

ARTICLE

OPEN

Tropical rainfall subseasonal-to-seasonal predictability types

Vincent Moron^{1,2*} and Andrew W. Robertson²

Tropical rainfall is mostly convective and its subseasonal-to-seasonal (S2S) prediction remains challenging. We show that state-of-art model forecast skill 3 + 4 weeks ahead is systematically lower over land than ocean, which is matched by a similar land-ocean contrast in the spatial scales of observed biweekly rainfall anomalies. Regional differences in predictability are then interpreted using observed characteristics of daily rainfall (wet-patch size, mean intensity as well as the strength of local S2S modes of rainfall variation), and classified into six S2S predictability types. Both forecast skill and spatial scales are reduced over the continents, either because daily rainfall patches are small and poorly organized by S2S modes of variation (as over equatorial and northern tropical Africa), or where the daily mean intensity is very high (as over South and SE Asia). Forecast skill and spatial scales are largest where daily rainfall is synchronized by intraseasonal (such as the Madden-Julian Oscillation) as well as interannual ocean-atmosphere modes of variation (such as El Niño-Southern Oscillation), especially over northern Australia and parts of the Maritime Continent, and over parts of eastern, southern Africa and northeast South America. The oceans exhibit the highest skill and largest spatial scales, especially where interannual (central equatorial Pacific) or intraseasonal (central and eastern Tropical Indian Ocean and Western Pacific) variability is largest. These results provide a relevant regional typology of the potential drivers and controls on S2S predictability of tropical rainfall, informing intrinsic limits and possible improvements toward useful S2S climate prediction at regional scale.

npj Climate and Atmospheric Science (2020)3:4; <https://doi.org/10.1038/s41612-020-0107-3>

INTRODUCTION

Tropical rainfall variability strongly impacts local societies, especially through the annual rhythm of the monsoons with distinct wet and dry seasons, which often undergo pronounced interannual and intraseasonal fluctuations.^{1,2} A major goal of current international research efforts is to improve forecasts and understanding on the subseasonal to seasonal (S2S) range, 2 weeks to a season ahead, filling the gap between medium range weather forecasts and seasonal forecasts.^{3–7} This timescale includes the meridional seasonal migration of the intertropical convergence zone (ITCZ) and the monsoons, together with the primary modes of tropical S2S variability, especially the intraseasonal Madden Julian Oscillation (MJO^{8–10}), and El Niño-Southern Oscillation (ENSO^{11,12}) on the seasonal to interannual scales.^{11–14} However, S2S prediction of tropical rainfall over land remains a major challenge, and the geographical distribution of its predictability and the factors that limit it are poorly understood.^{6,7,15,16} In particular, the S2S scale encompasses interactions between intrinsic monsoon dynamics and the MJO leading to monsoon intraseasonal oscillations typified by active and break phases and changes in the onset date.^{17–19}

The famous Koppen-Geiger classification of local mean climates was developed over 100 years ago²⁰ and is still in widespread use in atlases and beyond.²¹ However, an analogous typology for rainfall variability and predictability is still lacking. The goal of this paper is to make a relevant geographical classification of tropical rainfall predictability and its controlling factors from both S2S forecast models and observed data, thereby identifying the countries and regions that could benefit most from S2S prediction through development of effective weather and climate services.²²

Climate prediction is a signal-to-noise (SN) problem where the signal (noise) is the predictable (unpredictable) fraction of the climate variation at a given timescale.^{23–25} S2S predictions of weekly or biweekly average rainfall^{5,6} are possible where the

signal lent by predictable modes of climate (especially MJO and ENSO) exceeds the unpredictable noise, largely associated with sub-weekly weather,^{26–29} provided that the forecast model can capture the predictable signals. This competition between signal and noise can be estimated from S2S forecast model outputs alone, and by confronting the model output with observations over a reforecast period (model skill). On the other hand, S2S signals and noise can also be assessed empirically, purely from observed rainfall data. Larger spatial scales of biweekly average observed rainfall anomalies imply larger S2S predictability because the predictable climate drivers (MJO and ENSO) are large scale and persistent at this timescale, while the unpredictable weather noise in tropical convective rainfall is small scale and random in time and space. In this way, the spatial scales of rainfall time averages can empirically measure the upper bound of S2S predictability at local or gridbox scale from observed data alone.^{30,31}

Determining the potential for S2S prediction at local to regional scales requires mapping both signals and noise, as well as their underlying physical controls. We combine both model-based and observations-based estimates and focus here on biweekly rainfall amounts at a 2–4 week forecast lead time (15–28 days ahead, also referred to as forecast weeks 3 + 4). We first assess skill and ensemble spread from a 20-year ensemble of 11 runs of ECMWF model reforecasts. These model-based estimates of predictability are then interpreted geographically by clustering various empirical estimates of signal and noise, based on observed rainfall data. S2S predictable signals are defined empirically using three measures: the fraction of biweekly rainfall variance in the 7–20 day and 20–90 day intraseasonal, and interannual frequency bands. The noise is expressed empirically using four measures: the average area or “patch-size” of daily rainfall events, the mean daily rainfall intensity, spatial autocorrelation of biweekly rainfall anomalies, and its subweekly variance fraction.

¹Aix Marseille University, CNRS, IRD, INRAE, Coll. de France, CEREGE, Aix en Provence, France. ²IRI, Columbia University, Palisades, NY, USA. *email: moron@cerege.fr

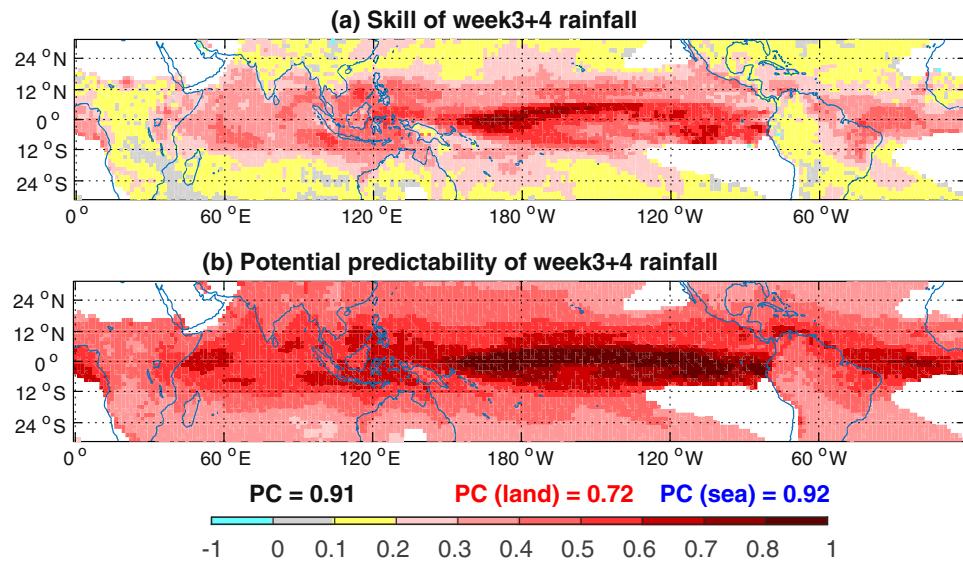


Fig. 1 **a** ECMWF skill of week 3 + 4 amounts of rainfall. The temporal units receiving <1 mm/day in mean over the 1998–2017 period (across the whole set of 11 runs) are not considered. The ensemble mean rainfall is computed and then normalized versus the mean annual cycle and the anomalies are lastly converted into ranks. The skill is computed as the Spearman correlation with the ranks of ECMWF reforecasts vs GPCP-1DD over the same time units considering the available time periods (and after having linearly interpolated the GPCP-1DD rainfall onto the 1.5° ECMWF grid). **b** ECMWF model potential predictability of week 3 + 4 amounts of rainfall. The potential predictability is defined as the Spearman correlation with the ranks of each run of ECMWF reforecasts vs the ensemble mean. The values below the map indicate the pattern correlations (PC) between the skill and the potential predictability for the whole domain (black), the landmasses only (red) and the oceans only (blue). In **a**, **b** the blank areas never reach a climatological daily mean of 1 mm/day and are thus not considered in the computations.

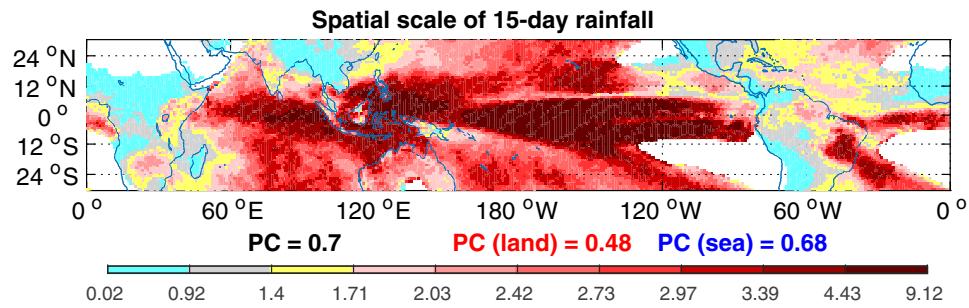


Fig. 2 **Spatial scale (in 10^6 sq. kilometers) of the running 15-day amount of observed rainfall from GPCP-1DD rainfall.** The spatial scales are defined as the area where the Spearman correlations are over $1/e$ ($=0.3679$) with the target grid-point. The 15-day windows when the climatological mean is below 1 mm/day are not considered. The colors are defined with the deciles of the spatial scales. The values below the map indicate the pattern correlations between the skill (Fig. 1a) and the spatial scale for the whole domain (black), the landmasses only (red) and the oceans only (blue).

RESULTS

Model skill and predictability and observed spatial scales both peak over the equatorial oceans

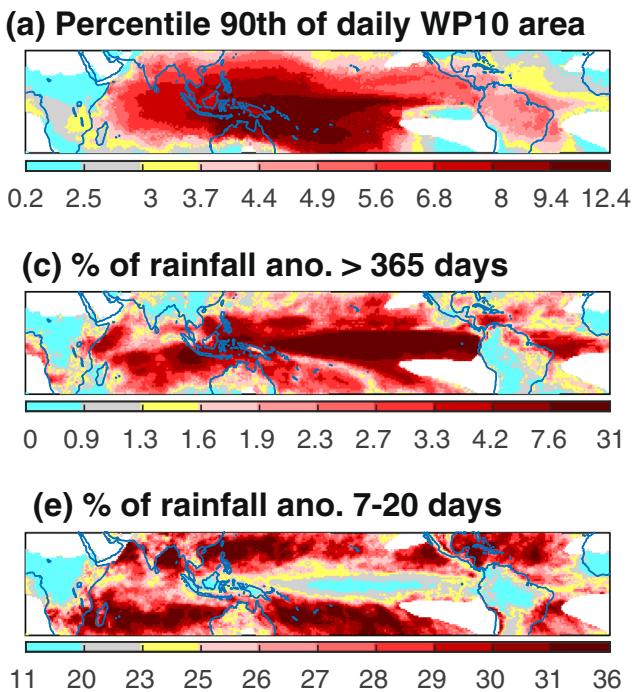
ECMWF model forecast skill—defined here by the Spearman correlation between the ensemble mean of the 11 runs and the observations—and potential predictability³² (PP)—defined here by the averaged Spearman correlation between each run vs the ensemble mean—for week 3 + 4 rainfall are both highest over the equatorial Pacific ENSO core region, and are smaller over land than over the ocean^{5,6} (Fig. 1). The pattern correlation (PC hereafter) between both maps equals 0.91. Over land, a relatively large PP is not always associated with high skill, such as over western India or Southern Africa, which may reflect poor simulation of predictable regional circulation patterns or a relatively large amplitude of internal atmospheric variations due to chaotic dynamics.^{33,34}

There is a striking similarity between the geographical pattern of model skill (Fig. 1a) and spatial scales (Fig. 2) since the largest spatial co-variations in skill and spatial scale ($PC = 0.7$) are

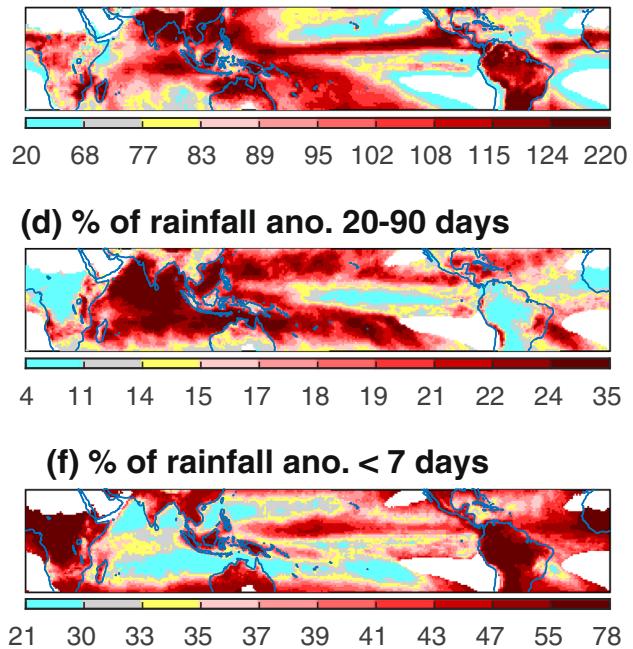
between land (relatively low skill and spatial scale) and ocean (relatively high skill and spatial scale). Spatial scales are significantly smaller across the landmasses, especially over northern and equatorial Africa, southern Asia between the Gangetic plain and Indochina, western South America and most of central America. Scales are larger over the equatorial Pacific and around the Maritime Continent. Pockets of larger spatial scales over eastern South America or eastern Africa (Fig. 2) usually coincide with higher skill in Fig. 1a. These results demonstrate that observed spatial scales of biweekly rainfall variability are usually larger where model forecast skill (and PP) are large and vice versa, with oceanic rainfall variations being more predictable and larger scale, than over land.

What explains these skill/scale geographical contrasts?

The basic unit of instantaneous convective rainfall is the convective cell $O(10\text{ km})$, often embedded in meso-scale convective clusters, tropical depressions and/or cyclones^{1,2,35–37}. These are sometimes accompanied or followed by more extensive



(b) Mean intensity of daily rainfall



(d) % of rainfall ano. 20-90 days

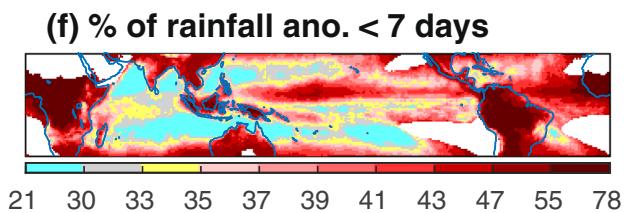
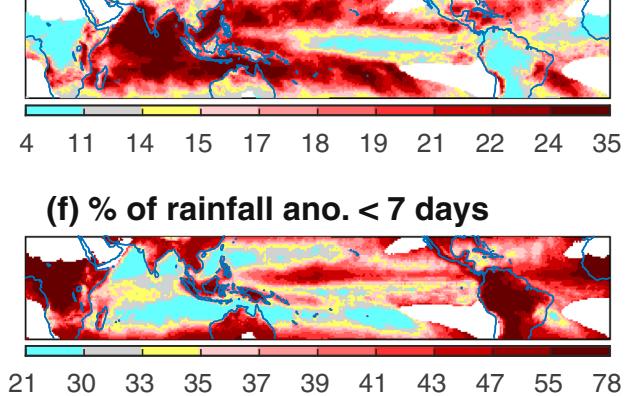


Fig. 3 Spatial variations of factors driving partly the skill/scales of rainfall amounts. The colors are defined with the 10th-90th percentiles of each variable. **a** 90th percentile of the area (in 10^6 sq. kilometers) of the daily wet patches receiving at least 10 mm. Wet patches are defined from the contiguous (by gridbox edge or corner) grid-points. The areas are attributed to the grid-points belonging to them. **b** Mean daily intensity in $1/10^6$ mm/day. The mean intensity is computed as the total amount of rainfall (over 1996-2017) divided by the total number of wet days ≥ 1 mm. **c-f** Percentage of the total variance of rainfall anomalies in 4 frequency bands. The daily rainfall amounts are firstly square-rooted to reduce the skewness. The daily anomalies are then computed as the difference between the square-rooted daily rainfall and the climatological daily mean of the square-rooted amounts giving more weight to the anomalies during the wettest periods of the year. The low-pass variations >365 days define the interannual variance in **c**. The band-pass variations between 20 and 90 days and between 7 and 20 days are shown in **d**, **e** respectively and the high-pass variations <7 days are shown in **f**. The same mask is used as in Fig. 2.

stratiform rains.^{38,39} A useful diagnostic of biweekly rainfall scales may thus be the size of daily wet patches (WP), given by the contiguous area receiving at least N mm of rain. Daily aggregates of rainfall will increase the patch-size somewhat through the integration and movement of rainfall within a day, and a threshold of $N=10$ mm (WP10) was chosen empirically from the daily rainfall data (Supplementary Fig. 1). A larger daily WP10 may be expected to lead to larger spatial scales of aggregated rainfall amounts and vice versa. A second factor relevant to biweekly rainfall anomaly spatial scale and predictability is the mean rainfall intensity. Since heavy convective rainfall events tend to be highly localized, a large mean intensity associated with such events will tend to be associated with smaller spatial scales of S2S rainfall anomalies, and potentially lower prediction skill.^{31,40,41} Finally, the essential source of S2S rainfall predictability is the impact of S2S phenomena, that modulate or “synchronize” small-scale instantaneous rainfall in space and time, such as the MJO^{10,42} and other convectively coupled equatorial waves,⁴³⁻⁴⁵ as well as the impact of slowly evolving surface boundary conditions including soil moisture and sea surface temperature (SST) anomalies, most prominently driven by ENSO.^{12,46}

Figure 3 shows the geographical distributions of these interrelated controls on S2S rainfall predictability, estimated from observed data (see Supplementary Table 1 for their PCs). The 90th percentile of daily WP10 shows a large maximum across the western and central equatorial Pacific, with very small values over Africa and the mean subsidence regions in the subtropics and eastern Pacific, linked to descending branches of the Hadley and Walker circulations (Fig. 3a). The spatial pattern of mean intensity (Fig. 3b) is similar to that of mean rainfall amount (not shown) but with an important difference: the mean intensity tends to peak over the landmasses, especially South and southeast Asia,

northern Australia and the Maritime Continent as well as most of South and Central America, while Tropical Africa is largely the exception, even if rather large values are observed over Madagascar and Mozambique, partly due to cyclonic events there. The synchronizing impact of S2S phenomena is broken down by timescale into interannual, and the 20–90 day and 7–20 day intraseasonal bands in Fig. 3c–e, in terms of the fraction of rainfall variance in each frequency band. Interannual variations (Fig. 3c) peak over the core ENSO region, and are generally larger over the oceans; smallest values occur over equatorial and northern Africa and western South America as well as the regions of mean subsidence. The intraseasonal variances (Fig. 3d, e) both exhibit large values over the subtropical oceans, likely associated with equatorial Rossby waves,^{43–45} while the 20–90 band also peaks over the equatorial Indian Ocean and western tropical Pacific MJO core region.^{8–10,42} Over land, the intraseasonal bands have moderate to strong impacts on South Asia, northern Australia, eastern and southern Africa, central America and eastern Brazil (Fig. 3d). Lastly, the <7 days variance provides a measure of weather noise which includes the impact of synoptic transient perturbations as well as the fastest CCEWs,^{44,45} and clearly shows a sea-land contrast with highest values over the landmasses (Fig. 3f). This is especially clear over Africa, South and SE Asia, the Maritime Continent and Australia. The values peak above 50 % of the total variance over equatorial and northern equatorial Africa and central and western South America.

A geographical classification of rainfall predictability characteristics

We next classify each land or ocean gridpoint into six geographical “S2S predictability types” according to the values

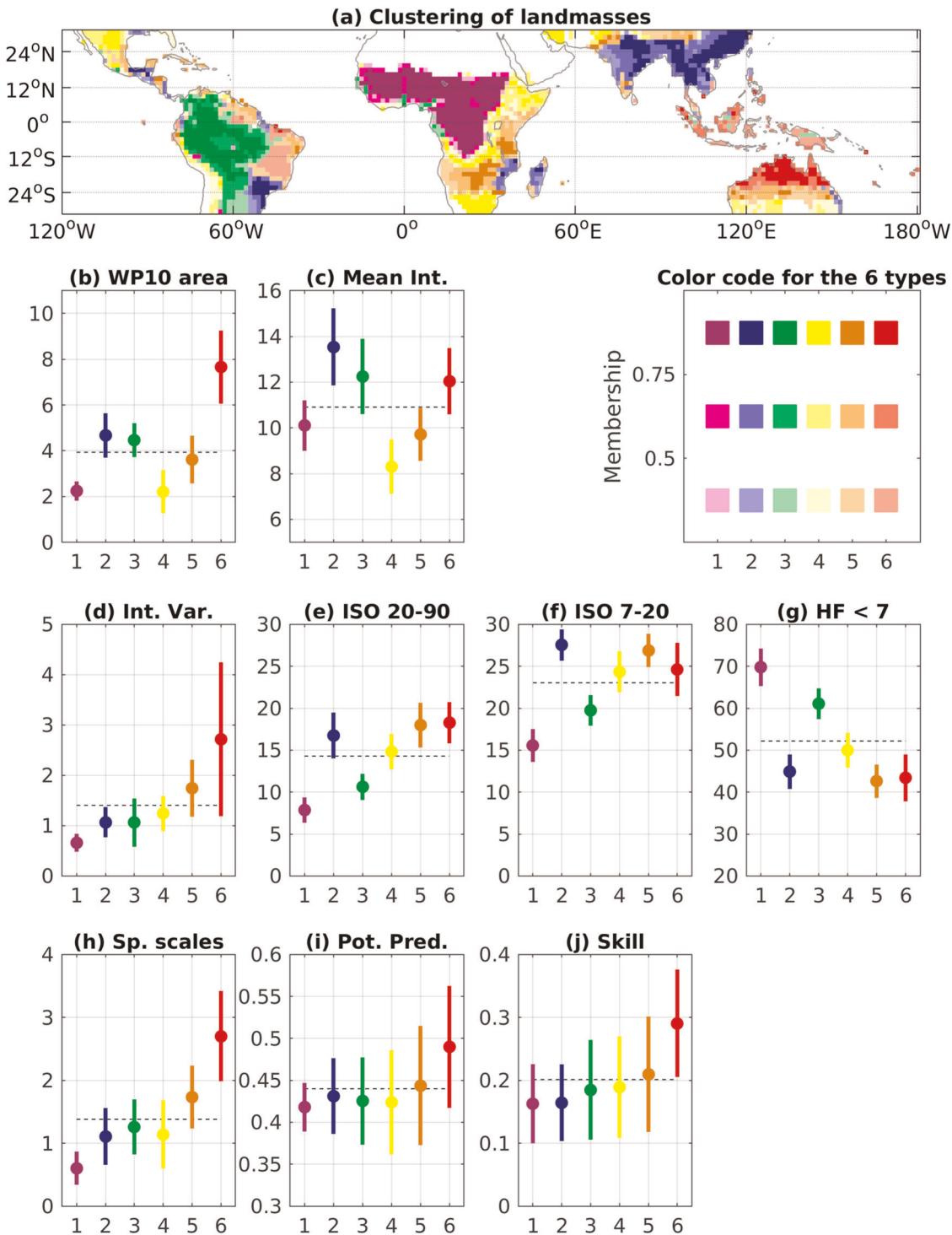


Fig. 4 **a** Clustering of land points according to the potential drivers of S2S rainfall predictability. The standardized anomalies of the variables shown in Figs. 2, 3 are clustered using a fuzzy k -means algorithm. Each grid-point is assigned to the cluster for which its membership is the highest. The lower panels **b–h** show the weighted average (filled circle) with ± 1 weighted standard deviation (vertical bar) of the variables shown in Figs. 2, 3, while the **i, j** show the associated ECMWF model potential predictability and skill shown in Fig. 1 for each cluster. In **b–j** the weights of the spatial average are membership and the cosine of the latitude of each grid-point belonging to a cluster, and the horizontal dashed black line is the weighted spatial average across all land points. The units are 10^6 km^2 in **b, h**, mm/day in **c**, percent of total temporal variations in **d–g** and dimensionless in **i, j**.

of its seven observed rainfall characteristics mapped in Figs. 2, 3. This classification was done using a fuzzy k -means cluster^{47,48} analysis, carried out for 2100 land and 6387 ocean points separately (Fig. 4 and Supplementary Fig. 2); no spatial location data was used. The six resulting predictability types (clusters of

gridpoints on Fig. 4a and Supplementary Fig. 2), derived only using observed rainfall characteristics, are ordered by the ECMWF model forecast skill at those gridpoints (Fig. 1a) which was not used in the cluster analysis. It is striking that a classification of locations based on observed quantities only, without any explicit

regression on various sources of skill (e.g., MJO time sequence or ENSO states) leads to a relevant geography of the model's skill (Fig. 1a). Note also that the clustering is rather robust versus the choice the included variables (Supplementary Figs. 3, 4). Removing either the fraction of interannual variance, or that of the fraction of subweekly variance, or considering only the three bands of interannual and both intra-seasonal bandwidths lead to a similar geography of the S2S predictability types.

Highest predictability over land (Type 6; red/pink) occurs over northern Australia, the Maritime Continent, areas of eastern and NE Brazil and few grid-points over coastal Ecuador and Galapagos. Interannual and 20–90 day ISO synchronization are largest for this type, highlighting the role of ENSO and the MJO, and spatial scales and daily wet-patch areas are both very large. Perhaps surprisingly, mean daily intensity is above the continental average, (horizontal dashed line), but this is counteracted by the larger spatial scales and S2S synchronization.

Land areas with moderate S2S skill are split into three clusters (Types 3–5). These extend over large parts of southern and eastern Africa, South and Central America, and subtropical Australia. The skill in Type 3 may be increased through relatively large WP10 while Types 4 and 5 exhibit relatively large synchronization at interannual and intraseasonal time scales.

Predictability types 1 and 2 have the lowest skill, located over large parts of west and central Africa, except coastal areas (Type 1), and most of South and SE Asia (except NW and southern India), SE Brazil and Uruguay and parts of central America, SE Africa and Madagascar (Type 2). These low predictability types are quite distinct: Type 2 has high ISO synchronization (especially 7–20 day), as well as slightly larger WP10 than the land average, all of which may increase predictability—but the mean intensity is very large, tending to decrease it. On the other hand, Type 1 (over Africa) lacks interannual or intraseasonal synchronization, and rainfall scales are very small. The impact of <7-day variance clearly peaks for this cluster.

The ocean rainfall predictability types (Supplementary Fig. 2) are organized around the central Pacific, where the skill is the highest, in almost concentric areas. Note the strong impact of 20–90-day band on cluster 3. The skill is lower across the Atlantic basin, but also over subtropical northern Pacific and western and southern Indian Ocean (Types 1 and 2), mostly in association with small WP10 and a poor synchronization at interannual time scale (Supplementary Fig. 2).

DISCUSSION

The empirically derived S2S rainfall predictability types over land demonstrate a one-to-one agreement between the spatial scales of observed biweekly rainfall anomalies and ECMWF model week 3 + 4 forecast skill. This is a remarkable agreement of empirical and model results. While models can be expected to improve in the future, this result underlines the expectation that S2S rainfall skill is fundamentally limited in several regions of the tropics, which either lack the predictable climate signals in the S2S band, or in which the inherent scales of rainfall are too small, or rainfall intensities are too large, to make biweekly averages usefully predictable. Most prominently, large parts of West and Central Africa and the Sahel (Type 1) lack S2S rainfall predictability, on average, due to the absence of strong intraseasonal and interannual dynamical controls, coupled with the small spatial scales of daily rainfall; the latter is consistent with the short duration of the convective (or MCS) events across this area⁴⁹ and the large amplitude of the diurnal cycle.⁵⁰ Note that the decadal time scale, which may be related with significant SST forcing across West Africa⁵¹ is not sampled enough by the period analyzed here from 1996.

In contrast, the other broad region of low S2S rainfall predictability (Type 2) over South and SE Asia does experience temporal synchronization of rainfall from both intraseasonal

bands, but the mean intensity (and thus noise) is very large.^{31,40} Note also that WP size and mean rainfall intensity are positively related (viz. Figs. 4b, c), consistent with the scaling properties of tropical rainfall, at least over the ocean.⁵² The impact of WP10 area on the signal-to-noise ratio thus represents a trade-off between the spatial integration (increasing the signal) and the maximum rainfall received near the center (increasing the noise). Improved and higher resolution models may improve skill here since dynamical sources of predictability and skill are present, as may the consideration of rainfall frequency as the predictand variable. The role of the mean intensity suggests that S2S predictability may decrease due to global warming where a general increase of rainfall intensity in wet areas/seasons is expected.^{53,54}

Regions of higher continental skill over SE Asia and Australia (Type 6), eastern and southern Africa (Types 4 and 5) and South America (Types 3 and 6) are mostly related to an increased amplitude of the temporal synchronization due to interannual and interannual bands, and where the relative influence of synoptic time scales <7 days is strongly reduced (over southern Africa). The locations of Types 5 and 6 across the landmasses (Fig. 4a) coincide rather well with areas where either seasonal predictions or impact of MJO have been established.^{9,13,23,29,55,56}

One caveat to the above classification is that the situation over the continents is complex and that the contribution of the analyzed drivers is only partial, as suggested by the intra-cluster spread in Fig. 4j. Our results pertain to the entire calendar year and predictability may be higher in specific parts of the calendar year, and especially during the early and late phases of the monsoon seasons including the onset and/or demise dates.^{57,58}

Over the oceans, the skill is larger and the contribution of the S2S climate drivers is stronger than over land. The predictability types are organized broadly in a symmetric way, around the tropical Indian and Pacific Ocean where the skill/scale peak. The large and strong dynamical control of interannual (central Pacific) and mostly 20–90-day (western Pacific and Indian ocean) increase the skill/scales there. As for Type 1 over the landmasses, Types 1 and 2 over the oceans (well extended over the tropical Atlantic and subtropical margins elsewhere) are penalized by the small size of the wet patches while the synchronization provided by interannual and 20–90-days is weak to neutral.

DATA AND METHODS

Daily rainfall data

Rainfall amount was extracted from the GPCP-1DD product (resolution $1^\circ \times 1^\circ$, daily, version 1.3) from 1 October 1996 to 31 December 2017. The GPCP dataset blends various satellite estimates and the daily amounts are calibrated with rain-gauges.^{59,60} The horizontal resolution already filters some of the noise since the minimal size of wet patches or observed spatial scales is roughly 11000 km^2 , which is far larger than the basic unit of convection ($\sim 10-100 \text{ km}^2$) and corresponds broadly to the lower bound of the meso-scale convective systems (MCS). Thus, we are unable to say if a wet cell is related to a single localized thunderstorm or an amalgamation provided by an MCS. The daily resolution is a limit of the analysis, but wet events usually last for more than one day.^{2,49,50} The calibration using rain-gauges, mostly located across the landmasses, may be a source of heterogeneity in the spatial variations of the characteristics defined using observed rainfall. But spatial variations of either WP10, spatial scales of biweekly rainfall anomalies, and fraction of bandwidth variations are very similar in uncalibrated Outgoing Longwave Radiation with the same resolution (not shown).

Daily ECMWF rainfall

Daily ECMWF model rainfall was obtained from an 11-member ensemble of S2S reforecasts for the 1998–2017 period (105 starting

dates each year). The ECMWF forecasting system is a merged 47-day ensemble system updated two times a week that has been run since 2002.¹⁵ The ECMWF runs are performed “on the fly”, with a 20-year reforecast set of runs made every time that a real-time forecast is issued. We use the model configuration, which was operational in 2018 (version CY43R3 for the forecast-issue dates from 1 January to 6 June and version CY45R1 for the starting dates from June 6 to December 31). The differences between both configurations are minor and are assumed to be negligible for the analyses performed in this study. Both versions use an atmospheric model at T62 (about 16 km) up to day 15 and T42 (about 32 km) horizontal resolution after day 15 with 91 vertical levels. The atmospheric model includes land surface interactions and is fully coupled to ocean and sea-ice models. The sea surface temperatures and continental surface characteristics are realistic at the starting dates. Daily rainfall accumulations are available on a 1.5° grid and the week 3 + 4 rainfall is computed from days 15 to 28. The skill is computed considering the same dates between the ensemble mean and GPCP-1DD rainfall after having linearly interpolated the GPCP-1DD rainfall onto the 1.5° grid. The potential predictability (PP) is defined as the mean Spearman correlation between anomalies of each run vs the ensemble mean. As for the GPCP-1DD rainfall, ranks are used instead of amount themselves and slots when the climatological mean is lower than 1 mm/day (from the climatology of the ensemble) are excluded from the computations.

Defining the spatial scales and wet patches areas

The daily rainfall data from GPCP-1DD are used to compute the spatial scales of 15-day rainfall anomalies (Fig. 2) and various characteristics of the rainfall variability (Fig. 3). The spatial scales are computed as the area where the correlations are above 1/e (~ 0.3679) around each grid-point.⁴⁹ The rainfall anomalies are computed as the differences between the running 15-day amounts and the climatological 15-day average. The periods when the climatological daily average (daily mean smoothed by a 1/90 c-p-d low-pass recursive filter) are not considered and the anomalies are converted into ranks to cancel the effect of very large positive anomalies. So, the spatial scales are estimated using Spearman’s instead of Pearson’s correlations. The wet patches are also defined from the GPCP-1DD fields. Wet patches (WP) are defined as the contiguous area (either by gridbox edge or corner) receiving at least 10 mm (WP10). A buffer zone of 10° is used to define the WPs on the 30°N–30°S domain. Considering a lower threshold of 1 or 5 mm leads to patchy, but very large, structures while higher thresholds usually limit the size of most of WP to few grid-points. A higher threshold tends to give very small samples for some areas. There are a total of ~ 730000 WP10 and more than 30% cover only a single grid-point. To compute the 90th percentiles of WP10, each WP10 area is attributed to each of the grid-points belonging to it, and the percentile is computed locally. In that context, a given WP10 is counted more than once as soon as it covers at least two contiguous grid-points.

Daily mean intensity

The mean daily intensity is simply computed as the total amount of rainfall across the whole set of days divided by the total frequency of wet days ≥ 1 mm. This characteristic mixes the instantaneous rain rate, that may depend mostly on the intrinsic nature (i.e. convective vs stratiform) of the rain-bearing clouds, and the intra-day duration of the wet event that may be mostly driven by the dynamical control of the rainy system (a longer duration is expected for MCS-TD system vs purely local thunderstorm).

Temporal decomposition of the rainfall variance

The fraction of variance belonging to four different frequency bands (Fig. 3) is computed from square-rooted daily rainfall.

Anomalies are computed from the climatological daily mean from these square-rooted daily rainfalls, and a recursive filter is used to extract the low-pass (>365 day), band-pass (20–90 day and 7–20 day) and high-pass (<7 days) variations. The 4 bands are expressed as their variance divided by the unfiltered variance. The sum of the 4 bands is usually close to 1.

Clustering of the rainfall drivers of the skill

Both a hard and fuzzy k -means cluster analysis^{47,48} are applied to the standardized anomalies of the 7 observed variables shown in Figs. 2, 3, separately for the 2100 (land) and 6387 (sea) grid-points. Dry areas which never reach a climatological daily mean of 1 mm/day are not considered in the computations. A pre-processing is done with a principal components analysis, retaining 90% of the variance in the four leading PCs. The goal of this clustering is to give a sound and interpretable regionalization (Fig. 4a and Supplementary Fig. 2a), explaining at least some of the spatial variation of the skill of Fig. 1a, 4j and Supplementary Fig. 2j). The hard k -means algorithm partitions the data into k clusters by iteratively maximizing the ratio of the sum of the between-group variance vs total variance, with the entire process repeated 1000 times from random initial seeds. The 2100 land grid points (and 6387 over ocean) are divided into k distinct clusters, where each gridpoint point can only belong to one cluster. A relatively large increase of this ratio occurs for $k = 3$ (land), $k = 6$ (land and ocean), and then for $k = 13$ (land) and $k = 9$ (ocean) (not shown). The latter solutions leads to far more complicated maps (not shown) than Fig. 4a and Supplementary Fig. 2a. A fuzzy k -means with $k = 6$ is then applied to the four principal components explaining 90% of the variance for land and ocean separately. When the value of the fuzzifier is >2 , all memberships tend to 1/6, and thus a meaningless clustering, while when it is <1.5 , it tends progressively to hard clusters, with all memberships tending to 1. The subjectively chosen value of 1.75 is thus relevant to distinguish core and marginal areas belonging to the six clusters.

The 7 variables included in the clustering are not independent to each other, e.g. the fraction of variability conveyed by the four complementary frequency bands are related by construction. However the pattern correlations between them (Supplementary Table 1) are mostly moderate, suggesting that they may thus contribute differently to the building of scales and skill of time-averaged rainfall anomalies. Nevertheless, the sensitivity of the clustering to the included variables should be tested. We present three alternative clusterings in Supplementary Fig. 3b–d for land and in Supplementary Fig. 4b–d for ocean grid points. The two first sensitivity tests omit the fraction of the variance conveyed by the interannual (Supplementary Figs. 3b, 4b) and high frequency (<7 days (Supplementary Figs. 3c, 4c) components respectively. Both resulting clusterings are very similar to those obtained with all seven variables (Supplementary Figs. 3a, 4a): without inter-annual variations (Supplementary Figs. 3b, 4b), 92% of grid-points over land and sea belong to the same cluster; without high frequency variations, 90% (79%) of grid-points over land (sea) (Supplementary Figs. 3c, 4c) belong to the same cluster. The order of the clusters from the point of view of spatially averaged skill used to rank the six types on Fig. 4a and Supplementary Fig. 2a sometimes switches between contiguous clusters (for example between clusters #1 and #2 or between clusters #3 and #4). The third sensitivity test is more stringent by considering only the 3 slower bands of variations in the clustering (Supplementary Figs. 3d, 4d). Even in this case, where the explicit measures of noise (high frequency <7 days and mean intensity) are excluded, the geographical pattern as well as the relative skill rankings (revealed by the colors from blue/purple for low skill to orange/red for high skill), are broadly similar. In fact, 58% (land) and 55% (sea) of the grid-points are in the same clusters as Fig. 4a and Supplementary Fig. 2a. Nevertheless, the map is slightly noisier and small pockets of different clusters within the regionally consistent ones appears,

such as over SE Asia and Central Amazonia. Overall, the sensitivity analysis results confirm that the clustering is rather robust and support including the seven variables as different facets of noise and signal (Supplementary Table 1). It is also possible that other relevant variables may be included, but is beyond the scope of this study.

DATA AVAILABILITY

The daily rainfall data from GPCCP 1DD are available free of charge (for example from <https://rda.ucar.edu/datasets/ds728.3/>). The daily rainfall from ECMWF S2S retrospective forecasts have been extracted from IRI library (<https://iridl.ideo.columbia.edu/SOURCES/ECMWF/S2S/ECMF/>).

CODE AVAILABILITY

The statistical analyses have been made using Matlab software and the corresponding codes that support the findings of this study are available upon request from the corresponding author.

Received: 18 July 2019; Accepted: 5 December 2019;

Published online: 30 January 2020

REFERENCES

- Moncrieff, M. W. The Multiscale Organization of Moist Convection and the Intersection of Weather and Climate. in *Climate Dynamics: Why Does Climate Vary?* (eds D. Sun and F. Bryan). <https://doi.org/10.1029/2008GM000838> (2013).
- Houze, J. et al. The variable nature of convection in the tropics and subtropics: a legacy of 16 years of the tropical rainfall measuring mission satellite. *Rev. Geophys.* **53**, 994–1012 (2015).
- Vitart, F., & Robertson, A. W. The Sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. *npj Clim. Atmos. Sci.* **1**, <https://doi.org/10.1038/s41612-018-0013-0>. (2018).
- Robertson, A. W., & Vitart, F. *Sub-seasonal to seasonal prediction*. (Elsevier, 2018).
- Li, S., & Robertson, A. W. Evaluation of submonthly precipitation forecast skill from global ensemble prediction systems. *Monthly Weather Rev.* **143**, 2871–2889 (2015).
- de Andrade, F. M., Coelho, C. A. S., & Cavalcanti, I. F. A. Global precipitation hindcast quality assessment of the Subseasonal to Seasonal (S2S) prediction project models. *Clim. Dyn.* **52**, 5451–5475 (2019).
- Moron, V., Robertson, A. W., & L. Wang. Weather within climate: sub-seasonal predictability of tropical daily rainfall characteristics. in Sub-seasonal to seasonal prediction (eds Robertson, A. W., & Vitart, F.) Sub-seasonal to seasonal prediction p 47–64 (Elsevier, 2019).
- Madden, R. A. & Julian, P. R. Observations of the 40–50 day tropical oscillation – a review. *Monthly Weather Rev.* **122**, 814–837 (1994).
- Waliser, D. E., Lau, K. M., Stern, W., & Jones, C. Potential predictability of the Madden-Julian oscillation. *Bull. Am. Meteo. Soc.* **84**, 33–50 (2003).
- Zhang, C. Madden-Julian oscillation. *Rev. Geophys.* **43**, 2004RG000158 (2005).
- Bjerknes, J. Atmospheric teleconnections from the Equatorial Pacific. *Monthly Weather Rev.* **97**, 163–172 (1969).
- Rasmusson, E. M., & Carpenter, T. H. Variations in Tropical Sea surface temperature and surface wind fields associated with the southern oscillation/El Niño. *Monthly Weather Rev.* **110**, 354–384 (1982).
- Barnston, A. G. et al. Long-lead seasonal forecasts – where do we stand? *Bull. Am. Meteo. Soc.* **75**, 2097–2114 (1994).
- Westra, S., & Sharma, A. An upper limit to seasonal rainfall predictability? *J. Clim.* **23**, 3332–3350 (2010).
- Vitart, F. Evolution of ECMWF sub-seasonal forecast skill scores. *Q. J. Meteorological Soc.* **140**, 1889–1899 (2014).
- Hoskins, B. The potential for skill across the range of the seamless weather-climate prediction problem: a stimulus for our science. *Q. J. Meteo. Soc.* **139**, 573–584 (2013).
- Annamalai, H., & Slingo, J. Active/break cycles: diagnosis of the intraseasonal variability of the Asian Summer Monsoon. *Clim. Dyn.* **18**, 85–102 (2001).
- Pai, D. S. et al. Impact of MJO on the intraseasonal variation of summer monsoon rainfall over India. *Clim. Dyn.* **36**, 41–55 (2011).
- World Meteorological Organization. Subseasonal to Seasonal prediction research implementation plan. https://library.wmo.int/doc_num.php?expnum_id=7632 (2012).
- Köppen, W. Die wärmezonen der erde, nach der dauer der heissen, gemässigten und kalten zeit und nach des wirkung der wärme auf die organische welt betrachtet. *Meteorologie Z.* **1**, 215–226 (1884).
- Peel, W. C., Finlayson, B. L., & MacMahon, T. A. Updated world map of the Köppen-Geiger climate classification. *Hydroclim. Earth Syst. Sci.* **11**, 1635–1644 (2007).
- Vaughan, C., & Dessai, S. Climate services for society: origins, institutional arrangements, and design elements for an evaluation framework. *Wiley Interdiscip. Rev.-Clim. Change* **5**, 587–603 (2014).
- Leith, C. E. Predictability of climate. *Nature* **276**, 352–355 (1978).
- Shukla, J. Predictability in the midst of chaos: a scientific basis for climate forecasting. *Science* **282**, 728–731 (1998).
- Von Storch, H., von Storch, J. S., & Müller, P. Noise in the Climate System – Ubiquitous, Constitutive and Concealing. in *Mathematics Unlimited — 2001 and Beyond*. (eds Engquist B., Schmid W.) (Springer, Berlin, Heidelberg, 2001).
- Stockdale, T. N. et al. Global seasonal rainfall forecasts using a coupled ocean-atmosphere model. *Nature* **392**, 370–373 (1998).
- Shukla, J. et al. Dynamical seasonal prediction. *Bull. Am. Meteo. Soc.* **81**, 2593–2606 (2000).
- Goddard, L. et al. Current approaches to seasonal-to-interannual climate predictions. *Int. J. Climatol.* **21**, 1111–1152 (2001).
- Kang, I. S., & Shukla, J. Dynamic seasonal prediction and predictability of the monsoon. in *The Asian Monsoon*. Springer Praxis Books. (Springer, Berlin, Heidelberg, 2006).
- Moron, V., Robertson, A. W., & Ward, M. N. Seasonal predictability and spatial coherence of rainfall characteristics in the tropical setting of Senegal. *Monthly Weather Rev.* **134**, 3248–8133262 (2006).
- Moron, V. et al. Spatial coherence of tropical rainfall at the regional scale. *J. Clim.* **20**, 5244–5263 (2007).
- Rowell, D. P. Assessing potential seasonal predictability with an ensemble of multidecadal GCM simulations. *J. Clim.* **11**, 109–120 (1998).
- Rodríguez-Iturbe, I. et al. Chaos in rainfall. *Water Resour. Res.* **25**, 1667–1675 (1989).
- Sivakumar, B. Rainfall dynamics at different temporal scales: a chaotic perspective. *Hydroclim. Earth Syst. Sci. Discuss.* **5**, 645–652 (2001).
- Orlanski, I. A rationale subdivision of scales for atmospheric processes. *Bull. Am. Meteo. Soc.* **56**, 527–530 (1975).
- Trenberth, K. E., Zhang, Y., & Gehne, M. Intermittency in precipitation: duration, frequency, intensity, and amounts using hourly data. *J. Hydrometeorol.* **18**, 1393–1412 (2017).
- Huang, X. et al. A long-term tropical mesoscale convective systems dataset based on a novel objective automatic tracking algorithm. *Clim. Dyn.* **51**, 3145–3159 (2018).
- Adler, R. F., & Negri, A. J. A satellite infrared technique to estimate tropical convective and stratiform rainfall. *J. Appl. Meteorol.* **27**, 30–51 (1988).
- Schumacher, C., & Houze, R. A. Stratiform rain in the tropics as seen by the TRMM precipitation radar. *J. Clim.* **16**, 1739–1756 (2003).
- Stephenson, D. B. et al. Extreme daily rainfall events and their impact on ensemble forecasts of the Indian monsoon. *Monthly Weather Rev.* **127**, 1954–1966 (1999).
- Moron, V., Robertson, A. W., & Pai, D. S. On the spatial coherence of sub-seasonal to seasonal Indian rainfall anomalies. *Clim. Dyn.* **49**, 3403–3423 (2017).
- Donald, A. et al. Near-global impact of the Madden-Julian oscillation on rainfall. *Geophys. Res. Lett.* **33**, L09704 (2006).
- Wheeler, M., & Kiladis, G. N. Convectively coupled equatorial waves: analysis of clouds and temperature in the wavenumber-frequency domain. *J. Atmos. Sci.* **56**, 374–399 (1999).
- Roundy, P. E., & Frank, W. M. A climatology of waves in the equatorial region. *J. Atmos. Sci.* **61**, 2105–2132 (2004).
- Lubis, S. W., & Jacobi, C. The modulating influence of convectively coupled equatorial waves (CECWs) on the variability of tropical precipitation. *Int. J. Climatol.* **35**, 1465–1483 (2015).
- Ropelewski, C. F., & Halpert, M. Global and regional scale precipitation patterns associated with the El Niño Southern Oscillation. *Monthly Weather Rev.* **115**, 1606–1626 (1987).
- McBratney, A. B., & Moore, A. W. Applications of fuzzy sets to climatic classification. *Agric. For. Meteorol.* **35**, 165–185 (1985).
- Moron, V., Camberlin, P., & Robertson, A. W. Extracting subseasonal scenarios: an alternative method to analyze seasonal predictability of regional-scale tropical rainfall. *J. Clim.* **26**, 2580–2600 (2013).
- Ricciardulli, L., & Sardeshmukh, P. D. Local time- and space- scales of organized tropical deep convection. *J. Clim.* **15**, 2775–2790 (2001).
- Yang, G. Y., & Slingo, J. The diurnal cycle in the tropics. *Monthly Weather Rev.* **129**, 784–801 (2001).
- Folland, C. K., Palmer, T. N., & Parker, D. E. Sahel rainfall and worldwide sea temperatures, 1901–85. *Nature* **320**, 602–607 (1986).
- Teo, C.-K. et al. The universal scaling characteristics of tropical oceanic rain clusters. *J. Geophys. Res. Atmos.* **122**, 5582–5599 (2017).
- Meehl, G. A., Arblaster, J. M., & Tebaldi, C. Understanding future patterns of increased precipitation intensity in climate model simulations. *Geophys. Res. Lett.* **32**, L18719 (2005).

54. Kendon, E. J. et al. Heavier summer downpours with climate change revealed by weather forecast resolution model. *Nat. Clim. Change* **4**, 570–576 (2014).
55. Wheeler, M. C. & Hendon, H. H. An all-season real-time multivariate MJO index: development of an index for monitoring and prediction. *Monthly Weather Rev.* **132**, 1917–1932 (2004).
56. Barnston, A. G. et al. Verification of the first 11 years of IRI's seasonal climate forecast. *J. Clim. Appl. Meteorol.* **49**, 493–520 (2010).
57. Bombardi, R. J. & al. Sub-seasonal predictability of the onset and demise of the rainy season over Monsoonal regions. *Front. Earth Sci.* **5**, <https://doi.org/10.3389/feart.2017.00014> (2017).
58. Vigaud, N. & Giannini, A. West-African convection regimes and their predictability from submonthly forecasts. *Clim. Dyn.* **52**, 7029–7048 (2019).
59. Huffman, G. J. et al. The global precipitation climatology project (GPCP) combined precipitation data set. *Bull. Am. Meteo. Soc.* **78**, 5–20 (1997).
60. Huffman, G. J. et al. Global precipitation at one-degree daily resolution from multisatellite observations. *J. Hydrometeorol.* **2**, 36–50 (2001).

ACKNOWLEDGEMENTS

A.W.R. acknowledges the support of a fellowship from Columbia University's Center for Climate and Life. The interaction between V.M. and A.W.R. benefitted from a 1-month visit during May 2019 granted to AWR by Aix-Marseille University under the FIR ("Fonds Incitatif pour la Recherche")-2018 contract. This work is based on S2S data, obtained from IRI Data Library. S2S is a joint initiative of the World Weather Research Programme (WWRP) and the World Climate Research Programme (WCRP). The original S2S database is hosted at ECMWF as an extension of the TIGGE database.

AUTHOR CONTRIBUTIONS

V.M. and A.W.R. designed the research. V.M. performed the statistical analyses. V.M. and A.W.R. contributed to the interpretation of the results and the writing of the paper.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at <https://doi.org/10.1038/s41612-020-0107-3>.

Correspondence and requests for materials should be addressed to V.M.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020