Subseasonal prediction performance for South American land–atmosphere coupling in extended austral summer

Received: 22 March 2021 | Revised: 12 November 2021 | Accepted: 18 November 2021

DOI: 10.1002/cli2.28

Climate Resilience and Sustainability
Interdisciplinary Approaches Towards Solutions for Climate Change
Open Access
RMetS

ORIGINAL ARTICLE

Subseasonal prediction performance for South American land–atmosphere coupling in extended austral summer

Amulya Chevuturi[1,2,3] | **Nicholas P. Klingaman**[1,2] | **Liang Guo**[1,2] | **Christopher E. Holloway**[2] | **Bruno S. Guimarães**[4] | **Caio A. S. Coelho**[4] | **Paulo Y. Kubota**[4] | **Matthew Young**[1,2] | **Emily Black**[1,2] | **Jessica C.A. Baker**[5] | **Pier Luigi Vidale**[1,2]

[1] National Centre for Atmospheric Science, University of Reading, Reading, United Kingdom

[2] Department of Meteorology, University of Reading, Reading, United Kingdom

[3] UK Centre for Ecology & Hydrology, Benson Lane, Crowmarsh Gifford, Wallingford, United Kingdom

[4] Centre for Weather Forecast and Climate Studies (CPTEC), National Institute for Space Research (INPE), Cachoeira Paulista, SP, Brazil

[5] School of Earth and Environment, University of Leeds, Leeds, United Kingdom

Correspondence
Amulya Chevuturi, UK Centre for Ecology & Hydrology, Benson Lane, Crowmarsh Gifford, Wallingford, OX10 8BB, United Kingdom.
Email: amuche@ceh.ac.uk

Funding information
Newton Fund, Grant/Award Number: CSSP Brazil DUBSTEP project; Fundação de Amparo à Pesquisa do Estado de São Paulo, Grant/Award Number: process 2015/50687-8 (CLIMAX Project); National Centre for Atmospheric Science, Grant/Award Number: ACREW Grant Number NE/R000034/1; European Research Council, Grant/Award Numbers: DECAF project, Grant agreement no. 771492; Global Challenges Research Fund, Grant/Award Number: ACREW Grant Number NE/R000034/1; Natural Environment Research Council, Grant/Award Number: Independent Research Fellowship (NE/L010976/1); Conselho Nacional de Desenvolvimento Científico e Tecnológico, Grant/Award Number: process 305206/2019-2

Abstract

Land–atmosphere feedbacks, through water and energy exchanges, provide subseasonal-to-seasonal predictability of the hydrological cycle. We analyse subseasonal land–atmosphere coupling over South America (SA) during extended austral summer for the soil moisture-to-precipitation and soil moisture-to-air temperature feedback pathways. We evaluate subseasonal hindcasts from global forecasting systems from the UK Met Office, the National Centers for Environmental Prediction (NCEP), the European Centre for Medium Range Weather Forecasts and the Center for Weather Forecast and Climate Studies (CPTEC), for the common period of 1999–2010, against two reanalyses. Biases in land–atmosphere states are established in the first week of hindcasts and increase with lead time. By Week 5, all the models only demonstrate good performance over northern, northeastern and southeastern SA for soil moisture and evapotranspiration and over tropical and subtropical SA for temperature. The hindcasts show stronger coupling at longer lead–lag between variables than reanalyses. Our results highlight possible deficiencies in feedbacks between soil moisture and precipitation in CPTEC and NCEP forecasts over the Amazon due to initial dry soil moisture biases, and in feedbacks between soil moisture and temperature for all four investigated models over southeastern SA due to erroneous representations of evapotranspiration.

KEYWORDS
land–atmosphere feedback, soil moisture, subseasonal forecasts

1 | INTRODUCTION

The land and atmosphere are connected through exchanges of water and energy (Dirmeyer et al., 2018). As the land surface usually evolves more slowly than the atmosphere, it modulates atmospheric variability on scales longer than synoptic (Koster & Suarez, 2001). Land–atmosphere interactions provide subseasonal-to-seasonal (S2S) predictability for the hydrological cycle on S2S timescales, influenced by the relatively slowly varying aspects of the land surface (Dirmeyer et al., 2015). The land surface can affect atmospheric S2S variability when three conditions are satisfied: (i) the atmosphere is sensitive to variations in the land surface state (coupling); (ii) the land surface state fluctuates (variability); and (iii) these fluctuations persist (memory).

Soil moisture is one of the most important land surface features for S2S predictability (National Academy of Sciences, 2016), as it significantly modulates evaporation in moisture-limited regimes and ultimately precipitation through local and regional water and energy cycles (Dirmeyer et al., 2018). Dirmeyer & Halder (2017) identified a significant relationship between S2S forecast skill and land–atmosphere coupling strength, over most of the world, through both water and energy pathways. Consequently, S2S prediction skill is significantly limited by model errors in land–atmosphere feedbacks (Dirmeyer et al., 2019). Other than model errors, loss of potential predictability may arise from poor representation of atmospheric conditions, poor initial soil states due to limited observations and erroneous coupling between land surface and atmospheric models (Dirmeyer & Halder, 2017).

Many studies show that some regions of South America (SA), like the Amazon and southeastern SA, are "hot spots" for land–atmosphere interactions (e.g. Koster et al., 2004; Orlowsky & Seneviratne, 2010; Sörensson & Menéndez, 2011; Ruscica et al., 2014, 2015; Spennemann et al., 2015). Advances in modelling soil and vegetation processes have improved the representation of SA summer precipitation in climate simulations (Ma et al., 2011). However, poor model prediction performance for soil moisture variability reduces the performance of the Climate Forecast System version 2 for precipitation over parts of SA (Koster et al., 2009; Dirmeyer et al., 2019). Thus, errors in SA land surface predictions may propagate to errors in SA precipitation forecasts (Koster et al., 2000). Improved representations of soil moisture control processes may improve model representations of seasonal to annual land-atmosphere interactions over SA (Baker et al., 2021).

Some S2S models have shown promising performance for SA rainfall forecasts three weeks ahead (Li & Robertson, 2015; Hirata & Grimm, 2018). Studies have also evaluated and proposed forecast verification methods for SA precipitation at the subseasonal timescale for autumn (Coelho et al., 2018) and summer (Klingaman et al., 2021). S2S forecasts for SA summer precipitation have substantial biases, with forecast performance for weekly rainfall extending up to three weeks over northern, northeastern and southeastern SA but only to the first week for southern Amazonia and the Andes (Klingaman et al., 2021). We expand on Klingaman et al. (2021) by comprehensively analysing the skill and performance of the SA land surface and atmosphere states and land–atmosphere coupling at subseasonal lead times in the same four global forecasting systems used in that study. We focus on the soil moisture–evaporation–precipitation and soil moisture–evaporation–air temperature feedback pathways (Seneviratne et al., 2010) to investigate the representation of the hydrological cycle over SA during extended austral summer (NDJFM) at lead times of one to five weeks. This work provides a holistic view of the performance of four S2S models for regional subseasonal land–atmosphere feedbacks over SA.

The models, hindcasts, reanalyses data (Section 2.1) and methods used (Section 2.2) are introduced in Section 2; verification of land and atmosphere state in models is discussed in Section 3; verification of land–atmosphere coupling is discussed in Section 4. We discuss the broader context and limitations of our results in Section 5 and conclude our findings in Section 6.

2 | DATA AND METHODS

2.1 | Data

We evaluate hindcasts (reforecasts) of three models from the S2S project database (Vitart et al., 2017) and from a new Brazilian model (Guimarães et al., 2020) over the common period of 1999–2010. The four hindcast datasets and associated modelling systems are described below and in Table 1.

- **CPTEC**: The Centre for Weather Forecast and Climate Studies Brazilian Global Atmospheric Model version 1.2 (BAM-1.2) is used to generate hindcasts twice per month

TABLE 1 Description of S2S hindcast model components and initialization along with ensemble and verification sizes

	Atmosphere		Land-Surface		Original ensemble size	Final ensemble size	Verification size
Dataset	**Model**	**Initialisation**	**Model**	**Initialisation**			
CPTEC	Centre for Weather Forecast and Climate Studies Brazilian Global Atmospheric Model version 1.2 (BAM-1.2) (Guimarães et al., 2020)	ERA Interim (Dee et al., 2011)	IBIS (Integrated Biosphere Simulator) CPTEC surface model (Kubota, 2012)	Seasonal climatology developed from Willmott et al. (1985)	11	11	110
ECMWF	Integrated forecasting system (IFS; Cycles 43R1 and 43R3) spectral atmospheric model (ECMWF, 2014)	ERA Interim reanalysis (Dee et al., 2011)	Revised Tiled ECMWF Scheme for Surface Exchanges over Land (HTESSEL) (Dutra et al., 2009)	ERA Interim/Land soil reanalysis (Balsamo et al., 2015)	11	33	220
NCEP	Climate Forecast System (CFS; version2) spectral atmospheric model (Saha et al., 2014)	Climate Forecast System Reanalysis (CFSR) (Saha et al., 2010)	4-layer NOAH Land surface model (version 2.7.1) (Ek et al., 2003)	CFSR and associated GLDAS analysis (Meng et al., 2012)	4	28	220
UKMO	UK Met Office Unified Model (UM) Global Atmosphere 6.0 (Hadley Centre Global Environment Model version 3; HadGEM3) (Walters et al., 2017)	ERA Interim (Dee et al., 2011)	UM Global Land 6.0 (Joint UK Land Environment Simulator; JULES) (Walters et al., 2017)	ERA Interim (Dee et al., 2011)	7	7	220

with 11 ensemble members (Guimarães et al., 2020). CPTEC performs a "fixed" (frozen) hindcast set. The initialization dates can be found in Table 1 of Guimarães et al. (2020).

- **ECMWF**: The European Centre for Medium Range Weather Forecast Integrated forecast system (IFS) cycles 43R1 and 43R3, with 11 ensemble members (Vitart, 2014). Hindcasts are run twice per week (Monday and Thursday). ECMWF generate hindcasts "on the fly" (i.e. in near-real time, alongside the operational forecasts), and we use hindcasts performed during May 2017–April 2018.
- **NCEP**: The National Centers for Environmental Prediction Climate Forecast System version 2 (CFSv2) is initialized daily to generate hindcasts with four ensemble members (Saha et al., 2014). NCEP also performs a fixed hindcast set.
- **UKMO**: The UK Met Office Global seasonal forecasting system configuration 2.0 (GloSea5-GC2), with seven ensemble members is initialized on the 1st, 9th, 17th and 25th of each month (MacLachlan et al., 2015). UKMO also generates hindcasts "on the fly", and we use hindcasts performed during May 2017–April 2018.

To compare models with different initialization frequencies and start dates, we use time-lagged ensemble means of the more frequently initialized NCEP and ECMWF hindcasts, to align with the less frequently initialized UKMO hindcasts (Klingaman et al., 2021). We use an 8-day window of forecast initialization dates for the NCEP and ECMWF lagged ensemble prior to and including the UKMO initialization date (Kirtman et al., 2014). The lagged ensemble size is the total number of ensemble members we use, for NCEP and ECMWF, to match the UKMO initialization date. We use the UKMO dates, rather than the CPTEC dates, because the approximately 15-day separation between CPTEC dates would lead to the loss of up to two weeks' lead time for ECMWF and NCEP. Thus, CPTEC has a smaller verification sample size than the other models (Table 1). Verification sample size is the lagged ensemble size multiplied by the number of initializations. The S2S ensembles have different final ensemble sizes in our study: CPTEC (11), ECMWF (33), NCEP (28) and UKMO (7). For CPTEC and UKMO, the final ensemble size is same as the original ensemble size, as they are not lagged.

The sample size and the window used for creating the lagged ensembles are very important for forecast performance at subseasonal timescales (e.g. DelSole et al., 2017; Trenary et al., 2017, 2018). Sensitivity tests comparing lagged ensembles of NCEP and ECMWF using an 8-day window to ensembles of similar size to UKMO (based on 2-day window for NCEP and one initialization for ECMWF) showed that using a longer window for lagged ensemble had slightly lower performance for rainfall for Weeks 1 and 2 and slightly higher performance for Weeks 3–5 (Klingaman et al., 2021). We use the larger lagged ensembles for ECMWF and NCEP in this study, as they perform slightly better beyond Week 2, which coincides with the lead times at which subseasonal forecasts are generally used.

We use weekly-mean ensemble hindcasts of 0–20 cm soil moisture (SM), evapotranspiration (ET), precipitation (PR) and 2m air temperature (T2) from the four models for extended austral summer (November–December–January-February–March; NDJFM). We analyse austral summer as it is the major wet season for most of SA, and land–atmosphere coupling strongly influences the SA monsoon (Sörensson & Menéndez, 2011). In winter, the land and atmosphere are mostly decoupled over most SA (Sun & Wang, 2012).

Hindcasts are verified against the fifth-generation ECMWF land reanalysis (ERA5L; Muñoz-Sabater et al., 2021) and NASA and NOAA's Global land data assimilation system version 2.0 analysis (GLDAS; Rodell et al., 2004; Rui & Beaudoing, 2018). ERA5L is generated using the Tiled ECMWF Scheme for Surface Exchanges over Land incorporating land surface hydrology (H-TESSEL) of the IFS cycle 45R1, but without coupling the IFS to the atmospheric module or the ocean wave model (Muñoz-Sabater et al., 2021). ERA5L has a spatial resolution of 9 km and is forced by the atmospheric output of the ERA5 system. It is a high-resolution "replay" of the land component of the ERA5 reanalysis. As ERA5L is forced with ERA5 surface fluxes, it is expected to have similar land–atmosphere feedbacks to ERA5, particularly on the regional-continental scales of interest here. In other words, the land–atmosphere feedbacks do not benefit from the higher resolution in ERA5L. We use ERA5L, over ERA5, to gain higher resolution land surface fields that are physically consistent with the ERA5 atmospheric reanalysis, not to capture higher resolution land-atmosphere feedbacks than those in ERA5. GLDAS integrates observations (from satellites and ground observations) using the four-layer NOAH land surface model without coupling to the atmosphere (i.e. offline). It generates a land surface analysis that is physically consistent with real-world input data, using a spatial resolution of 0.25° (Rui & Beaudoing, 2018). We use version 2.0, which is forced with the global meteorological dataset from Princeton University.

Scarce SM and ET observations over parts of SA create challenges to verify forecasts with observations alone (Spennemann et al., 2015). Thus, we must use reanalysis products (ERA5L and GLDAS) to verify the hindcasts. ERA5 (Baker et al., 2021) and GLDAS (Spennemann et al., 2015) represent well SM variability over SA, with general agreement between the datasets (Piles et al., 2019). Other studies (e.g. Spennemann et al., 2015) using reanalysis conclude that offline simulations not only avoid atmospheric model biases but also produce more consistent results. Further, there are no qualitative differences between results from ERA5 and ERA5L for our study (not shown), as ERA5L is a high-resolution "replay" of the land component of the ERA5 reanalysis. Alternative products are less reliable such as Modern-Era Retrospective Analysis (MERRA)-Land, which shows poor land–atmosphere coupling due to the non-stationarity in the satellite observations it assimilates, which artificially increases soil moisture variability (Dirmeyer, 2011). As the coupling results are strongly sensitive to the land surface model used (e.g. Zhang et al., 2008; Menéndez et al., 2019), which may skew our results in favour of one model, for example ECMWF if we only used ERA5L as the verification dataset. Thus, we use both datasets to not only understand the regions of agreement between reanalyses for land–atmosphere coupling but also regions with large uncertainties for forecast verification. We use reanalysis PR from ERA5L and GLDAS to calculate land–atmosphere coupling rather than observed PR, to have a physically consistent representation of land–atmosphere feedbacks, as in previous studies (e.g. Spennemann & Saulo, 2015). However, we compare reanalysis PR with observed PR from Climate Hazards Group

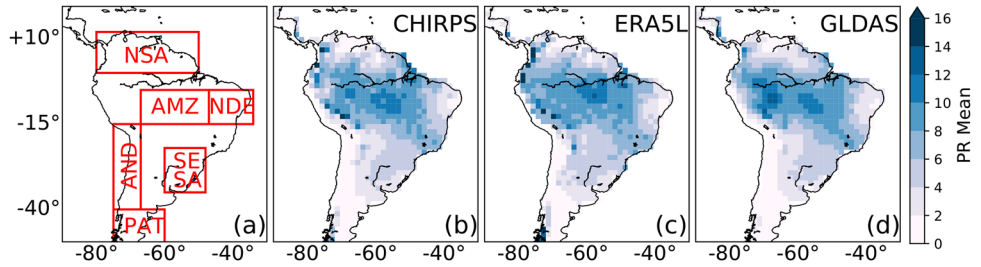

FIGURE 1 (a) The six regions of SA analysed in the study; (b–d) NDJFM mean PR (mm·day^{-1}) from (b) CHIRPS (observed), (c) ERA5L (reanalysis) and (d) GLDAS (reanalysis) over 1999–2010

InfraRed Precipitation with Station (CHIRPS) data (Funk et al., 2015) as a quality check.

As in Klingaman et al. (2021), we focus our evaluation across six regions of SA (Figure 1a), which have distinct precipitation regimes that are relatively homogeneous within the region:

1. NSA (0–12°N; 50–80°W) - Northern SA
2. AMZ (5–15°S; 47–67°W): Southern Amazonia
3. NDE (5–15°S; 34–47°W): Northeastern SA
4. SESA (22–35°S; 48–60°W): Southeastern SA
5. AND (15–40°S; 67–75°W): Central Andes
6. PAT (40–50°S; 60–75°W): Patagonia.

2.2 | Methods

We evaluate prediction performance and skill of weekly-mean values of three variables (SM, ET, T2) at lead times of one to five weeks over the extended austral summer (NDJFM) for 1999–2010. We interpolate the reanalyses and the hindcasts ensemble members to a common grid of 1.5° horizontal spatial resolution. For detailed analysis of the models' performance for PR, see Klingaman et al. (2021). We use deterministic (bias; root mean square error, RMSE; correlation coefficient, CC) and probabilistic (Ranked Probability Skill Score, RPSS) verification metrics to evaluate the hindcasts. Using multiple verification metrics allows for a holistic analysis of hindcast performance and skill. All deterministic verification metrics are calculated for the ensemble mean, which allows for a fair comparison of models with different ensemble sizes. We compute biases in the hindcast mean of forecast Week 1–5 for all three variables compared against ERA5L and GLDAS. The RMSE is calculated by comparing weekly anomalies of variables (Week 1–5) for the models and does not include contributions from the model mean bias. Anomalies for each year are calculated relative to the weekly 1999–2010 climatology, with the current year excluded to mimic an operational forecast process. We compute CC as correlation coefficients between weekly anomalies of reanalyses and hindcasts (Week 1–5). We consider CC values statistically significant at the 5% level. The CC critical value of statistical significance accounts for serial (lag-1) auto-correlation between consecutive forecast initializations at the same lead time, following the method of Zwiers & Von Storch (1995). The RPSS is computed for terciles (derived from the respective reanalysis or hindcast spread) of weekly mean variables and is evaluated against the climatological reference forecast, which has a probability of 1/3 for each tercile. For more details on the metrics used, see Klingaman et al. (2021).

To evaluate land–atmosphere feedbacks, we trace energy or moisture feedback pathways from the surface to the atmosphere using coupling metrics (Dirmeyer et al., 2014). This method is based on physical understanding of the factors that control land–atmosphere interactions, to identify areas of strong land–atmosphere coupling. The feedback pathway is broken into two "legs": the terrestrial leg, which measures the strength of coupling between a surface variable (SM) and a flux variable (ET); and the atmospheric leg, which measures the relationship between the flux variable (ET) and an atmospheric variable (PR or T2). These four variables allow us to understand the coupling strengths for the SM–ET–T2 and SM–ET–PR feedback pathways (Seneviratne et al., 2010).

The coupling strength for each leg of the pathway can be calculated using a coupling index (CI; Guo et al., 2006; Dirmeyer, 2011; Baker et al., 2021). We calculate CI for the terrestrial leg (T_{CI}; Equation 1) by calculating the regression slope between weekly anomalies of the surface variable (S'), that is SM, and the surface flux (F'), that is ET, multiplied by the standard deviation of the surface variable ($\sigma_{S'}$). For the atmospheric leg (A_{CI}; Equation 2), we calculate the regression slope between weekly anomalies of the surface flux (F'), that is ET, and the atmospheric variable (A'), that is PR or T2, multiplied by the standard deviation of the surface flux ($\sigma_{F'}$).

$$T_{CI} = \frac{dF'}{dS'} \times \sigma_{S'}, \tag{1}$$

$$A_{CI} = \frac{dA'}{dF'} \times \sigma_{F'}. \quad (2)$$

The CI will be weak if the linear association is strong, but the variability is low, or if the variability is high but the linear association is weak, as the index reflects the correlation magnitude as well as the variability of the dependent field. If the relationship between the land and atmosphere were linear, then the total feedback (TF) between S' and the A' could be calculated using CI only. However, as the relationships are non-linear, we calculate TF (Equation 3) between S' and A' using the two-legged metric (Dirmeyer et al., 2014; Spennemann & Saulo, 2015), as the product of CI for the terrestrial (T_{CI}) and atmospheric (A_{CI}) legs within the pathway.

$$TF = T_{CI} \times A_{CI}. \quad (3)$$

We also computed land–atmosphere coupling strengths with the Zeng's Gamma (Zeng et al., 2010) and temperature-evapotranspiration (Seneviratne et al., 2006) metrics, as it is essential to compare multiple coupling metrics for robustness (Lorenz et al., 2015; Spennemann et al., 2018). These metrics provide similar results to the CI (Figure S4 in the Supporting Information).

Land–atmosphere feedbacks are not always instantaneous due to the longer memory of the land surface and to intervening processes. To understand the effect of persistence of land–atmosphere fields on the covariability of their states, we use the feedback parameter (FP; Liu et al., 2006). The FP defines the lead–lag relationship between SM and PR, after removing the impact of auto-correlation from SM (Notaro, 2008). Using the Sun & Wang (2012) method, we calculate the FP as the lead–lag correlation between PR and SM in increments of forecast weeks (τ), divided by the auto-correlation in SM at the same lead or lag (Equation 4) for the six regions.

$$FP = \frac{corr[PR(t), SM(t \pm \tau)]}{corr[SM(t), SM(t \pm \tau)]}. \quad (4)$$

Using Equation (4), we calculate the correlation between the variables in a particular leg (e.g. SM and PR) for each increment of forecast week and divide it by the auto-correlation of the first variable (e.g. SM) for the same increment of the forecast week. We calculate FP for all legs of the land–atmosphere pathway (SM–ET, ET–PR and ET–T2) and for the full pathway (SM–PR and SM–T2). We calculate FP using the regional mean variables to represent the regional coupling feedback at different lead times for our six SA regions.

3 | LAND AND ATMOSPHERE STATE

We analyse the performance of four model hindcasts for the climatology and weekly variability of land surface and near-surface atmospheric fields (SM, ET and T2). The same models' performance for PR is analysed in detail by Klingaman et al. (2021) against CHIRPS. As this study uses PR from ERA5L and GLDAS rather than CHIRPS (see Section 2), we first discuss the differences between these datasets. ERA5L and GLDAS (Figure 1c,d) show a similar NDJFM climatology to CHIRPS (Figure 1b). There is a northwest–southeast-oriented band of high PR, with maxima in AMZ, the eastern slopes of Andes and parts of the Atlantic coast of Brazil, with minima over the northern NSA, NDE and the western slopes of Andes. GLDAS shows lower PR over AMZ, the eastern slopes of Andes and eastern PAT than CHIRPS and ERA5L. ERA5L shows higher PR over eastern NSA than CHIRPS and GLDAS.

The spatial pattern of climatological SM during NDJFM in ERA5L and GLDAS (Figure 2a,f) is similar to that of PR (Figure 1b,c), as expected. Regions with lower PR during this season have lower SM (NDE, southern and southwestern SA); and regions with higher PR show higher SM (AMZ, parts of Andes and Atlantic coast of Brazil). ET maxima are seen over parts of SESA, linked to the higher SM availability (Seneviratne et al., 2010); ET minima are seen over parts of AND, PAT, northernmost SA and NDE (Figure 2k,p). However, the regions with maximum SM, AMZ and its surrounding regions, concurrent with regions of maximum PR, show moderate ET as the region has an energy-limited ET regime rather than an SM-limited ET regime (also in Spennemann et al., 2015; Menéndez et al., 2016). T2 is lowest over regions with high orography (AND) and over the mid-latitudes (southern SA) (Figure 2u,z). Despite differences in magnitude, the spatial variability of all three variables is similar in ERA5L and GLDAS, which leads to similar spatial patterns of biases in S2S models (discussed below). GLDAS shows lower weekly climatological mean SM than ERA5L, which may lead to weaker ET and higher T2 in GLDAS than ERA5L. Despite the similarities in ERA5L and GLDAS, the differences between them lead to uncertainty in the "observed" state. At the end of Section 3, we discuss the implication of the reanalysis uncertainty and of verifying hindcasts against a reanalysis that uses the same physical model.

CPTEC and NCEP have strong dry SM biases, whereas ECMWF and UKMO have weak wet biases over most of SA against ERA5L (Figure 2b–e). The sign of SM biases against GLDAS is the same as those against ERA5L in all models; however, the magnitudes of the biases differ (Figure 2g–j) due to differences between ERA5L and GLDAS. Over eastern SA, the models show a positive bias

Climate Resilience and Sustainability
Interdisciplinary Approaches Towards Solutions for Climate Change
Open Access
RMetS

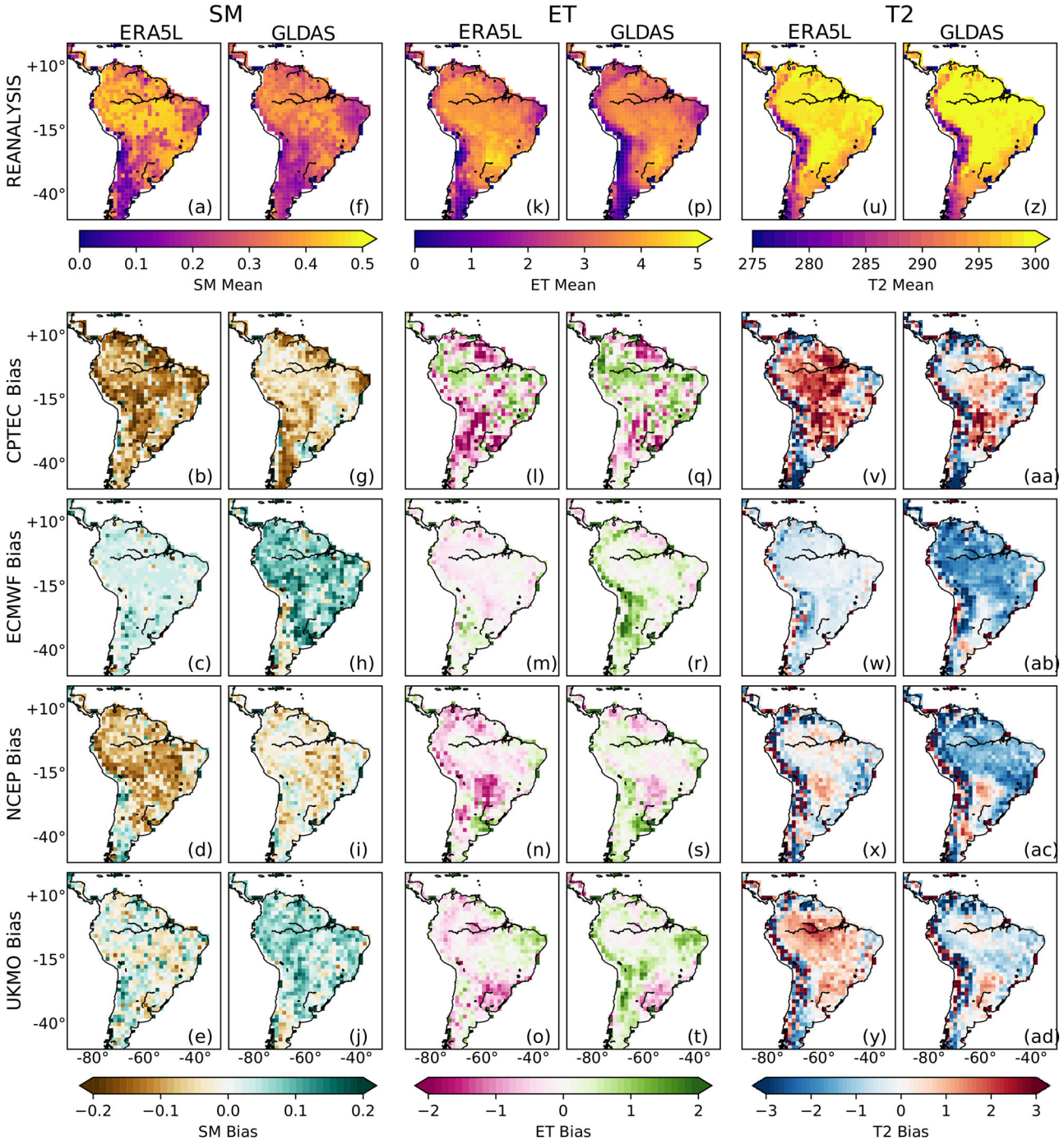

FIGURE 2 NDJFM climatological (a) SM ($m^3 \cdot m^{-3}$) from ERA5L and (b–e) SM biases ($m^3 \cdot m^{-3}$) against ERA5L from S2S hindcasts at Week 1 lead time from (b) CPTEC, (c) ECMWF, (d) NCEP and (e) UKMO over the 1999–2010 period. (f–j) is the same as (a–e) but against GLDAS as reanalysis. (k–t) is the same as (a–j) for mean ET ($mm \cdot day^{-1}$) and (u–ad) is the same as (a–j) but for mean T2 (K). Note that the first row uses different colourbars, placed below the respective panels; hindcast bias colourbars are placed at the bottom of each column

in ET, whereas over western parts of SA the models show a negative bias against ERA5L (Figure 2l–o). Over most of SA, biases against GLDAS are positive except over northern SA and parts of SESA (Figure 2q–t). All four models have a strong negative ET bias over SESA against both reanalyses (weakest in ECMWF and strongest in CPTEC), which is associated with the warm T2 bias over the same region (Figure 2v–y). SESA has warm and dry conditions in reanalyses and the models, but the models show a warmer T2 in comparison. The negative ET bias and the warm T2 bias might be influenced by errors in the representation of the SM–ET–T2 pathway, which is discussed further in Section 4. ECMWF and NCEP have cold T2 biases over most of SA against ERA5L, except in parts of SESA

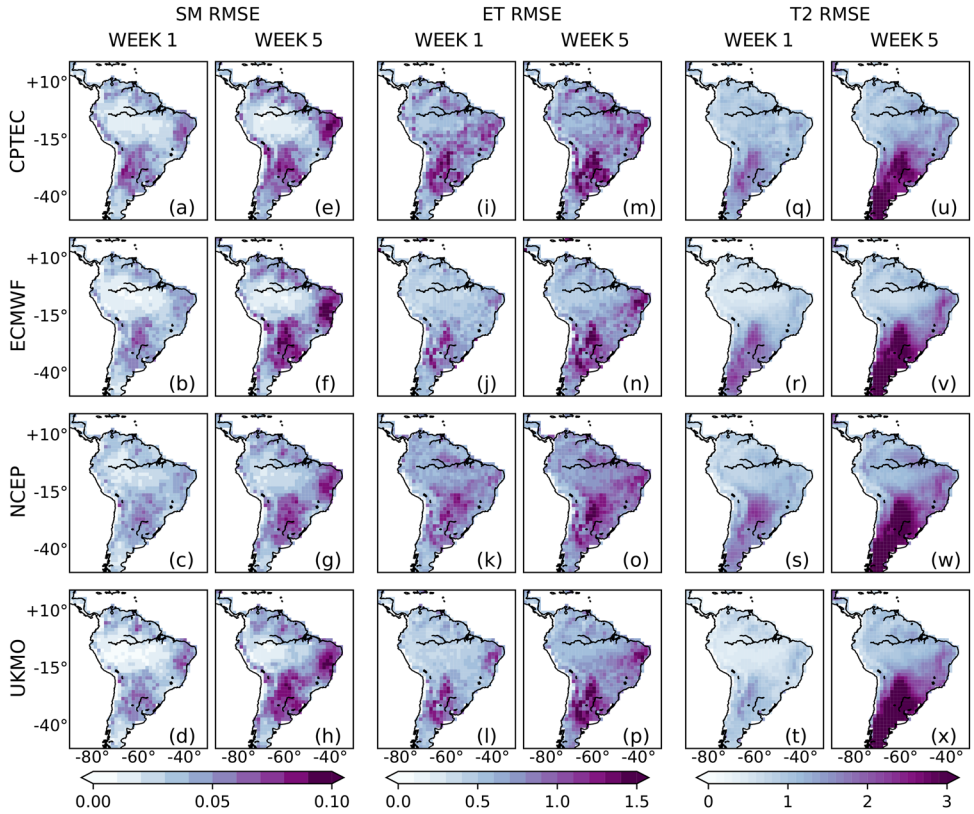

FIGURE 3 RMSEs in hindcast NDJFM weekly-mean SM anomalies ($m^3 \cdot m^{-3}$) against ERA5L for (a,e) CPTEC, (b,f) ECMWF, (c,g) NCEP and (d,h) UKMO at (a–d) Week 1 and (e–h) Week 5 lead times over the 1999–2010 period. (i–p) is the same as (a–h) but for weekly-mean ET anomalies ($mm \cdot day^{-1}$) and (q–x) is the same as (a–h) but for weekly-mean T2 anomalies (K). Note that each variable uses a separate colourbar, placed below the respective columns

and PAT, whereas CPTEC and UKMO have warm biases except over parts of NDE and NSA (Figure 2v–y). The signs of the biases remain the same against GLDAS, but the magnitudes increase in EMCWF and NCEP and reduce in CPTEC and UKMO (Figure 2 aa–ad). Over the Andes, all models except ECMWF show highly variable biases against both reanalyses, which may be associated with difficulties in modelling processes over complex topography. Similar to PR (Klingaman et al., 2021), the biases of all three variables are well established by Week 1 and show relatively little drift with lead time (not shown). The Week 1 biases in SM most probably stem from initial condition biases, as demonstrated in Dirmeyer & Halder (2017).

Among the six regions, NSA, NDE and SESA show the strongest RMSE for SM weekly anomalies against ERA5L for Week 1 in all S2S models (Figure 3a–d), as for PR (Klingaman et al., 2021). For ET (Figure 3i–l) and T2 (Figure 3q–t), the models also show high errors over NDE and SESA but not over NSA. The errors increase from Week 1–5 for all the three variables (Figure 3e–h, m–p, u–x), as expected. Models also show a steep increase with lead time in errors over southern SA for T2, associated with stronger mid-latitude temperature variability. NCEP shows the lowest errors in SM over all regions except over the AMZ, whereas ECMWF shows the lowest errors in ET and T2. The spatial pattern of errors in all models against GLDAS are the same as with ERA5L, with only small differences in the magnitude (Figure S2 in the Supporting Information). The models also perform well for CC (i.e. CC between models and reanalysis is significant at the 5% level) over most of SA for Week 1 against ERA5L. This performance remains significant to Week 5 over NSA, NDE, SESA (Figure 4). Models show poor performance (i.e. CC between models and reanalysis is not significant at the 5% level) by Week 3 for SM and ET over AMZ and central SA and by Week 4 for T2 over southern SA, with CPTEC showing poor performance by Week 2. As with RMSE, models show lower CC against GLDAS than ERA5L, with poor performance of SM and ET over AMZ and the surrounding regions even in Week 1 (Figure S3 in the Supporting Information).

Seneviratne et al. (2006) summarized that regions with lower SM have lower ET, which leads to higher T2 due to increases in the sensible heat flux, while regions with higher SM and higher ET tend to have lower T2. Furthermore, higher (lower) T2 leads to a higher (lower) vapour pressure deficit, which increases (decreases) ET and may

Climate Resilience and Sustainability
Interdisciplinary Approaches Towards Solutions for Climate Change
Open Access
RMetS

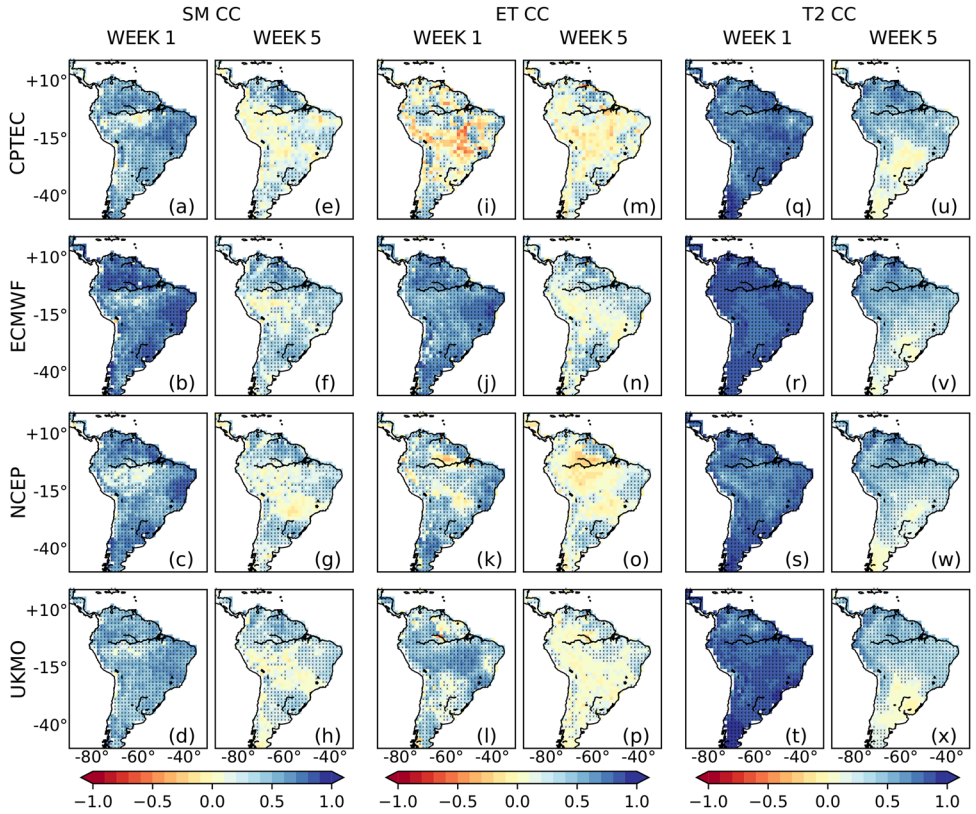

FIGURE 4 Same as Figure 3 but with CC (shaded) and statistical significance at the 5% level (stippling)

thus reduce (increase) SM. This might explain why the biases in all hindcasts for SM and T2 are strongly related (Figure 2) to regions with dry SM biases showing warm T2 biases and regions with wet SM biases showing cold T2 biases over most of SA for Week 1. ET biases do not show a strong relationship with SM over parts of SA, as the relationship between SM and ET depends highly on the regional moisture regime. Over SESA, there is a strong negative ET bias in all four hindcasts (strongest in CPTEC and weakest in ECMWF) and a related warm bias in T2. This may indicate an erroneously strong relationship between ET and T2 in models as compared to reanalyses. Parts of the AMZ region in all the models show poor performance (represented by insignificant CC) for SM and ET, whereas parts of southern SA and SESA show poor performance for T2 (Figure 4). All four investigated models show high errors for all three variables in almost the same regions of SESA even in Week 1, which worsens by Week 5 (Figure 3). The exception to this is T2, which shows high errors over southern SA for Week 5, compared to high errors in SM and ET over SESA, NDE and parts of NSA. Similarly, models also show poor skill (measured by CC) over the same regions (Figure 4), which may indicate erroneous coupling in the SM–ET–T2 pathway, which is discussed further in Section 4.

The forecast skill for tercile categories of SM is measured by RPSS and is compared against the climatological reference forecast. Models show a better RPSS than the climatological reference forecast over NSA, NDE and SESA for ECMWF and NCEP and worse for CPTEC and UKMO, with ECMWF having the highest skill (Figure 5a–f). Skill for all models is worse than the climatological reference forecast over AND, PAT and AMZ (except ECMWF) even for Week 1, which may be due to difficulty in modelling SM over complex topography (AND and PAT) and highly forested regions (AMZ). Skill for ET (Figure 5g–l) is worse than the climatological reference forecast for all models except ECMWF from Week 1 over the six regions, except for NCEP for NDE compared against ERA5L (Figure 5g–l). Forecast skill for T2 (Figure 5m–r) in all models is generally better than the climatological reference forecast over all six regions for Weeks 1 and 2, with ECMWF showing the highest and CPTEC showing the lowest skill, likely due to underdispersion of the CPTEC ensemble (Guimarães et al., 2020). The models have better skill than the climatological reference forecast over NSA even at Week 5, perhaps due to lower T2 variability in the tropics. Although models have higher skill in Week 1 over PAT for T2, models lose skill rapidly with lead time. Forecast skill for all three variables converges towards climatological skill for all models

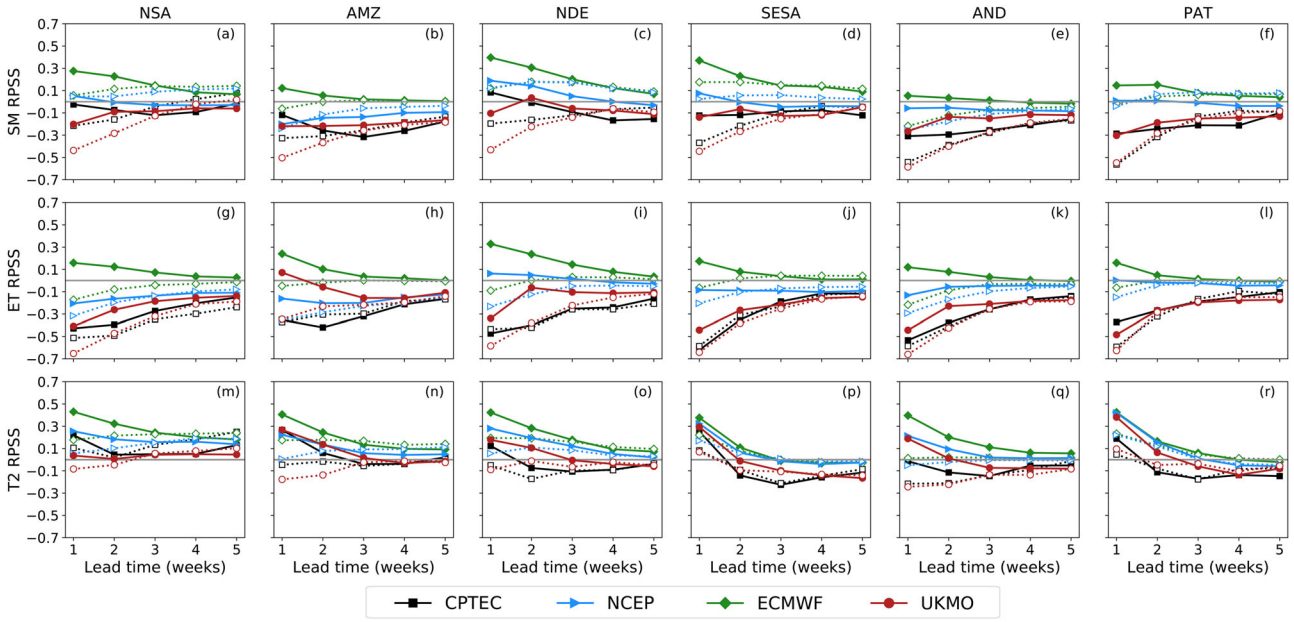

FIGURE 5 Regional-mean RPSS for NDJFM weekly-mean tercile categories of (a–f) SM for each model's hindcast distribution, computed relative to a climatological forecast over the 1999–2010 period, for (a) NSA, (b) AMZ, (c) NDE, (d) SESA, (e) AND and (f) PAT. Solid lines and filled markers show RPSS for hindcasts against ERA5L; dotted lines and unfilled markers show RPSS for hindcasts against GLDAS. (g–l) is same as (a–f) but for weekly-mean ET and (m–r) is same as (a–f) but for weekly-mean T2. Grey horizontal line shows RPSS equal to zero. RPSS is computed on the original 1.5° grid, then averaged over the region. The regions are shown in Figure 1a

by Week 5. As with PR (Klingaman et al., 2021), the forecast skill for tercile categories of the three variables in the S2S models is linked to higher skill in the above- and below-normal categories of variables, relative to skill for the normal (middle tercile) category (not shown).

Uncertainty in the verification data (reanalyses) leads to uncertainty in quantifying hindcast error and skill. For understanding and quantifying the uncertainty in the verification data, to have more confidence in our results, we verify against both GLDAS and ERA5L. Models generally show poorer skill when verified against GLDAS than ERA5L (Figure 5). However, ECMWF and UKMO have lower biases when compared to ERA5L, and CPTEC and NCEP have higher biases when compared against GLDAS (Figure 2). Similarly, ECMWF consistently performs best (with positive RPSS over all regions) against ERA5L for all three variables; however, this is not a fair comparison as the ECMWF hindcasts and ERA5L use the same physical model. ECMWF also has the highest skill when verified against GLDAS, but the RPSS values are negative for ET for all six regions (Figure 5g–l) and SM for AMZ and AND (Figure 5b,e). This shows that although we are confident that ECMWF outperforms the climatological reference forecast for SM and T2, ET skill is more uncertain due to disagreements between ERA5L and GLDAS. There is also higher confidence that NCEP skill for SM (Figure 5a–f) and T2 (Figure 5m–r) is better than the climatological reference forecast, as skill for NCEP verified against ERA5L and GLDAS are similar. Further, there is higher confidence that UKMO and CPTEC skill is worse than the climatological reference forecast for SM (Figure 5a–f) and ET (Figure 5m–r), against both verification datasets. We are also confident that models fail to represent the variables in specific regions (parts of AMZ for SM, AMZ and NSA for ET and southern SA for T2) by Week 5 as the errors are high (Figure 3 and Figure S2 in the Supporting Information) and CC insignificant (Figure 4 and Figure S3 in the Supporting Information) when compared against both ERA5L and GLDAS.

4 | LAND–ATMOSPHERE COUPLING

In this section, we discuss the land–atmosphere coupling in the hindcasts. We first look at the spatial variability of CI for the terrestrial (T_{CI}; Figure 6) and atmospheric (A_{CI}; Figure 7) legs and the total feedback pathways (TF; Figure 8) over SA. The regional coupling strengths are shown as matrices of lead–lag coupling (using FP; Figure 9). The matrices in Figure 9 show each FP relationship over a region, with one 2×9 matrix for reanalyses, and four 5×9 matrices, one for each model. The x-axis of the reanalyses matrix shows a column each for ERA5L and GLDAS. The x-axis of each model matrix represents the five lead-weeks (Week 1–5) of the hindcasts. The y-axis for all the matrices represents the FP relationship at coincidence (0), four lead

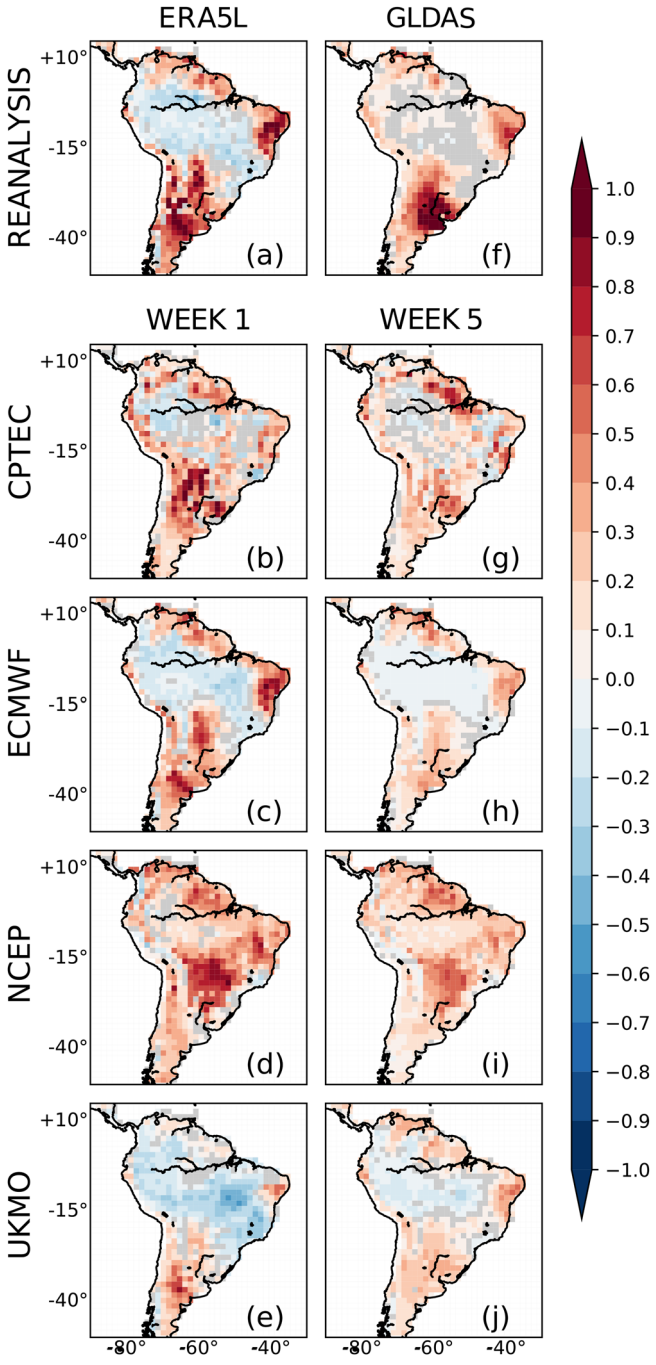

FIGURE 6 T_{CI} between SM and ET (mm·day^{-1}) for (a) ERA5L, (f) GLDAS and the S2S hindcasts: (b) CPTEC, (c) ECMWF, (d) NCEP, and (e) UKMO at (b–e) Week 1 and (g–j) Week 5 lead times over the 1999–2010 period. Insignificant CI ($p < 0.01$) is shown in grey shading

weeks (+1 to +4) and four lag weeks (−1 to −4). For SM–ET FP over AMZ (Figure 9a), the negative values on the *y*-axis represent ET leading SM and positive values on the *y*-axis represent SM leading ET, with a similar format for other legs (Figure 9b–e). The grid cells within each matrix represented with grey 'x' show the incremental week that cannot be calculated, as it does not exist. For brevity, we only discuss AMZ and SESA (Figure 9f–j) here, due to the large errors in hindcasts over these regions (refer to Figure S4, in the Supporting Information, for the other four regions).

SM–ET coupling (Figure 6) is the terrestrial leg of both the SM–PR and SM–T2 feedback pathways (Seneviratne et al., 2010). Over SA, SM–ET coupling in ERA5L is positive over all subregions except AMZ and central SA, where the relationship is negative (Figure 6a). These regions have negative SM–ET interactions as they are energy-limited regimes, where due to high SM, ET variability depends mainly on net radiation rather than SM. Positive SM–ET coupling in other regions of SA indicates that these are SM-limited ET regimes, where SM is the first-order constraint on ET. Regions with no significant SM–ET relationship are transition zones between the SM-limited and energy-limited ET regimes. Spennemann & Saulo (2015) and Baker et al. (2021) found that these results also hold on monthly scales; however, the transition zone (non-significant SM–ET interaction) is larger in this study as we analyse subseasonal timescales. It should be noted that the non-significant results for the SM–ET interaction will be sensitive to the significance level used. In GLDAS (Figure 6f), there is no energy-limited regime in SA, as AMZ has a positive SM–ET relationship, which may be associated with lower SM in GLDAS over AMZ. GLDAS and ERA5L show strong positive SM–ET coupling over SESA (as in Spennemann et al., 2018) and NDE, but GLDAS has much stronger coupling over SESA than ERA5L.

All models except NCEP show negative SM–ET coupling over AMZ and central SA and positive coupling over the rest of SA for Week 1 (Figure 6b–e) like in ERA5L. While CPTEC shows transition zones over parts of AMZ, ECMWF shows no transition zones, and UKMO shows stronger negative SM–ET coupling than ERA5L and weaker positive SM–ET coupling than both reanalyses. NCEP shows positive SM–ET coupling over most of SA, with a coupling pattern similar to GLDAS, which may be a result of using the same land surface model in both datasets and a dry SM bias. Despite the coupling errors, the models represent the maxima over NDE and SESA, even though the positions are slightly displaced in all models. These coupling errors in the models may be linked to the strong SM dry bias in CPTEC and NCEP, the strong ET biases in CPTEC, NCEP and UKMO (Figure 2) and the biases in SM and ET variability (Figure 4). Coupling strength in the models declines from Week 1–5, especially over the regions of maxima (Figure 6g–j), further shown with FP analysis (Figure 9), which may be associated with the models' skill for SM and ET (Figure 5a–l) converging to the climatological reference forecast by Week 5 as discussed in Section 3.

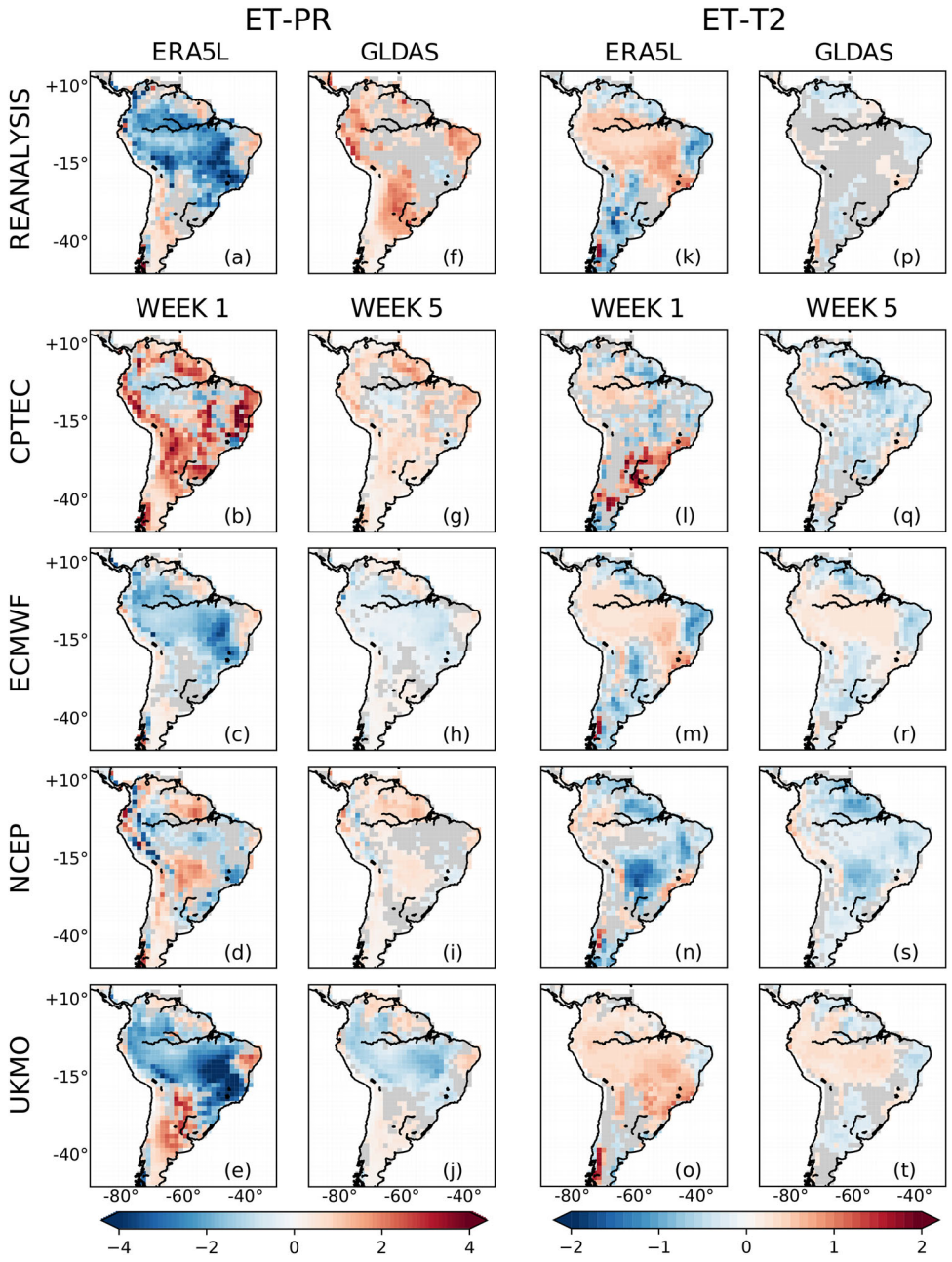

FIGURE 7 Same as Figure 6 but (a–j) for A_{CI} between ET and PR (mm·day^{-1}) and (k–t) for A_{CI} between ET and T2 (K)

Over AMZ, ERA5L shows negative coupling at lag 0 and ET leading SM at a one-week lag, whereas GLDAS does not show significant coupling at any lag between SM–ET (Figure 9a). ECMWF and UKMO represent the negative coupling similar to ERA5L, but the relationship persists longer, which shows that increasing ET reduces SM. For NCEP and CPTEC, the sign of coupling is positive for all lags, which disagrees with both reanalyses and suggests that SM and ET biases strongly reinforce each other. Over SESA, both reanalyses show positive SM–ET coupling at all leads and lags, with GLDAS showing stronger coupling than ERA5L (Figure 9f). Over an SM-limited region like SESA, SM is a persistent constraint for ET variability, but ET also influences SM variability. All models show positive coupling at all lags, similar to the reanalyses but the coupling in CPTEC, NCEP and GLDAS is stronger than in ECMWF, UKMO and ERA5L.

The atmospheric ET–PR interaction (Figure 7a,f) is positive over parts of NSA, NDE, AND and PAT in both ERA5L and GLDAS. Over SESA, ERA5L shows insignificant coupling (Figure 7a) and GLDAS shows positive coupling (Figure 7f). For the rest of SA, ERA5L shows strong negative coupling whereas GLDAS shows insignificant coupling. Negative coupling between ET–PR over AMZ and

Climate Resilience and Sustainability
Interdisciplinary Approaches Towards Solutions for Climate Change
Open Access
RMetS

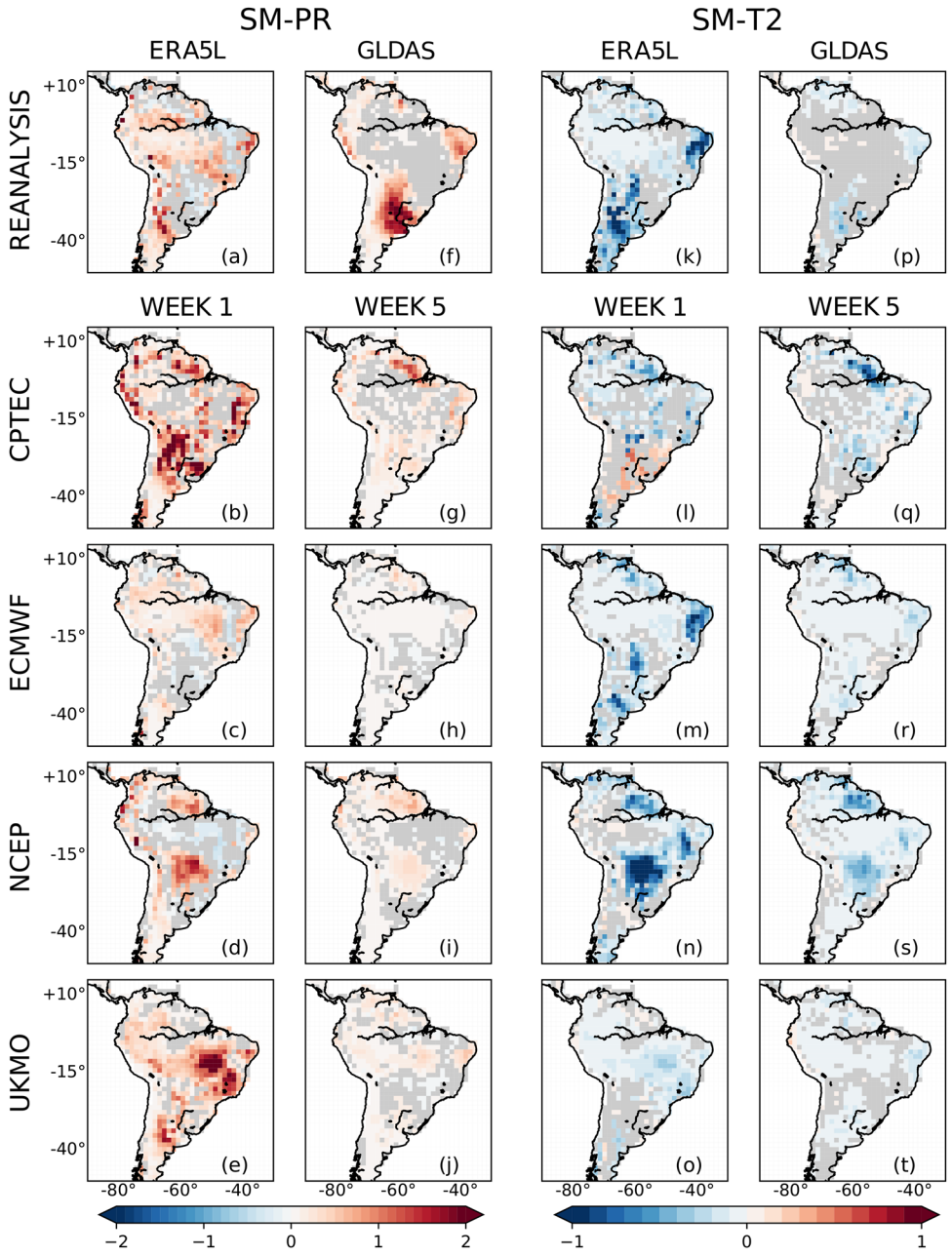

FIGURE 8 Same as Figure 6 but (a–j) for TF between SM and PR ($mm^2 \cdot day^{-2}$) and (k-t) for TF between SM and T2 ($mm \cdot K \cdot day^{-1}$)

positive coupling over SESA is also found at monthly scales (Baker et al., 2021). ET–PR coupling is highly uncertain and involves multiple complex processes like convective instability, moisture recycling, boundary-layer instability and land surface heterogeneity (e.g. Eltahir, 1998; Seneviratne et al., 2010; Tawfik et al., 2015). The models show positive ET–PR coupling over NSA, NDE and PAT (i.e. peripheral regions of SA), and negative coupling over the rest of SA for Week 1 (Figure 7b–e). However, the strength of coupling is overestimated in CPTEC (in the positive ET–PR regions) and UKMO (in the negative ET–PR regions). Similar to SM–ET coupling, only NCEP shows no coupling over parts of AMZ, like in GLDAS, and negative coupling over the central parts of AMZ. Compared to the reanalyses, ECMWF shows weaker ET–PR coupling over PAT.

ET–PR regional coupling (Figure 9b,g) has a similar sign for SM–ET coupling in ERA5L and the models. For ET–PR in AMZ (Figure 9b), the models show significant coupling with PR leading ET, whereas coupling with ET leading PR is significant only up to a one-week lag. As with SM–ET coupling, CPTEC and NCEP show the wrong sign of coupling for ET–PR. For SESA (Figure 9g), ET–PR coupling is positive in reanalyses and the S2S models, with no coupling of ET leading PR in ERA5L, CPTEC, ECMWF and UKMO.

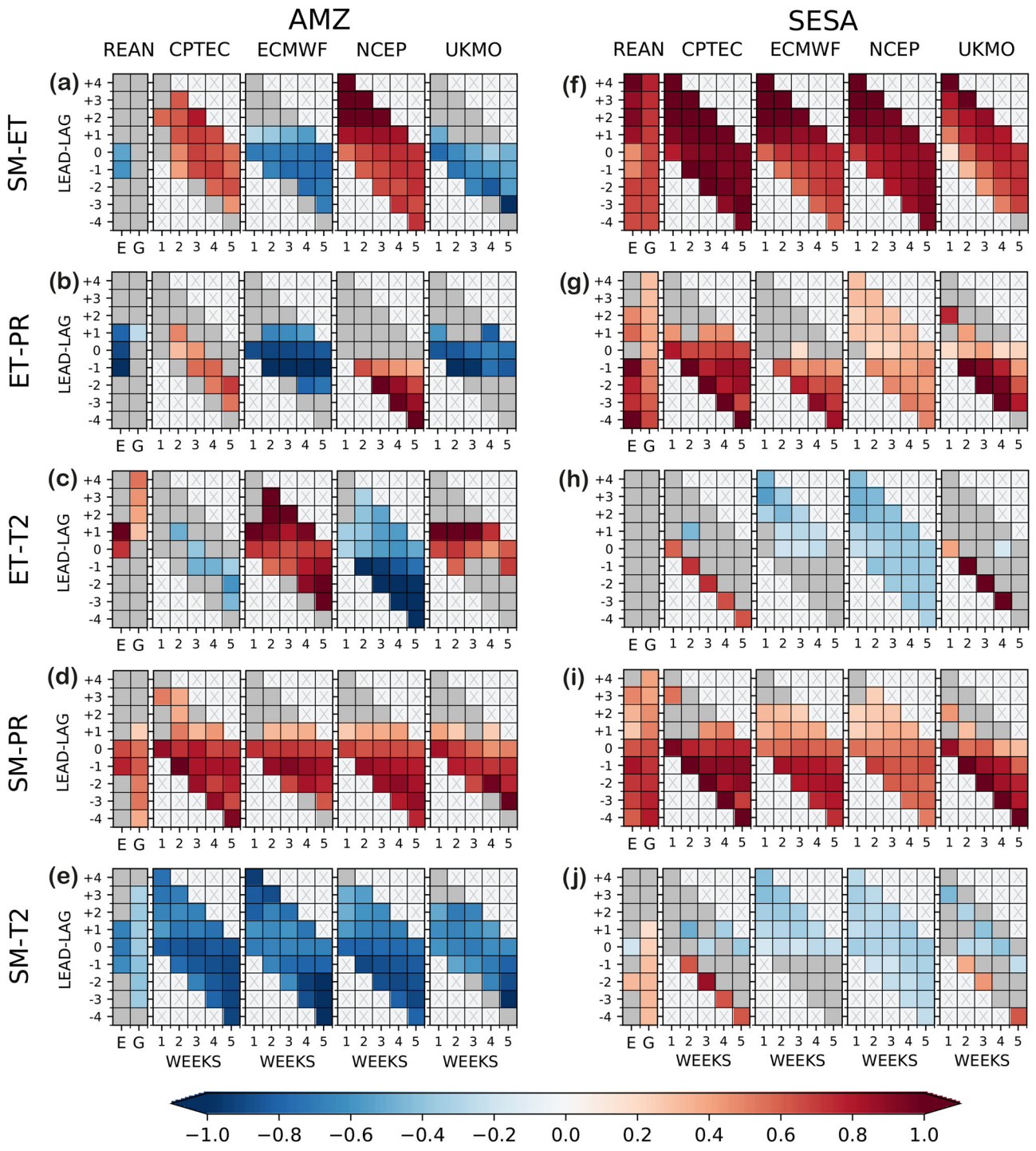

FIGURE 9 (a) Lead–lag regional-mean FP between SM and ET (unitless) for reanalyses (REAN; ERA5L and GLDAS marked as 'E' and 'G', respectively) and S2S hindcasts (CPTEC, ECMWF, NCEP and UKMO) at different lead weeks for AMZ over the 1999–2010 period. The reanalyses have one 2×9 matrix, and the four models have one 5×9 matrix each. The *x*-axis of the reanalyses matrix shows a column each for ERA5L and GLDAS. The *x*-axis of each model matrix represents the five lead-weeks (Week 1–5) of the hindcasts. The *y*-axis for all the matrices represents the FP relationship at coincidence (0), four lead weeks (+1 to +4) and four lag weeks (−1 to −4). The negative values on the *y*-axis represent ET leading SM and positive values on the *y*-axis represent SM leading ET. Insignificant FP ($p < 0.01$) is shown in grey shading. Grey crosses mark the points where lead–lag FP cannot be calculated. (b–e) are same as (a) but (b) shows regional-mean FP between ET and PR (unitless), (c) shows regional-mean FP between ET and T2 (unitless), (d) shows regional-mean FP between SM and PR (unitless) and (e) shows regional-mean FP between SM and T2 (unitless). (f–j) is same as (a–e) but for SESA. FP is calculated for regional-mean variables

However, the same feedback is well established at all lags for GLDAS and NCEP (Figure 9g), which again reiterates the similarities in these two datasets.

Positive ET–T2 coupling indicates regions under 'atmospheric control', that is T2 influencing ET in energy-limited regimes; negative coupling indicates regions under 'land surface control', that is ET influencing T2 in SM-limited regimes (Spennemann & Saulo, 2015). Thus, ET–T2 (Figure 7k,p) shows similar patterns but opposite signs to SM–ET (Figure 6). The sign of ET–T2 is also opposite to that of ET–PR over the same regions in ERA5L, but GLDAS shows insignificant ET–T2 over most of SA. For ET–T2, the models' representation is inconsistent for Week 1 (Figure 7l–o). CPTEC shows positive ET–T2 over the regions surrounding the Amazon River like in ERA5L, and no coupling over some parts of SA like GLDAS. However, CPTEC shows very strong positive coupling over SESA unlike any other reanalysis or model, which may be associated with the strong positive T2 bias (Figure 2v,aa) or from the negative ET bias (Figure 2l,q) over the same region. NCEP, like GLDAS, shows limited regions of positive coupling around AMZ but has a strong negative ET–T2 hotspot over central SA, as also seen in SM–ET (Figure 6d) and ET–PR (Figure 7d), which may stem from the negative ET bias in NCEP (Figure 2n,s). UKMO shows weak negative coupling and ECMWF has stronger negative coupling compared to the reanalyses over NSA, NDE and PAT. As with SM–ET, the coupling strength for the atmospheric legs in the S2S models weakens from Week 1–5, except for ET–T2 for CPTEC and UKMO over SESA, where the positive coupling in Week 1 changes to negative coupling by Week 5.

ET–T2 in ERA5L shows positive coupling only with no lag or ET leading T2 by one week; however, GLDAS shows positive coupling only with ET leading T2 up to lag of four weeks over AMZ (Figure 9c). In the models, ECMWF and UKMO show the correct positive ET–T2, whereas CPTEC and NCEP show erroneous negative ET–T2. For SESA, ERA5L and GLDAS show no coupling at any lag, whereas ECMWF and NCEP show negative coupling (Figure 9h). CPTEC and UKMO mostly show insignificant coupling, except for strong positive coupling with T2 leading ET at one-week lag. Figure 9 also indicates that coupling in the models weakens with lead time.

Total feedbacks for the pathways SM–PR and SM–T2 summarize the interactions between the terrestrial and atmospheric legs (Figure 8). SM–PR is positive over most of SA (Figure 8a–j), while SM–T2 is negative (Figure 8k–t). In these models, there are strong regional variations in the strength of the feedbacks, as also shown by Orlowsky & Seneviratne (2010). For AMZ (Figure 9d,e), ERA5L shows stronger but less persistent SM–PR and SM–T2 than GLDAS, whereas all the models show stronger as well as more persistent coupling than both reanalyses. CPTEC and NCEP also represent the correct sign of the SM–PR and SM–T2 pathways, despite incorrectly representing the atmospheric and terrestrial legs of these pathways. For SESA, SM–PR (Figure 9i) magnitude is well represented by the models but the feedback of SM leading PR is insignificant, whereas the models fail to represent SM–T2 (Figure 9j), due to errors in the representation of ET–T2 (Figure 9h).

5 | DISCUSSION

Errors in analysis datasets, for variables like SM and ET, due to sparse observations and poor representation of physical processes in background models make it challenging to verify and improve operational forecasts over regions like AMZ. It is important to use multiple verification products, particularly in data-sparse regions and in cases where the verification dataset uses the same physical model as the forecast. The two reanalyses show similar spatial patterns of SM, ET and T2, and, thus, the pattern of biases in the S2S models compared against ERA5L and GLDAS is similar. There is higher confidence in results when models perform similarly against both verification datasets. For example, we have high confidence that ECMWF has the highest skill amongst the four models for SM and T2, as this result holds verifying against both ERA5L and GLDAS.

However, the intensity of the biases and coupling strength differs between reanalyses and this leads to uncertainty in the model verification results and fidelity of the reanalyses. At shorter lead times, the errors in the variables are smaller and thus the associated coupling strengths would be more realistic, as expected. At shorter lead times (e.g. Week 1; Figure 3), when the model state is closer to the initial conditions, errors in land–atmosphere variables are smaller, and thus the associated coupling strengths are more realistic, as expected (Figure 6). At longer lead times (e.g. Week 5; Figure 5), when the model state has drifted closer to the model climatology, the errors increase and the coupling strengths weaken (Figure 6). Previous research shows that initial biases in land surface variables are associated with the errors in the initialization (Dirmeyer & Halder, 2017) and that land–atmosphere interactions depend strongly on the specific model physics (Zhang et al., 2008). Similarities in feedbacks between ECMWF versus ERA5L and NCEP versus GLDAS may be linked to the use of the same physical models. ECMWF and ERA5L use different versions of the same coupled model (IFS) and the same land surface model (HTESSEL); and NCEP and GLDAS use the same land surface model (NOAH). Similar SM biases in ECMWF and UKMO may also be attributed to initialization with the same dataset (Table 1). For this reason, we evaluated all models against two

reference datasets and have greatest confidence in our findings when model performance validated against both reference datasets was similar.

As with PR (Klingaman et al., 2021), the model biases for SM, ET and T2 are established by Week 1. The Week 1 biases for SM are possibly associated with the biases in the initialization of SM, as also shown in Dirmeyer & Halder (2017). Biases and errors in SM show a clear connection to T2 in Week 1, possibly through the changes in sensible heat flux and vapour pressure deficit (Seneviratne et al., 2006). NSA, NDE and SESA have relatively moderate mean PR, compared to the heavier PR in AMZ. These three regions are considered SM-limited regimes, where land surface features can significantly influence the atmosphere (Seneviratne et al., 2010). Good skill for SM in the models over NSA and PAT, as shown by low RMSE and significant CC, is not reflected in skill for PR (Klingaman et al., 2021). According to Dirmeyer et al. (2018), this mismatch in skill between SM and PR might stem from errors in convective parameterization. The only SM available in the S2S datasets is the value averaged over the top 20 cm soil. This SM does not include moisture from deeper layers, which can also influence the land–atmosphere coupling. It should also be noted that the SM–PR coupling we discuss in this work represents the statistical relationship between SM and PR, which primarily represents the local recycling of moisture, which only provides 9–10% of all moisture for precipitation over SA; the majority of moisture comes from remote sources via advection (Zemp et al., 2014). Thus, errors in PR from moisture advection likely play a leading role in the model forecasts; PR errors in forecasts cannot be explained only by errors in land–atmosphere feedbacks. Further research linking PR error to errors in moisture transport and convergence would provide more insights about PR errors in the S2S models.

Hindcasts show poor performance for all three variables over regions with relatively higher climatological PR, like AMZ and parts of central SA, which have humid tropical or monsoonal subtropical climates. Over these regions, during NDJFM, the land surface may not substantially influence PR due to short SM memory (Dirmeyer et al., 2019) and high SM, making these regions energy limited rather than moisture limited (Seneviratne et al., 2010). ERA5L represents this region as energy limited, whereas the coupling strength in GLDAS is insignificant. This difference between the reanalysis datasets leads to higher uncertainty for skill in the investigated models. ECMWF and UKMO show similar energy-limited regimes as ERA5L over AMZ and central SA. NCEP and CPTEC show an energy-limited regime only over the regions surrounding the Amazon River, with a moisture-limited regime over the southeastern Amazon that may stem from dry SM biases in both models. NCEP shows larger regions with negligible coupling over AMZ and central SA like GLDAS. NCEP also has a strong hotspot of land–atmosphere coupling over central SA, which may be linked to errors in ET and T2. Due to this uncertainty and differences in the representation of the land–atmosphere state and coupling, there is a need to identify the sources of error in the modelled processes. To diagnose the source of errors in the land–atmosphere feedback, more in-depth analysis of the model process parameterizations is required (e.g. Guo et al., 2006).

6 | CONCLUSIONS

Errors in the representation of land–atmosphere feedbacks may limit prediction performance in models (Dirmeyer et al., 2019). Models are known to represent well the seasonal variation in the land–atmosphere moisture feedbacks over most of SA, except over the Amazon, which may be due to biases in PR and SM control processes (Baker et al., 2021). Previous research shows that S2S models have moderate performance at forecasting weekly rainfall over SA at subseasonal lead times (de Andrade et al., 2019; Klingaman et al., 2021). In this study, we identify inter-model and inter-region differences in the simulated land–atmosphere state and coupling in the same four models (CPTEC, ECMWF, NCEP and UKMO) compared against two reanalyses (ERA5L and GLDAS). We focus our analysis on SM–ET–PR and SM–ET–T2 feedback pathways, discussed in Seneviratne et al. (2010). This study contributes to understanding of the representation and predictability of the hydrological cycle over SA, during extended austral summer (NDJFM), at lead times of one to five weeks.

Biases in the land–atmosphere states are well established by Week 1 and increase with lead time. All models demonstrate the good prediction performance (measured by significant CC at Week 5) for SM, ET only over NSA, NDE and SESA, and for T2 in tropical and subtropical SA. Model skill for SM, ET and T2 persists for longer than the skill for PR shown in Klingaman et al. (2021). The model skill by Week 5 degrades to be similar to that of climatological reference forecast, which may be associated with weaker coupling. The errors for SM and ET are slower to develop in the models than the errors for PR due to the slowly varying nature of the land surface. The SM biases in models (stemming from the SM initialization) define the hydrological regime in the model. The dry SM biases in CPTEC and NCEP lead to an erroneous simulated SM-limited regime in AMZ. These errors in the terrestrial leg for CPTEC and NCEP also lead to errors in both the atmospheric legs of the pathways over AMZ. However, all four models represent the area-averaged feedbacks over AMZ in the overall SM–PR and SM–T2 pathways well, but CPTEC and NCEP have incorrect feedbacks in the individual legs.

CPTEC and NCEP do not show the correct feedbacks for the SM–PR and SM–T2 pathways over AMZ at the grid-point scale, which may be due to the poor representation of model fields at a finer scale. Over SESA, all four models represent the SM–PR pathway well, but fail to reproduce the atmospheric leg in the SM–T2 pathway. This leads models to have a direct and negative relationship between SM and T2 biases over SESA, which is not seen in the reanalyses. The strength of the feedbacks between all the land—atmosphere pathways in S2S models generally weakens with increasing lead time, due to accumulation of errors with lead times that may lead to errors in the result of physical processes.

The results in the current study are influenced by the model and initialization data used, as discussed before and in Koster et al. (2004). Our results may also be sensitive to limiting our analysis to the common hindcast period of the models (1999–2010) and to the creation of lagged ensembles for ECMWF and NCEP. Further research into the impact of model initialization and intervening processes of the models' feedback pathways are required to understand the error sources in land surface variables and simulated land—atmosphere feedbacks.

ACKNOWLEDGEMENTS

This study was supported by the Climate Science for Services Partnership (CSSP) Brazil Diagnosing and Understanding Brazilian Subseasonal Tropical and Extratropical Processes (DUBSTEP) project, funded by the Newton Fund. NPK was supported by an Independent Research Fellowship from the Natural Environment Research Council (NE/L010976/1). NPK, MY and EB are grateful for the support of NCAS and the Global Challenges Research Fund (GCRF), via the Official Development Assistance programme Atmospheric Hazard in Developing Countries: Risk Assessment and Early Warning (ACREW) (grant number NE/R000034/1). CASC thanks CNPq, process 305206/2019-2, and Fundação de Amparo á Pesquisa do Estado de São Paulo (FAPESP), process 2015/50687-8 (CLIMAX Project) for the support received. JCAB acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (DECAF project, Grant agreement no. 771492).

DATA AVAILABILITY STATEMENT

The S2S database can be accessed online at http://apps.ecmwf.int/datasets/data/s2s. ERA5L data is available at Copernicus Climate Change Service Climate Data Store (https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=form). GLDAS data is available at NASA's Goddard Earth Sciences Data and Information Services Center (https://disc.gsfc.nasa.gov/datasets/GLDAS_NOAH10_3H_2.1/summary?keywords=GLDAS). Source for CHIRPS dataset is https://data.chc.ucsb.edu/products/CHIRPS-2.0/global_daily/netcdf/p05/. The code for the diagnostics used in this work is available at https://github.com/nick-klingaman/dubstep/tree/master/coupling_strength.

ORCID

Amulya Chevuturi https://orcid.org/0000-0003-2815-7221
Nicholas P. Klingaman https://orcid.org/0000-0002-2927-9303
Liang Guo https://orcid.org/0000-0002-6004-3884
Christopher E. Holloway https://orcid.org/0000-0001-9903-8989
Caio A. S. Coelho https://orcid.org/0000-0002-9695-5113
Emily Black https://orcid.org/0000-0003-1344-6186
Jessica C.A. Baker https://orcid.org/0000-0002-3720-4758
Pier Luigi Vidale https://orcid.org/0000-0002-1800-8460

REFERENCES

Baker, J.C., de Souza, D.C., Kubota, P., Buermann, W., Coelho, C.A., Andrews, M.B., Gloor, M., Garcia-Carreras, L., Figueroa, S.N. & Spracklen, D.V. (2021) An assessment of land-atmosphere interactions over South America using satellites, reanalysis and two global climate models. *Journal of Hydrometeorology*, 22(40), 905–922.

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., Dee, D., Dutra, E., Muñoz-Sabater, J., Pappenberger, F. et al. (2015) ERA-Interim/Land: a global land surface reanalysis data set. *Hydrology and Earth System Sciences*, 19(1), 389–407.

Coelho, C.A., Firpo, M.A., de Andrade, F.M. et al. (2018) A verification framework for South American sub-seasonal precipitation predictions. *Meteorologische Zeitschrift*, 27(6), 503–520.

de Andrade, F.M., Coelho, C.A. & Cavalcanti, I.F. (2019) Global precipitation hindcast quality assessment of the subseasonal to seasonal (S2S) prediction project models. *Climate Dynamics*, 52(9), 5451–5475.

Dee, D.P., Uppala, S.M., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, D.P. et al. (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656), 553–597.

DelSole, T., Trenary, L. & Tippett, M. (2017) The weighted-average lagged ensemble. *Journal of Advances in Modeling Earth Systems*, 9(7), 2739–2752.

Dirmeyer, P.A. (2011) The terrestrial segment of soil moisture–climate coupling. *Geophysical Research Letters*, 38(16), L16702.

Dirmeyer, P.A., Gentine, P., Ek, M.B. & Balsamo, G. (2019) Land surface processes relevant to sub-seasonal to seasonal (S2S) prediction. In: Robertson, A.W. & Vitart, F. (Eds.) *Sub-seasonal to seasonal prediction*. Amsterdam: Elsevier, pp. 165–181.

Dirmeyer, P.A. & Halder, S. (2017) Application of the land–atmosphere coupling paradigm to the operational Coupled

Forecast System, version 2 (CFSv2). *Journal of Hydrometeorology*, 18(1), 85–108.

Dirmeyer, P.A., Halder, S. & Bombardi, R. (2018) On the harvest of predictability from land states in a global forecast model. *Journal of Geophysical Research: Atmospheres*, 123(23), 13–111.

Dirmeyer, P., Peters-Lidard, C. & Balsamo, G. (2015) Land-atmosphere interactions and the water cycle. In: Brunet, G., Jones, S. & Ruti, P. (Eds.) *Seamless prediction of the Earth system: From minutes to months*. Geneva, Switzerland: World Meteorological Organization, pp. 145–154.

Dirmeyer, P.A., Wang, Z., Mbuh, M.J. & Norton, H.E. (2014) Intensified land surface control on boundary layer growth in a changing climate. *Geophysical Research Letters*, 41(4), 1290–1294.

Dutra, E., Balsamo, G., Viterbo, P., Miranda, P., Beljaars, A., Schär, C. & Elder, K. (2009) New snow scheme in HTESSEL: Description and offline validation. ECMWF. Available at: https://www.ecmwf.int/node/9167 [Accessed 17th September 2019].

ECMWF (2014) IFS documentation CY40R1. IFS Documentation. ECMWF. Available at: https://www.ecmwf.int/node/9203 [Accessed 3rd July 2019].

Ek, M., Mitchell, K., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G. & Tarpley, J. (2003) Implementation of NOAH land surface model advances in the National Centers for Environmental Prediction operational mesoscale ETA model. *Journal of Geophysical Research: Atmospheres*, 108(D22), 8851.

Eltahir, E.A. (1998) A soil moisture–rainfall feedback mechanism: 1. Theory and observations. *Water Resources Research*, 34(4), 765–776.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A. et al. (2015) The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data*, 2(1), 1–21.

Guimarães, B.S., Coelho, C.A., Woolnough, S.J., Kubota, P.Y., Bastarz, C.F., Figueroa, S.N., Bonatti, J.P. & de Souza, D.C. (2020) Configuration and hindcast quality assessment of a Brazilian global subseasonal prediction system. *Quarterly Journal of the Royal Meteorological Society*, 146, 1067–1084.

Guo, Z., Dirmeyer, P.A., Koster, R.D., Sud, Y., Bonan, G., Oleson, K.W., Chan, E., Verseghy, D., Cox, P., Gordon, C. et al. (2006) GLACE: the global land–atmosphere coupling experiment. Part II: analysis. *Journal of Hydrometeorology*, 7(4), 611–625.

Hirata, F.E. & Grimm, A.M. (2018) Extended-range prediction of South Atlantic convergence zone rainfall with calibrated CFSv2 reforecast. *Climate Dynamics*, 50(9–10), 3699–3710.

Kirtman, B.P., Min, D., Infanti, J.M., Kinter III, J.L., Paolino, D.A., Zhang, Q., Van Den Dool, H., Saha, S., Mendez, M.P., Becker, E. et al. (2014) The North American multimodel ensemble: Phase-1 seasonal-to-interannual prediction; Phase-2 toward developing intraseasonal prediction. *Bulletin of the American Meteorological Society*, 95(4), 585–601.

Klingaman, N.P., Young, M., Chevuturi, A., Guimarães, B., Guo, L., Woolnough, S.J., Coelho, C.A.S., Kubota, P.Y. & Holloway, C.E. (2021) Sub-seasonal prediction performance for austral summer South American rainfall. *Weather and Forecasting*, 36(1), 147–169.

Koster, R.D., Dirmeyer, P.A., Guo, Z., Bonan, G., Chan, E., Cox, P., Gordon, C., Kanae, S., Kowalczyk, E., Lawrence, D. et al. (2004) Regions of strong coupling between soil moisture and precipitation. *Science*, 305(5687), 1138–1140.

Koster, R.D., Guo, Z., Yang, R., Dirmeyer, P.A., Mitchell, K. & Puma, M.J. (2009) On the nature of soil moisture in land surface models. *Journal of Climate*, 22(16), 4322–4335.

Koster, R.D. & Suarez, M.J. (2001) Soil moisture memory in climate models. *Journal of Hydrometeorology*, 2(6), 558–570.

Koster, R.D., Suarez, M.J. & Heiser, M. (2000) Variance and predictability of precipitation at seasonal-to-interannual timescales. *Journal of Hydrometeorology*, 1(1), 26–46.

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

Li, S. & Robertson, A.W. (2015) Evaluation of submonthly precipitation forecast skill from global ensemble prediction systems. *Monthly Weather Review*, 143(7), 2871–2889.

Liu, Z., Notaro, M., Kutzbach, J. & Liu, N. (2006) Assessing global vegetation–climate feedbacks from observations. *Journal of Climate*, 19(5), 787–814.

Lorenz, R., Pitman, A.J., Hirsch, A.L. & Srbinovsky, J. (2015) Intraseasonal versus interannual measures of land–atmosphere coupling strength in a global climate model: GLACE-1 versus GLACE-CMIP5 experiments in ACCESS1. 3b. *Journal of Hydrometeorology*, 16(5), 2276–2295.

Ma, H.-Y., Mechoso, C., Xue, Y., Xiao, H., Wu, C.-M., Li, J.-L. & De Sales, F. (2011) Impact of land surface processes on the South American warm season climate. *Climate Dynamics*, 37(1–2), 187–203.

MacLachlan, C., Arribas, A., Peterson, K., Maidens, A., Fereday, D., Scaife, A., Gordon, M., Vellinga, M., Williams, A., Comer, R. et al. (2015) Global Seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system. *Quarterly Journal of the Royal Meteorological Society*, 141(689), 1072–1084.

Menéndez, C.G., Giles, J., Ruscica, R., Zaninelli, P., Coronato, T., Falco, M., Sörensson, A., Fita, L., Carril, A. & Li, L. (2019) Temperature variability and soil–atmosphere interaction in South America simulated by two regional climate models. *Climate Dynamics*, 53(5), 2919–2930.

Menéndez, C.G., Zaninelli, P.G., Carril, A.F. & Sánchez, E. (2016) Hydrological cycle, temperature, and land surface atmosphere interaction in the La Plata Basin during summer: response to climate change. *Climate Research*, 68(2-3), 231–241.

Meng, J., Yang, R., Wei, H., Ek, M., Gayno, G., Xie, P. & Mitchell, K. (2012) The land surface analysis in the NCEP Climate Forecast System Reanalysis. *Journal of Hydrometeorology*, 13(5), 1621–1630.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D.G., Piles, M., Rodríguez-Fernández, N.J., Zsoter, E., Buontempo, C. & Thépaut, J.-N. (2021) ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, 13(9), 4349–4383.

National Academy of Sciences (2016) *Next generation earth system prediction: Strategies for subseasonal to seasonal forecasts*. Washington, DC: National Academies Press.

Notaro, M. (2008) Statistical identification of global hot spots in soil moisture feedbacks among IPCC AR4 models. *Journal of Geophysical Research: Atmospheres*, 113(D9), D09101.

Orlowsky, B. & Seneviratne, S.I. (2010) Statistical analyses of land–atmosphere feedbacks and their possible pitfalls. *Journal of Climate*, 23(14), 3918–3932.

Piles, M., Ballabrera-Poy, J. & Muñoz-Sabater, J. (2019) Dominant features of global surface soil moisture variability observed by the SMOS satellite. *Remote Sensing*, 11(1), 95.

Rodell, M., Houser, P., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M. et al. (2004) The global land data assimilation system. *Bulletin of the American Meteorological Society*, 85(3), 381–394.

Rui, H. & Beaudoing, H. (2018) *README document for NASA GLDAS version 2 data products*. NASA's Greenbelt, MD: Goddard Space Flight Center. Available at: https://hydro1.gesdisc.eosdis.nasa.gov/data/GLDAS/README_GLDAS2.pdf

Ruscica, R., Sörensson, A. & Menendez, C.G. (2014) Hydrological links in Southeastern South America: soil moisture memory and coupling within a hot spot. *International Journal of Climatology*, 34(14), 3641–3653.

Ruscica, R.C., Sörensson, A.A. & Menéndez, C.G. (2015) Pathways between soil moisture and precipitation in southeastern South America. *Atmospheric Science Letters*, 16(3), 267–272.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D. et al. (2010) The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, 91(8), 1015–1058.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H.-y., Iredell, M. et al. (2014) The NCEP climate forecast system version 2. *Journal of Climate*, 27(6), 2185–2208.

Seneviratne, S.I., Corti, T., Davin, E.L., Hirschi, M., Jaeger, E.B., Lehner, I., Orlowsky, B. & Teuling, A.J. (2010) Investigating soil moisture–climate interactions in a changing climate: a review. *Earth-Science Reviews*, 99(3-4), 125–161.

Seneviratne, S.I., Lüthi, D., Litschi, M. & Schär, C. (2006) Land–atmosphere coupling and climate change in Europe. *Nature*, 443(7108), 205–209.

Sörensson, A.A. & Menéndez, C.G. (2011) Summer soil—precipitation coupling in South America. *Tellus A: Dynamic Meteorology and Oceanography*, 63(1), 56–68.

Spennemann, P.C., Rivera, J.A., Saulo, A.C. & Penalba, O.C. (2015) A comparison of GLDAS soil moisture anomalies against standardized precipitation index and multisatellite estimations over South America. *Journal of Hydrometeorology*, 16(1), 158–171.

Spennemann, P.C., Salvia, M., Ruscica, R., Sörensson, A., Grings, F. & Karszenbaum, H. (2018) Land-atmosphere interaction patterns in southeastern South America using satellite products and climate models. *International Journal of Applied Earth Observation and Geoinformation*, 64, 96–103.

Spennemann, P.C. & Saulo, A.C. (2015) An estimation of the land-atmosphere coupling strength in South America using the Global Land Data Assimilation System. *International Journal of Climatology*, 35(14), 4151–4166.

Sun, S. & Wang, G. (2012) The complexity of using a feedback parameter to quantify the soil moisture-precipitation relationship. *Journal of Geophysical Research: Atmospheres*, 117(D11), D11113.

Tawfik, A.B., Dirmeyer, P.A. & Santanello Jr, J.A. (2015) The heated condensation framework. Part II: Climatological behavior of convective initiation and land–atmosphere coupling over the conterminous United States. *Journal of Hydrometeorology*, 16(5), 1946–1961.

Trenary, L., DelSole, T., Tippett, M. & Pegion, K. (2017) A new method for determining the optimal lagged ensemble. *Journal of Advances in Modeling Earth Systems*, 9(1), 291–306.

Trenary, L., DelSole, T., Tippett, M. & Pegion, K. (2018) Monthly ENSO forecast skill and lagged ensemble size. *Journal of Advances in Modeling Earth Systems*, 10(4), 1074–1086.

Vitart, F. (2014) Evolution of ECMWF sub-seasonal forecast skill scores. *Quarterly Journal of the Royal Meteorological Society*, 140(683), 1889–1899.

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., Déqué, M., Ferranti, L., Fucile, E., Fuentes, M. et al. (2017) The subseasonal to seasonal (S2S) prediction project database. *Bulletin of the American Meteorological Society*, 98(1), 163–173.

Walters, D., Brooks, M., Boutle, I., Melvin, T., Stratton, R., Vosper, S., Wells, H., Williams, K., Wood, N., Allen, T. et al. (2017) The Met Office Unified Model Global Atmosphere 6.0/6.1 and JULES Global Land 6.0/6.1 configurations. *Geoscientific Model Development*, 10(4), 1487–1520.

Willmott, C.J., Rowe, C.M. & Mintz, Y. (1985) Climatology of the terrestrial seasonal water cycle. *Journal of Climatology*, 5(6), 589–606.

Zemp, D., Schleussner, C.-F., Barbosa, H., Van der Ent, R., Donges, J.F., Heinke, J., Sampaio, G. & Rammig, A. (2014) On the importance of cascading moisture recycling in South America. *Atmospheric Chemistry and Physics*, 14, 13337–13359.

Zeng, X., Barlage, M., Castro, C. & Fling, K. (2010) Comparison of land–precipitation coupling strength using observations and models. *Journal of Hydrometeorology*, 11(4), 979–994.

Zhang, J., Wang, W.-C. & Wei, J. (2008) Assessing land-atmosphere coupling using soil moisture from the Global Land Data Assimilation System and observational precipitation. *Journal of Geophysical Research: Atmospheres*, 113(D17), D17119.

Zwiers, F.W. & Von Storch, H. (1995) Taking serial correlation into account in tests of the mean. *Journal of Climate*, 8(2), 336–351.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Chevuturi, A., Klingaman, N.P., Guo, L., Holloway, C.E., Guimarães, B.S., Coelho, C.A.S., Kubota, P.Y., Young, M., Black, E., Baker, J.C.A., Vidale, P.L. Subseasonal prediction performance for South American land–atmosphere coupling in extended austral summer. *Climate Resilience and Sustainability*. 2021;1–19. https://doi.org/10.1002/cli2.28