Tropical rainfall predictions from multiple seasonal forecast systems

Adam A. Scaife1,2 | Laura Ferranti3 | Oscar Alves4 | Panos Athanasiadis5 | Johanna Baehr6 | Michel Dequé7 | Tina Dippe8 | Nick Dunstone1 | David Fereday1 | Richard G. Gudgel9 | Richard J. Greatbatch8 | Leon Hermanson1 | Yukiko Imada10 | Shipra Jain11 | Arun Kumar12 | Craig MacLachlan1 | William Merryfield13 | Wolfgang A. Müller14 | Hong-Li Ren15 | Doug Smith1 | Yuhei Takaya10 | Gabriel Vecchi16 | Xiaosong Yang9,17

1Met Office Hadley Centre, Met Office, Exeter, UK
2College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, UK
3European Centre for Medium-Range Weather Forecast (ECMWF), Reading, UK
4Bureau of Meteorology, Melbourne, Australia
5Centro Euro-Mediterraneo sui Cambiamenti Climatici, Bologna, Italy
6Center for Earth System Research and Sustainability, Institute of Oceanography, Universität Hamburg, Hamburg, Germany
7Centre National de Recherches Météorologiques (UMR 3589), Toulouse, France
8GEOMAR Helmholtz Centre for Ocean Research Kiel Düsternbrooker Weg 20, Kiel, Germany
9Geophysical Fluid Dynamics Laboratory, National Oceanic and Atmospheric Administration, Princeton, New Jersey
10Climate Research Department, Meteorological Research Institute, Japan Meteorological Agency, Tsukuba, Japan
11National Centre for Medium Range Weather Forecasting (NCMRWF), Ministry of Earth Sciences, Noida, India
12NOAA/National Centers for Environmental Prediction, College Park, Maryland
13Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada, University of Victoria, Victoria, British Columbia, Canada
14Max Planck Institute for Meteorology, Hamburg, Germany
15Laboratory for Climate Studies, National Climate Center, China Meteorological Administration, Beijing, China
16Geosciences Department and Princeton Environmental Institute, Princeton University, Princeton, New Jersey
17University Corporation for Atmospheric Research, Boulder, Colorado

We quantify seasonal prediction skill of tropical winter rainfall in 14 climate forecast systems. High levels of seasonal prediction skill exist for year-to-year rainfall variability in all tropical ocean basins. The tropical East Pacific is the most skilful region, with very high correlation scores, and the tropical West Pacific is also highly skilful. Predictions of tropical Atlantic and Indian Ocean rainfall show lower but statistically significant scores.

We compare prediction skill (measured against observed variability) with model predictability (using single forecasts as surrogate observations). Model predictability matches prediction skill in some regions but it is generally greater, especially over the Indian Ocean. We also find significant inter-basin connections in both observed and predicted rainfall. Teleconnections between basins due to El Niño–Southern Oscillation (ENSO) appear to be reproduced in multi-model predictions and are responsible for much of the prediction skill. They also explain the...
relative magnitude of inter-annual variability, the relative magnitude of predictable rainfall signals and the ranking of prediction skill across different basins.

These seasonal tropical rainfall predictions exhibit a severe wet bias, often in excess of 20% of mean rainfall. However, we find little direct relationship between bias and prediction skill. Our results suggest that future prediction systems would be best improved through better model representation of inter-basin rainfall connections as these are strongly related to prediction skill, particularly in the Indian and West Pacific regions. Finally, we show that predictions of tropical rainfall alone can generate highly skillful forecasts of the main modes of extratropical circulation via linear relationships that might provide a useful tool to interpret real-time forecasts.

KEYWORDS
ensemble, ENSO, NAO, PNA, seasonal prediction, tropical rainfall

1 | INTRODUCTION

While significant progress has been made in recent years, there is a limit of around 2 weeks to deterministic weather forecast skill for daily rainfall, and most of the skill falls away by the second week (e.g., Li and Robertson, 2015; Stern and Davidson, 2015). In addition, numerous studies point to limitations of general circulation models in the simulation of rainfall in the Tropics. The poor reproduction of the Madden–Julian Oscillation (e.g., Ahn et al., 2017) and the challenges of accurately parametrizing tropical convection (e.g., Arakawa, 2004) are commonly cited reasons.

Despite these well-known limitations on shorter time-scales, the long-range prediction of seasonal rainfall in the Tropics is highly skilful, and seasonal mean skill scores a few months ahead far exceed the skill of weather forecasts in the extratropics just days ahead (Stockdale et al., 1998; Dequé, 2001; Kumar et al., 2013; Molteni et al., 2015; Scaife et al., 2017). This capability is of great interest for early warning of the risk of tropical drought and flooding (e.g., Dutra et al., 2014; Li et al., 2016). However, we also note that tropical rainfall has an influence on the extratropics via Rossby waves (e.g., Hoskins and Karoly, 1981; Simmonds et al., 1983; Li et al., 2014). Recent advances in winter seasonal prediction of extratropical circulation (e.g., Riddle et al., 2013; Kang et al., 2014; Scaife et al., 2014; Yang et al., 2015; Athanasiadis et al., 2017) are strongly linked to these teleconnections from the Tropics (Greatbatch et al., 2012; Molteni et al., 2015; Kumar and Chen, 2017; Scaife et al., 2017) as are some inter-annual (Dunstone et al., 2016) and even decadal variations (Trenberth et al., 2014; Smith et al., 2016) and so our study is focussed on the winter season.

Previous single model studies (Kumar et al., 2013; Molteni et al., 2015) have documented high tropical skill and linear inverse modelling suggests that seasonal forecasts may already be near the predictability limit in the East Pacific (Newman and Sardeshmukh, 2017). We therefore investigate tropical rainfall predictability in multiple seasonal prediction systems to document the variation of skill across different models and also across different tropical regions (section 3). We also compare the skill in predicting observed rainfall variability with the level of predictability inherent in the models (section 4) and show how well the dominant influence of El Niño–Southern Oscillation (ENSO) and inter-basin connections are reproduced in current prediction systems in section 5. Given that much climate model development is focused on improving the mean state, in section 6 we investigate whether there is a relationship between the magnitude of mean state errors (i.e., forecast drift) and seasonal forecast skill. Finally, in section 7 we show that using tropical rainfall forecasts alone can provide highly skilful predictions of extratropical inter-annual variability in the winter Pacific North American pattern and the North Atlantic Oscillation.

2 | SEASONAL PREDICTION SYSTEMS

We analyse winter (December–January–February) mean predictions of tropical rainfall from 14 seasonal prediction systems over the period 1992/1993 to 2011/2012 where available. These retrospective predictions are all initialized on or close to November 1 and the November data are excluded to prevent contamination from medium range predictability. Brief details and references for further details of each system follow:

U.K. Met Office predictions are from the operational global seasonal prediction system GloSea (Arribas et al., 2011). Data used here are from GloSea5—the fifth generation of this forecast system which has relatively high resolution and uses coupled ocean, sea-ice and land surface model
components (MacLachlan et al., 2015). Ensemble generation is through a combination of lagged start dates and stochastic physics perturbations to produce an ensemble of 24 member forecasts for each winter, initialized around early November, approximately 1 month ahead of winter as described in MacLachlan et al. (2015). The atmospheric resolution of the model is $0.83^\circ$ longitude by $0.55^\circ$ latitude with 85 quasi-horizontal atmospheric levels and an upper boundary at 85 km. Ocean resolution is $0.25^\circ$ globally with 75 quasi-horizontal levels.

Canadian Climate Centre models CanCM3 and CanCM4 share the same coupled ocean, land and sea-ice components. The land component is version 2.7 of the Canadian Land Surface Scheme (CLASS). Sea-ice dynamics are governed by cavitating fluid rheology, and thermodynamics by a simple energy balance model. The CanCM3 atmospheric component is CanAM3 (Scinocca et al., 2008), whereas the CanCM4 atmospheric component is CanAM4 (Von Salzen et al., 2013). Upgraded physical parameterizations in CanAM4 include fully prognostic cloud and aerosol schemes, improved radiation schemes and a parameterization of shallow convection. Ten ensemble members are generated for both CanCM3 and CanCM4 using initial conditions from 10 separate simulations constrained by observations. Horizontal resolution in both cases is T63 or about 2.8°. CanAM3 has 31 vertical levels and CanAM4 35 levels, both extending to 1 hPa. Horizontal resolution of the ocean component is approximately 100 km, with 40 vertical levels. Further details are provided in Merryfield et al. (2013).

European Centre for Medium Range Weather Forecasting (ECMWF) seasonal forecasts are from ECMWF system 4 (Molteni et al., 2011) with atmospheric model IFS cycle 36r4 coupled to the HTESSEL land surface model and the Nucleus for European Modelling of the Ocean (NEMO) ocean model (Madec, 2008) via the OASIS3 coupler (Valcke, 2013). An ensemble of 15 forecasts was made from November 1 for each winter using stochastic physics perturbations. Ozone is a prognostic variable and is radiatively active. Time-variation of greenhouse gases and solar cycle are specified, although solar variability is not spectrally resolved. Volcanic aerosols are included based on the estimated distribution in the month prior to the start of the forecast, and then follow damped persistence. There is no dynamical sea-ice in this system. For the first 10 days, the forecast persists the initial sea-ice analysis; then there is a transition towards specified ice conditions derived from the previous 5 years. This sea-ice configuration captures the main trend in sea-ice and gives a representation of the uncertainty in sea-ice conditions. The atmospheric model has a spectral truncation T255 (~80 km resolution) and 91 levels in the vertical. The ocean model has 42 vertical levels and a horizontal resolution of about $1^\circ$ in the extratropics with an equatorial refinement to $1/3^\circ$ latitude.

Météo-France predictions are from version 5 of the Meteo-France seasonal forecast system. It is based on the Centre National de Recherches Météorologiques Coupled Model version 5 (CNRM-CM5; Voldoire et al., 2013). The four components of the model are Arpege 6.0 (atmosphere), Surfex 7.3 (continental surfaces), Nemo 3.2 (ocean) and Gelato 5.1 (sea ice). Each forecast ensemble is initialized on November 1 and contains 15 ensemble members generated by stochastic perturbations (Batté and Déqué, 2016). This model also uses coupled, prognostic ozone, in this case initialized from climatology. The atmosphere resolution is TL255 ($0.7^\circ$) with 91 vertical levels. The ocean resolution is $1^\circ$ and has 42 vertical levels.

The Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC) data used here are from the CMCC Seasonal Prediction System version 1.5 (hereafter referred to as CMCC, Materia et al., 2014), while a general description of the model components is given in Alessandri et al. (2010). The CMCC-SPS-v1.5 uses the ECHAM5 atmospheric model coupled to OPA8.2 ocean model and the SILVA land surface model and sea-ice initialized from climatology. The ensemble consists of nine members covering the period 1983–2011 and initialized on November 1 each year using lagged initial conditions. The horizontal resolution in the atmosphere is T63 with 19 vertical levels up to 10 hPa. The ocean resolution is around $2^\circ$ with 31 vertical levels.

Beijing Climate Centre (BCC) seasonal prediction data are from the BCC/China Meteorological Administration (CMA) operational system 2, which is based on the BCC Climate System Model version 1.1 m (BCC_CSM1.1m; Wu et al., 2013). The BCC Atmospheric GCM is coupled to the BCC Atmosphere and Vegetation Interaction Model version 1.0, the Geophysical Fluid Dynamics Laboratory (GFDL) Modular Ocean Model version 4 (Griffies et al., 2005) and the Sea Ice Simulator (Winton, 2000). Forecasts are initialized for each calendar month from the four-time daily NCEP/NCAR R1 data and the oceanic initial values from ocean temperature of the NCEP Global Oceanic Data Assimilation System (GODAS), using a nudging scheme with timescale of 2 days while sea ice is interactive but not initialized (Liu et al., 2015). The BCC ensemble includes 24 forecast members initialized near the beginning of November: 9 are perturbed by an empirical singular vector (Cheng et al., 2010) and 15 are generated from the lagged average of the atmospheric states on the first 5 days of each month and the ocean states on the first 3 days which are combined to generate 15 further members. The ocean model resolution is $1^\circ$ with 40 levels and the atmospheric horizontal resolution is T106 with 26 vertical hybrid sigma/pressure levels (Wu et al., 2010).

National Centers for Environmental Prediction (NCEP) data are from the Climate Forecast System version (CFSv2; Saha et al., 2014). CFSv2 is a coupled ocean–atmosphere–land dynamical seasonal prediction system. The oceanic
component is the Geophysical Fluid Dynamics Laboratory Modular Ocean Model version 4 including the sea-ice simulator (MOM4; Griffies et al., 2005). CFSv2 forecasts are initialized from the Climate Forecast System Reanalysis (CFSR; Saha et al., 2010). For CFSv2 forecasts, there is one forecast at 00Z, 06Z, 12Z and 18Z every fifth day of the year and the 12 forecast members from October 18, 23, 28 are used here. The atmospheric component is at horizontal resolution of T126 (~100 km) with 64 vertical levels and the ocean is at 0.25° with 10° of the equator, tapering to 0.5° poleward of 30° latitude.

The Kiel Climate Model (KCM) couples the atmospheric model ECHAM5 (Roeckner et al., 2003) with interactive land surface to the NEMO-based ocean model OPA9 (Madec et al., 1998; Madec, 2008) with coupled LIM2 sea-ice model, using the OASIS3 coupler (Valcke, 2013). For more details see Park et al. (2009). An ensemble of nine members was run from November 1 for each winter using the nine different combinations of ocean and atmosphere initial states from three assimilation runs, where the model was run in partially coupled mode to minimize equatorial initialization shock (Ding et al., 2013; Thoma et al., 2015). The ocean and sea-ice components in the assimilation runs were forced with observed wind stress anomalies from ERA-Interim, added to the model’s native wind stress climatology. Radiative forcing was constant in time. Here we used the KCM in a coarse resolution: ECHAM5 at T31 with 19 vertical levels, OPA9 with the ORAC2 horizontal grid (roughly 1.3° horizontal resolution, refined to 0.5° at the equator) and 31 vertical levels.

The hindcasts from GFDL were produced from the Forecast-oriented Low Ocean Resolution model (FLOR; Vecchi et al., 2014). The atmosphere and land components of FLOR are taken from the GFDL Coupled Model version 2.5 (CM2.5; Delworth et al., 2012), whereas the ocean and sea-ice components are based on the GFDL Coupled Model version 2.1 (CM2.1; Delworth et al., 2006; Wittenberg et al., 2006). FLOR is an operational seasonal forecast model in the North American Multi-Model Ensemble for seasonal prediction (Kirtman et al., 2014). Twelve-member ensemble forecasts are initialized in November from 1991 to 2012. The initial conditions of the ocean and ice components are from the GFDL ensemble coupled data assimilation (ECDA) system (Zhang et al., 2007; Chang et al., 2013). Initial conditions for the atmosphere and land are from a set of AMIP simulations with time-varying observed SST (Reynolds et al., 2002) and radiative forcing. FLOR has a spatial resolution of ~50 km in the atmosphere and land, ~100 km in the ocean and 32 (50) vertical levels in the atmosphere (ocean). Further details of FLOR and its initialization can be found in Vecchi et al. (2014).

The Australian Bureau of Meteorology POAMA seasonal forecast system is based on a coupled ocean-atmosphere model and data assimilation system (Hudson et al., 2013). The land surface component is a simple bucket model for soil moisture (Manabe and Holloway, 1975) and has three soil levels for temperature (Hudson et al., 2011). The ocean model is the Australian Community Ocean Model version 2 (ACOM2; Schiller et al., 1997; 2002) and is based on the Geophysical Fluid Dynamics Laboratory (GFDL) Modular Ocean Model (MOM version 2). The atmosphere and ocean models are coupled using the Ocean Atmosphere Sea Ice Soil (OASIS) coupling software (Valcke, 2013). Forecasts are initialized from assimilated atmospheric and oceanic states using ocean initial conditions from the POAMA Ensemble Ocean Data Assimilation System (Yin et al., 2011a). Climatological sea-ice is imposed in the simulations and the atmosphere and land initial conditions are taken from the atmosphere–land initialization scheme (Hudson et al., 2011). A 10 member ensemble was initialized on November 1 using a coupled-model breeding scheme (Yin et al., 2011b; Hudson et al., 2013). The atmospheric model has T47 horizontal resolution with 17 vertical levels and the ocean grid resolution is 2° in longitude and 0.5° in latitude at the Equator, increasing to 1.5° near the poles.

The Max-Planck Institute Earth System Model version 1.0 in low resolution (MPI-ESM-LR; Giorgetta et al., 2013) was used to perform an ensemble of hindcasts (Baehr et al., 2015). This configuration consists of the atmospheric component ECHAM6 (Stevens et al., 2013). The ocean component consists of the Max-Planck Institute Ocean Model (MPIOM; Jungclaus et al., 2013). The oceanic and atmospheric components are coupled through the Ocean–Atmosphere–Sea-Ice coupler (Valcke, 2013). The initial conditions are obtained by Newtonian relaxation (“nudging”) of the atmosphere, ocean and sea-ice to ERA-Interim, ORAS4 and NSIDC, respectively (see Baehr et al., 2015 for details). Bred vectors in the oceanic component of the model are used to generate initial perturbations for the ensemble (following Baehr and Piontek, 2013). An ensemble of 10 forecasts was made from November 1. The atmospheric component is spectrally resolved with a truncation at wavenumber 63 (~200 km), and with physics represented on a regular Gaussian grid in the horizontal and 47 vertical levels and the ocean model uses a bi-polar grid at nominal 1.5° horizontal resolution.

Japan Meteorological Agency (JMA) predictions are from the operational global seasonal prediction system JMA/MRI-CPS2 (Takaya et al., 2017). The sea-ice component is incorporated in the coupled model. Data analysed here are from a lagged ensemble of 10 member forecasts for each winter that started from initial dates in the second half of October as described in Takaya et al. (2017). The model was initialized using the atmosphere and land analysis from the Japanese 55-year Reanalysis (JRA-55; Kobayashi et al., 2015), and ocean and sea-ice analysis from the Multivariate Ocean Variational Estimation/Meteorological Research Institute Community Ocean Model-Global version 2 (MOVE/
The largest mean rainfall and the largest inter-annual variability in seasonal rainfall totals occurs in the Tropics (Figure 1). The intense year-to-year variability of boreal winter rainfall is also well connected to extratropical predictions via teleconnections that are thought to be mediated by Rossby waves (Hoskins and Karoly, 1981). Following Scaife et al. (2017), we examine rainfall predictions for four tropical regions that show high variability and are connected to the extratropical winter circulation in models and observations: the tropical Indian Ocean (TIO: 45°–100°E, 5°S–10°N), tropical West Pacific (TWP: 110°–140°E, 5°S–25°N), tropical East Pacific (TEP: 200 –90 W, 5°S–10°N) and tropical Atlantic (TA: 60°–0 W, 5°S–5°N) shown by the black boxes in Figure 1. Prediction skill for each of the models and each of the four ocean basin regions is measured by ensemble mean correlations with GPCP data and illustrated in Figure 2. Raw ensemble mean rainfall predictions are plotted in each case to illustrate any bias and the ensemble mean variability from year to year. It is immediately clear that most models and regions are biased wet, as is typically found in climate models (Mueller and Seneviratne, 2014). For all of these regions, scores calculated using different sized ensembles converge quickly with ensemble sizes of 10–15 members being enough for correlation scores to converge (not shown). Consistent with this, the ensemble mean variance is also comparable to that in the observations, suggesting that a large proportion of observed rainfall variability is predictable.

Figure 2 shows that tropical West Pacific rainfall is well predicted by current seasonal forecast systems, with correlations ranging from 0.68 to 0.90 with seasonal mean GPCP rainfall observational data. However, a clear wet bias of around 30% is again present in predicted rainfall. Interestingly, although it does not have higher skill in this region, the BCC model shows a very small bias compared to other models.

Skilful seasonal predictions of tropical Indian Ocean rainfall are produced by most prediction systems (Figure 2), although results are varied and correlations range from small positive values to almost 0.7. Models are again generally too wet with around 30% too much rainfall in most cases. Interestingly the BCC model again shows only a very small mean bias, but this does not relate to its prediction skill which is in the middle of the range from other systems. Many models correctly predict a dip in the 1997/1998 winter and peaks in the 1998/1999 and 2006/2007 winters.

Predictions of tropical Atlantic rainfall show encouraging skill (Figure 2) but the levels vary between systems, with scores ranging from 0.49 to 0.83. All are highly significant, but the skill is lower here than in the tropical West Pacific in almost all prediction systems. Mean biases are more mixed in the Atlantic than the Indian or Pacific basins and over this region biases are generally smaller, with a range of wet and dry biases across different prediction systems.

The prediction systems show near perfect skill scores for the tropical East Pacific region (Figure 2) due mainly to ENSO as we show below. All systems correctly predict the peaks during El Niño and troughs during La Niña but here again a wet bias is evident in almost all systems, although biases are small in the CMCC and KCM models. The BCC and MIROC systems also show suppressed inter-annual anomalies compared to other models. Nevertheless, for this region the very high correlation scores suggest that almost all inter-annual variability is predictable on seasonal timescales.

We should note that these results are also lower bounds on the predictability of tropical rainfall due to the finite ensemble sizes (see model descriptions), inevitable errors in the individual forecast systems and errors in the observations.
used for initialization and verification. It is common practice to increase ensemble size and remove some of these errors by cancellation by taking multi-model mean forecasts (dashed lines in Figure 2). In our case this results in an ensemble size of over 100 members and this multi-model average shows the best overall scores (Figure 2, inset scores in black). However, the multi-model mean still shows similar skill levels to the best single model in each region. This similar skill, despite the much larger multi-model ensemble size, is consistent with rapid convergence of skill for tropical rainfall (Kumar and Chen, 2015). As in the single model cases, there is a clear ranking of skill across the different regions with TEP > TWP > TA > TIO. We offer an explanation for this ranking below when we consider teleconnections to the El Niño–Southern Oscillation.

4 | MODELLLED PREDICTABILITY

It is important to distinguish the predictability of the real climate system from that in prediction systems, as these may not always be the same (Kumar et al., 2014a). Neither is modelled predictability an upper estimate of the real-world predictability as is often assumed. A few studies have tried to estimate seasonal predictability from analysis of observations alone (e.g., Keeley et al., 2009; Feng et al., 2012) but we can estimate the predictability of the real world directly by taking correlations between ensemble mean forecasts and observations. Similarly, we can estimate the predictability of the model by substituting single ensemble members for the observations and correlating with the mean of the remaining members. Of course, if our prediction systems (and the climate models they contain) were perfect then these two measures would be statistically identical. This is the crucial assumption in so-called “perfect model” studies where modelled predictability is used to estimate real-world predictability. However, this is not always the case, and examples have been found where the predictability of the model is either higher or lower than that of the real world (e.g., Eade et al., 2014; Scaife et al., 2014; Seviour et al., 2014; Weisheimer and Palmer, 2014; Kumar et al., 2014a; Dunstone et al., 2016; Kumar and Chen, 2017; Saito et al., 2017).

Figure 3 compares the predictability of the models with their skill in predicting the observations. One value is plotted for each model and each region. Perfect models would lie close to the diagonal line where the model skill in predicting one of its own ensemble members equals its skill of predicting the real world. For the TEP region we can see this is almost the case as predictions are near perfect on this timescale, although even here, prediction systems are slightly better at predicting themselves than the observations. In the TWP, modelled predictability is again close to prediction skill, with similar high (typically ~0.8) correlations when the ensemble mean is compared with single ensemble members or observations alike. For the tropical Atlantic rainfall almost all models show a greater correlation with their own ensemble members than with the observations, suggesting they are
overconfident, or equivalently, that they are better at predicting themselves than the real-world rainfall variability. Finally, the Indian Ocean rainfall shows the largest errors: almost all models are overconfident and correlations of the ensemble mean with single model ensemble members are around twice as large as the correlations with observations. As noted by other studies, this overconfidence can present serious problems with the use of forecasts (Weisheimer and Palmer, 2014). Depending on its cause, overconfidence may also indicate a potential for high skill if systematic errors can be corrected. For example, a systematic error in the spatial structure of a predictable teleconnection could lead to overconfident forecasts. If this teleconnection were improved in future climate models and hence future seasonal forecast systems, then skill could in principle rise. However, we also note that if the overconfidence of rainfall forecasts for the Indian Ocean is due to the absence of \textit{unpredictable} “noise” in models, from weak model representation of the Madden–Julian Oscillation for example, then there may actually be limited potential for improvement.

5 | INTER-BASIN CONNECTIONS AND THE EFFECTS OF ENSO

Atmospheric variability over the tropical oceans is correlated across different ocean basins (e.g., Camberlin et al., 2004; Kumar et al., 2014b; Molteni et al., 2015; Scaife et al., 2017). These links arise primarily due to changes in atmospheric circulation that can bridge land regions and create remote teleconnections, often due to ENSO (e.g., Giannini et al., 2001; Dong et al., 2006; Toniazzo and Scaife, 2006; Smith et al., 2010). If we are to extract the maximum prediction skill from globally important sources of predictability such as ENSO, it is therefore important to have good teleconnections between rainfall over different tropical ocean basins.
The skill of seasonal predictions of tropical rainfall is found to be generally high in our analysis (Figure 2). However, inter-basin connections are more uncertain because, unlike deterministic prediction skill which involves the ensemble mean, model inter-basin relationships can only be meaningfully compared to observations using the correlation across basins in individual ensemble members, which will therefore contain significant unpredictable internal variability. Some of these statistics can therefore vary substantially across the GPCP observational record due to sampling variability. To try and address this sampling variability we therefore examine inter-basin connections from all the forecast systems by calculating the correlation between rainfall in pairs of tropical regions using single ensemble members (c.f., Johnson et al., 2017). Table 1 shows the strength of these relationships in observations and single model ensemble members. In observations, the strongest relationship is found in the anti-correlation between rainfall in the tropical East and West Pacific; as would be expected given the east-west seesaw in Pacific rainfall due to ENSO. Strong anti-correlation is also found between the tropical East Pacific and the neighbouring tropical Atlantic. Consistent with these results, a positive correlation occurs between rainfall in the West Pacific and that in the tropical Atlantic. Finally, East Pacific/West Pacific rainfall is positively/negatively correlated with that in the Indian Ocean in this season and there is only a weak relationship between Atlantic and Indian Ocean rainfall variations. Table 1 also shows the corresponding range of values from ensemble members and the number of individual models whose member correlations span the value found in observations. It is immediately clear from the range of correlations found in ensemble members that the observed inter-basin relationships are captured by the multi-model ensemble as a whole. A very broad range of correlations is generated and so it is hard to argue that inter-basin connections are misrepresented in the multi-model ensemble. However, it is possible to show that the correlations found in members from individual models (Table 1, third row) do not span the value found in observations. It is interesting to note that despite the very high skill in predicting tropical East Pacific rainfall, its inter-basin connections are least well represented and are often weaker than observed, particularly between the East and West Pacific, but also elsewhere, and so this is an important area where seasonal prediction systems might be improved in future.

Many studies note that seasonal predictions are more skillful during periods when ENSO is active (e.g., Arribas et al., 2011; Kim et al., 2012; Lu et al., 2017) so an obvious question is whether the ENSO influence on the different basins is properly represented in current forecast systems. The top left panel of Figure 4 shows the strength of predicted year-to-year variations in tropical rainfall in each of our four regions, calculated as the standard deviation of the multi-model ensemble mean. In order of magnitude, the largest predictable signals are found in the East Pacific, West Pacific, Atlantic and Indian Ocean, which shows the smallest predictable signal. Interestingly, the same ranking is found in the magnitude of observed inter-annual variability (Figure 4, top right) and also in the skill of our multi-model predictions (Figure 2). Figure 4, bottom panels show the signal from a typical ENSO event, obtained by regressing the Niño3.4 SST index against observed and predicted rainfall and scaling to a 2 K ENSO event. The strength of the ENSO signals is lower than observed but within the observational uncertainty in all basins. Note that the strength of ENSO effects follows exactly the same ranking as the skill and variance and is strongest in the East Pacific.

| TABLE 1 | Observed and modelled inter-basin rainfall correlations. Observed inter-basin correlations (top row), range of inter-basin correlations from all models and all ensemble members (middle row) and number of models whose individual member correlations span the observed correlation (bottom row) |
|----------|------------------|----------------|-----------------|-----------|----------------|------------------|
| Observed | −0.89            | −0.62          | +0.41           | +0.61     | −0.30         | −0.07            |
| Modelled | −0.96, −0.24     | −0.85, 0.16    | −0.67, 0.75     | −0.09, 0.83 | −0.65, 0.64  | −0.69, 0.75      |
| No. models spanning obs. | 6/14 | 10/14 | 9/14 | 12/14 | 11/14 | 11/14 |
Inter-annual variability of different basins. The observed variability, the size of predictable signals and the relative skill of seasonal predictions for different basins can therefore be explained by the relative strength of the ENSO influence on each basin. Indeed, when we removed the ENSO influence from forecasts for each basin using linear regression on Niño3.4, only small and statistically insignificant (but positive) levels of skill remained (not shown).

6 | WHAT GOVERNS PREDICTION SKILL?

So far we have examined anomalies from the climatological mean of forecasts for many years, after linearly correcting away the mean bias in each of the models. However, initialized predictions from different systems drift to a greater or lesser degree (e.g., Hermanson et al., 2017) and if the resulting bias from the real world is large enough, then it could be that this affects the skill of predictions (e.g., Magnusson et al., 2013; Smith et al., 2013; Vecchi et al., 2014; Kim et al., 2017). We therefore examined the prediction skill for different regions as a function of model bias.

The skill of rainfall predictions from each forecast system for the same four regions was compared with the corresponding mean rainfall bias in each system. For the tropical West Pacific, tropical East Pacific and tropical Indian Ocean, correlations between skill and mean bias across the systems were 0.0, 0.1 and −0.1, respectively, indicating no simple systematic relationship between mean rainfall bias and the skill of predictions. This is a simple test for a link between mean bias and skill and it could be that mean biases in other variables control both the mean rainfall and prediction skill (c.f., Magnusson et al., 2013; Richter, 2015; Mulholland et al., 2017). Nevertheless, it suggests that if we were to focus exclusively on improving the mean rainfall biases in these regions in future models, the skill of our seasonal climate predictions may not be improved. In the tropical Atlantic there is a weak relationship between rainfall bias and
prediction skill \( (r = -0.3) \). Although it is not statistically significant, this inverse relationship between bias and prediction skill in different systems is what would be expected if model bias were having a detrimental effect on forecast skill. Alleviating model bias in this region might therefore yield higher prediction skill. This possibility is supported by Ding et al. (2015). Using the KCM with a correction to surface heat fluxes to reduce SST biases, they improved the fidelity of their data assimilation runs in boreal summer (JJA) and showed a greater role for ocean dynamics in the variability of the tropical Atlantic (Dippe et al., 2017).

If a small mean bias is not sufficient for high prediction skill, then what is? Given the strong influence of ENSO on observed inter-annual variability, forecast inter-annual variability and the ranking of skill across different basins identified above, we now test whether prediction skill is related to the strength of teleconnections to the East Pacific (ENSO) region. Figure 5 shows the relationship between prediction skill and the strength of inter-basin teleconnections to the East Pacific rainfall for each of the other three basins. Most models underestimate the strength of inter-basin teleconnections and clear relationships are now found with prediction skill, which increases as the strength of the teleconnections to the East Pacific increases. The skill of rainfall predictions in the tropical West Pacific and the Indian Ocean are strongly related to the strength of their relationship with the East Pacific.

At this point, we need to be careful about what we conclude from the relationships in Figure 5. For example, if the TIO variability originates from two components: one being the ENSO signal, and the other being poorly predicted but nearly uncorrelated with ENSO, then the correlation skill of predicted TIO rainfall would necessarily be an increasing function of the variance explained by ENSO, as seen in Figure 5. However, we note that the observed relationship between the TIO or TWP and ENSO is larger than the model correlation in almost all cases, even though model data are from the ensemble mean and observed data are single realizations. This is surprising given that we might expect larger correlations from the ensemble mean data and is consistent with the model under-representing the observed inter-basin correlation, at least over this period. In summary, while it is also important to enhance the predictive skill for non-ENSO-related variability where possible, our analysis suggests that errors in inter-basin connections may limit current prediction skill.

7 | LINKS TO EXTRATROPICAL CLIMATE

Seasonal predictions for the extratropics generally show much less skill than for the Tropics (e.g., Kim et al., 2012) but robust skill has previously been established for the Pacific North American (PNA) pattern via skilful predictions of ENSO (e.g., Derome et al., 2005; Athanasiadis et al., 2014) and more recently for the North Atlantic Oscillation (NAO) via skilful predictions of the Tropics (e.g., Greatbatch et al., 2012; Scaife et al., 2014; 2017). International projects are now focusing on these mechanisms as a route to improved climate predictions for the whole globe (e.g., Merryfield et al., 2017).

We determine the potential for skilful prediction of the main modes of extratropical inter-annual variability using our multi-model tropical rainfall predictions alone. The PNA is defined using the index of Wallace and Gutzler (1981) applied to upper tropospheric (200 hPa) geopotential height and the NAO is defined according to the sea level pressure difference between Iceland and the Azores (see Figures 6 and 7, respectively, and figure captions for detailed definitions). Multiple linear regression was then carried out between rainfall in our four tropical regions and these observed PNA and NAO indices (c.f., Scaife et al., 2017).

Figures 6 and 7, middle panels show the strong relationship between observed winter rainfall variations and associated variability in the PNA (Figure 6) and the NAO (Figure 7). Using observed winter mean rainfall suggests that perfect advance knowledge of rainfall in our four regions would allow highly skilful predictions of both the
PNA \( (r = 0.87) \) and the NAO \( (r = 0.70) \). Of course in practice, the observed winter rainfall is not known in advance but some skill can even be derived from knowledge of November rainfall, with significant, though lower, correlations of 0.39 and 0.49 for the PNA and NAO, respectively. Multi-model predictions of tropical rainfall improve on these empirical forecast scores with correlations of 0.82 (PNA) and 0.55 (NAO), and reach even higher values in some single models, confirming that skilful predictions of tropical rainfall are likely to be crucial for the prediction of major modes of extratropical winter variability.

8 | CONCLUSIONS

We find encouraging levels of tropical winter rainfall prediction skill on seasonal timescales in different forecast systems. High levels of skill are found in tropical rainfall predictions in all basins and a very similar pattern emerges across different forecast systems: the East Pacific is the most predictable region, followed by the West Pacific, then the Atlantic and finally the Indian region. Our multi-model-mean predictions suggest that the proportion of observed rainfall variance that is predictable in each basin is at least:

\[
\text{TEP}(94\%) > \text{TWP}(85\%) > \text{TA}(67\%) > \text{TIO}(42\%).
\]

Furthermore, we find an identical ranking in the observed levels of inter-annual variability, in the magnitude of predicted ensemble mean signals and in the influence of ENSO on each of these regions. This striking similarity, and the fact that only small levels of skill remain after ENSO variability is linearly removed from the forecasts, suggests that the skill of tropical rainfall in current seasonal forecast systems is largely driven by ENSO. However, the imperfect correlations between rainfall in the East Pacific and other regions, the imperfect correlation between the inferred NAO and PNA forecasts shown here, and the (small) but positive residual skill in each basin after ENSO is linearly removed all suggest that a single ENSO index is not sufficient to explain all tropical rainfall variability.

Unlike the extratropics, the signal to noise ratio in tropical rainfall forecasts is very large and relatively few ensemble members (<15) are needed to realize almost all of the available prediction skill (Kumar and Chen, 2015). Modelled predictability is generally higher than the skill in predicting
observed variations, suggesting overconfidence, especially over the Indian Ocean, which showed the lowest skill despite remote teleconnections that have been identified for Indian rainfall (e.g., Kucharski et al., 2009; Molteni et al., 2015). This contrasts sharply with the extratropics in the Atlantic for example, where modelled predictability can be lower than the skill of predicting the observations, and models can be underconfident (Eade et al., 2014; Scaife et al., 2014; Stockdale et al., 2015; Dunstone et al., 2016; Kumar and Chen, 2017; Saito et al., 2017).

Potential for improvement appears to exist in the Indian Ocean basin, where modelled predictability far exceeds current prediction skill. However, we emphasize that high levels of model predictability does not always indicate potential for improvement and an important caveat here is the lack of realistic MJO activity in models which could itself be unpredictable on seasonal timescales and thereby prevent future improvement in forecast skill. There is also potential for improvement in the Atlantic where model bias may be degrading skill. This is perhaps not surprising given large model biases in the tropical Atlantic, and nearby SST biases that are often large enough to reverse the zonal SST gradient (Richter, 2015). However, no significant relationship was detectable between local biases and prediction skill in the regions we examined. Instead, we point to the very clear relationship between the strength of inter-basin rainfall connections and the skill of model predictions. These connections are often misrepresented in models and focused model development to improve inter-basin rainfall connections would likely yield improved climate predictions.

Overall, our results demonstrate high levels of seasonal predictability in tropical rainfall in all basins and across current seasonal prediction systems. Finally, although there may be some multidecadal variability in skill levels (e.g., Weisheimer et al. 2015; Kumar and Chen, 2017), links between tropical rainfall and major modes of extratropical variability such as the PNA and NAO indicate that these highly skilful seasonal predictions of tropical rainfall can drive highly skilful predictions for the extratropics if models can accurately represent the mechanisms of tropical–extratropical interaction.

ACKNOWLEDGEMENTS
This work was supported by the UK-China Research & Innovation Partnership Fund through the Met Office Climate Science for Service Partnership (CSSP) China as part of the Newton Fund and the Joint DECC/Defra Met Office Hadley Centre Climate Programme (GA01101). This paper is also an outcome of the “Interaction/teleconnection between tropics and extratropics” initiative of the World Climate Research Programme’s Working Group on Subseasonal to Seasonal Prediction (WGSIP). We thank the WMO-WCRP database through the WGSIP. W.A.M. was supported by the German Ministry of Education and Research (BMBF) under the MiKlip project FLEXFORDEC (Grant No. 01LP1519A). J.B. was supported by the Cluster of Excellence CliSAP (EXC177), Universität Hamburg, funded through the German Science Foundation (DFG). T.C. and R.J.G. acknowledge support from the German Ministry of Education and Science (BMBF) through MiKlip2 subproject ATMOS-MODIN (Grant No. 01LP1517D) and SACUS (Grant No. 03G0837A) and by the European Union 7th Framework Programme (FP7 2007–2013) under Grant agreement 603521 PREFACE project.

ORCID
Adam A. Scaife http://orcid.org/0000-0002-5189-7538

REFERENCES
Camberlin, P., Chauvin, F., Douville, H. and Zhao, Y. (2004) Simulated ENSO–tropical rainfall teleconnections in present-day and under enhanced


