



## RESEARCH ARTICLE

10.1029/2019MS001845

## Key Points:

- We identify WRF physics configurations that perform well with CABLE
- Simulation of latent heat fluxes over Australia are more skillful with WRF-LIS-CABLE than WRF-Noah

## Supporting Information:

- Supporting Information S1

## Correspondence to:

A. L. Hirsch,  
a.hirsch@unsw.edu.au

## Citation:

Hirsch, A. L., Kala, J., Carouge, C. C., De Kauwe, M. G., Di Virgilio, G., Ukkola, A. M., et al (2019). Evaluation of the CABLEv2.3.4 Land Surface Model Coupled to NU-WRFv3.9.1.1 in Simulating Temperature and Precipitation Means and Extremes Over CORDEX AustralAsia Within a WRF Physics Ensemble. *Journal of Advances in Modeling Earth Systems*, 11, 4466–4488. <https://doi.org/10.1029/2019MS001845>

Received 30 JUL 2019

Accepted 25 NOV 2019

Accepted article online 11 DEC 2019

Published online 18 DEC 2019

# Evaluation of the CABLEv2.3.4 Land Surface Model Coupled to NU-WRFv3.9.1.1 in Simulating Temperature and Precipitation Means and Extremes Over CORDEX AustralAsia Within a WRF Physics Ensemble

Annette L. Hirsch<sup>1,2</sup>, Jatin Kala<sup>1,3</sup>, Claire C. Carouge<sup>1,2</sup>, Martin G. De Kauwe<sup>1,2</sup>, Giovanni Di Virgilio<sup>2</sup>, Anna M. Ukkola<sup>4</sup>, Jason P. Evans<sup>1,2</sup>, and Gab Abramowitz<sup>1,2</sup>

<sup>1</sup>ARC Centre of Excellence for Climate Extremes, University of New South Wales, Sydney, New South Wales, Australia,

<sup>2</sup>Climate Change Research Centre, University of New South Wales, Sydney, New South Wales, Australia, <sup>3</sup>Environmental and Conservation Sciences, Murdoch University, Perth, Western Australia, Australia, <sup>4</sup>ARC Centre of Excellence for Climate Extremes, The Australian National University, Canberra, ACT, Australia

**Abstract** The Community Atmosphere Biosphere Land Exchange (CABLE) model is a third-generation land surface model (LSM). CABLE is commonly used as a stand-alone LSM, coupled to the Australian Community Climate and Earth Systems Simulator global climate model and coupled to the Weather Research and Forecasting (WRF) model for regional applications. Here, we evaluate an updated version of CABLE within a WRF physics ensemble over the COordinated Regional Downscaling EXperiment (CORDEX) AustralAsia domain. The ensemble consists of different cumulus, radiation and planetary boundary layer (PBL) schemes. Simulations are carried out within the NASA Unified WRF modeling framework, NU-WRF. Our analysis did not identify one configuration that consistently performed the best for all diagnostics and regions. Of the cumulus parameterizations the Grell-Freitas cumulus scheme consistently overpredicted precipitation, while the new Tiedtke scheme was the best in simulating the timing of precipitation events. For the radiation schemes, the RRTMG radiation scheme had a general warm bias. For the PBL schemes, the YSU scheme had a warm bias, and the MYJ PBL scheme a cool bias. Results are strongly dependent on the region of interest, with the northern tropics and southwest Western Australia being more sensitive to the choice of physics options compared to southeastern Australia which showed less overall variation and overall better performance across the ensemble. Comparisons with simulations using the Unified Noah LSM showed that CABLE in NU-WRF has a more realistic simulation of evapotranspiration when compared to GLEAM estimates.

## 1. Introduction

Regional climate models (RCMs) are the ideal tool for investigating regional climate and extremes at spatial scales more resolved than the typical global climate model. While several RCMs are available, the Weather Research and Forecasting (WRF) model (Skamarock et al., 2008) has a large worldwide community and is used for both operational purposes and research (Skamarock et al., 2019). A key advantage of WRF is the ability to configure different physical parameterizations to create physics ensembles (e.g., Evans et al., 2012). The WRF model is modular by design. Indeed, there are several parameterization options available for radiation (7), planetary boundary layer (13), cumulus convection (13), microphysics (22), and land surface (7) processes leading to an excess of 180,000 possible configurations. However, a substantial limitation that arises from this capability is that there is limited guidance on which combination of physical parameterizations are suitable and whether there is a dependence on the climate regime of the simulated domain. Therefore, we examine whether the physics configuration is dependent on the regional climate regime and whether the same configuration can be used for multiple purposes including the simulation of mean climate and extreme events.

There are enhanced versions of WRF intended for particular applications that may involve cyclone simulation, data assimilation, and coupled hydrological modeling. We use a version developed by the National

©2019. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Aeronautics and Space Administration (NASA) called the NASA-Unified WRF (NU-WRF) modeling system (Peters-Lidard et al., 2015) that enables the user to couple more sophisticated land surface schemes via the NASA Land Information System (LIS; Kumar et al., 2006; Peters-Lidard et al., 2007). Through NUWRF-LIS we have coupled the Community Atmosphere Biosphere Land Exchange (CABLE) land surface model (LSM; Kowalczyk et al., 2006; Wang et al., 2011).

CABLE simulates the fluxes of energy, water, and carbon at the land surface, and in recent years, the model has undergone significant developments. These developments include new parameterizations around photosynthetic (Haverd et al., 2018) and stomatal optimality theory (De Kauwe et al., 2015; Kala, De Kauwe, et al., 2015), improvements in the representation of soil physics, soil moisture dynamics, and runoff parameterization (Decker, 2015; Haverd et al., 2016), and a new soil evaporation parameterization (Decker, Or, et al., 2017). Collectively, these process developments have all contributed to reducing the overestimation of evapotranspiration by CABLE. CABLE is also a key component of the Australian Community Climate and Earth Systems Simulator (ACCESS; Bi et al., 2013; Law et al., 2017), which is a fully coupled climate model used as part of the fifth and sixth phases of the Coupled Model Intercomparison Project (CMIP). One of the main benefits of using CABLE within NU-WRF as compared to the more widely used unified Noah LSM (Chen & Dudhia, 2001) is that these enhancements in CABLE are not available in the unified version of the Noah LSM. Hereafter, we refer to CABLE coupled to NU-WRF as WRF-LIS-CABLE.

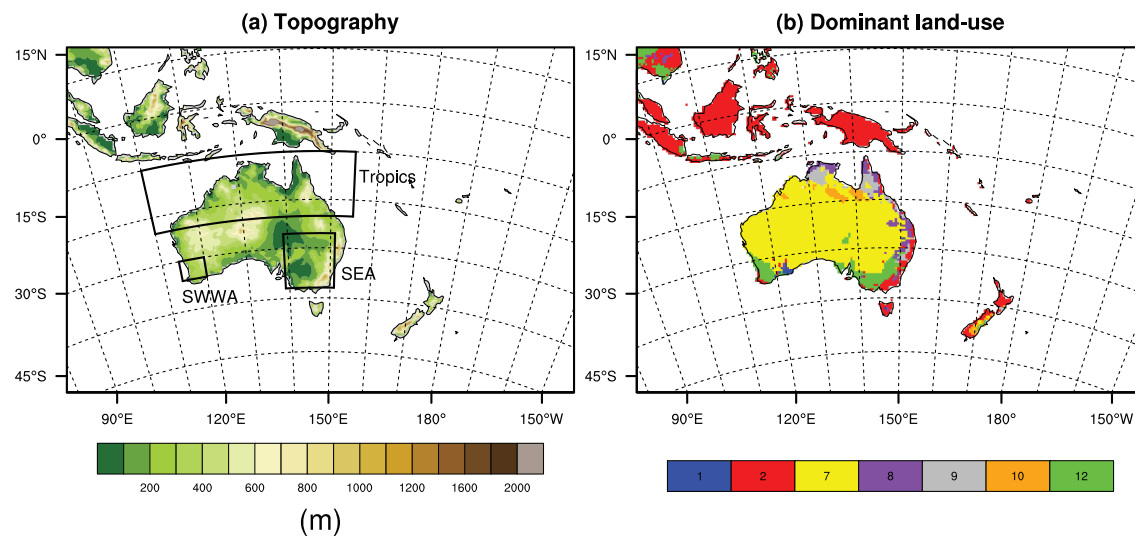
To date, testing of CABLE process improvements has only been in offline simulations at individual sites, globally with prescribed meteorology from the Global Soil Wetness Project (Dirmeyer et al., 1999) and in coupled land-atmosphere simulations with CABLE coupled to the ACCESS model. For coupled regional applications, previous work with WRF-LIS-CABLE used an older version of CABLE (version 1.4b), which does not incorporate any of these recent developments. Additionally, these studies only focused on the Austral summer period, as they investigated heat extremes (Hirsch, Kala, et al., 2014; Hirsch, Pitman, Seneviratne, et al., 2014; Hirsch, Pitman, & Kala, 2014; Hirsch et al., 2015). Given the WRF model is well documented to be highly sensitive to the choice of physical parameterizations (e.g., Evans et al., 2012; Kala, De Kauwe, et al., 2015), and previous work with WRF-LIS-CABLE has been limited to four different WRF configurations only, a comprehensive sensitivity analysis of the WRF model to different parameterization options for Australia is limited. Evans et al. (2012) is the only study to have conducted a real WRF physics ensemble, but only focused on extreme precipitation events over southeast Australia. The WRF sensitivity analysis carried out by Kala, Andrys, et al. (2015) was not a complete ensemble and only focused on southwest Western Australia. A comprehensive sensitivity analysis across the Australian continent has therefore not been carried out to date.

This paper therefore evaluates WRF-LIS-CABLE using the latest developments in CABLE, and extends the analysis to all seasons rather than just the Austral summer period. We also undertake a more in-depth analysis of the performance of WRF-LIS-CABLE within an ensemble of model configurations/parameterization options across the entire continent. Our overarching aim is to provide guidance on configurations of WRF-LIS-CABLE that perform reasonably and would form a basis for future studies targeting specific science questions.

## 2. Materials and Methods

### 2.1. Model Description

The version of CABLE used in this study is 2.3.4 (hereafter CABLEv2.3.4). A comprehensive description of the earlier version of CABLE (version 1.4b) used in all previous work using WRF-LIS-CABLE can be found in Kowalczyk et al. (2006) and Wang et al. (2011). The key differences between versions 1.4b and 2.3.4 as used in this study are as follows: (i) a new stomatal conductance scheme with an explicit parameterization of plant water use strategy by plant functional type (De Kauwe et al., 2015; Kala, De Kauwe, et al., 2015); (ii) a groundwater model and subgrid-scale runoff parameterization (Decker, 2015); and (iii) a new pore-scale model-based formulation for soil evaporation (Decker, Or, et al., 2017). We note that we do not test these new developments against the older version 1.4b in this manuscript as these new developments are the new default settings for CABLE. Furthermore, the new stomatal conductance scheme has been tested offline (De Kauwe et al., 2015) and in ACCESS (Kala, De Kauwe, et al., 2015), and the groundwater model and subgrid-scale runoff parameterization and new pore-scale formulation for soil evaporation has been tested



**Figure 1.** Simulation domain, (a) topography and (b) dominant land use category (1 = Evergreen needleleaf forest, 2 = evergreen broadleaf forest, 7 = Open shrublands, 8 = Woody Savanna, 9 = Savanna, 10 = Grassland, 12 = Cropland). The boxes in (a) denote the Tropics, South East Australia (SEA), and South West Western Australia (SWWA) regions.

offline (Decker 2015; Decker, Or, et al., 2017) and in WRF (Decker, Ma, et al., 2017). The focus of this work is to test the sensitivity of WRF-LIS-CABLE to a collection of available atmospheric physics options in WRF.

Parameter data sets used by WRF-LIS-CABLE include the United States Geological Survey Global 30 arcsec digital elevation (GTOPO30) data for describing surface elevation. Slope parameters required for the groundwater and subgrid-scale runoff scheme use 1 km elevation data from GTOPO. A hybrid 16-category soil texture map that describes the silt, clay, and sand fractions of each soil category is based on STATSGO/FAO (PSU, 2006). Leaf area index (LAI) is prescribed in CABLE using monthly climatological values from the MODerate resolution Imaging Spectroradiometer (MODIS; Yuan et al., 2011). To obtain daily values a simple linear interpolation algorithm is applied within LIS. Vegetation is specified by the Terra-MODIS sensor based on the International Geosphere-Biosphere Programme land classification map (Friedl et al., 2002) that has been modified to exclude permanent wetlands and cropland/natural vegetation mosaic. The vegetation map is aggregated from a 1 km resolution to the model resolution to determine the grid cell fraction of each vegetation class.

In this study, CABLEv2.3.4 was coupled to NU-WRFv9-3.9.1.1. NU-WRF differs from standard WRF in several aspects. NU-WRF allows a user to run their LSM offline using a variety of meteorological input forcings for model spin-up of land surface quantities, such as soil moisture and temperature, before running coupled land-atmosphere simulations. The offline spin-up is carried out on the same computational domain used for the coupled simulations, allowing for the best possible initial estimates of the land surface state (e.g., soil moisture), and this has been shown to be critical in many regions for more accurate simulations (Santanello, Peters-Lidard, et al., 2013; Santanello, Kumar, et al., 2013), especially for warm extremes in Australia (e.g., Hirsch, Kala, et al., 2014). NU-WRF also includes several additional tools as plug-ins via LIS (Kumar et al., 2008), such as parameter estimation (e.g., Santanello, Kumar, et al., 2013), data assimilation (e.g., Kumar et al., 2018), and uncertainty analysis (e.g., Harrison et al., 2012), which are not routinely available in standard WRF.

The computational domain is the COordinated Regional Downscaling EXperiment (CORDEX) AustralAsia domain (spatial resolution of approximately 0.44°) illustrated in Figure 1 showing the model topography and dominant land use categories. This domain was used to enable comparison of WRF-LIS-CABLE outputs to existing CORDEX AustralAsia simulations using WRF (e.g., Di Virgilio et al., 2019) and to reduce the computational and storage costs associated with running finer-scale resolution domains.

In the model simulations, subgrid vegetation tiling was used with each grid cell having up to a maximum of eight tiles. A minimum cutoff of 5% was imposed on the vegetation fraction for a given vegetation class to be included. Coupled land-atmosphere simulations were carried out over the period August 2008 to October

**Table 1**  
*WRF Atmospheric Parameterization Schemes Tested in WRF-LIS-CABLE*

Shortwave/Longwave radiation	Planetary boundary layer (PBL)	Cumulus
Dudhia (Dudhia, 1989) /RRTM (Mlawer et al., 1997)	Yonsei University (YSU; Hong et al., 2006)	Kain-Fritsch (KF; Kain, 2004)
RRTMG (Iacono et al., 2008)	Mellor-Yamada-Janjic (MYJ; Janjic, 1996, 2002)	Betts-Miller-Janjic (BMJ; Janjic, 1994, 2000)
New Goddard (Chou et al., 2001; Chou & Suarez, 1999)	MYNN (Nakanishi & Niino, 2006, 2009)	Grell-Freitas (Grell & Freitas, 2014) New Tiedtke (Zhang & Wang, 2017)

2010 with 6-hourly lateral boundary conditions and sea surface temperatures from the ERA-Interim reanalysis (Dee et al., 2011). This period corresponds to the termination of the Millennium drought in Australia (van Dijk et al., 2013). This was chosen as it includes periods of extreme heat, dry as well as relatively wet conditions corresponding to the termination of the drought, allowing us to assess the relative strengths of each configuration under a range of extreme conditions.

Offline model spin-up was carried out over a period of 4 years, starting in August 2004, with the Princeton meteorological forcing at 3-hourly frequency (Sheffield et al., 2006), consistent with our previous work using NU-WRF (Hirsch, Kala, et al., 2014; Hirsch, Pitman, Seneviratne, et al., 2014; Hirsch, Pitman, & Kala, 2014; Hirsch et al., 2015). We note that while it is theoretically possible to run the offline spin-up with ERA-Interim, for consistency with the lateral boundary conditions used during the coupled simulation, 6-hourly meteorological forcings are not frequent enough for offline simulations, hence our choice of Princeton forcing. Additionally, the first 2 months of the coupled simulations were not used to allow for sufficient spin-up of the atmosphere.

To create a WRF physics ensemble, we choose a combination of options that are commonly used and combined these with several new schemes. All configurations are run with the WRF Single-Moment 5-class (WSM5) microphysics scheme as it has five classes (cloud water, cloud ice, rain, snow, and vapor) without the added computation costs associated with double moment schemes. The ensemble consists of three shortwave/longwave radiation, three planetary boundary layer (PBL) and four cumulus options, resulting in a total of 36 coupled land-atmosphere simulations, summarized in Table 1, and all simulations used the same offline spin-up. The selection of the WRF physics configurations is based on schemes that are often used by the WRF user community (e.g., Cohen et al., 2015; Evans et al., 2012; Gbode et al., 2019) in addition to some recent advanced schemes that are relatively efficient without compromising on accuracy. The Community Atmosphere Model (CAM) radiation scheme was also tested but found to produce significant cold temperature biases ( $> 5^{\circ}\text{C}$ ) making this radiation scheme unsuitable for use with CABLE. Similar cold biases were identified by Di Virgilio et al. (2019) over the CORDEX AustralAsia domain for ERA-Interim driven simulations using WRF-Noah with CAM radiation. We do not recommend using the CAM radiation schemes for a Southern Hemisphere domain. Therefore, our analysis provides a large ensemble to identify WRF physics configurations that perform reasonably with CABLE.

## 2.2. Observations and Statistical Methods

To evaluate the skill of the WRF-LIS-CABLE model simulations we use the Australian Gridded Climate Data (AGCD) daily precipitation and temperature data sets (Jones et al., 2009) which have a spatial resolution of  $0.05^{\circ}$ . AGCD now comes with an error estimate based upon quality control information embedded with the station level data on the Australian Data Archive for Meteorology. Regions where the station density is scarce over central and western Australia have been masked for the AGCD data according to King et al. (2013). Further data sets used include the Global Land surface Evaporation: the Amsterdam Methodology (GLEAM) data sets (Martens et al., 2017; Miralles et al., 2011), which has a resolution of  $0.25^{\circ}$ , to evaluate the simulation of daily latent heat fluxes. To estimate the observational uncertainty in the latent heat flux, we also use estimates from van Dijk et al. (2018; hereafter vD18). vD18 is a daily model-data fusion product that assimilates satellite information into a hydrological model and is available globally at 5-km resolution. Both GLEAM and vD18 are model-derived estimates of the latent heat flux, and therefore we also compare WRF-LIS-CABLE estimates to Australian flux tower data (Beringer et al., 2016) corresponding to the simulation period in the supporting information. However, there are limitations to comparing 50-km resolution WRF-LIS-CABLE outputs to station data. To enable direct comparison, we interpolate all data sets to the model resolution using a first-order conservative remapping algorithm.

To evaluate the simulation skill, we use the metrics of the International Land Model Benchmarking system (Collier et al., 2018). This involves evaluating the simulation skill between the observational data ( $O$ ) and the model output ( $M$ ) for different attributes of the simulated climate. These metrics include the relative bias to evaluate the skill in simulating the variable amplitude:

$$\epsilon_{\text{bias}} = (\overline{M} - \overline{O}) / \sqrt{\frac{1}{N} \sum_{\text{time}} (O_t - \overline{O})^2} \quad (1)$$

The relative root-mean-square error (RMSE) score, to evaluate the skill in capturing the temporal variability is:

$$\epsilon_{\text{rmse}} = \sqrt{\frac{1}{N} \sum_{\text{time}} ((M_t - \overline{M}) - (O_t - \overline{O}))^2} / \sqrt{\frac{1}{N} \sum_{\text{time}} (O_t - \overline{O})^2} \quad (2)$$

Note that  $\overline{X}$  denotes the time mean from either the observations or model output and  $N$  denotes the number of days over which the calculation is made (e.g., 365 for annual; 90 for DJF). To compute normalized scores,  $\epsilon_{\text{bias}}$  and  $\epsilon_{\text{rmse}}$  are passed through the exponential function such that the skill scores are within the interval [0 1]:

$$s = e^{-\epsilon} \quad (3)$$

The phase shift in the annual or seasonal maxima is calculated as

$$s_{\text{phase}} = \frac{1}{2} \left( 1 + \cos \left( \frac{2\pi(\max(M) - \max(O))}{N} \right) \right) \quad (4)$$

Note that  $\max(X)$  is the day to which the maximum value occurs with  $N$  corresponding to the number of days for the period of interest. The  $s_{\text{bias}}$ ,  $s_{\text{rmse}}$ , and  $s_{\text{phase}}$  skill scores are evaluated at each grid cell ( $i,j$ ) and aggregated to a field scalar score using an area-weighting approach. Finally, the spatial distribution skill score is calculated as:

$$s_{\text{dist}} = 2(1 + R) / \left( \sigma + \frac{1}{\sigma} \right)^2 \quad (5)$$

Where  $R$  is the spatial correlation between the time average of the observations and model output and  $\sigma$  is the normalized standard deviation:

$$\sigma = \frac{\text{stdev}(\overline{M})}{\text{stdev}(\overline{O})} \quad (6)$$

By using both the spatial correlation and the normalized standard deviation in  $s_{\text{dist}}$  skill is reduced when either  $R$  or  $\sigma$  deviate from a value of 1. Note that  $s_{\text{dist}}$  essentially measures the skill in capturing the maxima and minima spatially, whereas  $s_{\text{phase}}$  evaluates this from a temporal perspective.

These four skill scores are then combined as a weighted average to provide the overall skill score:

$$S_{\text{overall}} = (s_{\text{bias}} + 2 \times s_{\text{rmse}} + s_{\text{phase}} + s_{\text{dist}}) / 5 \quad (7)$$

Where  $s_{\text{rmse}}$  is given a larger weight than the other skill metrics to emphasize its importance. To assist in identifying the relative merits of a given physics configuration, we present the relative skill scores which have been calculated from the absolute values as:

$$S_{\text{relative}} = \frac{S_i - \mu}{\sigma} \quad (8)$$

Where  $\mu$  is the mean skill score across all physics configurations for a given variable and  $\sigma$  is the standard deviation of the skill scores across all physics configurations for a given variable. Note that for all absolute skill scores, values are within the interval of [0,1] however for the relative skill scores, values can exceed  $\pm 1$ . Note that values close to +1 are indicative of greater skill, values that are zero or negative indicate poor skill.

**Table 2**  
*The Subset of ETCCDI Extreme Indices Evaluated in This Paper*

Name	Acronym	Description
5th percentile maximum Temperature	TX <sub>5</sub>	Over a given period, the 5th percentile of the daily maximum 2 m air temperature
95th percentile maximum Temperature	TX <sub>95</sub>	Over a given period, the 95th percentile of the daily maximum 2 m air temperature
Hottest day	TX <sub>x</sub>	Maximum value of the daily maximum 2 m air temperature
5th percentile minimum Temperature	TN <sub>5</sub>	Over a given period, the 5th percentile of the daily minimum 2 m air temperature
95th percentile minimum Temperature	TN <sub>95</sub>	Over a given period, the 95th percentile of the daily minimum 2 m air temperature
Coldest night	TN <sub>n</sub>	Minimum value of the daily minimum 2 m air temperature
Diurnal temperature range	DTR	Difference between the daily maximum 2 m air temperature and daily minimum 2 m air temperature
Maximum 1-day precipitation	RX <sub>1DAY</sub>	Maximum 1-day precipitation amount over a defined period
Maximum consecutive 5-day precipitation	RX <sub>5DAY</sub>	5 consecutive day precipitation amounts. Maximum value of these 5-day values
Number of extreme wet days	R <sub>10mm</sub>	Number of days over a period when daily precipitation exceeds 10 mm
Maximum dry spell length	CDD	Maximum number of consecutive days with precipitation less than 1 mm/day

These skill scores are calculated for the variables: daily precipitation totals (PR), daily maximum 2 m air temperature (TX), daily minimum 2 m air temperature (TN), and daily mean latent heat flux (QE) and aggregated to either annual or seasonal averages. Further spatial aggregation is carried out for three distinctive regions of Australia that encapsulate different climatic regimes: tropical Australia, South East Australia (SEA), and South West Western Australia (SWWA; Figure 1a).

In addition to evaluating simulation skill of annual and seasonal climate, we include 11 of the extremes indices defined by the Expert Team on Climate Change Detection and Indices (ETCCDI; <https://www.climdex.org/learn/indices/>) that can be derived from daily TX, TN, and precipitation (Table 2).

### 3. Results

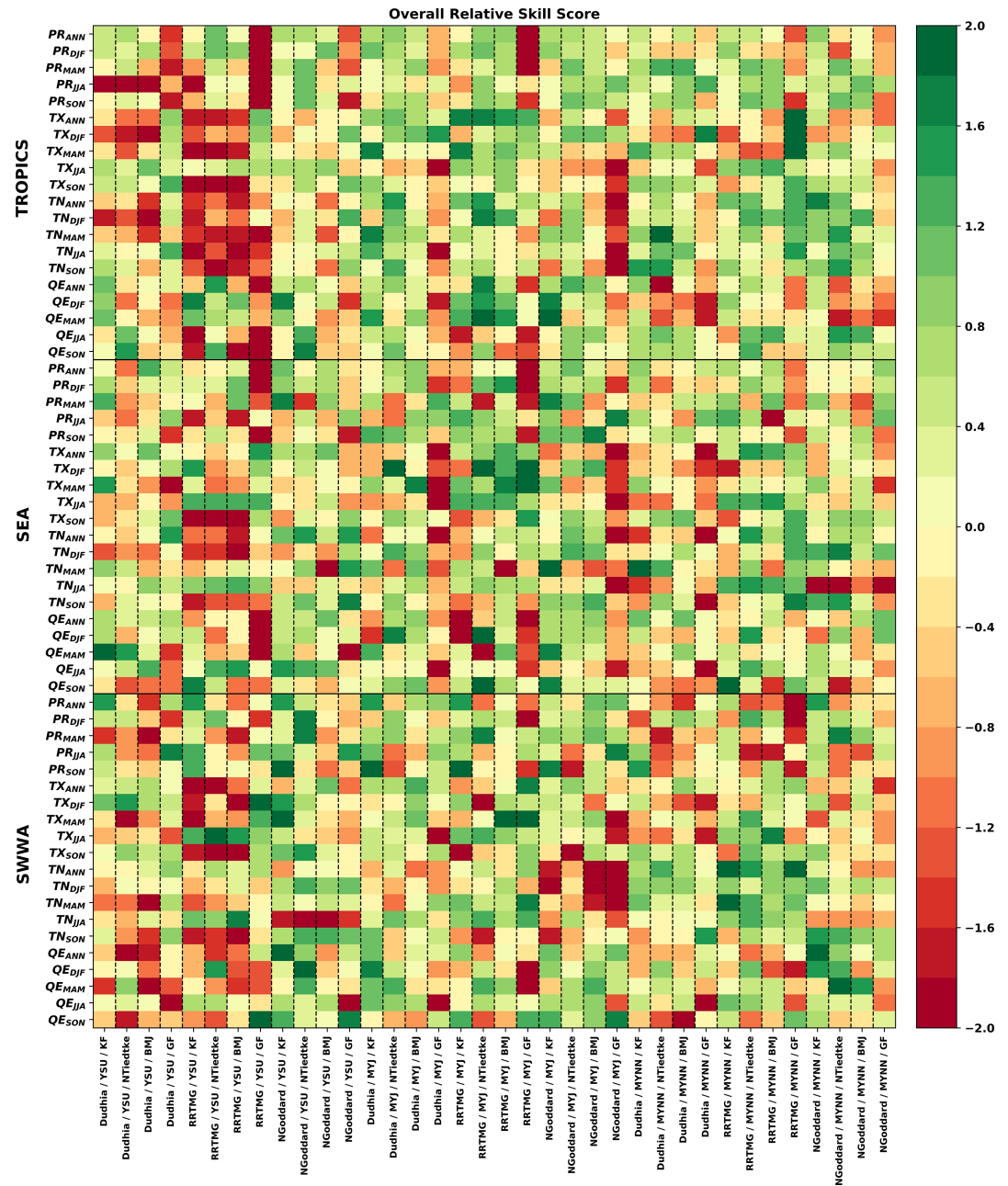
Given the large size of the ensemble (36 members), we first present a high-level skill summary of the overall and relative scores (equations (7) and (8)) over the entire simulation period October 2008 to October 2010 (excluding the spin-up period). Based on this, we then show more detailed analyses of particular ensemble members during periods of extreme heat, dry, and wet conditions.

#### 3.1. Overall Relative Skill for Climate and Extremes

Figure 2 shows the overall relative skill score for precipitation (PR), maximum temperature (TX), minimum temperature (TN), and latent heat flux (QE), for each of the four seasons (DJF-Austral summer, MAM-Autumn, JJA-winter, SON-spring), and annually. The scores are computed for three key regions: the Tropics, southeast Australia (SEA), and Southwest West Western Australia (SWWA). Figure 3 is the same as Figure 2, except that it shows results for the precipitation and temperature extremes defined in Table 2.

Figure 2, shows overall that each variable has configurations that are more skillful than others. Examining performance between the three PBL schemes, configurations using the YSU scheme tend to show poorer performance (relative  $S_{\text{overall}} \sim < -0.6$ ) as compared to MYJ and MYNN (relative  $S_{\text{overall}} \sim > +0.5$ ), particularly for the summer months. This is also evident for TX<sub>95</sub> (Figure 3). Between the four different cumulus schemes, configurations using GF tend to have relatively poorer skill in simulating seasonal and annual precipitation totals and extreme precipitation compared to the KF, BMJ, and Tiedtke schemes. In particular, time series analysis (Figure S1) revealed that the simulations using GF tend to trigger precipitation more frequently than the other schemes which also affects the land water balance, with corresponding higher estimates in the latent heat flux coinciding with anomalously high-precipitation events. From Figure 2, there are five configurations that are, relatively, the least skillful, this includes all configurations with the YSU PBL paired with RRTMG radiation (i.e., irrespective of the cumulus scheme) and the New Goddard radiation/MYJ PBL/GF cumulus configuration.

The absolute skill scores for climate (Figure S2) and extremes (Figure S3) show that overall, model performance for precipitation is often worse than maximum temperature, and minimum temperature generally has better scores than maximum temperature. This is particularly evident in the extremes indices (Figure S3). Poor performance in precipitation is also generally reflected by poor performance in the latent heat flux, which is also to be expected, as latent heating is influenced by the water availability supplied by



**Figure 2.** Relative overall skill score for climate variables per successful physics configuration evaluated according to annual and seasonal means and aggregated over the three regions denoted in Figure 1. The absolute skill scores are available in the supporting information. Interpretation: Green indicates relative high skill while red indicates relatively poor skill. Note that evaluation of the latent heat flux is against GLEAM.

precipitation (and vice versa). More specifically, Figure S2 reveals that most physics configurations have more skill at simulating SEA seasonal and extreme temperatures (generally  $>0.6$ ) as compared to the Tropics ( $0.4\text{--}0.6$ ). Model skill is also slightly higher ( $\sim 0.1$ ) for seasonal and extreme precipitation in SEA than the Tropics. Corresponding figures of the components of the absolute skill score are available in the supporting information (bias skill score  $s_{\text{bias}}$ : Figure S4; RMSE score  $s_{\text{rmse}}$  Figure S5; and spatial distribution score  $s_{\text{dist}}$  Figure S6). These absolute skill scores reveal that many of the physics configurations are comparable in skill. In particular, for precipitation  $s_{\text{bias}}$  values are often  $>0.7$  but  $s_{\text{rmse}}$  values are  $0.2\text{--}0.3$  indicating that precipitation timing has a larger impact on simulation skill. In contrast

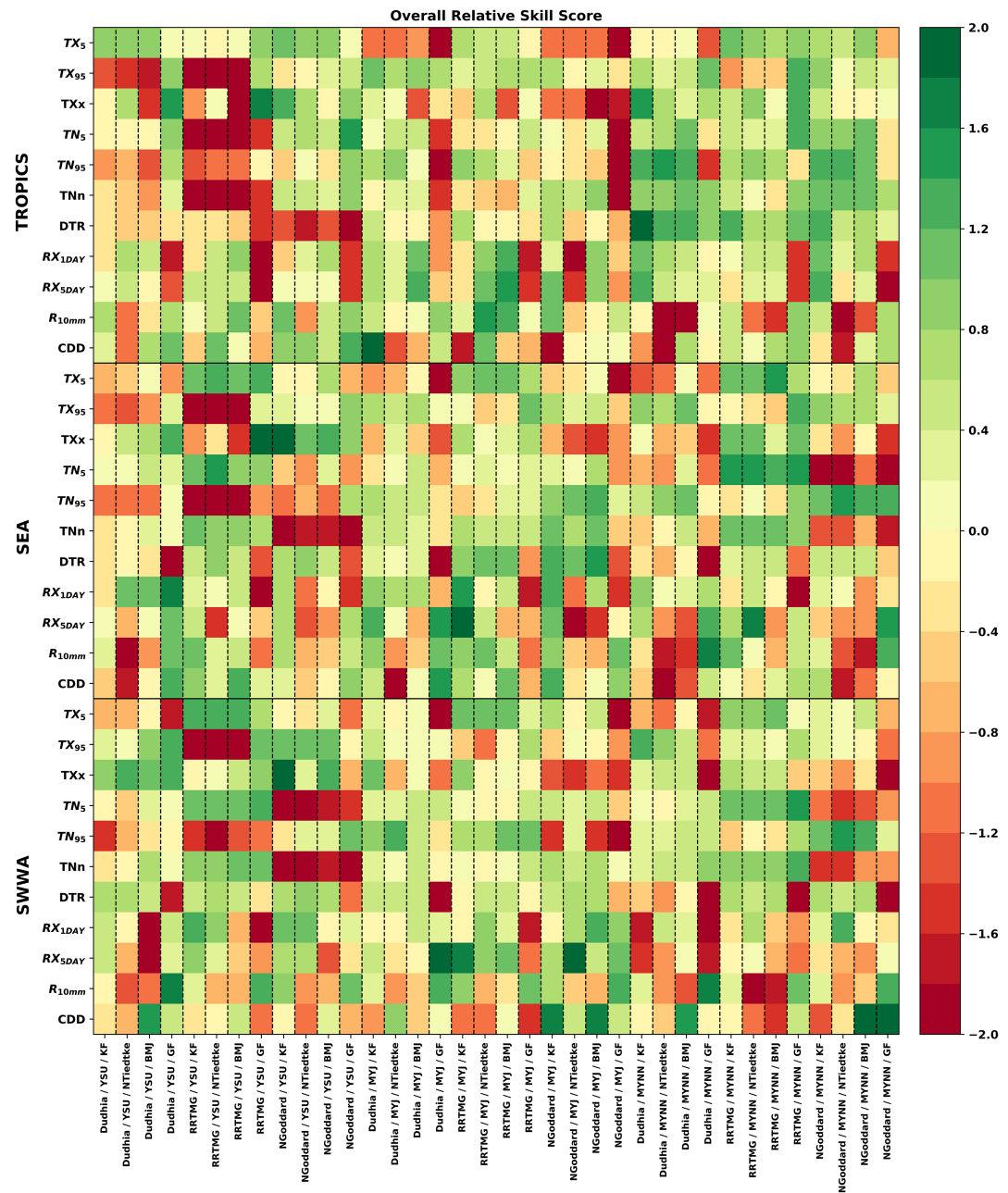


Figure 3. As per Figure 2 but for the extremes indices described in Table 2 on an annual, rather than monthly, interval.

for TX and TN, most configurations have greater skill in simulating the daily variations for temperature than precipitation (Figure 3). Incidentally,  $s_{\text{dist}}$  in Figure S6 demonstrates that most configurations have higher skill in capturing the spatial variability of all variables for SEA compared to the Tropics and SWWA. This reflects the sensitivity of WRF to regions which experience markedly different climatology, that is, tropical/monsoon for the Tropics and Mediterranean climates for SWWA (hot and dry summers and cool and wet winters), versus SEA which does not have a distinct dry season as compared to SWWA and the Tropics. This suggests that for analysis domains that include the Tropics or SWWA, greater care in the selection of the WRF physics configuration is necessary compared to domains that just focus over SEA.

To provide further guidance on the relative strengths of the different WRF physics configurations, we used the relative skill metrics (Figures 2 and 3) to calculate the total score across all variables by either region,



**Table 3**  
*Top Five Physics Configurations Determined by Summing the Relative Skill Scores by Region and the Whole of Australia, for Climate or Extremes, and Both*

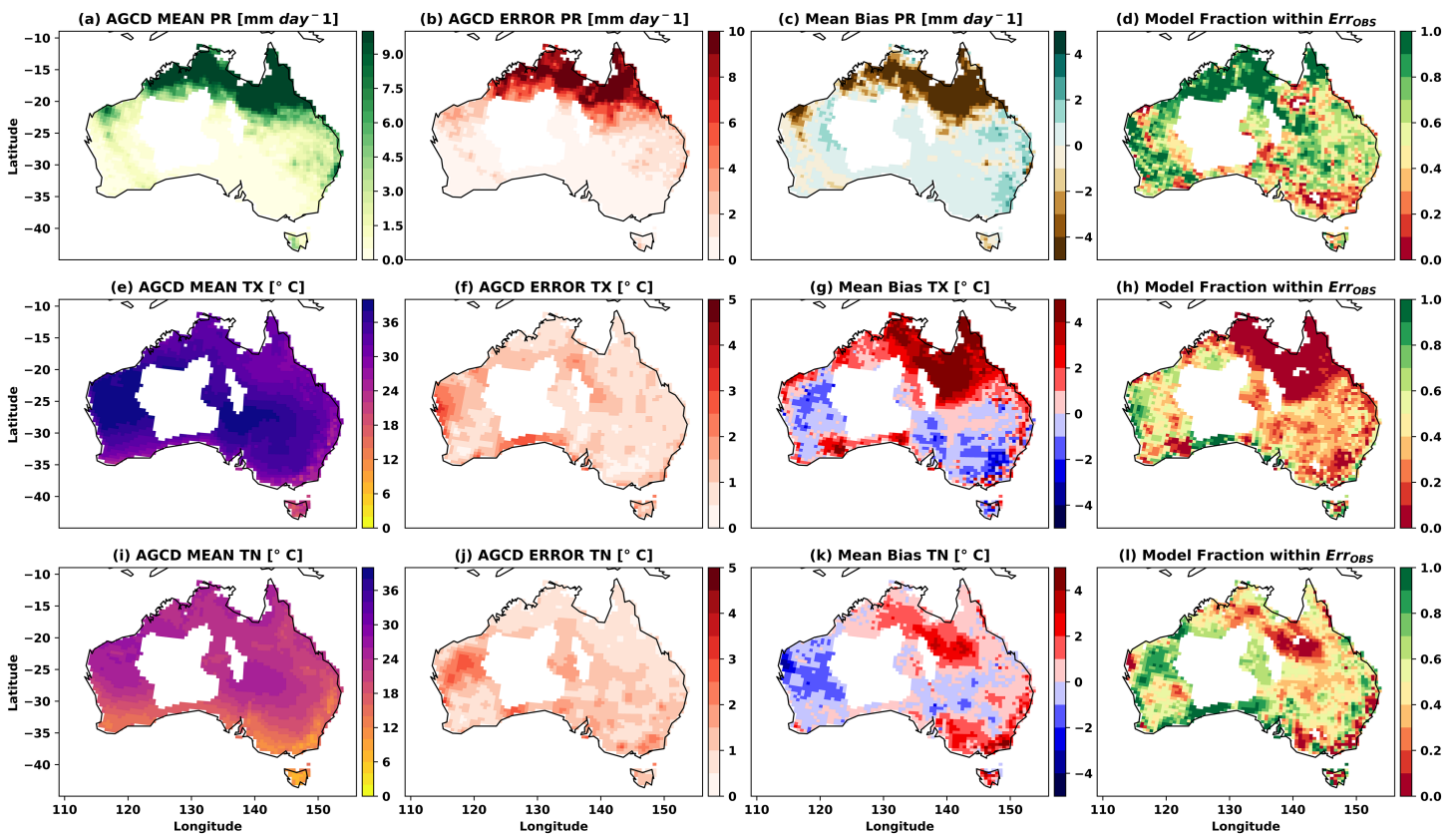
	Climate	Extremes	Total
Tropics	RRTMG/MYJ/NTiedtke (17)	NGoddard/MYNN/KF (9)	RRTMG/MYJ/NTiedtke (23)
	Dudhia/MYJ/NTiedtke (13)	Dudhia/MYNN/KF (8)	NGoddard/MYNN/KF (19)
	Dudhia/MYJ/KF (12)	RRTMG/MYJ/NTiedtke (6)	Dudhia/MYJ/KF (17)
	RRTMG/MYJ/BMJ (11)	NGoddard/MYNN/BMJ (6)	Dudhia/MYNN/KF (17)
	NGoddard/MYJ/NTiedtke (10)	Dudhia/MYNN/BMJ (5)	RRTMG/MYJ/BMJ (14)
SEA	NGoddard/MYJ/KF (12)	RRTMG/MYJ/KF (8)	NGoddard/MYJ/KF (19)
	RRTMG/MYNN/NTiedtke (11)	NGoddard/MYJ/KF (8)	RRTMG/MYNN/NTiedtke (18)
	RRTMG/MYJ/BMJ (11)	RRTMG/MYNN/NTiedtke (7)	RRTMG/MYJ/BMJ (15)
	Dudhia/MYJ/BMJ (10)	RRTMG/MYNN/KF (6)	RRTMG/MYJ/NTiedtke (13)
	RRTMG/MYJ/NTiedtke (9)	RRTMG/MYNN/GF (5)	RRTMG/MYNN/GF (12)
SWWA	NGoddard/YSU/KF (14)	RRTMG/MYNN/KF (4)	Dudhia/MYJ/KF (16)
	Dudhia/MYJ/KF (12)	RRTMG/MYJ/KF (4)	NGoddard/YSU/KF (15)
	NGoddard/YSU/NTiedtke (12)	RRTMG/MYJ/BMJ (4)	RRTMG/MYNN/KF (11)
	NGoddard/MYNN/KF (10)	Dudhia/MYJ/KF (4)	RRTMG/MYJ/BMJ (10)
	RRTMG/MYNN/KF (7)	NGoddard/MYJ/NTiedtke (3)	RRTMG/MYJ/KF (9)
Total	RRTMG/MYJ/BMJ (28)	RRTMG/MYNN/KF (15)	RRTMG/MYJ/BMJ (39)
	RRTMG/MYJ/NTiedtke (26)	Dudhia/MYJ/KF (13)	RRTMG/MYNN/NTiedtke (37)
	NGoddard/YSU/NTiedtke (25)	RRTMG/MYNN/NTiedtke (12)	RRTMG/MYJ/NTiedtke (37)
	RRTMG/MYNN/NTiedtke (24)	RRTMG/MYJ/KF (12)	Dudhia/MYJ/KF (36)
	Dudhia/MYJ/KF (23)	RRTMG/MYNN/GF (11)	NGoddard/MYNN/KF (30)

*Note.* Configurations with the highest total relative skill score are listed first with the score rounded to the nearest whole number listed in brackets.

climate or extremes, and aggregating over both groups (Table 3). These revealed challenges to the appropriate selection of a WRF physics configuration depending on the intended purpose. In particular, if the focus was on skillful simulation of climate, the same configuration may not be as skillful at the simulation of extremes. There were exceptions to this, with greater skill using the RRTMG/MYJ/New Tiedtke configuration for both climate and extremes over the Tropics. Similarly, the New Goddard/MYJ/KF configuration is skillful for both climate and extremes over SEA. Furthermore, there are regional differences regarding skillful configurations. Rarely is there a configuration that is highly skillful for both the Tropics and SEA with the exception of the RRTMG/MYJ/BMJ for simulating climate but not extremes. Therefore, in selecting a configuration for a large domain that encompasses different climate regimes, as is presented here, some trade-off in skill is unavoidable.

### 3.2. Spatial Identification of Systematic Biases

The skill assessment provides context on the relative performance of the different ensemble members; however, it is also necessary to identify any systematic biases in WRF-LIS-CABLE from both spatial and temporal perspectives. The simulation period was selected on the basis that there are several extreme events to comprehensively evaluate the skill of WRF-LIS-CABLE. During 2009 to 2010, Australia experienced several extreme heat events (Bureau of Meteorology, 2009). Figure 4 illustrates the mean climate observed and model bias over one of these extreme heat periods corresponding to January and February 2009. During this period the monsoon was active over tropical Australia contributing to the larger mean daily precipitation that, on average, was in excess of 10 mm/day (Figure 4a) with a corresponding large observational uncertainty (>5 mm/day) compared to the rest of the continent (Figure 4b). The ensemble mean precipitation bias (Figure 4c) illustrates that WRF-LIS-CABLE has a systematic dry bias across tropical Australia of the order of 4 mm/day; however, the majority of ensemble members are within the observation error estimate (Figure 4d). There are regions where few ensemble members are within the observational uncertainty. For precipitation, these largely coincide with regions with low observational error (Figure 4b) such as southern Australia, where the precipitation bias is within 1 mm/day (Figure 4c). Most of Australia had a mean daily maximum temperature in excess of 30 °C over the period (Figure 4e). WRF-LIS-CABLE has a persistent hot bias of 2–5 °C over tropical Australia (Figure 4g). The hot bias coincides with the region where there was a dry precipitation bias. The TX observational error is within 2 °C over most of Australia (Figure 4f) with the hot bias over tropical Australia an attribute of nearly all ensemble members (Figure 4h). The corresponding



**Figure 4.** Assessment of physics ensemble skill against the AGCD for a period of extreme heat during January-February 2009. With the observed mean (a, e, and i), observational error (b, f, and j), mean multiphysics bias (c, g, and k), and the fraction of ensemble members within the observational error (d, h, and l). For precipitation (PR, mm/day, a–d), daily maximum 2 m air temperature (TX, °C, e–h) and daily minimum 2 m air temperature (TN, °C, i–l).

results for TN also indicate that WRF-LIS-CABLE is too warm ( $\sim 2^\circ\text{C}$ ) over the Tropics during this period (Figure 4k) and is also a persistent attribute of the ensemble (Figure~4l). There is a cool bias of 1–2 °C for TX over southeastern Australia and TN over Western Australia. These regions with a cool temperature bias over southeastern Australia coincide with regions with a slight wet bias in precipitation of  $\sim 1$  mm/day.

In 2009, March to November was marked by extremely dry conditions across most of Australia (Figure 5a). The ensemble mean bias is within 0.5 mm/day for most of Australia with the exception of Tasmania (Figure 5c). Our results show that the diurnal temperature range for WRF-LIS-CABLE is too narrow particularly over eastern Australia during this period with a cool bias of 1–2 °C in TX (Figure 5g) and a warm bias of 2 °C in TN (Figure 5k). For PR and TN, however, most WRF-LIS-CABLE ensemble members are within the observational error (Figure 5d and 5l) while TX biases exceed the observational error (Figure 5h).

The beginning of 2010 saw the termination of the drought with extensive precipitation across most of SEA (Figure 6a). Again, WRF-LIS-CABLE underpredicts precipitation amounts (2–4 mm/day, Figure 6c), although the observational error was high (Figure 6b). TX (Figure 6g), and TN (Figure 6k) both have a warm bias of 2–3 °C over Northern and Eastern Australia, whereas there was a cool bias 2–3 °C over Western Australia.

The spatial analysis reveals that in general there is a systematic dry bias across all of the physics configurations. This may contribute to the warm temperature bias of the ensemble through decreased land water availability that subsequently influences the surface energy partitioning. Further investigation of the cause of the dry bias is presented in section 4.

### 3.3. Diagnosis of Bias Tendencies by Physics Parameterization

To highlight the role of different parameterization choices on the simulated climate, Figure 7 presents box-plots derived from the spatial distribution of the model biases for the three regions and climate variables.

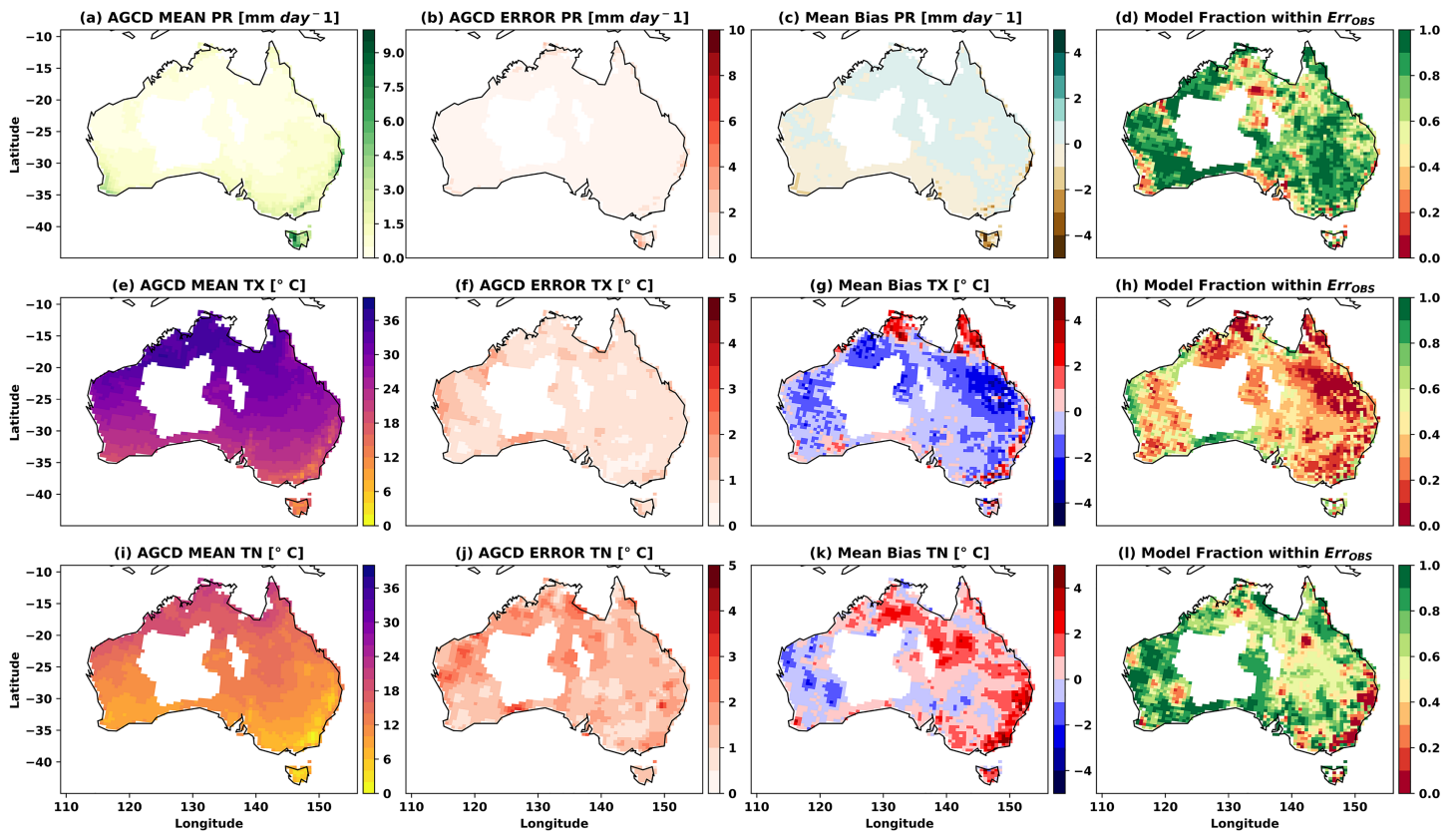


Figure 5. As per Figure 4 but for an example of extreme dry conditions over the period March to November 2009.

Each box represents the distribution of model biases for all ensemble members using a given parameterization. For precipitation (Figures 7a–7c) it is evident that most configurations have a dry bias (median  $\sim -0.2$  mm/day) for all regions with the exception of those configurations using the GF cumulus parameterization (median  $\sim +0.2$  mm/day). Configurations using the New Tiedtke cumulus parameterization appear to have the largest dry bias in the Tropics (median  $\sim -0.4$  mm/day). The precipitation bias is less skewed for configurations using the RRTMG radiation schemes over the Tropics and SEA (median of 0 mm/day). The QE biases (Figure 7j–7l) reflect similar attributes as the precipitation biases. Specifically, there was generally a negative QE bias (median of 2–8 mm/day) across the parameterizations and regions. A slight positive QE bias for configurations using GF, the largest negative bias when using the New Tiedtke (median of 6–8 mm/day) and less skewed biases for RRTMG (median  $\sim 0$  mm/day). The TX and TN biases show a moderate response to the wet precipitation biases. In particular, configurations using GF that have a wet precipitation bias and positive QE bias also have a cool TX bias (median of 1–2 °C, Figures 7d–7f) and less skewed TN (median of 0–0.5 °C, Figures 7g–7i) bias for all regions. However, in general across the configurations, there was a warm TX bias for the Tropics (median  $\sim 1$  °C, Figure 7d) and all regions have a warm TN bias (median 1–2 °C, Figures 7g–7i). Configurations using the YSU PBL or RRTMG radiation tended to run warmer than configurations with other PBL or radiation scheme choices. Configurations with the MYJ PBL scheme tend to have cool TX bias (median of 1 °C) in the Tropics that appear independent of the precipitation and QE biases. The New Goddard radiation scheme also has a cool bias for TX (median  $\sim 1$  °C) in all regions. When integrated over the entire simulation period, median biases are generally within an order of magnitude of zero for most surface climate variables (e.g., precipitation, TX and TN) and regions. However, some of the physics configurations tested here, as already noted, have a stronger bias tendency than others. Therefore, the selection of the physics configuration where possible should account for the potential that those with a strong bias tendency may influence results. This is perhaps more important in selecting configurations for running climate projections.

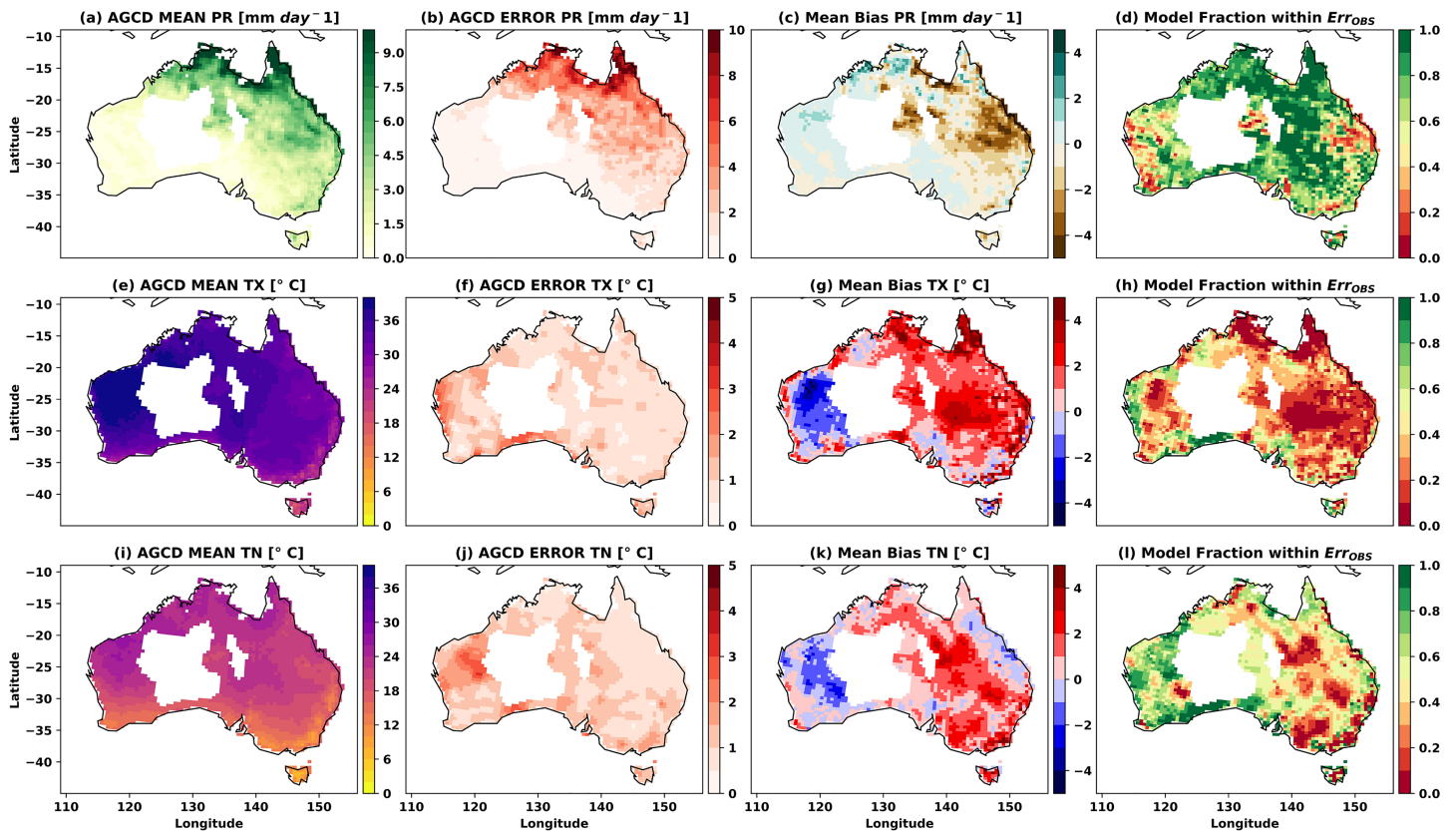
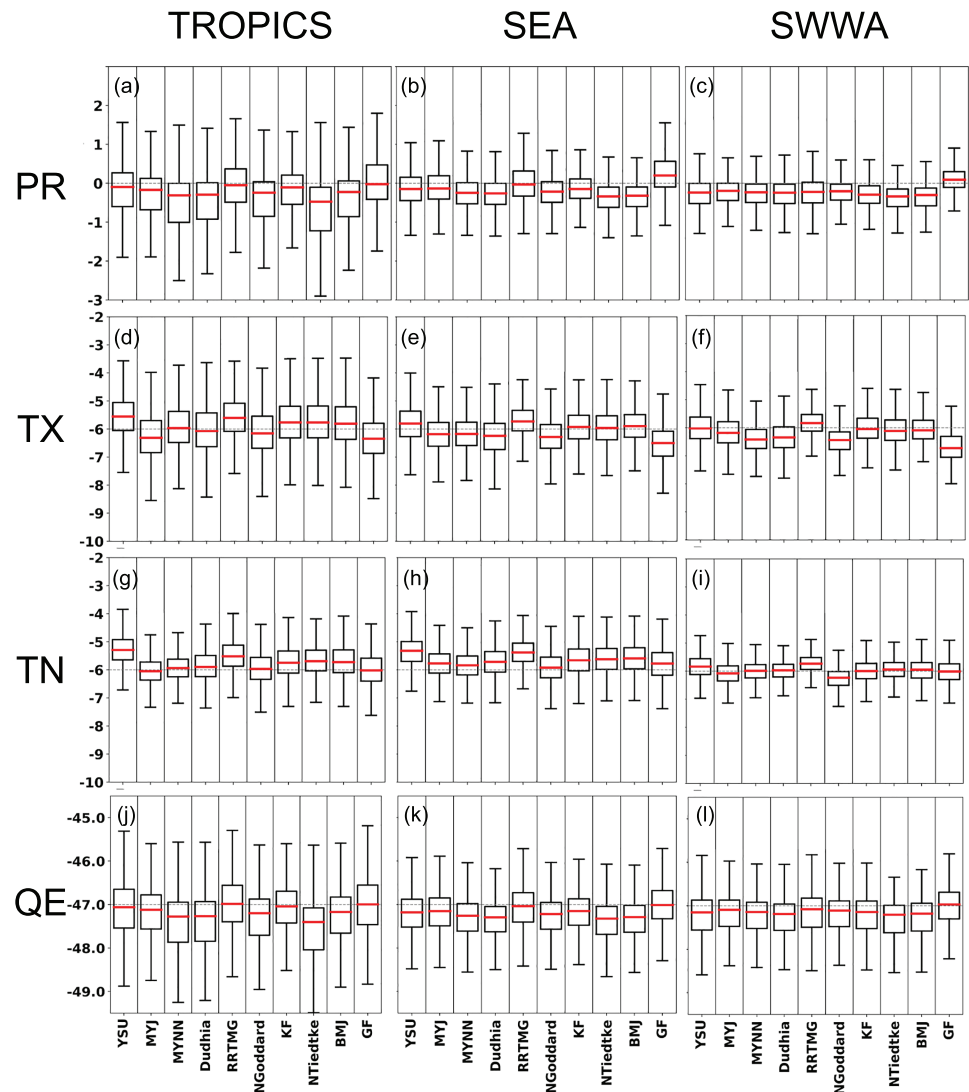


Figure 6. As per Figure 4 but for an example of extreme wet conditions over SEA for the period January to March 2010.

### 3.4. Assessment of Temporal Variability

To demonstrate skill in the simulation of temporal variability, regionally aggregated time series of key variables for the Tropics (Figure 8) and SEA (Figure 9) are presented with the full ensemble as the gray shaded region, and configurations with greater skill as colored lines. Precipitation time series show that some configurations are more skillful at simulating the timing of events. In particular, the GF scheme tended to have more frequent precipitation events than the observations (Figure S1) and the New Tiedtke scheme was one of the more accurate schemes regarding the timing of events but not necessarily the magnitude. This is consistent with Figure 7. TX and TN time series show that WRF-LIS-CABLE is skillful at capturing the variability of the observations. For TX more than TN, time series can vary from a positive to negative bias. Finally, the QE time series are the most telling regarding sensitivity to choice of WRF physics configuration with an ensemble range of  $50 \text{ W/m}^2$  for the warmer months of the simulation period of both regions. Even the configurations considered more skillful overall are subject to significant biases. For the Tropics, dry season QE is overestimated and during the wet season underestimated. The peaks in QE often coincide with precipitation events and then decay. For SEA, October to April QE is underestimated yet still responding to precipitation events. It is worth acknowledging that gridded observational estimates of QE are themselves a model product. Therefore, to evaluate whether this has an impact on evaluating WRF-LIS-CABLE skill in simulating QE, we also include estimates from vD18 in all QE time series. Estimates from vD18 are generally larger than GLEAM on heavy precipitation days and decline faster than GLEAM on the days following these events. However, both QE estimates suggested that while the revised hydrology of CABLE has addressed several of the deficiencies of the old scheme, there remains a tendency to underpredict QE. As noted earlier, both GLEAM and vD18 are model-derived estimates of QE and therefore we provide a comparison of WRF-LIS-CABLE QE estimates to six flux tower sites that had data covering the simulation period. Generally, WRF-LIS-CABLE can simulate the seasonality across the flux tower sites, but as in the GLEAM and vD18 comparison, often underestimates the magnitude of the flux. This is associated with differences between

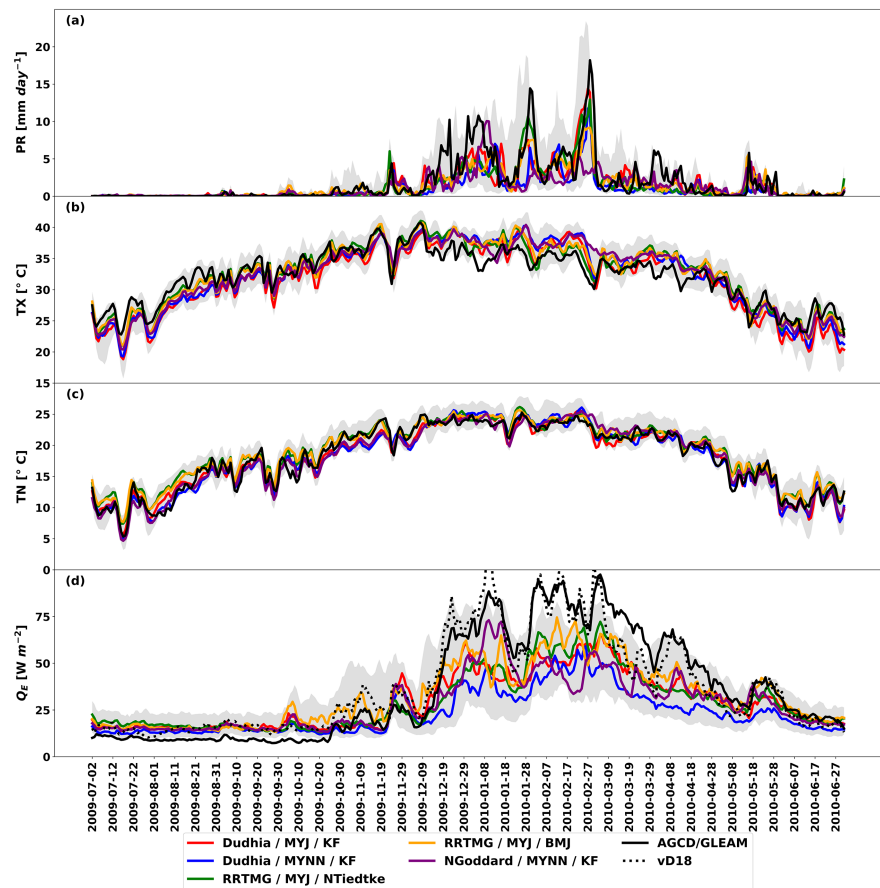


**Figure 7.** Boxplots of the ensemble bias for daily precipitation (PR, mm/day, a–c), daily maximum 2 m air temperature (TX, °C, d–f), daily minimum 2 m air temperature (TN, °C, g–i), and daily mean latent heat flux (QE,  $W/m^2$ , j–l). Each box is constructed from all the configurations that use a given parameterization scheme. Split according to the regions: Tropics (a, d, g, and j), SEA (b, e, h, and k), and SWWA (c, f, i, and l). Note that evaluation of the latent heat flux is again GLEAM.

the vegetation and soil characteristics of the model grid cell to which the flux tower data are compared among other factors.

#### 4. Discussion

In this paper we have evaluated the simulation skill of a 36-member ensemble of the WRF-LIS-CABLE regional climate model over the CORDEX AustralAsia domain encompassing a variety of climate regimes over the period October 2008 to October 2010 when several extreme events were experienced. The 36-member ensemble consists of different configurations of the WRF atmospheric parameterizations, with the main aim to identify those configurations that are more skillful when run with the CABLE LSM. We evaluate skill in simulating both temporal and spatial characteristics of Australian climate and extremes. However, we identify systematic dry biases over the Tropics during the monsoon season for most configurations. In the following sections, we discuss the following: (i) the implications of this bias, (ii) potential sources of bias, and (iii) a brief comparison between WRF-LIS-CABLE and existing CORDEX AustralAsia



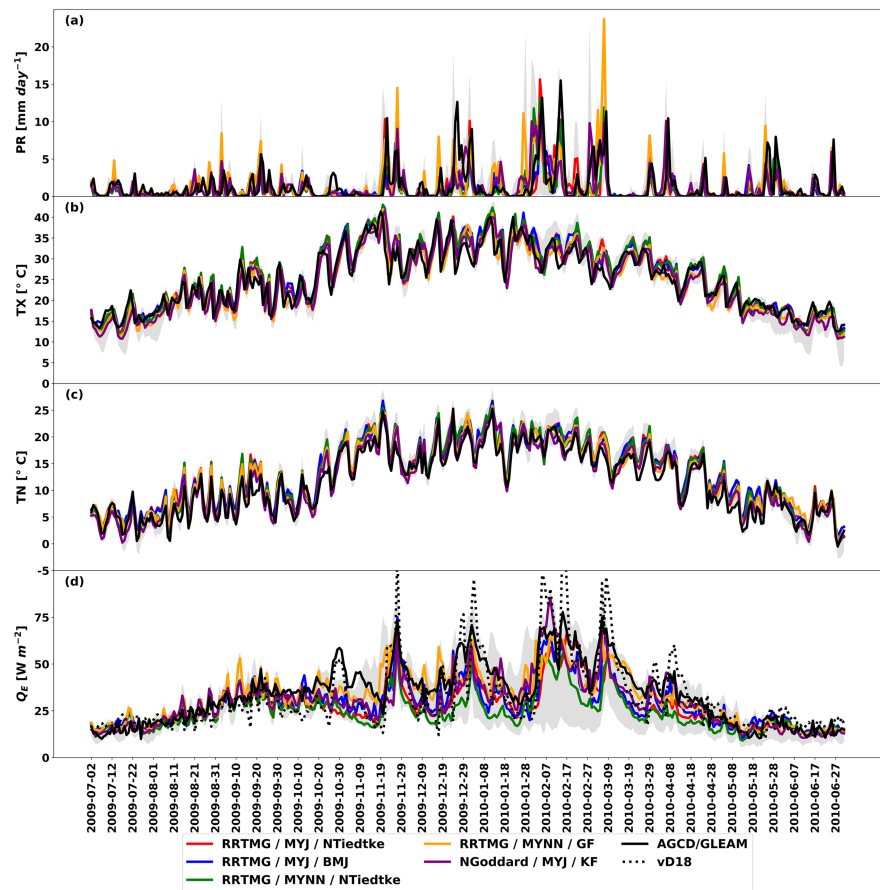
**Figure 8.** Area-averaged time series from July 2009 to June 2010 over the Tropics region: Daily precipitation (PR, mm/day, a), daily maximum 2 m air temperature (TX, °C, b), daily minimum 2 m air temperature (TN, °C, c), and daily mean latent heat flux (QE,  $W/m^2$ , d). For observations in black, physics ensemble range in gray shading and selection of physics configurations with highest relative skill scores across both climate and extremes. To gauge the observational uncertainty in QE, estimates from van Dijk et al. (2018) are also included as the dashed black line.

simulations using standard WRF with the Unified Noah LSM (hereafter referred to WRF-Noah). Finally, we provide a brief conclusion with recommendations on WRF configurations for use with CABLE.

#### 4.1. Summary of Systematic Biases and Their Implications

Of the 36 ensemble members, the nine configurations using the GF cumulus scheme had a higher precipitation frequency than the observational record (Figure S1). As a consequence, these configurations do not have the same systematic dry bias during the tropical monsoon season, as observed for the other configurations. Instead, this wet bias, contributes to a cool temperature bias. Therefore, it is likely that the GF configurations, avoided the dry bias for the wrong reasons, instead due to problems with the timing and magnitude of precipitation events. Previous evaluations of the different cumulus parameterizations over North America (e.g., Hu et al., 2018) also find that the GF scheme is wetter than the other cumulus schemes that we have tested. Note that the GF scheme performs better at finer-scale resolutions, as suggested by Fowler et al. (2016).

Regarding the selection of the PBL scheme, the temperature biases (e.g., Figure 7) of the YSU (warm;  $N = 12$ ) and MYJ (cold;  $N = 12$ ) configurations are broadly consistent with those reported by Cohen et al. (2015). The warm bias of the YSU PBL scheme may be associated with deeper boundary layers that contributes to drier conditions near the surface (Cohen et al., 2015). The MYNN configurations ( $N = 12$ ) appear to have the least skewed temperature biases. Despite the temperature biases showing sensitivity to PBL choice, when paired with specific cumulus and radiation schemes, these biases are not detrimental to overall simulation skill. In

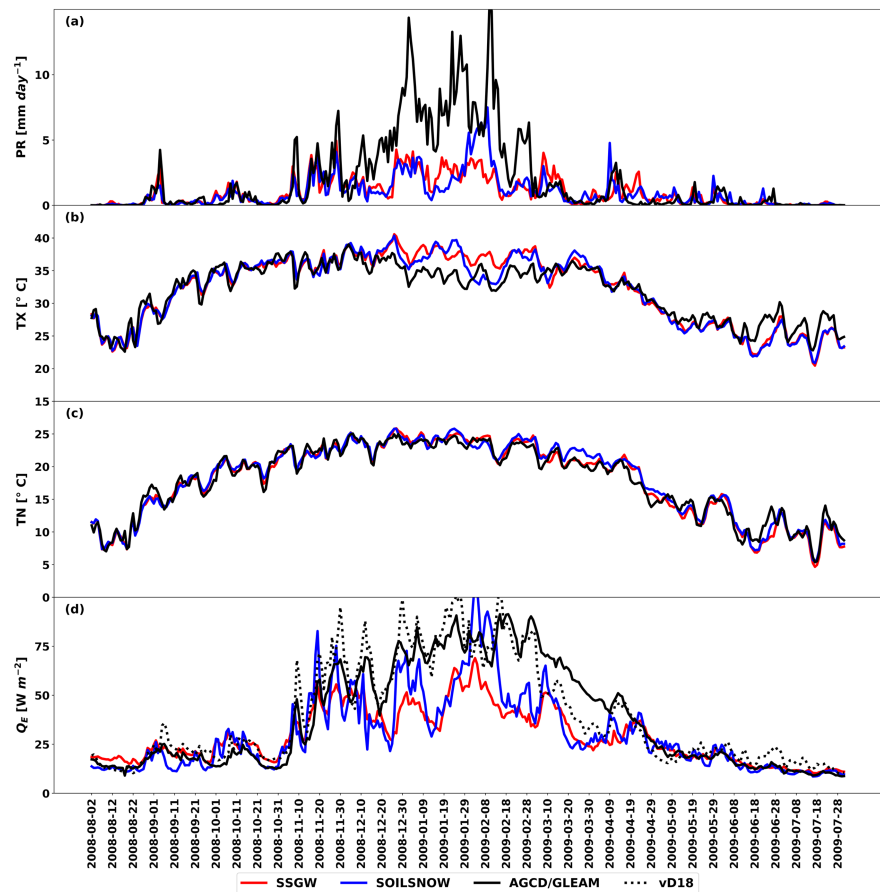


**Figure 9.** As per Figure 8 but for South East Australia.

particular, one of the configurations that had a higher overall total skill score, the RRTMG radiation/MYJ PBL/New Tiedtke cumulus scheme, uses physics that tend to have systematic biases (e.g., MYJ cold temperature bias and New Tiedtke dry precipitation bias; Figure 7). Incidentally, this configuration is also the first physics suite recommended on the WRF user community webpage ([http://www2.mmm.ucar.edu/wrf/users/ncar\\_convection\\_suite.php](http://www2.mmm.ucar.edu/wrf/users/ncar_convection_suite.php)).

There are 27 configurations that had a systematic dry bias (5 mm/day) during the tropical monsoon season that had subsequent consequences on the simulation of surface temperatures (+5 °C) due to the underprediction of latent heating (i.e., evaporative cooling). There are, however, legitimate concerns about what implications these biases may have on subsequent research or climate adaptation planning because of the impact this has on the predictability of heat extremes and their underlying mechanisms. Systematic evapotranspiration biases have been identified in offline models (e.g., Whitley et al., 2016) and in global coupled models (e.g., Ukkola et al., 2018). Therefore, they are not likely to be specific to these simulations. However, we do acknowledge that biases are likely to change for a different simulation domain, resolution, simulation period, land cover scenario or boundary conditions. Should the same configurations be employed to down-scale future climate projections or estimate the likelihood of particular climate extremes, these systematic biases would likely have an impact on conclusions drawn from those simulations. Therefore, we recommend future users of WRF-LIS-CABLE consider the potential influence of any systematic bias when interpreting their results.

Most of the 36 ensemble members are comparable in the absolute skill scores for simulating the Australian climate (Figure S2) with greater differentiation between the schemes when it comes to the simulation of extremes (Figure S3). Therefore, it is recommended that configuration selection occurs in the context of the intended purpose.



**Figure 10.** As per Figure 8 for the period August 2008 to July 2009, including the coupled spin-up period, Tropical region but comparing the CABLE hydrology schemes with the newer SSGW scheme (red), the older SOILSNOW scheme (blue), and observed (black).

## 4.2. Potential Land Surface Limitations

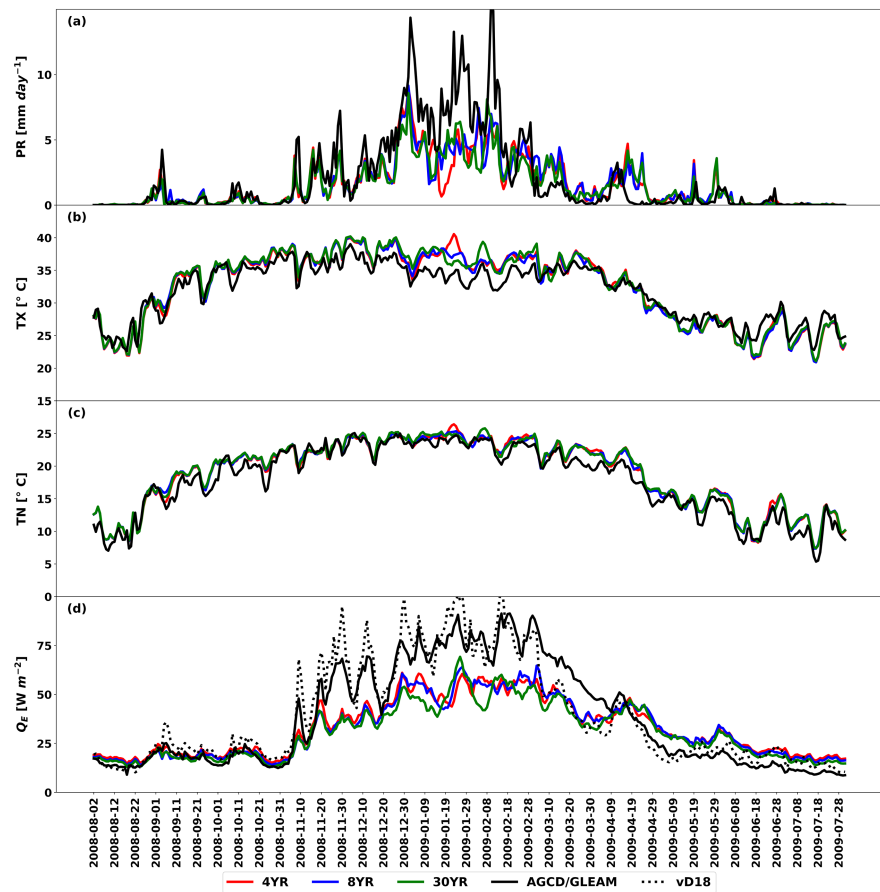
Over the simulation period, WRF-LIS-CABLE has a persistent dry bias over the Tropics (5 mm/day; Figure 4c) that propagates through the coupled model to contribute to the warm temperature bias via the underestimation of the latent heat flux. To establish whether this is a feature of coupling CABLE to NUWRF, the one component that is common to all ensemble members, we tested two potential sources: the first is the CABLE soil hydrology and the second is the CABLE offline spinup length. The former provides a basis to compare the impact of the new hydrology in CABLE while the latter considers whether the 4-year spin-up period was sufficient for the newer hydrology parameterizations.

### 4.2.1. Impact of CABLE Hydrology Configuration

Recent developments of CABLE have significantly improved the representation of soil hydrology (Decker, 2015; Decker, Or, et al., 2017), that has had an impact on simulated QE which forms one of the lower boundary conditions for the WRF model. In particular, the new soil hydrology has improved problems with excessive evapotranspiration (e.g., Decker, 2015) but this may contribute to the dry precipitation bias by limiting the land surface moisture source to the atmosphere. Therefore, to examine whether this improvement in the soil hydrology has had an impact on the coupled simulations, we compare (Figure 10) two simulations with the New Goddard radiation/MYNN PBL/BMJ cumulus configuration but with different soil hydrology schemes that either use the newer scheme (Subgrid Soil Ground Water: SSGW) or its predecessor (SOILSNOW).

Clearly changes in the soil hydrology have the greatest impact on the simulation of land surface conditions with significant differences in QE (Figure 10d). In particular, SSGW has greater overall skill (Figure S8) in QE (0.5–0.6 for SSGW compared to 0.35–0.55 for SOILSNOW). Although SOILSNOW appears to achieve





**Figure 11.** As per Figure 8 for the period August 2008 to July 2009, including the coupled spin-up period, for the Tropical region but comparing a 4 year (red), 8 year (blue), and 30 year (green) offline spin-up of CABLE to observations.

higher QE magnitudes ( $s_{\text{bias}}$  of  $\sim 0.75$  for SOILSNOW,  $\sim 0.65$  for SSGW) the skill (relative to GLEAM) at capturing the variability is considerably poorer ( $s_{\text{rmse}}$  of  $\sim 0.2$  for SOILSNOW,  $\sim 0.4$  for SSGW). The differences in QE do have some impact on the surface climate. More specifically, for precipitation (Figure 10a) SOILSNOW on occasion has considerably higher daily precipitation but still under predicts observed values. The warm TX and TN biases (Figures 10b and 10c) are also not necessarily resolved by changing to SOILSNOW. For SEA (not shown) the higher QE values often contribute to a cool TX bias. Therefore, although the soil hydrology parameterization can have a considerable influence on the simulated land surface conditions, the differences are not sufficient to resolve the dry bias in the WRF-LIS-CABLE simulations.

#### 4.2.2. Impact of CABLE Offline Spin-Up Length

Another potential source for the systematic model biases may stem from the initialization of the coupled simulations where the offline spin-up of 4 years covers the latter half of the Millennium Drought. Certainly, emerging research on decadal predictability of mean temperature and precipitation indicates that initializing coupled simulations during extended periods of extreme dry or wet conditions may have an impact on skill in the first few years (e.g., Liu et al., 2019). In particular, the groundwater scheme can take longer to stabilize where the recharge rates can only be suitably estimated over a period that captures the considerable climate variability of the region. Therefore, we conduct another two simulations with a longer CABLE offline spin-up of 8 and 30 years to evaluate whether the length of the offline spin-up has an impact on the coupled simulation skill. Here, we use the WRF atmospheric configuration with RRTMG radiation/MYJ PBL/New Tiedtke cumulus parameterization. Time series analysis (Figure 11) indicates that there are subtle differences in QE with differences of order  $1 \text{ W/m}^2$  for the first several months of the coupled simulation period. After the first 4 months the QE time series for both longer spin-up simulations is

smoother indicating that the land surface is less reactive to precipitation events. However, in the context of the surface climate the impact of a longer CABLE offline spin-up is marginal. At the peak of the monsoon season, there remains a dry precipitation bias ( $\sim 5$  mm/day), although not always as dry as the 4-year spin-up simulation. TX and TN biases ( $\sim 2$  °C) still reflect a close relationship to the QE biases ( $\sim 25$  W/m<sup>2</sup>), with warm biases coinciding with a negative QE bias and cool bias when there is a positive QE bias. An examination of the skill metrics (Figure S9) also suggest that some skill improvement is possible but does not resolve the dry precipitation bias over the tropics. Therefore, the CABLE offline spin-up length is not likely the main cause of the persistent dry bias but can contribute some skill improvement.

#### 4.2.3. Other Potential Sources for Systematic Biases

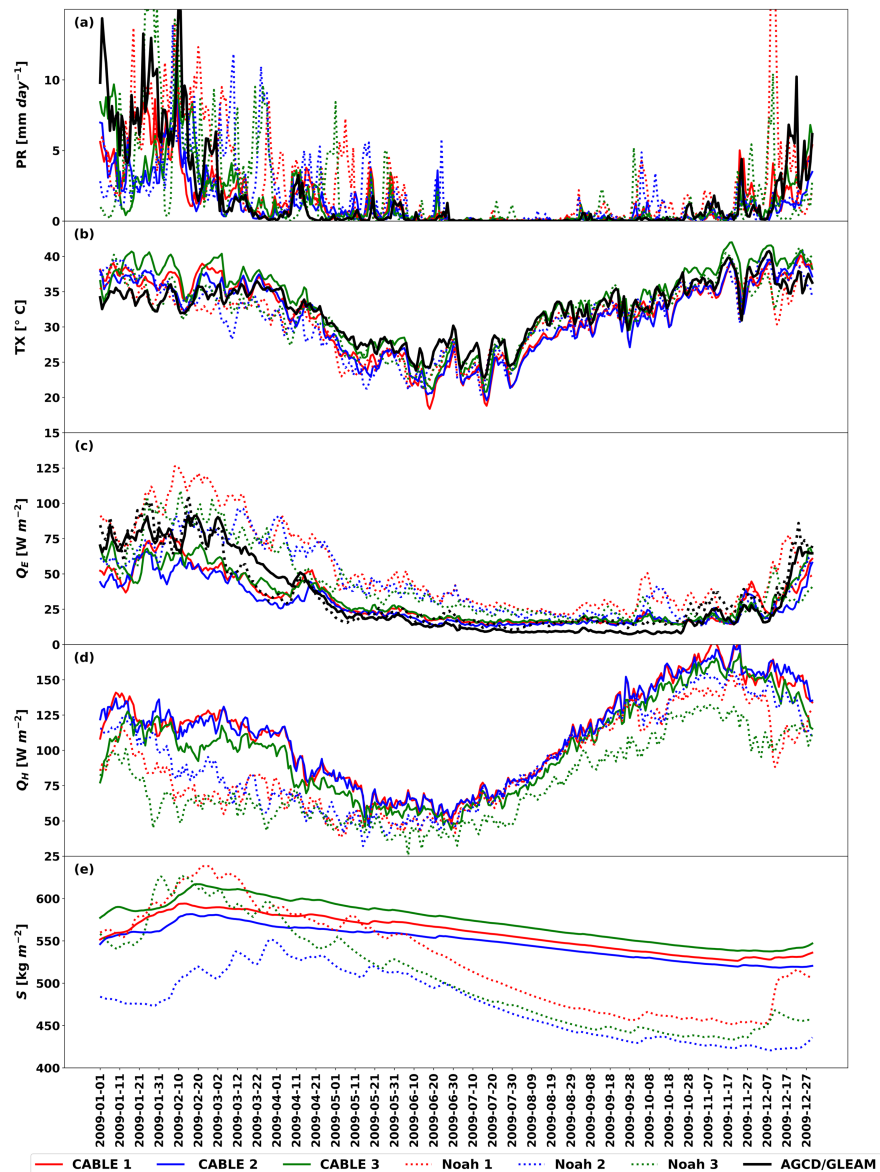
Although limitations with the land surface representation remains a possible source for the systematic biases we identify with WRF-LIS-CABLE, other sources cannot be excluded without further testing. In particular, as the coupled simulations are initialized toward the end of an extended period of drought, the dry bias may stem from the boundary conditions that establish the large-scale synoptic features of the domain. To evaluate this, one could either test with a different data source for the boundary conditions or select a wetter time period to run WRF-LIS-CABLE. Both of these are beyond the scope of the present study where the aim was to evaluate the simulation skill of WRF-LIS-CABLE for a period coinciding with the termination of prolonged drought conditions.

Previous research suggests that the ability to simulate tropical climate is linked to the resolution of the model and in our case, perhaps also the forcing data sets. Tropical cyclones and closed tropical low-pressure systems make an important contribution to annual rainfall in northern Australia (Lavender & Abbs, 2013). Previous research suggests there is a relationship between model spatial resolution and the simulated rate of tropical cyclone formation (e.g., Walsh et al., 2013). More specifically, models with a finer resolution are able to better resolve the pattern of cyclone formation. This raises the possibility that the driving data, which has a resolution of  $\sim 80$  km, might not adequately capture aspects of these phenomena, such as their frequency of occurrence, and this may also contribute to the dry patterns noted in this study. Furthermore, it is unlikely that the resolution of our simulations ( $0.44^\circ$ ) is able to appropriately resolve the convective precipitation associated with these weather systems. However, the resolution of our simulations enables comparison to existing CORDEX simulations (see section 4.3) at a reasonable computational expense given the large ensemble size.

The underestimation of monsoon season evapotranspiration (ET; Figure 8d) is not unique to WRF-LIS-CABLE, and many daily ET data sets are model derivatives themselves (e.g., GLEAM and vD18). In particular, it appears to be a common feature of many different types of LSMs, particularly for tropical Savanna ecosystems (e.g., Whitley et al., 2016). More specifically, the underestimation of ET is mostly constrained to the wet season and not the dry season (Figure 8d). Whitley et al. (2016) showed that ET biases resulted from deficiencies in the representation of seasonal variation in vegetation characteristics such as grass phenology and rooting depth. This hypothesis has some credence as these are known limitations in the current representation of vegetation within CABLE (e.g., Whitley et al., 2016). Therefore, the systematic bias over the tropical monsoon season likely stems partially from limitations in both the LSM CABLE and in the WRF physics experimental setup. Further decomposition of the respective contributions will be the focus of future research.

#### 4.3. Comparison Between WRF-LIS-CABLE and WRF-Noah

To provide context on the potential advantages of using WRF-LIS-CABLE we compare the output from three of the tested configurations to existing simulations run with standard WRF with the Unified-Noah LSM for CORDEX AustralAsia (Di Virgilio et al., 2019). Both WRF-LIS-CABLE and WRF-Noah simulations are run on the CORDEX AustralAsia domain with the same ERA-Interim boundary conditions. Note that here, the same PBL, cumulus and radiation configurations are compared and only for the year 2009, the common period between both sets of simulation experiments. There are, however, a few differences between the WRF-LIS-CABLE and WRF-Noah experimental setups. The first is in the WRF versions, 3.9.1.1. (WRF-LIS-CABLE) and 3.6.0 (WRF-Noah). Second are differences in the simulation setup. For WRF-LIS-CABLE, land surface variables had a 4-year offline spin-up and a coupled atmospheric spin-up of 2 months. In contrast for WRF-Noah, there is no offline spin-up but the simulations start in 1980 and therefore the effective spin-up is comparatively longer. The third main difference concerns the microphysics parameterization, we use the



**Figure 12.** Area-averaged time series from January to December 2009 over the Tropics region: Daily precipitation (PR, mm/day, a), daily maximum 2 m air temperature (TX, °C, b), daily mean latent heat flux (QE,  $W/m^2$ , c), daily mean sensible heat flux (QH,  $W/m^2$ , d), and total column soil moisture (S,  $kg/m^2$ , e). Observations are in black for PR, TX, and QE. Solid colored lines correspond to WRF-LIS-CABLE simulations and dashed lines correspond to WRF-Noah simulations. Physics configurations are all with the Dudhia/RRTM radiation schemes and either MYJ-KF (red), MYJ-BMJ (blue), or YSU-KF (green) PBL/Cumulus scheme combinations. Note that the YSU-KF configurations both use WSM5 microphysics. To gauge the observational uncertainty in QE, estimates from van Dijk et al. (2018) are also included as the dashed black line.

WSM5 scheme for all WRF-LIS-CABLE simulations. For the WRF-Noah simulations, two of the simulations use the WRF Double Moment 5-class (WDM5) while one uses WSM5 (with YSU PBL/KF cumulus). Note that although key differences in the simulation setup exist between WRF-LIS-CABLE and WRF-Noah, this comparison provides a basis for the respective merits of each modeling system in the simulation of Australian climate.

The impact of the LSM selection (Figure 12) is evident in both surface climate and land surface conditions. Coincidentally, the precipitation rates from WRF-Noah are considerably larger and not synchronized with the observations. Maximum temperatures during March-April (Figure 12b) show that WRF-LIS-CABLE

has a warm bias, while WRF-Noah has a cool bias but otherwise in the latter half of 2009 both LSMs show similar biases. Perhaps the largest impact of the LSM is in the surface energy partitioning between latent (QE) and sensible (QH) heating. In particular, WRF-Noah consistently overpredicts QE (Figure 12c), with subsequently lower estimates for QH (Figure 12d) that likely explains the cool temperature bias at the beginning of 2009 for the WRF-Noah simulations. In contrast, WRF-LIS-CABLE simulations appear more skillful in the estimation of QE, despite the underestimation during the wetter months. The total column soil moisture ( $S$ , Figure 12e) reveals that there is a stronger decrease over 2009 in WRF-Noah than in WRF-LIS-CABLE. WRF-Noah soil moisture certainly shows more variability than WRF-LIS-CABLE; however, there could be multiple reasons for this, which is beyond the scope of this paper. Overall, precipitation and latent heat time series for the Tropics (Figures 12a and 12c) and also for SEA (Figure S10) suggest that WRF-LIS-CABLE has superior skill over Australia. The skill metrics (Figure S11) also confirm that WRF-LIS-CABLE has greater skill than WRF-Noah, particularly for the bias and RMSE skill scores associated with the simulation of precipitation and latent heating.

Given the excessive tropical wet season precipitation in WRF-Noah this suggests that the dry bias observed for WRF-LIS-CABLE is a symptom of two possible causes: (i) coupled initialization during a drought, and (ii) land surface representation. While it is possible that the microphysics selection could also be a contributing factor, we exclude this on the basis that the precipitation time series comparison (Figure 12a) suggests that the impact of the LSM is greater than differences in the microphysics selection at the model resolution tested here. Specifically, comparing the green configurations that use YSU PBL/KF cumulus/WSM5 microphysics the difference between WRF-LIS-CABLE and WRF-Noah are comparable to the difference between these models where the microphysics also differs. Similar conclusions about the impact of microphysics selection at resolutions  $>10$  km also conclude that this has minimal impact on precipitation amounts compared to the selection of PBL and cumulus parameterization schemes (e.g., Evans et al., 2012; Hu et al., 2018).

## 5. Conclusions

In this paper we have evaluated the simulation skill of a 36-member ensemble with the WRF-LIS-CABLE model covering the period October 2008 to October 2010. This period corresponds to the transition period out of the Millennium Drought, which affected Eastern Australia during the 2000s, and was marked by several extreme events, providing a basis to evaluate the spatial and temporal skill of WRF-LIS-CABLE in the context of both climate and extremes.

By varying the WRF atmospheric physics configurations we are able to provide recommendations on the atmospheric physics configuration of WRF-LIS-CABLE. Most simulations were comparable in their skill with some notable exceptions. These include an overprediction of the precipitation frequency when using the GF cumulus scheme, irrespective of which other physics parameterizations it is paired with. The New Tiedtke scheme was one of the better cumulus schemes when it came to simulating precipitation timing. Of the radiation schemes, RRTMG has a warm bias while the other two schemes (Dudhia and New Goddard) did not have a significant bias tendency. Regarding PBL choice, the YSU has a warm bias and MYJ a cool bias, consistent with previous research. The choice of cumulus and radiation scheme to which the PBL schemes are paired with can moderate these biases through their interaction. In particular, configurations using MYJ and RRTMG were often identified as top performers, perhaps because together their opposing temperature bias tendencies canceled. Although we identify WRF physics configurations that have a relatively higher skill, we see limited consistency in the selection of configurations across regions and between climate versus extremes. This is likely due to challenges in capturing both the temporal variability and magnitude of all of the variables examined here. For example, a configuration may perform well at simulating extremes but not the climate due to a systematic bias. Finally, it is possible that should a different domain, resolution or boundary conditions be used, these results may not apply. Future research, however, can be informed by the results presented here, particularly on the potential trade-offs in skill when selecting particular configurations.

Simulations over tropical Australia and SWWA were generally less skillful than that over South East Australia. This largely stems from a significant dry bias during the monsoon season over tropical Australia that leads to a hot temperature bias associated with the underestimation of latent cooling. Further investigation of the dry bias demonstrated the limited role of the offline LSM spin-up length and

soil hydrology scheme. It is likely that the dry bias is a result of a combination of initializing the coupled simulations during a period of extended drought, resolving tropical systems in the boundary conditions and limitations in the representation of savanna ecosystems. Resolving this bias is the focus of future research. Our results also highlight the sensitivity of WRF options in different regions of Australia. Future studies focusing on the Tropics and SWWA should pay more attention to the choice of physics options as compared to SEA. We also compared WRF-LIS-CABLE to existing WRF-Noah simulations with the same radiation/PBL/cumulus configuration to provide context of the potential advantages of using WRF-LIS-CABLE. These include (i) more realistic representation of evapotranspiration, and (ii) gains in computational efficiency through offline land surface spinup ideal for running case study simulations of extreme events. This makes running WRF-LIS-CABLE a very attractive option to target further science questions, especially those related to the role of groundwater, which is not resolved by the unified Noah model in WRF but it is the Noah-MP LSM.

Finally, our analysis does not clearly identify a single ideal best-performing model configuration of WRF-LIS-CABLE for all regions and variables of interest. However, we have clearly demonstrated which WRF physics configurations are likely to provide the best results for a given variable and region, and for extremes versus mean climate. This in turn provides a basis for future studies to develop physics ensembles, depending on two key factors: (i) the region of interest within the CORDEX AustralAsia domain, and (ii) whether the focus is the climate means or extremes or both, and this is highly valuable.

#### Acknowledgments

This research was undertaken with the assistance of resources and services from the National Computational Infrastructure (NCI), which is supported by the Australian Government. ERA-Interim data used in this study/project have been provided by ECMWF and obtained from the ECMWF Data Server. This project is supported through funding from the Australian Research Council (ARC) Centre of Excellence for Climate Extremes (CE170100023). J. Kala is supported by an ARC Discovery Early Career Researcher Grant (DE170100102). All data will be published as an NCI data collection and all software scripts to process the data and reproduce the analysis are available at: [https://zenodo.org/record/3563154#\\_XfGAL79S88Z](https://zenodo.org/record/3563154#_XfGAL79S88Z) (doi:10.5281/zenodo.3563154).

#### References

- Beringer, J., Hutley, L. B., McHugh, I., Arndt, S. K., Campbell, D., Cleugh, H. A., et al. (2016). An introduction to the Australian and New Zealand flux tower network—OzFlux. *Biogeosciences*, *13*, 5895–5916. <https://doi.org/10.5194/bg-13-5895-2016>
- Bi, D., M. Dix, S. J. Marsland, S. O'Farrell, H. A. Rashid, P. Uotila, et al. (2013). The ACCESS coupled model: Description, control climate and evaluation, *Australian Meteorological and Oceanographic Journal*, *63*, 41–64. <https://doi.org/10.22499/2.6301.004> [http://www.bom.gov.au/climate/annual\\_sum/2009/AnClimSum09\\_HR1.1.pdf](http://www.bom.gov.au/climate/annual_sum/2009/AnClimSum09_HR1.1.pdf)
- Bureau of Meteorology (2009). *Annual Climate Summary 2009*. Canberra, Australia: Commonwealth of Australia. ISSN 1837-8854. [http://www.bom.gov.au/climate/annual\\_sum/2009/AnClimSum09\\_HR1.1.pdf](http://www.bom.gov.au/climate/annual_sum/2009/AnClimSum09_HR1.1.pdf)
- Chen, F., & Dudhia, J. (2001). Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 Modeling System. Part I: Model implementation and sensitivity. *Monthly Weather Review*, *129*, 569–585. [https://doi.org/10.1175/1520-0493\(2001\)129<0569:CAALSH>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2)
- Chou, M.-D., & Suarez, M. J. (1999). A solar radiation parameterization for atmospheric studies, NASA Tech. Memo. NASA/TM-1999-104606, 15, 40 pp. <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19990060930.pdf>
- Chou, M. D., Suarez, M. J., Liang, X. Z., & Yan, M. M. H. (2001). A thermal infrared radiation parameterization for atmospheric studies, NASA Tech. Memo. NASA/TM-2001-104606, 19, 68 pp. <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20010072848.pdf>
- Cohen, A. E., Cavallo, S. M., Coniglio, M. C., & Brooks, H. E. (2015). A review of planetary boundary layer parameterization schemes and their sensitivity in simulating southeastern U.S. cold season severe weather environments. *Weather and Forecasting*, *30*, 591–612. <https://doi.org/10.1175/WAF-D-14-001105.1>
- Collier, N., Hoffman, F. M., Lawrence, D. M., Keppel-Aleks, G., Koven, C. D., Riley, W. J., et al. (2018). The International Land Model Benchmarking (ILAMB) system: Design, theory, and implementation. *Journal of Advances in Modeling Earth Systems*, *10*, 2731–2754. <https://doi.org/10.1029/2018MS001354>
- De Kauwe, M. G., Kala, J., Lin, Y. S., Pitman, A. J., Medlyn, B. E., Duursma, R. A., et al. (2015). A test of an optimal stomatal conductance scheme within the CABLE land surface model. *Geoscientific Model Development*, *8*, 431–452. <https://doi.org/10.5194/gmd-8-431-2015>
- Decker, M. (2015). Development and evaluation of a new soil moisture and runoff parameterization for the CABLE LSM including subgrid-scale processes. *Journal of Advances in Modeling Earth Systems*, *7*, 1788–1809. <https://doi.org/10.1002/2015MS000507>
- Decker, M., Ma, S., & Pitman, A. (2017). Local land-atmosphere feedbacks limit irrigation demand. *Environmental Research Letters*, *12*, 054003. <https://doi.org/10.1088/1748-9326/aa65a6>
- Decker, M., Or, D., Pitman, A., & Ukkola, A. (2017). New turbulent resistance parameterization for soil evaporation based on a pore-scale model: Impact on surface fluxes in CABLE. *Journal of Advances in Modeling Earth Systems*, *9*, 220–238. <https://doi.org/10.1002/2016MS000832>
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., 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. <https://doi.org/10.1002/qj.828>
- Di Virgilio, G., Evans, J. P., Di Luca, A., Olson, R., Argüeso, D., Kala, J., et al. (2019). Evaluation of ERA-Interim-driven CORDEX regional climate model simulations over Australia. *Climate Dynamics*, *53*(5–6), 2985–3005. <https://doi.org/10.1007/s00382-019-04672-w>
- Dirmeyer, P. A., Dolman, A. J., & Sato, N. (1999). The global soil wetness project: A pilot project for global land surface modeling and validation. *Bulletin of the American Meteorological Society*, *80*, 851–878. [https://doi.org/10.1175/1520-0477\(1999\)080<0851:TPPOTG>2.0.CO;2](https://doi.org/10.1175/1520-0477(1999)080<0851:TPPOTG>2.0.CO;2)
- Dudhia, J. (1989). Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *Journal of the Atmospheric Sciences*, *46*, 3077–3107. [https://doi.org/10.1175/1520-0469\(1989\)046<3077:NSOCOD>2.0.CO;2](https://doi.org/10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2)
- Evans, J. P., Ekström, M., & Ji, F. (2012). Evaluating the performance of a WRF physics ensemble over south-east Australia. *Climate Dynamics*, *39*, 1241–1258. <https://doi.org/10.1007/s00382-011-1244-5>
- Fowler, L. D., Skamarock, W. C., Grell, G. A., Freitas, S. R., & Duda, M. G. (2016). Analyzing the Grell-Freitas convection scheme from hydrostatic to nonhydrostatic scales within a global model. *Monthly Weather Review*, *144*, 2285–2306. <https://doi.org/10.1175/MWR-D-15-0311.1>

- Friedl, M. A., McIver, D. K., Hodges, J. C. F., Zhang, X. Y., Muchoney, D., Strahler, A. H., et al. (2002). Global land cover mapping from MODIS: Algorithms and early results. *Remote Sensing of Environment*, *83*, 287–302. [https://doi.org/10.1016/S0034-4257\(02\)00078-0](https://doi.org/10.1016/S0034-4257(02)00078-0)
- Gbode, I. E., Dudhia, J., Ogunjobi, K. O., & Ajayi, V. O. (2019). Sensitivity of different physics schemes in the WRF model during a West African monsoon regime. *Theoretical and Applied Climatology*, *136*, 733–751. <https://doi.org/10.1007/s00704-18-2538-x>
- Grell, G. A., & Freitas, S. R. (2014). A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. *Atmospheric Chemistry and Physics*, *14*, 5233–5250. <https://doi.org/10.5194/acp-14-5233-2014>
- Harrison, K. E., Kumar, S. V., Santanello, J. A., & Peters-Lidard, C. D. (2012). Quantifying the change in soil moisture modeling uncertainty from remote sensing observations using Bayesian inference techniques. *Water Resources Research*, *48*, W11514. <https://doi.org/10.1029/2012WR012337>
- Haverd, V., Cuntz, M., Nieradzki, L. P., & Harman, I. N. (2016). Improved representations of coupled soil–canopy processes in the CABLE land surface model (Subversion revision 3432). *Geoscientific Model Development*, *9*, 3111–3122. <https://doi.org/10.5194/gmd-9-3111-2016>
- Haverd, V., Smith, B., Nieradzki, L., Briggs, P. R., Woodgate, W., Trudinger, C. M., et al. (2018). A new version of the CABLE land surface model (Subversion revision r4601) incorporating land use and land cover change, woody vegetation demography, and a novel optimisation-based approach to plant coordination of photosynthesis. *Geoscientific Model Development*, *11*, 2995–3026. <https://doi.org/10.5194/gmd-11-2995-2018>
- Hirsch, A. L., Kala, J., Pitman, A. J., Carouge, C., Evans, J. P., Haverd, V., & Mocko, D. (2014). Impact of land surface initialisation approach on sub-seasonal forecast skill: A regional analysis in the southern hemisphere. *Journal of Hydrometeorology*, *15*, 300–319. <https://doi.org/10.1175/JHM-D-13-05.1>
- Hirsch, A. L., Pitman, A. J., & Kala, J. (2014). The role of land cover change in modulating the soil moisture–temperature land–atmosphere coupling strength over Australia. *Geophysical Research Letters*, *41*, 5883–5890. <https://doi.org/10.1002/2014GL061179>
- Hirsch, A. L., Pitman, A. J., Kala, J., Lorenz, R., & Donat, M. G. (2015). Modulation of land use change impacts on temperature extremes via land–atmosphere coupling over Australia. *Earth Interactions*, *19*, 1–24. <https://doi.org/10.1175/EI-D-15-0011.1>
- Hirsch, A. L., Pitman, A. J., Seneviratne, S. I., Evans, J. P., & Haverd, V. (2014). Summertime maximum and minimum temperature asymmetry over Australia determined using WRF. *Geophysical Research Letters*, *41*, 1546–1552. <https://doi.org/10.1002/2013GL059055>
- Hong, S.-Y., Noh, Y., & Dudhia, J. (2006). A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review*, *134*, 2318–2341. <https://doi.org/10.1175/MWR3199.1>
- Hu, X.-M., Xue, M., McPherson, R. A., Martin, E., Rosendahl, D. H., & Qiao, L. (2018). Precipitation dynamical downscaling over the Great Plains. *Journal of Advances in Modeling Earth Systems*, *10*, 421–447. <https://doi.org/10.1002/2017MS001154>
- Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research*, *113*, D13103. <https://doi.org/10.1029/2008JD009944>
- Janjic, Z. I. (1994). The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer and turbulence closure schemes. *Monthly Weather Review*, *122*, 927–945. [https://doi.org/10.1175/1520-0493\(1994\)122<0927:TSMECM>2.0.CO;2](https://doi.org/10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2)
- Janjic, Z. I. (1996). The surface layer in the NCEP Eta Model, Eleventh Conference on Numerical Weather Prediction, Norfolk, VA, 19–23 August; Amer. Meteor. Soc., Boston, MA, 354–355. [http://www2.mmm.ucar.edu/wrf/users/phys\\_refs/SURFACE\\_LAYER/eta\\_part3.pdf](http://www2.mmm.ucar.edu/wrf/users/phys_refs/SURFACE_LAYER/eta_part3.pdf)
- Janjic, Z. I. (2000). Comments on “Development and Evaluation of a Convection Scheme for Use in Climate Models”. *Journal of the Atmospheric Sciences*, *57*, 3686–3686. [https://doi.org/10.1175/1520-0469\(2000\)057<3686:CODAEO>2.0.CO;2](https://doi.org/10.1175/1520-0469(2000)057<3686:CODAEO>2.0.CO;2)
- Janjic, Z. I. (2002). Nonsingular implementation of the Mellor–Yamada Level 2.5 scheme in the NCEP meso model, NCEP Oce Note, No. 437, 61 pp., <http://citeserx.ist.psu.edu/viewdoc/download?doi=10.1.1.459.5434&rep=rep1&type=pdf>
- Jones, D., Wang, W., & Fawcett, R. (2009). High-quality spatial climate data-sets for Australia. *Australian Meteorological Magazine*, *58*, 233–248.
- Kain, J. S. (2004). The Kain–Fritsch convective parameterization: An update. *Journal of Applied Meteorology*, *43*, 170–181. [https://doi.org/10.1175/1520-0450\(2004\)043<0170:TKCPAU>2.0.CO;2](https://doi.org/10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2)
- Kala, J., Andrys, J., Lyons, T. J., Foster, I. J., Evans, B. J. (2015). Sensitivity of WRF to driving data and physics options on a seasonal time-scale for the southwest of Western Australia. *Climate Dynamics*, *44*, 633–659. <https://doi.org/10.1007/s00382-014-2160-2>
- Kala, J., De Kauwe, M. G., Pitman, A. J., Lorenz, R., Medlyn, B. E., Wang, Y.-P., et al. (2015). Implementation of an optimal stomatal conductance scheme in the Australian Community Climate Earth Systems Simulator (ACCESS1.3b). *Geoscientific Model Development*, *8*, 3877–3889. <https://doi.org/10.5194/gmd-8-3877-2015>
- King, A. D., Alexander, L. V., & Donat, M. G. (2013). The efficacy of using gridded data to examine extreme rainfall characteristics: A case study for Australia. *International Journal of Climatology*, *33*, 2376–2387. <https://doi.org/10.1002/joc.3588>
- Kowalczyk, E. A., Wang, Y. P., Law, R. M., Davies, H. L., McGregor, J. L., & Abramowitz, G. (2006). The CSIRO atmosphere biosphere land exchange (CABLE) model for use in climate models and as an offline model, CSIRO Marine and Atmospheric Research paper 013 (p. 37), [http://www.cmar.csiro.au/e-print/open/kowalczykea\\_2006a.pdf](http://www.cmar.csiro.au/e-print/open/kowalczykea_2006a.pdf)
- Kumar, S. V., Jasinski, M., Mocko, D. M., Rodell, M., Borak, J., Li, B., et al. (2018). NCA-LDAS land analysis: Development and performance of a multisensor, multivariate land data assimilation system for the National Climate Assessment. *Journal of Hydrometeorology*. <https://doi.org/10.1175/JHM-D-17-0125.1>
- Kumar, S. V., Peters-Lidard, C. D., Eastman, J. L., & Tao, W.-K. (2008). An integrated high resolution hydrometeorological modeling testbed using LIS and WRF. *Environmental Modelling and Software*, *23*, 169–181. <https://doi.org/10.1016/j.envsoft.2007.05.012>
- Kumar, S. V., Peters-Lidard, C. D., Tian, Y., Houser, P. R., Geiger, J., Olden, S., et al. (2006). Land Information System—An interoperable framework for high resolution land surface modeling. *Environmental Modelling and Software*, *21*, 1402–1415. <https://doi.org/10.1016/j.envsoft.2005.07.004>
- Lavender, S. L., & Abbs, D. J. (2013). Trends in Australian rainfall: Contribution of tropical cyclones and closed lows. *Climate Dynamics*, *40*(1–2), 317–326. <https://doi.org/10.1007/s00382-012-1566-y>
- Law, R., Ziehn, T., Matear, R., Lenton, A., Chamberlain, M., Stevens, L., et al. (2017). The carbon cycle in the Australian Community Climate and Earth System Simulator (ACCESS-ESM 1)—Part 1: Model description and pre-industrial simulation. *Geoscientific Model Development*, *10*(7), 2567–2590. <https://doi.org/10.5194/gmd-10-2567-2017>
- Liu, Y., Donat, M. G., Rust, H. W., Alexander, L. V., & England, M. H. (2019). Decadal predictability of temperature and precipitation means and extremes in a perfect-model experiment. *Climate Dynamics*, *53*(7–8), 3711–3729. <https://doi.org/10.1007/s00382-019-04734-z>
- Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., et al. (2017). GLEAM v3: Satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development*, *10*(5), 1903–1925. <https://doi.org/10.5194/gmd-10-1903-2017>

- Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. A. C., & Dolman, A. J. (2011). Global land-surface evaporation estimated from satellite-based observations. *Hydrology and Earth System Sciences*, *15*, 453–469. <https://doi.org/10.5194/hess-15-453-2011>
- Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997). Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research*, *102*, 16663–16682. <https://doi.org/10.1029/97JD00237>
- Nakanishi, M., & Niino, H. (2006). An improved Mellor-Yamada level 3 model: Its numerical stability and application to a regional prediction of advecting fog. *Boundary-Layer Meteorology*, *119*, 397–407. <https://doi.org/10.1007/s10546-005-9030-8>
- Nakanishi, M., & Niino, H. (2009). Development of an improved turbulence closure model for the atmospheric boundary layer. *Journal of the Meteorological Society of Japan*, *87*, 895–912. <https://doi.org/10.2151/jmsj.87.895>
- Pennsylvania State University, PSU (2006). PSU 1 km Multi-layer soil characteristics dataset. Version 1.0. UCAR/NCAR Earth Observing Laboratory, <https://data.eol.ucar.edu/dataset/8.43>.
- Peters-Lidard, C. D., Houser, P. R., Tian, Y., Kumar, S. V., Geiger, J., Olden, S., et al. (2007). High-performance Earth system modeling with NASA/GSFC's Land Information System. *Innovations in Systems and Software Engineering*, *3*, 157–165. <https://doi.org/10.1007/s11334-007-0028-x>
- Peters-Lidard, C. D., Kemp, E. M., Matsui, T., Santanello, J. A. Jr., Kumar, S. V., Jacob, J. P., et al. (2015). Integrated modeling of aerosol, cloud, precipitation and land processes at satellite-resolved scales. *Environmental Modelling and Software*, *67*, 149–159. <https://doi.org/10.1016/j.envsoft.2015.01.007>
- Santanello, J. A. Jr., Kumar, S. V., Peters-Lidard, C. D., Harrison, K., & Zhou, S. (2013). Impact of land model calibration on coupled land-atmosphere prediction. *Journal of Hydrometeorology*, *14*, 1373–1400. <https://doi.org/10.1175/JHM-D-12-0127.1>
- Santanello, J. A. Jr., Peters-Lidard, C. D., Kennedy, A., & Kumar, S. J. (2013). Diagnosing the nature of land-atmosphere coupling: A case study of dry/wet extremes in the U.S. Southern Great Plains. *Journal of Hydrometeorology*, *14*, 3–24. <https://doi.org/10.1175/JHM-D-12-023.1>
- Sheffield, J., Goteti, G., & Wood, E. F. (2006). Development of a 50-yr high-resolution global dataset of meteorological forcings for land surface modeling. *Journal of Climate*, *13*, 3088–3111. <https://doi.org/10.1175/JCLI3790.1>
- Skamarock W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., et al. (2008). A description of the advanced research WRF Version 3. NCAR Tech Note NCAR/TN-475 + STR. NCAR. Boulder, CO, <https://opensky.ucar.edu/islandora/object/technotes%3A500/datastream/PDF/view>
- Skamarock W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., et al. (2019). A description of the Advanced Research WRF Version 4. NCAR Tech Note NCAR/TN-566 + STR. NCAR. Boulder, CO, [http://www2.mmm.ucar.edu/wrf/users/docs/technote/v4\\_technote.pdf](http://www2.mmm.ucar.edu/wrf/users/docs/technote/v4_technote.pdf).
- Ukkola, A. M., Pitman, A. J., Donat, M. D., De Kauwe, M. G., & Angéilil, O. (2018). Evaluating the contribution of land-atmosphere coupling to heat extremes in CMIP5 models. *Geophysical Research Letters*, *45*, 9003–9012. <https://doi.org/10.1029/2018GL079102>
- van Dijk, A. I. J. M., Beck, H. E., Crosbie, R. S., de Jeu, R. A. M., Liu, Y. Y., Podger, G. M., et al. (2013). The millennium drought in southeast Australia (2001-2009): Natural and human causes and implications for water resources, ecosystems, economy, and society. *Water Resources Research*, *49*, 1040–1057. <https://doi.org/10.1002/wrcr.20123>
- van Dijk, A. I. J. M., Schellekens, J., Yebra, M., Beck, H. E., Renzullo, L. J., Weerts, A., & Donchyts, G. (2018). Global 5 km resolution estimates of secondary evaporation including irrigation through satellite data assimilation. *Hydrology and Earth System Sciences*, *22*, 4959–4980. <https://doi.org/10.5194/hess-22-4959-2018>
- Walsh, K., Lavender, S., Scoccimarro, E., & Murakami, H. (2013). Resolution dependence of tropical cyclone formation in CMIP3 and finer resolution models. *Climate Dynamics*, *40*(3-4), 585–599. <https://doi.org/10.1007/s00382-012-1298-z>
- Wang, Y. P., Kowalczyk, E., Leuning, R., Abramowitz, G., Raupach, M. R., Pak, B., et al. (2011). Diagnosing errors in a land surface model (CABLE) in the time and frequency domains. *Journal of Geophysical Research*, *116*, G01034. <https://doi.org/10.1029/2010JG001385>
- Whitley, R., Beringer, J., Hutley, L. B., Abramowitz, G., De Kauwe, M. G., Duursma, R. A., et al. (2016). A model inter-comparison study to examine limiting factors in modeling Australian tropical savannas. *Biogeosciences*, *13*(11), 3245–3265. <https://doi.org/10.5194/bg-13-3245-2016>
- Yuan, H., Dai, Y., Xiao, Z., Ji, D., & Shangguan, W. (2011). Reprocessing the MODIS leaf area index products for land surface and climate modeling. *Remote Sensing of Environment*, *115*, 1171–1187. <https://doi.org/10.1016/j.rse.2011.01.001>
- Zhang, C., & Wang, Y. (2017). Projected future changes of tropical cyclone activity over the western North and South Pacific in a 20-km-mesh regional climate model. *Journal of Climate*, *30*, 5923–5941. <https://doi.org/10.1175/JCLI-D-16-0597.1>