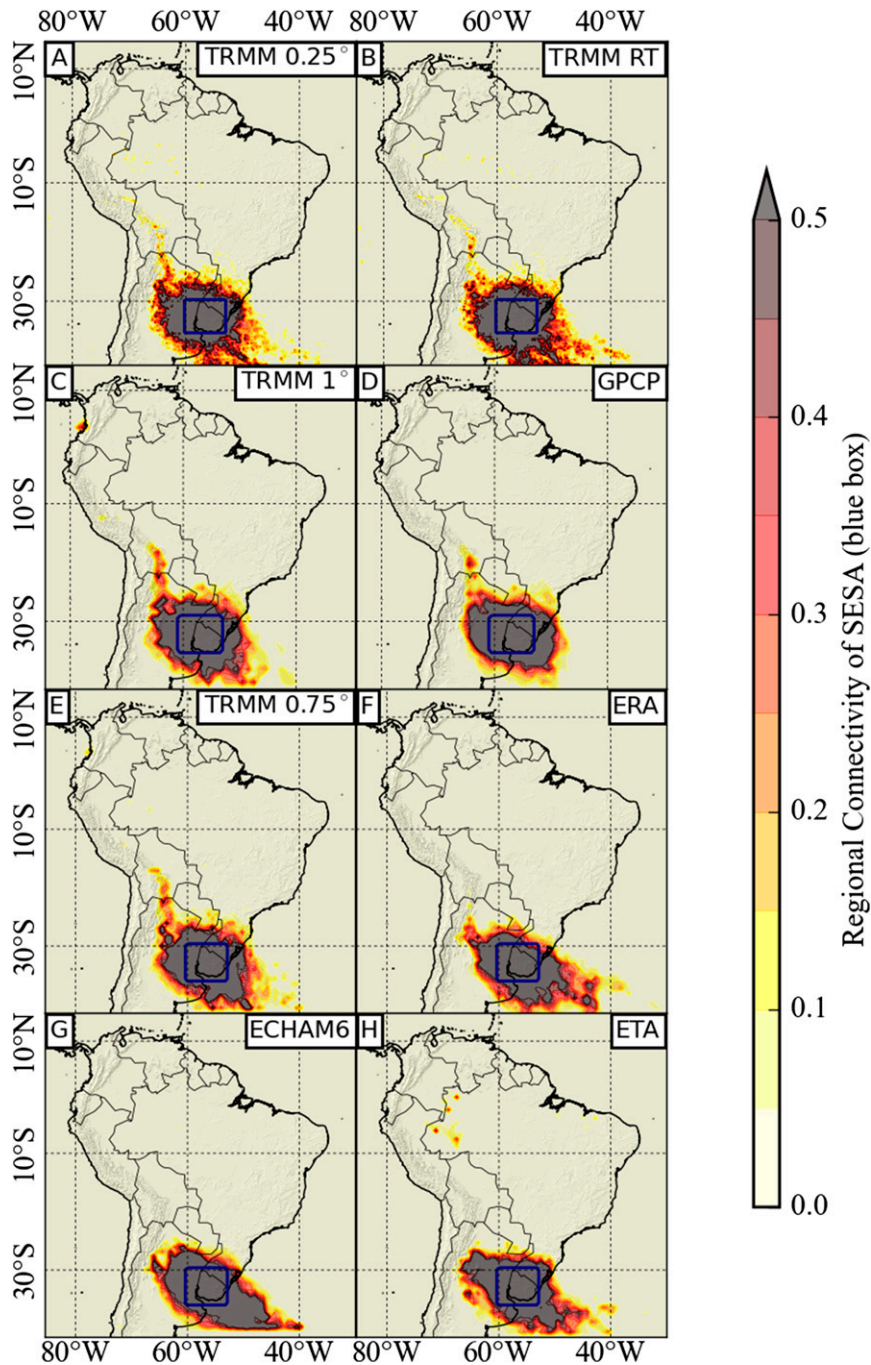


FIG. 13. Network measure regional connectivity for SESA (blue box) calculated for the 90th percentile of all DJF seasons for TRMM 3B42 V7 at (a) 0.25°, (c) 1°, and (e) 0.75° spatial resolutions and for (b) TRMM 3B42 V7 RT, (d) GPCP 1DD V1.2, (f) ERA-Interim, (g) ECHAM6, and (h) Eta.

(although less concise), none of the remaining three datasets yields a signature corresponding to these special MCS.

(iii) SACZ

The SACZ region is connected only locally with very few links pointing out of it for TRMM (Fig. 14). All five

remaining datasets quite accurately reproduce this rather localized signature.

5) SPATIAL AVERAGES OF NETWORK MEASURES

The four regions indicated as Amazon, SALLJ, MCS, and ITCZ in Fig. 1 are of particular relevance for some

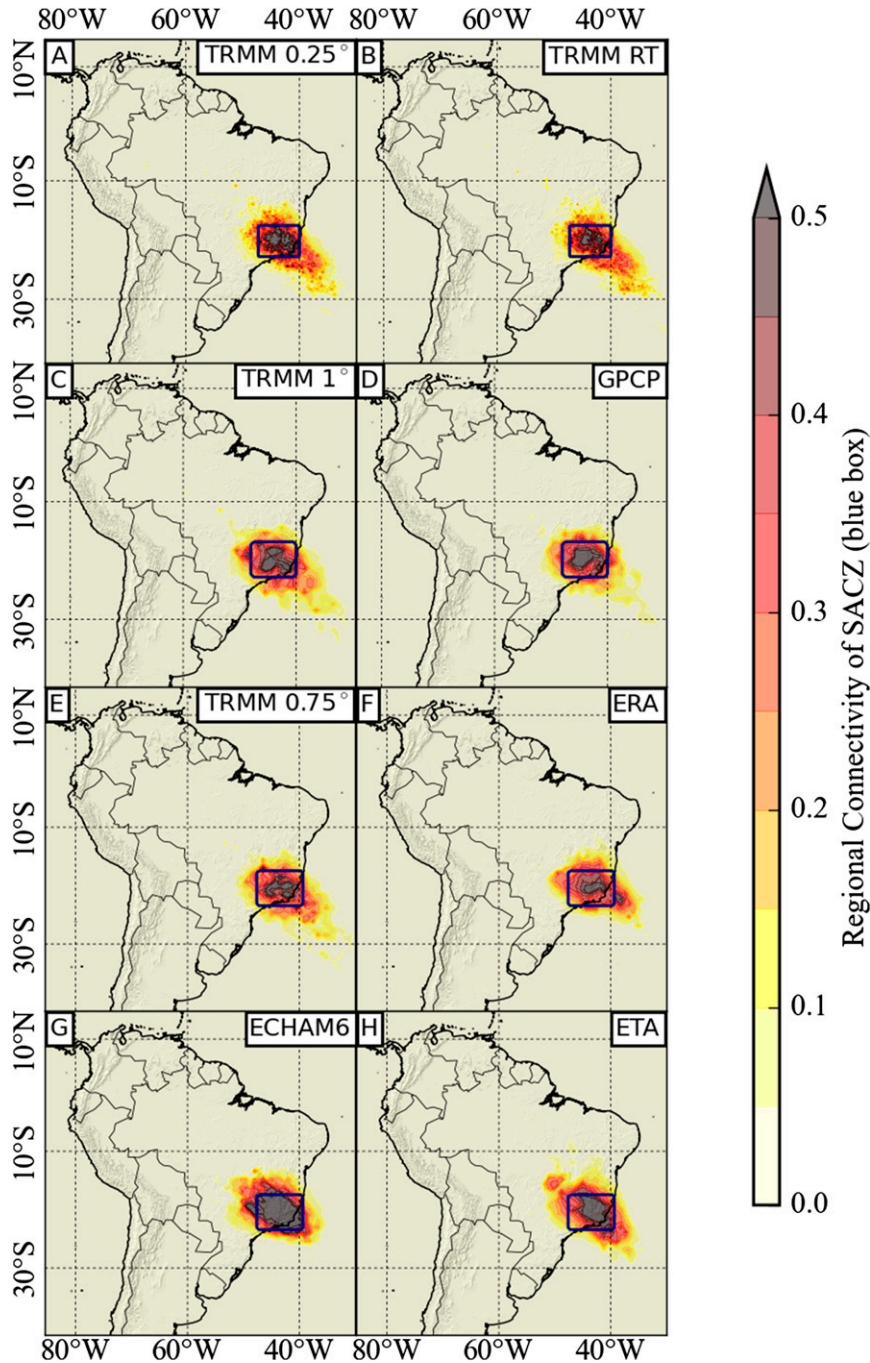


FIG. 14. Network measure regional connectivity for the SACZ (blue box) calculated for the 90th percentile of all DJF seasons for TRMM 3B42 V7 at (a) 0.25°, (c) 1°, and (e) 0.75° spatial resolutions and for (b) TRMM 3B42 V7 RT, (d) GPCP 1DD V1.2, (f) ERA-Interim, (g) ECHAM6, and (h) Eta.

key features of the SAMS: As explained in [section 2](#), the ITCZ, the Amazon, and the area east of the central Andes in the vicinity of the SALLJ play important roles for the large-scale propagation of extreme events. On the other hand, the MCS region is frequently exposed to

exceptionally large convective complexes, contributing large fractions of total seasonal rainfall. In addition to the pattern analysis carried out in the previous sections, we therefore compute spatial averages of the network measures DG, BC, and CL for these four regions ([Fig. 15](#)),

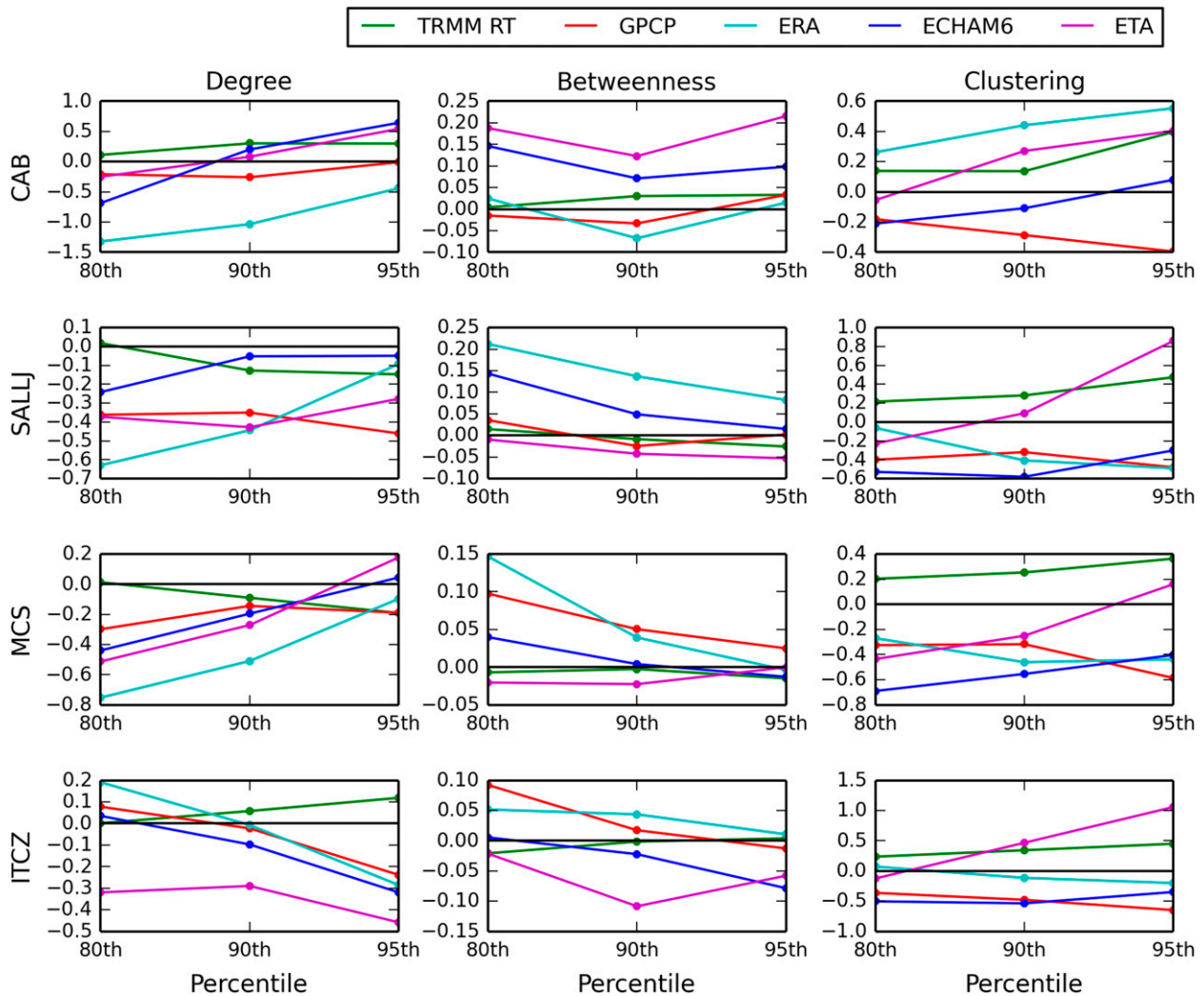


FIG. 15. Differences of network measures degree, betweenness centrality, and clustering to the TRMM reference dataset for the spatial boxes CAB, SALLJ, MCS, and ITCZ (see Fig. 1 for definitions of these geographical regions). Values are shown for rainfall events above the 80th, 90th, and 95th percentiles.

computed from networks derived from events above the 80th, 90th, and 95th percentiles, respectively. As above, we shall consider TRMM as reference and discuss deviations from the values found for TRMM for the remaining five datasets.

(i) CAB

Because of its central role for the propagation of extreme rainfall over the entire continent, the CAB is expected to exhibit high DG and BC (Fig. 15, top). For TRMM RT and GPCP, the local connectivity (as expressed by DG) and the long-ranged centrality (as expressed by BC) of the CAB are reproduced well for 80th, 90th, and 95th percentile events. For ERA-Interim, DG is strongly suppressed for all three event thresholds, but BC is reproduced reasonably well. Both

ECHAM6 and Eta show too low DG for 80th percentile events and too high DG for 95th percentile events, while DG for 90th percentile events is accurate. BC of the latter two datasets is too high for all three event types.

The crucial role of the Amazon basin for the distribution and propagation of extreme rainfall is thus captured by all datasets except ERA-Interim, with the Eta and ECHAM6 models even overestimating the long-ranged centrality as expressed by BC.

(ii) SALLJ

Both the local connectivity (DG) and the long-range centrality (BC) of extreme events in the SALLJ region are accurately represented by the TRMM RT dataset. GPCP strongly underestimates DG for all three event types, possibly because of lack of station data at the

Andean slopes. Interestingly, however, BC for the GPCP data is in good agreement with TRMM, indicating that the long-range character of extreme-event propagation in this region is still captured well, despite the shortcomings in local connectivity. For the most extreme events, ERA-Interim performs well with respect to the SALLJ, but for 80th and 90th percentile events, DG is strongly underestimated. In contrast, BC is substantially higher than for TRMM for all three event thresholds. ECHAM6 shows too low DG and too high BC for events above the 80th percentile, but both measures are in good agreement with TRMM for higher percentiles. The Eta model strongly underestimates DG in the SALLJ region, and also BC is too low for all three percentiles.

Together with the spatial patterns of DG and BC described above, our results indicate that the influence of the SALLJ on the propagation of extreme rainfall and associated orographic lifting effects are not adequately represented by either the station-based GPCP or the model-derived ERA-Interim and Eta datasets. In particular, for the Eta model, this is somewhat surprising, since the different vertical coordinate scheme is expected to improve the representation of rainfall over complex and steeply ascending terrain. As noted above, however, this only yields a better representation of mean daily rainfall when compared to ERA-Interim and ECHAM6, while the tail of the distribution at the eastern Andean slopes is underestimated by Eta. In view of the shortcomings with respect to DG and BC, we suggest that Eta has problems not only with the magnitudes but also with the synchronization structure of these events.

(iii) MCS

As mentioned in [section 2](#), the region labeled MCS in [Fig. 1](#) is exposed to exceptionally large mesoscale convective systems. According to our interpretation of CL, this should lead to high values of this measure.

The TRMM RT dataset exhibits too high CL values over the MCS region for all three percentile thresholds. While the absolute values of CL are too low for GPCP, its spatial pattern described above ([Fig. 11](#)) does exhibit relatively high values over the MCS region. In contrast, ERA-Interim and ECHAM6 underestimate the clustering of events above all three percentile thresholds in this region. This suggests that these models fail to reproduce the clustering in the spatial synchronization structure caused by mesoscale convective systems in this area. The problems of global and regional models to implement the amplitude and spatiotemporal patterns of large thunderstorms in this region are well known ([Solman et al. 2008](#); [Chou et al. 2012](#); [Solman et al. 2013](#)). It should therefore be emphasized that, while it

underestimates CL for events above the 80th and 90th percentiles, the Eta dataset is the only one for which we observe CL values comparable to TRMM for the most extreme events (above the 95th percentile).

(iv) ITCZ

The ITCZ plays an important role as source of extreme rainfall on the South American continent, with emphasis on long-ranged propagation. We therefore expect it to exhibit high BC, as is found for TRMM.

For the TRMM RT data, BC over the ITCZ region is in excellent agreement with TRMM. GPCP overestimates BC for events above the 80th percentile but also yields accurate values for events above the 90th and 95th percentiles. ERA-Interim overestimates BC over the ITCZ for all three event thresholds, while ECHAM6 is accurate for events above the 80th and 90th percentiles. The Eta model strongly underestimates the ITCZ's role for the long-range connectivity of all three event thresholds.

6. Conclusions

The remote sensing-derived, gauge-calibrated TRMM 3B42 V7 has been previously found to provide a reasonable dataset for delineating South American rainfall ([Carvalho et al. 2012](#)), in particular with respect to the spatiotemporal patterns of extreme events ([Boers et al. 2013](#)). In our study, we used TRMM 3B42 V7 as a reference dataset, to which we compare the real-time satellite product TRMM 3B42 V7 RT; the remote sensing and station-based product GPCP 1DD V1.2; the model-derived reanalysis product ERA-Interim; and data output from the global circulation model ECHAM6 and the regional climate model Eta, which is driven by ERA-Interim. In our analysis, we perform two comparisons: First, we evaluate dataset differences with a classical statistical approach. Second, we rely on complex network theory to analyze spatial patterns of extreme-event co-variability in the six datasets.

In the first part, we discussed classical statistical values. Overall best agreement with TRMM is found for the TRMM RT and the GPCP product. This is little surprising, because GPCP partly uses the same calibration scheme as TRMM. Most notably in this context, while mean daily rainfall values agree well with some discrepancies over the Amazon basin and the South Atlantic convergence zone, all three model-derived datasets underestimate the scores of the 90th and 95th percentiles when compared to TRMM 3B42 V7. Depending on the geographic area and the specific dataset, percentile scores remain in good agreement with TRMM up to the 65th to 85th percentiles. Above these percentiles, ERA-Interim, the Eta, and also the ECHAM6 model (in the South

Atlantic convergence zone) underestimate the frequency of extreme rainfall events. The right-hand tail of the daily rainfall distribution, representing pronounced low-frequency but high-magnitude events, is thus not reproduced well by any of the model-derived datasets.

In the second part, we have applied a recently introduced methodology based on complex network theory to analyze the spatial characteristics of extreme rainfall synchronicity. More traditional methods based on eigenvalue techniques (e.g., PCA) are not applicable for this purpose because of the binary structure of extreme-event time series but also because of general problems of PCA-like techniques with non-Gaussian data distributions.

We have used several complex network measures in order to quantify different aspects of this internal spatial structure of extreme rainfall synchronicity and, indeed, found that they reveal important features of the South American monsoon system. In particular, we have employed two additional complex network measures, which on the one hand support earlier results (degree) but also add new insights on climatic linkages between different locations (regional connectivity).

Our analysis suggests that TRMM RT's and GPCP's representations of these large-scale propagation patterns are the ones closest to the patterns found by TRMM. We observe that ERA-Interim data do not correctly reproduce key features of the South American monsoon system from the perspective of local synchronizations (degree). These include the effects of deep convection over the Amazon basin or the orographic barrier of the eastern Andes and their role for large-scale moisture transport. However, the model-derived ERA-Interim dataset incorporates these features accurately when taking into account long-ranged connections over several steps in the network (betweenness centrality). Still, the global model ECHAM6 and the regional model Eta perform better in reproducing these features. The propagation pathway from the western Amazon toward the subtropics along the eastern slopes of the Andes, which is associated with the South American low-level jet, is well represented by the ECHAM6 model, while the Eta model fails to produce this feature.

Concerning large convective systems, in contrast to TRMM and TRMM RT, ERA-Interim and ECHAM6 fail to reproduce any spatially organized rainfall clusters in regions where they are known to frequently occur, such as in southeastern South America. Eta exhibits some spatial patterns possibly corresponding to large organized systems but also not in this specific area where these would be most relevant. While the overall spatial organization for GPCP differs considerably from TRMM, mesoscale convective systems in southeastern South America are detected reasonably well.

Linkages of synchronous extreme rainfall from the central Amazon basin to the eastern slopes of the Andes observed for TRMM and TRMM RT are only resembled by the ECHAM6 model but not by the ERA-Interim, Eta, and GPCP datasets.

Regarding teleconnections from two regions important for the aforementioned rainfall dipole (southeastern South America and the South Atlantic convergence zone), all six datasets coherently find a rather localized pattern with little connections to other regions. However, a linkage from southeastern South America to the eastern slopes of the southern central Andes is only present for the TRMM, TRMM RT, and GPCP datasets.

Acknowledgments. This paper was developed within the scope of the IRTG 1740/TRP 2011/50151-0, funded by the DFG/FAPESP. J.M. was supported by the FAPESP Project Go Amazon, 2013/50538-7. J.K. acknowledges financial support from the Government of the Russian Federation (Agreement 14.Z50.31.0033).

Most Computations were performed with the IBM iDataPlex Cluster at the Potsdam Institute for Climate Impact Research. The ECHAM6 simulation was performed at the DKRZ (German Climate Computer Center) within the German consortium project STORM.

The authors thank Caroline Mourão for providing access to the Eta model data and Henrique Barbosa for the valuable comments and discussion.

REFERENCES

- Anabor, V., D. J. Stensrud, and O. L. L. de Moraes, 2008: Serial upstream-propagating mesoscale convective system events over southeastern South America. *Mon. Wea. Rev.*, **136**, 3087–3105, doi:10.1175/2007MWR2334.1.
- Boers, N., B. Bookhagen, N. Marwan, J. Kurths, and J. Marengo, 2013: Complex networks identify spatial patterns of extreme rainfall events of the South American monsoon system. *Geophys. Res. Lett.*, **40**, 4386–4392, doi:10.1002/grl.50681.
- , —, H. M. J. Barbosa, N. Marwan, J. Kurths, and J. A. Marengo, 2014a: Prediction of extreme floods in the eastern central Andes based on a complex network approach. *Nat. Commun.*, **5**, 5199, doi:10.1038/ncomms6199.
- , R. V. Donner, B. Bookhagen, and J. Kurths, 2014b: Complex network analysis helps to identify impacts of the El Niño Southern Oscillation on moisture divergence in South America. *Climate Dyn.*, doi:10.1007/s00382-014-2265-7, in press.
- Bookhagen, B., and M. R. Strecker, 2008: Orographic barriers, high-resolution TRMM rainfall, and relief variations along the eastern Andes. *Geophys. Res. Lett.*, **35**, L06403, doi:10.1029/2007GL032011.
- Carvalho, L. M. V., C. Jones, and B. Liebmann, 2002: Extreme precipitation events in southeastern South America and large-scale convective patterns in the South Atlantic convergence zone. *J. Climate*, **15**, 2377–2394, doi:10.1175/1520-0442(2002)015<2377:EPEISS>2.0.CO;2.
- , —, and —, 2004: The South Atlantic convergence zone: Intensity, form, persistence, and relationships with intraseasonal

- to interannual activity and extreme rainfall. *J. Climate*, **17**, 88–108, doi:10.1175/1520-0442(2004)017<0088:TSACZI>2.0.CO;2.
- , —, A. N. D. Posadas, R. Quiroz, B. Bookhagen, and B. Liebmann, 2012: Precipitation characteristics of the South American monsoon system derived from multiple datasets. *J. Climate*, **25**, 4600–4620, doi:10.1175/JCLI-D-11-00335.1.
- Chen, S., and Coauthors, 2013: Evaluation of the successive V6 and V7 TRMM multisatellite precipitation analysis over the continental United States. *Water Resour. Res.*, **49**, 8174–8186, doi:10.1002/2012WR012795.
- Chou, S. C., and Coauthors, 2012: Downscaling of South America present climate driven by 4-member HadCM3 runs. *Climate Dyn.*, **38**, 635–653, doi:10.1007/s00382-011-1002-8.
- Cohen, J., M. S. Dias, and C. Nobre, 1995: Environmental conditions associated with Amazonian squall lines: A case study. *Mon. Wea. Rev.*, **123**, 3163–3174, doi:10.1175/1520-0493(1995)123<3163:ECAWAS>2.0.CO;2.
- Coppus, R., and A. C. Imeson, 2002: Extreme events controlling erosion and sediment transport in a semi-arid sub-Andean valley. *Earth Surf. Processes Landforms*, **27**, 1365–1375, doi:10.1002/esp.435.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, doi:10.1002/qj.828.
- Donges, J. F., Y. Zou, N. Marwan, and J. Kurths, 2009: The backbone of the climate network. *Europhys. Lett.*, **87**, 48007, doi:10.1209/0295-5075/87/48007.
- Durkee, J. D., and T. L. Mote, 2010: A climatology of warm-season mesoscale convective complexes in subtropical South America. *Int. J. Climatol.*, **30**, 418–431, doi:10.1002/joc.1893.
- , —, and J. M. Shepherd, 2009: The contribution of mesoscale convective complexes to rainfall across subtropical South America. *J. Climate*, **22**, 4590–4605, doi:10.1175/2009JCLI2858.1.
- Eltahir, E. B., and R. L. Bras, 1994: Precipitation recycling in the Amazon basin. *Quart. J. Roy. Meteor. Soc.*, **120**, 861–880, doi:10.1002/qj.49712051806.
- Garreaud, R., 2000: Intraseasonal variability of moisture and rainfall over the South American altiplano. *Mon. Wea. Rev.*, **128**, 3337–3346, doi:10.1175/1520-0493(2000)128<3337:IVOMAR>2.0.CO;2.
- Hertwig, E., J.-S. von Storch, D. Handorf, K. Dethloff, I. Fast, and T. Krismer, 2014: Effect of horizontal resolution on ECHAM6-AMIP performance. *Climate Dyn.*, doi:10.1007/s00382-014-2396-x, in press.
- Huffman, G. J., R. F. Adler, M. M. Morrissey, D. T. Bolvin, S. Curtis, R. Joyce, B. McGavock, and J. Susskind, 2001: Global precipitation at one-degree daily resolution from multisatellite observations. *J. Hydrometeorol.*, **2**, 36–50, doi:10.1175/1525-7541(2001)002.0036:GPAODD.2.0.CO;2.
- , and Coauthors, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.*, **8**, 38–55, doi:10.1175/JHM560.1.
- Liebmann, B., G. N. Kiladis, C. S. Vera, A. C. Saulo, and L. M. V. Carvalho, 2004: Subseasonal variations of rainfall in South America in the vicinity of the low-level jet east of the Andes and comparison to those in the South Atlantic convergence zone. *J. Climate*, **17**, 3829–3842, doi:10.1175/1520-0442(2004)017<3829:SVORIS>2.0.CO;2.
- Maddox, R. A., 1980: Mesoscale convective complexes. *Bull. Amer. Meteor. Soc.*, **61**, 1374–1387, doi:10.1175/1520-0477(1980)061<1374:MCC>2.0.CO;2.
- Malik, N., B. Bookhagen, N. Marwan, and J. Kurths, 2012: Analysis of spatial and temporal extreme monsoonal rainfall over South Asia using complex networks. *Climate Dyn.*, **39**, 971–987, doi:10.1007/s00382-011-1156-4.
- Marengo, J. A., 2006: On the hydrological cycle of the Amazon basin: A historical review and current state-of-the-art. *Rev. Bras. Meteorol.*, **21**, 1–19.
- , W. R. Soares, C. Saulo, and M. Nicolini, 2004: Climatology of the low-level jet east of the Andes as derived from the NCEP–NCAR reanalyses: Characteristics and temporal variability. *J. Climate*, **17**, 2261–2280, doi:10.1175/1520-0442(2004)017<2261:COTLJE>2.0.CO;2.
- , and Coauthors, 2012: Recent developments on the South American monsoon system. *Int. J. Climatol.*, **32**, 1–21, doi:10.1002/joc.2254.
- , M. Valverde, and G. Obregon, 2013a: Observed and projected changes in rainfall extremes in the metropolitan area of São Paulo. *Climate Res.*, **57**, 61–72, doi:10.3354/cr01160.
- , and Coauthors, 2013b: Simulation of rainfall anomalies leading to the 2005 drought in Amazonia using the CLARIS LPB regional climate models. *Climate Dyn.*, **41**, 2937–2955, doi:10.1007/s00382-013-1919-1.
- Matsuyama, H., J. A. Marengo, G. O. Obregon, and C. A. Nobre, 2002: Spatial and temporal variabilities of rainfall in tropical South America as derived from Climate Prediction Center Merged Analysis of Precipitation. *Int. J. Climatol.*, **22**, 175–195, doi:10.1002/joc.724.
- Mesinger, F., 1984: A blocking technique for representation of mountains in atmospheric models. *Riv. Meteor. Aeronaut.*, **44**, 195–202.
- , and T. L. Black, 1992: On the impact on forecast accuracy of the step-mountain (eta) vs. sigma coordinate. *Meteor. Atmos. Phys.*, **50**, 47–60, doi:10.1007/BF01025504.
- , and Coauthors, 2012: An upgraded version of the Eta model. *Meteor. Atmos. Phys.*, **116**, 63–79, doi:10.1007/s00703-012-0182-z.
- Moreiras, S. M., 2005: Climatic effect of ENSO associated with landslide occurrence in the central Andes, Mendoza Province, Argentina. *Landslides*, **2**, 53–59, doi:10.1007/s10346-005-0046-4.
- Negrón Juárez, R. I., W. Li, R. Fu, K. Fernandes, and A. de Oliveira Cardoso, 2009: Comparison of precipitation datasets over the tropical South American and African continents. *J. Hydrometeorol.*, **10**, 289–299, doi:10.1175/2008JHM1023.1.
- O’Hare, G., and S. Rivas, 2005: The landslide hazard and human vulnerability in La Paz City, Bolivia. *Geogr. J.*, **171**, 239–258, doi:10.1111/j.1475-4959.2005.00163.x.
- Paegle, J. N., and K. C. Mo, 2002: Linkages between summer rainfall variability over South America and sea surface temperature anomalies. *J. Climate*, **15**, 1389–1407, doi:10.1175/1520-0442(2002)015<1389:LBSRVO>2.0.CO;2.
- Pesquero, J. F., S. C. Chou, C. A. Nobre, and J. A. Marengo, 2010: Climate downscaling over South America for 1961–1970 using the Eta model. *Theor. Appl. Climatol.*, **99**, 75–93, doi:10.1007/s00704-009-0123-z.
- Quiroga, R. Q., T. Kreuz, and P. Grassberger, 2002: Event synchronization: A simple and fast method to measure synchronicity and time delay patterns. *Phys. Rev.*, **66E**, 041904, doi:10.1103/PhysRevE.66.041904.
- Rheinwalt, A., N. Marwan, J. Kurths, P. Werner, and F.-W. Gerstengarbe, 2012: Boundary effects in network measures of spatially embedded networks. *Europhys. Lett.*, **100**, 28002, doi:10.1209/0295-5075/100/28002.

- Salio, P., M. Nicolini, and E. J. Zipser, 2007: Mesoscale convective systems over southeastern South America and their relationship with the South American low-level jet. *Mon. Wea. Rev.*, **135**, 1290–1309, doi:10.1175/MWR3305.1.
- Schuster, R. L., D. A. Salcedo, and L. Valenzuela, 2002: Overview of catastrophic landslides of South America in the twentieth century. *Catastrophic Landslides: Effects, Occurrence, and Mechanisms*, S. G. Evans and J. V. DeGraff, Eds., Reviews in Engineering Geology, Vol. 15, Geological Society of America, 1–33.
- Silva, V. B. S., V. E. Kousky, and R. W. Higgins, 2011: Daily precipitation statistics for South America: An intercomparison between NCEP reanalyses and observations. *J. Hydrometeorol.*, **12**, 101–117, doi:10.1175/2010JHM1303.1.
- Solman, S. A., M. N. Nuñez, and M. F. Cabré, 2008: Regional climate change experiments over southern South America. I: Present climate. *Climate Dyn.*, **30**, 533–552, doi:10.1007/s00382-007-0304-3.
- , and Coauthors, 2013: Evaluation of an ensemble of regional climate model simulations over South America driven by the ERA-Interim reanalysis: Model performance and uncertainties. *Climate Dyn.*, **41**, 1139–1157, doi:10.1007/s00382-013-1667-2.
- Stevens, B., and Coauthors, 2013: Atmospheric component of the MPI-M Earth system model: ECHAM6. *J. Adv. Model. Earth Syst.*, **5**, 146–172, doi:10.1002/jame.20015.
- Stolbova, V., P. Martin, B. Bookhagen, N. Marwan, and J. Kurths, 2014: Topology and seasonal evolution of the network of extreme precipitation over the Indian subcontinent and Sri Lanka. *Nonlinear Processes Geophys.*, **21**, 901–917, doi:10.5194/npg-21-901-2014.
- Tsonis, A., K. Swanson, and S. Kravtsov, 2007: A new dynamical mechanism for major climate shifts. *Geophys. Res. Lett.*, **34**, L13705, doi:10.1029/2007GL030288.
- Urrutia, R., and M. Vuille, 2009: Climate change projections for the tropical Andes using a regional climate model: Temperature and precipitation simulations for the end of the 21st century. *J. Geophys. Res.*, **114**, D02108, doi:10.1029/2008JD011021.
- Vera, C., and Coauthors, 2006: Toward a unified view of the American monsoon systems. *J. Climate*, **19**, 4977–5000, doi:10.1175/JCLI3896.1.
- Xue, X., Y. Hong, A. S. Limaye, J. J. Gourley, G. J. Huffman, S. I. Khan, C. Dorji, and S. Chen, 2013: Statistical and hydrological evaluation of TRMM-based Multi-satellite Precipitation Analysis over the Wangchu basin of Bhutan: Are the latest satellite precipitation products 3B42V7 ready for use in ungauged basins? *J. Hydrol.*, **499**, 91–99, doi:10.1016/j.jhydrol.2013.06.042.
- Yamasaki, K., A. Gozolchiani, and S. Havlin, 2008: Climate networks around the globe are significantly affected by El Niño. *Phys. Rev. Lett.*, **100**, 228501, doi:10.1103/PhysRevLett.100.228501.
- Yin, L., R. Fu, E. Shevliakova, and R. E. Dickinson, 2013: How well can CMIP5 simulate precipitation and its controlling processes over tropical South America? *Climate Dyn.*, **41**, 3127–3143, doi:10.1007/s00382-012-1582-y.
- Zhou, J., K. M. Lau, J. O. F. Climate, and G. Space, 1998: Does a monsoon climate exist over South America? *J. Climate*, **11**, 1020–1040, doi:10.1175/1520-0442(1998)011<1020:DAMCEO>2.0.CO;2.
- Zipser, E. J., C. Liu, D. J. Cecil, S. W. Nesbitt, and D. P. Yorty, 2006: Where are the most intense thunderstorms on Earth? *Bull. Amer. Meteor. Soc.*, **87**, 1057–1071, doi:10.1175/BAMS-87-8-1057.
- Zulkafli, Z., W. Buytaert, C. Onof, B. Manz, E. Tarnavsky, W. Lavado, and J.-L. Guyot, 2014: A comparative performance analysis of TRMM 3B42 (TMPA) versions 6 and 7 for hydrological applications over Andean–Amazon River basins. *J. Hydrometeorol.*, **15**, 581–592, doi:10.1175/JHM-D-13-094.1.