Modelling evapotranspiration during precipitation deficits: identifying critical processes in a land surface model

A.M. Ukkola[1], A.J. Pitman[1,2], M. Decker[1,2], M.G. De Kauwe[3], G. Abramowitz[1,2], J. Kala[1,4] and Y.-P. Wang[5]

[1]ARC Centre of Excellence for Climate System Science, University of New South Wales, Kensington, NSW 2052, Australia

[2]Climate Change Research Centre, University of New South Wales, Kensington, NSW 2052, Australia

[3]Department of Biological Sciences, Macquarie University, Balaclava Road, North Ryde, NSW 2109, Australia

[4]Murdoch University, School of Veterinary and Life Sciences, Environmental and Conservation Sciences, Murdoch, 6150, WA, Australia

[5]CSIRO Ocean and Atmosphere, Aspendale, Victoria 3195, Australia

Correspondence to: A.M. Ukkola (amukkola@gmail.com)

Abstract

Surface fluxes from land surface models (LSMs) have traditionally been evaluated against monthly, seasonal or annual mean states. The limited ability of LSMs to reproduce observed evaporative fluxes under water-stressed conditions has been previously noted, but very few studies have systematically evaluated these models during rainfall deficits. We evaluated latent heat fluxes simulated by the Community Atmosphere Biosphere Land Exchange (CABLE) LSM across 20 flux tower sites at sub-annual to inter-annual time scales, in particular focusing on model performance during seasonal-scale rainfall deficits. The importance of key model processes in

capturing the latent heat flux was explored by employing alternative representations of hydrology, leaf area index, soil properties and stomatal conductance. We found that the representation of hydrological processes was critical for capturing observed declines in latent heat during rainfall deficits. By contrast, the effects of soil properties, LAI and stomatal conductance were highly site-specific. Whilst the standard model performs reasonably well at annual scales as measured by common metrics, it grossly underestimates latent heat during rainfall deficits. A new version of CABLE, with a more physically consistent representation of hydrology, captures the variation in the latent heat flux during seasonal-scale rainfall deficits better than earlier versions but remaining biases point to future research needs. Our results highlight the importance of evaluating LSMs under water-stressed conditions and across multiple plant functional types and climate regimes.

1 Introduction

Droughts are expected to increase in frequency and intensity (Allen et al., 2010; Trenberth et al., 2014) in some regions due to the effects of climate change (Collins et al., 2013). This would have profound implications for affected regions and their socio-economic systems. Land surface models (LSMs) are a key tool for understanding the evolution of historical droughts and predicting future water scarcity when coupled to global climate models (Dai, 2012; Prudhomme et al., 2014). LSMs have been extensively evaluated for simulated water, energy and carbon fluxes, typically at seasonal to inter-annual time scales (Abramowitz et al., 2007; Best et al., 2015; Blyth et al., 2011; Dirmeyer, 2011; Zhou et al., 2012), and been found to perform reasonably well under well-watered conditions (e.g Best et al., 2015). However, recent studies have indicated that the ability of current LSMs to simulate these fluxes during water-stressed conditions is limited (De Kauwe et al., 2015b; Powell et al., 2013). LSMs have been shown to poorly characterise the magnitude, duration and frequency of droughts when evaluated against site- and catchment-scale observations of latent heat (Q_E) and streamflow (De Kauwe et al., 2015b; Li et al., 2012; Powell et al., 2013; Prudhomme et al., 2011; Tallaksen and Stahl, 2014). Similarly, LSM projections of future drought occurrence has been shown to be highly model dependent, with greater uncertainty in future projections arising from

differences between LSMs than from the climate model projections used to force them (Prudhomme et al., 2014). Further, changes in soil moisture in the future are also linked with changes in the probabilities and intensities of other extremes including heatwaves (Seneviratne et al., 2010). Clearly, a better understanding of limitations in LSMs under more extreme conditions, and ultimately improved performance by these models, is necessary for improving the future projections of drought and other land-surface influenced extremes by climate models.

We investigate the performance of the Australian Community Atmosphere-Biosphere Land Exchange (CABLE) in simulating observed declines in latent heat during rainfall deficits. CABLE is the LSM used within the Australian Community Climate and Earth Systems Simulator (ACCESS; Bi et al., 2013), a global climate model which has participated in the 5th assessment report of the International Panel on Climate Change (IPCC, 2013) and is used for numerical weather prediction research in Australia (Puri et al., 2013). In common with other LSMs (Prudhomme et al., 2011; Tallaksen and Stahl, 2014), CABLE has been found to poorly simulate the evolution of droughts, systematically underestimating site-scale Q_E during seasonal-scale droughts (De Kauwe et al., 2015b; Li et al., 2012). There could be many reasons for this systematic error. For example, unrealistic representation of plant drought responses has been identified as a major limitation in LSM simulations of drought (Egea et al., 2011; Powell et al., 2013), including CABLE (De Kauwe et al., 2015b). Recent studies have revised vegetation drought responses in CABLE but not fully resolved existing model biases (De Kauwe et al., 2015b; Li et al., 2012). In this paper, we examine CABLE in the context of another major area of uncertainty: how hydrological processes are parameterised and how associated parameters are selected. A better representation of hydrological processes, particularly soil moisture, has been identified as necessary for improving LSM simulations of drought (Tallaksen and Stahl, 2014) but this has not been widely explored. The parameterisations of soil hydrology and stomatal conductance (g_s) have recently been revised in CABLE and shown to improve seasonal to annual scale simulations of Q_E in CABLE (De Kauwe et al., 2015a; Decker, 2015). We explore if changes to these hydrological processes can also improve simulations of Q_E during dry periods and to guide development of more realistic drought mechanisms in LSMs.

We quantify the uncertainty arising from key model parameters: soil properties and leaf area index (LAI) inputs. CABLE-simulated Q_E has been shown to be sensitive to these parameters but they remain uncertain at both site and large scales (De Kauwe et al., 2011; Kala et al., 2014; Koster et al., 2009; Zhang et al., 2013). Quantifying the sensitivity of CABLE to LAI and soil properties is useful for separating parameter uncertainties from inadequate model parameterisations. Where the LSM cannot capture the observations, despite variations in LAI and soil parameters, systematic errors in the model's representation of physical processes are probable (assuming negligible errors in flux tower data used to drive and evaluate the model). While other parameters, including additional vegetation characteristics such as rooting depth (Li et al., 2012) are also potentially important, soil properties and leaf area index can be constrained from readily available global-scale datasets widely used in large-scale LSM applications.

We therefore explore CABLE performance at 20 flux tower sites distributed globally. We contrast model behaviour at annual to sub-seasonal scales to explore uncertainties in hydrological processes and parameters under conditions ranging from wet to dry. We concentrate on the ability of the model to capture the onset of drought in a drying phase as a pre-requisite for capturing the magnitude and intensity of droughts.

2 Methods

2.1 Flux tower sites

The flux tower data were collated as part of the Protocol for the Analysis of Land Surface models (PALS; Abramowitz, 2012) Land sUrface Model Benchmarking Evaluation pRoject (PLUMBER; Best et al., 2015), originally obtained through the Fluxnet LaThuile Free Fair-Use subset (fluxnet.ornl.gov). The PLUMBER sites represent a broad range of vegetation and climate types, and were also selected to maximise the length of measurement records (Best et al., 2015). Here, we focus on the results for six sites with a pronounced period of low precipitation (Fig. 1 and Table S1), each representing a different climate and vegetation type, but provide results for all study sites as Supplementary Information (Fig. S1-S4).

1

2 The 20 flux tower sites provide meteorological and flux measurements at 30-minute

3 resolution. The observed meteorological data (precipitation, short- and long-wave

4 radiation, surface air pressure, air temperature, specific humidity and wind speed)

5 were used to drive CABLE simulations. Observed Q_E was then used to evaluate

6 simulations because it is the variable that links the land surface energy, water and

7 carbon budgets (Pitman, 2003). It is also one of the variables supplied by a LSM to

8 the atmosphere and is therefore important to a climate model. We note that it would

9 also be desirable to evaluate soil moisture outputs from LSMs. Ultimately this is

10 problematic (Koster et al., 2009) as site measurements at depths which reflect the

11 plants' root-zone access (i.e. deeper than the top few centimetres) are not readily

12 available

13

14 **2.2 Description of the CABLE LSM and model parameterisations**

15

16 **2.2.1 General description**

17

18 The Community Atmosphere-Biosphere Land Exchange (CABLE) model is used to

19 simulate energy, water and carbon fluxes and the partitioning of net radiation into

20 latent and sensible heat fluxes. It can be employed offline with prescribed

21 meteorology, as in this paper, or within the Australian Community Climate Earth

22 System Simulator coupled climate model (ACCESS; Bi et al., 2013). It has been used

23 widely in coupled (Cruz et al., 2010; Lorenz and Pitman, 2014) and offline (Haverd et

24 al., 2013; Huang et al., 2015; Zhou et al., 2012) simulations and has been extensively

25 evaluated against flux site (De Kauwe et al., 2015b; Li et al., 2012; Wang et al., 2011;

26 Williams et al., 2009) and regional to global-scale observations (De Kauwe et al.,

27 2015a; Decker, 2015). Previous model inter-comparisons have shown that simulated

28 latent and sensible heat fluxes perform comparably to other LSMs (Best et al., 2015).

29

30 CABLE consists of sub-models for radiation, canopy, soil and ecosystem carbon.

31 Canopy processes are represented with a two-leaf model, which calculates

32 photosynthesis, stomatal conductance and leaf temperature separately for sunlit and

33 shaded leaves (Leuning, 1995; Wang and Leuning, 1998). The soil module simulates

34 the transfer of water within the soil and snowpack following the Richards equation.

CABLE has 11 plant functional types (PFT). A detailed description of model components can be found in Wang et al. (2011).

We used CABLE version 2.0 forced with site-specific meteorological data at 30-minute time steps. Site PFT was determined by matching site vegetation (fluxnet.ornl.gov) to CABLE PFTs. PFT parameters were taken from a standard look-up table provided with CABLE 2.0 and were not calibrated to match site characteristics. The model was run using two alternative hydrological modules, two stomatal conductance parameterisations, three soil types and three LAI time series. The new hydrological scheme implements a topographic slope parameter, which was varied between two values in additional simulations. This parameter controls the drainage rate and can in principle be constrained from high-resolution elevation data. We vary the slope parameter between 0 and 5 degrees, broadly coinciding with the observed range of 0-6 degrees at the flux sites as derived from the approximately 1 km spatial resolution Global 30-Arc Second Elevation (GTOPO30) elevation dataset (https://lta.cr.usgs.gov/GTOPO30).

CABLE was run with all parameterisation combinations, resulting in 18 simulations using the default and 36 using the new hydrological scheme. This enabled the quantification of individual parameter and/or parameterisation uncertainties on model simulations and accounts for interactions between different parameterisations. The individual model parameterisations varied in this study are detailed below.

2.2.2 Hydrological parameterisation

We use two different representations of hydrology. The two schemes are fully detailed in Decker (2015) but we will briefly describe the main differences here. The default soil hydrological scheme in CABLE simulates the exchange of water and heat based on six soil layers and up to three snow layers. The default parameterisation for soil moisture processes was developed by Kowalczyk et al. (1994) and later revised by Gordon et al. (2002) and is described in detail in Kowalczyk et al. (2006) and Wang et al. (2011). The default scheme only generates infiltration excess surface runoff when the top three soil layers are ≥95% saturated and otherwise lacks an explicit runoff generation scheme (Decker, 2015). It does not simulate saturated and

un-saturated top soil fractions separately or consider groundwater aquifer storage. The default scheme solves the vertical redistribution of soil water using the 1-D Richards equation. The bottom boundary condition for the solution of the 1-D Richards equation is given as

$$q_{sub} = C_{drain}\theta_n \tag{1}$$

where q_{sub} is the subsurface drainage (mm s^{-1}), θ_n the soil moisture content of the bottom soil layer (mm^3 mm^{-3}) and C_{drain} a tunable parameter (mm s^{-1}) (Decker, 2015). The scheme thus assumes a free draining lower boundary and water below the model domain (e.g. groundwater) cannot recharge the water content of the above soil column.

Soil evaporation (E_{soil}; W m^{-2}) is given by

$$E_{soil} = \beta_s \frac{L_v \rho_a (q_{sat}(T_{srf}) - q_a)}{r_g} \tag{2}$$

where L_v is the latent heat of vaporisation (J kg^{-1}), ρ_a the air density (kg m^{-3}), $q_{sat}(T_{srf})$ the saturated specific humidity at the surface temperature (kg kg^{-1}), q_a the specific humidity of air (kg kg^{-1}) and r_g (s m^{-1}) the aerodynamic resistance term. β_s is a dimensionless scalar (varying between 0 and 1) used to reduce E_{soil} when soil moisture is limiting and is given by the linear function

$$\beta_s = \frac{\theta_1 - 0.5\theta_w}{\theta_{fc} - 0.5\theta_w} \tag{3}$$

where θ_1 the soil moisture content of the first soil layer (m^3 m^{-3}), θ_w the wilting point (m^3 m^{-3}) and θ_{fc} the field capacity (m^3 m^{-3}).

Decker (2015) developed an improved representation of sub-surface hydrological processes similar to that implemented in the Community Land Model (Lawrence and Chase, 2007; Oleson et al., 2008). The new scheme explicitly simulates saturation-

1 and infiltration-excess runoff generation and has a dynamic groundwater component

2 with aquifer water storage. The scheme solves the vertical redistribution of soil water

3 (θ) following the Modified Richards Equation (Zeng and Decker, 2009):

4

5 $$\frac{\partial \theta}{\partial t} = -\frac{\partial}{\partial z} K \frac{\partial}{\partial z}(\Psi - \Psi_E) - F_{soil} \qquad (4)$$

6

7 where K (mm s^{-1}) is the hydraulic conductivity, Ψ (mm) the soil matric potential, Ψ_E

8 (mm) the equilibrium soil matric potential, z is soil depth (mm) and F_{soil} (mm s^{-1}) is

9 the sum of subsurface runoff and transpiration (Decker, 2015). An unconfined aquifer

10 is located below the six soil layers and is presented with a simple water balance

11 model:

12

13 $$\frac{dW_{aq}}{dt} = q_{re} - q_{aq,sub} \qquad (5)$$

14

15 where W_{aq} is the mass of water in the aquifer (mm), $q_{aq,sub}$ the subsurface runoff

16 removed from aquifer (mm s^{-1}) and q_{re} the water flux between the aquifer and the

17 bottom soil layer, given from Darcy's law as

18

19 $$q_{re} = K_{aq} \frac{(\Psi_{aq} - \Psi_n) - (\Psi_{E,aq} - \Psi_{E,n})}{z_{wtd} - z_n} \qquad (6)$$

20

21 where z_{wtd} is the water table depth (mm), K_{aq} is the hydraulic conductivity of the

22 aquifer (mm s^{-1}), and z_n is the depth of the lowest model layer (mm). The bottom

23 boundary condition is given as

24

25 $$q_{out} = 0 \qquad (7)$$

26

27 as the scheme assumes that the groundwater aquifer sits above an impermeable layer

28 of rock (Decker, 2015). Subsurface runoff (q_{sub}) is calculated from

29

30 $$q_{sub} = \sin\frac{dz}{dl}\hat{q}_{sub}e^{-\frac{z_{wtd}}{f_P}} \qquad (8)$$

1

where dz/dl is the mean slope, $\hat{q}_{sub}$ the maximum rate of subsurface drainage for a

fully saturated soil column (mm s^{-1}) and f_p is a tunable parameter (Decker, 2015). q_{sub}

is removed from the bottom three soil layers (which account for 4.366 m of the total

soil thickness of 4.6 m) by weighting the amount of water removed from each layer

based on the mass of liquid water present in each layer (Decker, 2015).

Sub-grid scale heterogeneity in soil moisture is permitted and a modified soil

evaporation formulation reflects this. At point scales the runoff generation from sub-

grid heterogeneity in soil moisture is neglected as the saturated fraction of the grid

cell is assumed to be equal to zero. Soil evaporation is given as (here assuming a

saturated grid cell fraction (F_{sat}) of 0):

$$E_{soil} = F_{sat}E^*_{soil} + (1 - F_{sat})\beta_s E^*_{soil} \tag{9}$$

where E^*_{soil} is the soil evaporation prior to soil moisture limitation (mm s^{-1}). β_s is

calculated using a non-linear function following Sakaguchi and Zeng (2009):

$$\beta_s = 0.25\left(1 - \cos\left(\pi\frac{\theta_{unsat}}{\theta_{fc}}\right)\right)^2 \tag{10}$$

where θ_{unsat} is the first layer soil moisture content in the unsaturated portion (the entire

soil layer in this study) (m^3 m^{-3}).

Both schemes calculate transpiration using the same method and, in common with

many LSMs (Verhoef and Egea, 2014), limit gas exchange during low soil moisture

using a dimensionless scalar (β) varying between 0 and 1:

$$\beta = \sum_{i=1}^{n} f_{root,i}\frac{\theta_i - \theta_w}{\theta_{fc} - \theta_w} \tag{11}$$

where θ_i is the soil moisture content of soil layer i (m^3 m^{-3}) and $f_{root,i}$ the root mass

fraction of soil layer i.

The default version of CABLE tends to overestimate Q_E at annual to seasonal scales when used coupled with the ACCESS climate model (Lorenz et al., 2014), but significantly underestimates Q_E during soil moisture deficits across six European flux tower sites (De Kauwe et al., 2015b) in uncoupled experiments. Decker (2015) showed that the new model reduced overestimations of Q_E by 50–70% compared to the default scheme and yielded an improved simulation of seasonal cycles when evaluated against observations from large river basins. The new scheme was also shown to better capture total water storage anomalies (an integral over depth of soil moisture changes) from the Gravity Recovery and Climate Experiment (GRACE; http://grace.jpl.nasa.gov) than the default scheme. It is not known if these improvements will also allow CABLE to better capture observed Q_E during dry-down and we explore this here.

2.2.3 Stomatal conductance parameterisation

We use two alternative parameterisations for stomatal conductance (g_s). The default CABLE currently implements an empirical g_s formulation following Leuning (1995):

$$g_s = g_0 + \frac{\alpha_1 \beta A}{(C_s - \Gamma)\left(1 + \dfrac{D}{D_0}\right)} \tag{12}$$

where A is the net assimilation rate (μmol m^{-2} s^{-1}), Γ (μmol mol^{-1}) is the CO_2 compensation point of photosynthesis, C_s (μmol mol^{-1}) and D (kPa) are the CO_2 concentration and the vapour pressure deficit at the leaf surface, respectively. g_0 (mol m^{-2} s^{-1}), D_0 (kPa) and α_1 are fitted constants representing the residual stomatal conductance when $A=0$, the sensitivity of stomatal conductance to D and the sensitivity of stomatal conductance to assimilation, respectively. Although the g_s formulation following Leuning (1995) (or equivalent Ball-Berry model; Ball et al., 1987) are widely used in LSMs, the model parameters are empirical. Thus, we cannot attach any theoretical distinction as to how parameters vary across data sets or among species (De Kauwe et al., 2015a; Medlyn et al., 2011). Consequently, as is common with many LSMs (e.g. Community Land Model version 4.5 (Oleson et al., 2013) and

1 the ORganizing Carbon and Hydrology in Dynamic EcosystEms model (Krinner et

2 al., 2005)), the default scheme only varies stomatal conductance parameters between

3 photosynthetic pathways (C_3 vs. C_4), rather than among PFTs.

4

5 As an alternative, we also ran CABLE using the g_s model following Medlyn et al.

6 (2011), a theoretical formulation based on the premise of optimal stomatal behaviour:

7

8 $$g_s = g_0 + 1.6\left(1 + \frac{g_1\beta}{\sqrt{D}}\right)\frac{A}{C_s}$$ (13)

9

10 where g_1 ($kPa^{0.5}$) is a fitted parameter representing the sensitivity of conductance to

11 the assimilation rate. Unlike the α_1 parameter in the Leuning model, g_1 has biological

12 meaning, representing a plant's water use strategy. Values of g_1 were derived

13 previously for each of CABLE's PFTs (De Kauwe et al., 2015a) based on a global

14 synthesis of stomatal behaviour (Lin et al., 2015). Further details and associated

15 parameter values can be found in De Kauwe et al. (2015a).

16

17 The Medlyn g_s model has been shown to improve existing CABLE biases,

18 particularly overestimations of Q_E in evergreen needleleaf and C_4 biomes (De Kauwe

19 et al., 2015a). We will explore if the re-parameterisation of g_s also improves the

20 simulation of dry-down in CABLE.

21

22 **2.2.4 Soil parameterisation**

23

24 The soil parameters were derived from a dataset provided by CABLE developers

25 (https://trac.nci.org.au/trac/cable/wiki; Global Soil Data Task Group, 2000; Zobler,

26 1999). The dataset consists of nine soil classes; here the two classes with the highest

27 sand and clay contents were used. The coarse sandy soil has a 83% sand content and

28 the fine clay soil a 67% clay content. The soil classes have eight associated

29 parameters for soil hydraulic and thermal capacities, fully detailed in Table S2. In

30 addition, an arbitrary "medium" soil class was created with equal fractions of sand,

31 silt and clay, with other soil parameters set as the median of the coarse sand and fine

32 clay soil classes (Table S2). CABLE was run with these three alternative soil classes,

33 fixing the soil parameters across all sites to generate a range in soil parameter values.

The default hydrological scheme uses all soil parameters directly, whereas the new scheme calculates the eight parameters governing hydraulic properties from sand, silt and clay fractions using the Clapp and Hornberger (1978) pedotransfer functions. The soil parameter values used by both schemes are detailed in Table S2.

2.2.5 Leaf Area Index

Leaf area index (LAI) plays an important role in the surface energy balance in CABLE, scaling sunlit and shaded leaf-level fluxes of photosynthesis, g_s and latent heat flux to the canopy. LAI was obtained from 8-daily gridded Moderate Resolution Imaging Spectroradiometer (MODIS) data at 1 km resolution (Yuan et al., 2011). The data were averaged to monthly time steps to smooth the time series and subsequently three alternative LAI time series were created for each site to take some account of uncertainties in LAI inputs. The first time series was constructed by extracting the grid cells that contained each site ("centre" time series). Two alternative time series were created using the minimum and maximum LAI value of the grid cell and its immediate neighbours ("minimum" and "maximum" time series, respectively). Time-varying LAI was used for years where the flux observations and MODIS data overlap (i.e. after 2000); a monthly climatology of common years was used otherwise. The minimum and maximum time series differ from the centre time series by 30% on average but the range varies between sites. The alternative LAI time series are shown in Fig. S5.

2.3 Analysis methods

We analyse CABLE's performance across three time scales: the whole observational, annual and sub-annual periods. As the observational records are generally short for characterising hydrological extremes (~5 years on average; Table S1), we have not adopted a formal statistical method for identifying periods of rainfall anomalies and thus do not refer to them as "droughts". We also note no one definition for droughts exists; instead, various indices have been employed based on, for example, precipitation, streamflow, soil moisture and measures of evaporative demand (Sheffield and Wood, 2011). In this study, the dry periods were defined based on precipitation as this allowed the use of available observations, but we note the

1   simulated fluxes will also depend on other processes such as soil moisture

2   availability. For the majority of sites (Howard Springs, Palang, and all supplementary

3   sites), we selected the year with the lowest precipitation total as the one-year period,

4   whilst for Amplero, Blodgett, Tumbarumba and UniMich, we selected a year when

5   the default CABLE significantly underestimated latent heat fluxes during a rainfall

6   deficit ("dry-down") period. The dry-down period generally coincides with the

7   maximum and the following minimum observed latent heat flux during the one-year

8   period but has been adjusted using expert judgment for some sites to best demonstrate

9   typical model behaviour (Fig. S6). Observed and simulated data were averaged to 14-

10   day running means for all analyses.

11

12   We follow Abramowitz et al. (2007) and the PALS protocol for calculating model

13   metrics. We use the normalised mean error (NME) to evaluate general model

14   performance:

15

16   $$NME = \frac{\sum_{i=1}^{n}|M_i - O_i|}{\sum_{i=1}^{n}|\overline{O} - O_i|}$$ (14)

17

18   where M represents the model values and O the observations. NME accounts for

19   mean model biases and the temporal coincidence and magnitude of variability, but

20   does not distinguish between them (Best et al., 2015). An NME of 0.0 represents

21   perfect agreement and a value of 1.0 represents model performance equal to that

22   expected from a constant value equal to the mean of all observations.

23

24   We examine mean bias error (MBE) to estimate absolute biases in CABLE

25   simulations; it is simply the difference between the mean modelled and observed

26   values:

27

28   $$MBE = \overline{M} - \overline{O}$$ (15)

29

30

31   **3 Results**

3.1 Whole time period

We first evaluated Q_E simulated by CABLE during the whole data period available for each flux site (ranging from 2 to 7 years for selected sites; Table S1). CABLE, using the default hydrological parameterisation, captures the general features, such as the timing and magnitude of seasonal cycles, in observed Q_E across the different sites (Fig. 2, left column). CABLE including the new hydrological scheme also captures these general features (Fig. 2, right column). Quantifying the performance of these two versions of CABLE over the full length of record does not indicate that there is a significant difference between the versions in either NME (Fig. 3) or MBE (Fig. 4). The average NME for all sites and parameter choices was 0.90 for the old scheme and 0.75 for the new scheme, and the average MBE was −1 and 6 W m^{-2}, respectively. The NME metric is <1.0 for the majority of sites using the new scheme, regardless of the choice of g_s, LAI or soil parameterisation. We note that the magnitude of Q_E for the evergreen broadleaf sites (Palang and Tumbarumba and supplementary site Espirra) is poorly captured (Fig. 4). Overall, both hydrological parameterisations tend to overestimate peak Q_E (Fig. 2); this tendency for excessive evapotranspiration has also been demonstrated in global applications of CABLE in both offline (De Kauwe et al., 2015a) and coupled (Lorenz et al., 2014) simulations. Furthermore, both schemes systematically overestimate Q_E in spring, particularly at cooler temperate sites such as UniMich (Fig. 2; also see deciduous broadleaf and needleleaf supplementary sites (Fig. S1)), and over-predict the short-term variability in Q_E (see e.g. Amplero in Fig. 2). Despite clear biases in simulated fluxes, the MBE metric approaches zero at most sites when evaluated at inter-annual time scales. While encouraging, this is due to compensating errors, such that early season overestimations in Q_E are counteracted by underestimations during the dry-down periods (see e.g. Blodgett and Tumbarumba in Figure 5). This is particularly evident with the default hydrology scheme. We therefore focus the remaining analyses on shorter time periods where compensating biases are less likely to hide weaknesses in the model performance.

3.2 Annual and dry-down period

1 CABLE-simulated Q_E was then evaluated during annual and seasonal dry-down

2 periods to explore model performance during rainfall deficits. The default scheme

3 demonstrates a range of major biases (Fig. 5). The model dries down too quickly at

4 Amplero, Blodgett, Palang, Tumbarumba and UniMich sites. At these sites, and at

5 Howard Springs, Q_E drops too low and drops to that minimum too early in the year.

6 At several sites, including Blodgett, Tumbarumba and UniMich, CABLE

7 systematically overestimates Q_E in spring. These characteristics of CABLE are not

8 dependent on the choice of LAI or soil inputs or g_s parameterisations; the range in Q_E

9 fails to overlap the observations irrespective of how these properties are varied. This

10 suggests parameterisation error as distinct from parameter choices as the cause of the

11 model weaknesses.

12

13 The new hydrological scheme demonstrates clear improvements at Amplero, Howard

14 Springs and Palang. At Blodgett, Tumbarumba and UniMich, the observations are

15 within the uncertainty due to the choice of g_s, LAI or soil parameters in the second

16 half of the year, but the excessive Q_E during spring and early summer remains a

17 problem. While there are obviously remaining errors, the new hydrological scheme

18 clearly improves the simulation of Q_E over the annual cycles (Fig. 5). Assessing the

19 overall performance at annual time scales also highlights clear improvements with the

20 new hydrology. Fig. 3 shows that for NME, the new hydrology scheme in CABLE

21 performs as well as, or better than the default at every site, with an average NME

22 across all sites of 0.68 compared to 0.90 for the default scheme. This is true also of

23 MBE (Fig. 4) for all sites except Tumbarumba.

24

25 Assessing the performance of the two schemes over the dry-down period using NME

26 is shown in Fig. 5. Using the default hydrology leads to worse performance on this

27 shorter time scale at Amplero, Blodgett, Palang and to a lesser degree at Howard

28 Springs and Tumbarumba compared to annual and inter-annual scales. In contrast,

29 CABLE with the new hydrology performs similarly to the longer ($\geq$ 1 year) time

30 scales at Blodgett, Palang, Howard Springs, Tumbarumba and UniMich and only

31 marginally poorer at Amplero. Comparing NME over this dry down period shows that

32 the new scheme strongly outperforms the default parameterisation (Fig. 3; the average

33 NME is 0.68 and 1.27 for new and default schemes, respectively). A similar

34 conclusion is reached using MBE (on average −4 and −22 W m^{-2} for the new and

default schemes, respectively). In short, the new hydrology does not dramatically improve the performance of CABLE on the long term (i.e. inter-annual scales) (Fig. 2) due to compensating biases in the default CABLE. These include overestimated spring and early summer Q_E, and consequently, at least in part, underestimated Q_E during the dry-down. Once we focus on shorter, sub-annual timescales that lack these compensating biases, CABLE with the new hydrology strongly outperforms the default version in the simulation of Q_E.

3.3 Impact of varying LAI, g_s and soil parameters

We now explore the individual contributions from soil parameters, g_s and LAI to uncertainties in simulated Q_E. Fig. 6 and Fig. 7 show the uncertainty in model simulations due to soil parameters, g_s and LAI using the default and new hydrological schemes, respectively. Both hydrological schemes are sensitive to soil parameters during the dry-down period but show smaller variations due to soil during other parts of the year (see Amplero, Blodgett, Howard Springs and Palang in Fig. 6 and Fig. 7). This transition from low to high sensitivity occurs as soil moisture stores begin to deplete and Q_E becomes increasingly limited by moisture supply (Figure S8). Both schemes show a similar sensitivity to g_s and LAI variations, which is generally smaller compared to soil variations, although the new scheme is more sensitive to g_s at Blodgett, Howard Springs and Palang, and to LAI in Amplero and Palang during dry-down.

While the new hydrological parameterisation systematically improved model performance across most sites (Fig. 3 and 4), the effect of LAI, g_s and soil parameters on the mean magnitude of simulated fluxes is highly site-specific during the annual and dry-down periods (Fig. 8). In agreement with (De Kauwe et al., 2015a), the choice of g_s scheme generally has a larger effect in needleleaf (Blodgett) and C$_4$ grass (Howard Springs) sites. Some sites, such as Howard Springs, are sensitive to multiple parameters, whilst others such as UniMich only respond minimally to parameter perturbations (Fig. 8). Whilst there is no *a priori* expectation that this should be the case, it highlights the importance of investigating model uncertainties and performance across multiple sites to capture full range of model sensitivities to parameter perturbations.

The results have so far assessed CABLE with the new hydrology using a 0° slope parameter as this enables a direct comparison with the default hydrology. The slope parameter, which can be derived from high-resolution elevation data, is scale dependent and was introduced by Decker (2015) to capture large-scale hydrological processes that are affected by landscape geometry. The slope parameter affects the rate of subsurface drainage and represents a key difference between the new and default schemes. With the exception of the UniMich site, Fig. 8 and 9 show that the model is highly sensitive to choice of the slope parameter across all sites, particularly during the dry-down period. The slope appears more critical for simulation of Q_E than the other parameterisations investigated here and has a strong effect on the magnitude of fluxes primarily during the dry-down (see e.g. Howard Springs and Palang in Fig. 9). Whilst this highlights the need to carefully set the slope parameter, it is unclear how well it can be constrained at the site scale. The surface slope derived from elevation data may not reflect large-scale features, such as subsurface geology, which can affect drainage rates and thus water availability for Q_E in highly site-specific ways.

4 Discussion

4.1 Simulation of dry-down

We have shown that the default version of CABLE significantly underestimates Q_E during rainfall deficits (Fig. 5). We have also shown that it is unlikely that uncertainties in key model soil and vegetation (LAI) inputs account for these biases (Fig. 6). The observations used to drive and evaluate the model themselves include errors, notably lack of energy balance closure (Leuning et al., 2012). However, given the systematic and large seasonal biases in CABLE simulations, any errors in the flux tower measurements are unlikely to explain poor model performance during dry-down. Instead, our results point to deficiencies in the representation of hydrological processes in the default version of CABLE (Fig. 3 and 4). The default CABLE has been shown to perform similarly to other LSMs in Best et al. (2015) and indeed in other model evaluation studies (Abramowitz et al., 2008). Hence, it is likely that the

errors of the kind identified here may be common among other models, as model benchmarking rarely examine sub-annual behaviour. The poor simulation of dry-down periods is important: if LSMs in general struggle to simulate this period they will fail to correctly capture water fluxes when serious soil moisture deficits are established. A model that simulates dry-down too fast will enter drought early and will tend to simulate longer, deeper and more frequent droughts than a model that enters drought too slowly. We suggest that systematic evaluation of LSMs during dry down periods would lead to the identification of major limitations in some models that are hidden by compensating errors over longer timescales. Resolution of those problems has the potential to improve the simulation of drought in climate models.

We also showed that the effect of individual parameterisations was magnified during dry periods (Fig. 6 and 7). Whilst the new hydrological scheme did not present a significant improvement on the annual and inter-annual timescales analysed here, it had an increasingly large positive impact on shorter time scales and in particular during the dry-down periods (Fig. 3,4 and 5). Similarly, the contribution of soil (Fig. 6a and 7a), g_s (Fig. 6b and 7b) and LAI (Fig. 6c and 7c) parameterisations to model uncertainties was generally larger during the dry-down. We will discuss each of these points below.

4.1.1. Soil and LAI inputs

We evaluated the uncertainty in Q_E simulations arising from inputs of soil properties and LAI. These variables are generally obtained from gridded datasets in LSM simulations and remain uncertain at the site (and larger) scale (De Kauwe et al., 2011; Koster et al., 2009).

Soil parameters feature in many hydrological model components and our results show that the range in simulated Q_E due to the choice of soil parameters is largest during the dry-down for both hydrological schemes (Fig. 6 and 7). Parameters for wilting point (θ_w) and field capacity (θ_{fc}) are particularly important during drying conditions as they determine how evapotranspiration is reduced as soil moisture becomes limiting (following Eq. (3), (10) and (11)). The model is also sensitive to value of the matric potential at saturation (Table 2). The vertical diffusive flux of soil water

between adjacent soil layers is proportional to the saturated matric potential. This control on the rate of vertical water movement alters the profile of vertical soil water during the dry-down impacting the water available for transpiration in a given soil layer. Using the default hydrological scheme, the observed Q_E could only be captured by varying the soil properties at Howard Springs. Elsewhere, the model underestimated observed Q_E during dry-down regardless of how the soil properties were varied (Fig. 6). This suggests that uncertainties in soil parameters cannot account for the poor simulation of dry-down by the default model.

Similarly, LAI, as it was varied here, could not explain the underestimation of Q_E. The range in LAI varied by site (30% on average) according to the remotely sensed data but was not lower at the drought sites. The model was generally not sensitive to changes in LAI during dry-down regardless of the choice of hydrological scheme (Fig. 6 and Fig. 7). This implies that the correct characterisation of canopy structure is probably not critical for the simulation of Q_E in CABLE during dry-down or that the scale of the errors in the simulations are too large to see any more subtle impact of these LAI variations. Nevertheless, we do note that leaf drop during drought events could lead to an increased or compensatory reflectance signal from deeper in the canopy profile, resulting in erroneous estimates of LAI from optically remote sensed products (cf. Amazon drought studies; Samanta et al., 2010).

4.1.2. Hydrological schemes

The new hydrological scheme was shown to improve CABLE simulations of Q_E during dry-down (Fig. 3 and 4). This results from higher soil moisture content simulated by the new scheme compared to the default model, particularly in the bottom soil layers (Fig. S7). This allows higher ET fluxes to be maintained during dry periods, mainly due to higher transpiration rates (Fig. S8).

The alternative hydrological schemes make different assumptions about subsurface drainage and how this is treated upon exiting the bottom soil column boundary. The default model assumes a free draining boundary for solving vertical water flow (Eq. (1)), such that the bottom soil layer essentially acts as a sink for the rest of the soil column as it can only remove water from the column. Conversely, the new scheme

simulated an unconfined groundwater aquifer below the bottom soil layer that is assumed to sit on an impermeable layer of rock so that no water is lost from the aquifer through downward flow (Eq. (7)). Soil moisture content of above soil layers can then be replenished through recharge from the aquifer to maintain higher soil moisture during dry periods (given a water table depth near the bottom of the soil column; Eq. (6); Fig. S7). Zeng and Decker (2009) demonstrated that assuming a free draining lower boundary requires an unrealistically high precipitation rate to maintain a relatively wet soil moisture content that allows vegetation to transpire without encountering soil moisture stress. Using a hypothetical example, the authors estimate that a minimum precipitation rate of 17.2 mm/day is required to maintain non-water-stressed conditions (a value much higher than is observed in most environments), implying overly dry soil conditions in many cases. Our results therefore suggest that the replacement of a constant drainage assumption in the original model with a physically based, dynamic bottom boundary condition for the soil column is important for improving Q_E fluxes in CABLE under water stressed conditions.

Whilst it was not possible to evaluate these simulations against soil moisture data due to a lack of observations for soil depths used in CABLE, Decker (2015) showed that the new scheme could better capture total soil column water anomalies and evapotranspiration (two variables that strongly depend on the correct simulation of soil moisture content) in comparison to the default scheme at river basin scales. This gives us confidence that the higher soil moisture levels simulated by the new scheme are supported by some observations. This result should be evaluated in future work against locations where deep soil moisture measurements are made available, or efforts to obtain observed soil moisture coincident with tower measurements of the fluxes should be encouraged.

4.1.3. Stomatal conductance schemes

Our results showed that CABLE is generally not sensitive to the choice of g_s scheme during dry-down at most sites (Fig. 6, 7 and 8), with the exception of Howard Springs (a C_4 grass site) and Blodgett (an evergreen needleleaf site). This result is largely to be expected: during drought both schemes are limited in the same fashion, with β (Eq. (11)) reducing the slope that relates g_s to photosynthesis. The noted differences

between schemes at the C_4 grass and evergreen needleleaf sites are consistent with results from De Kauwe et al. (2015a). At Howard Springs, De Kauwe et al. (2015a) found that the high g_0 value assumed in the Leuning model (0.04 mol m^{-2} leaf s_{-1}) accounted for the difference between schemes when g_s approached zero (for example during a drought). Differences between schemes at Blodgett stem, at least in part, from the use of a parameterisation of a conservative water use, found in evergreen needleleaf forests (Lin et al., 2015).

We note that the two stomatal schemes have different sensitivities to vapour pressure deficit (see De Kauwe et al. (2015a) for details). However, under current climatic conditions this assumption only results in a small difference between schemes, although this effect could be amplified in the future with expected increased in vapour pressure deficit in a warmer world.

4.2 Overestimation of soil evaporation

We identified systematic biases in the simulation of peak and spring Q_E, particularly at forested sites (e.g. Tumbarumba and Blodgett) (Fig. 2 and S7). The biases in the timing and magnitude of spring and peak fluxes not only have implications for the correct simulation of seasonal cycles, but can also affect the magnitude of dry-down simulated by the model. The excessive spring and early summer Q_E may reduce soil moisture levels prior to the dry-down, leading to the simulation of more severe reductions in Q_E during dry periods. Both hydrological schemes showed a tendency to significantly overestimate these fluxes. The reason for the overestimation of peak fluxes is not clear but is not resolved by the new hydrological scheme despite this parameterising many of the relevant processes differently. At many sites, the high Q_E in spring is associated with excessive soil evaporation and is not linked to transpiration, which closely follow the observed seasonal cycle (see e.g. Bugac, Harvard, Howland and Hyytiälä in Fig. S9).

There are a number of possible causes and solutions to this excessive soil evaporation. Insufficient drainage, and consequently overestimated surface soil moisture, and/or insufficient reduction of soil evaporation during soil drying may explain the excess spring Q_E. The default scheme uses a linear function to reduce soil evaporation when

soil moisture is limiting following Eq. (3). This is replaced with a non-linear function presented in Eq. (10) in the new scheme. The non-linear function provides a much stronger limitation on soil evaporation as soil moisture declines but, based on these results, this approach is not sufficient for resolving the excessive soil evaporation.

Haverd and Cuntz (2010) showed the inclusion of litter layer dynamics in an earlier version of CABLE improved the simulated timing of spring Q_E at Tumbarumba by suppressing soil evaporation but this was not implemented in the current study. Adding a litter layer may resolve excessive soil evaporation at some sites by adding an additional resistance to evaporation, but it is unclear that this approach would resolve errors at all PFTs. Errors in the timing of spring green-up at deciduous sites in the LAI inputs (e.g. Fisher and Mustard, 2007) may also contribute to excessive spring evaporation, whereby a delayed green-up would allow excessive radiation to reach the ground surface in early spring, increasing soil evaporation rates. We encourage researchers to make use of the Best et al. (2015) experimental protocol to fully explore this problem. Using multi-LSM simulations should be able to identify where CABLE is anomalous, and ideally implement the model parameterisations used in other LSMs that do not simulate excessive spring soil evaporation.

4.3 Further model uncertainties

In this study, we explored and quantified model uncertainties due to LAI, g_s, hydrological and soil parameters, limiting our analysis to parameters that can be derived from observationally based global datasets (despite considerable uncertainties). Other model processes, particularly more realistic representations of vegetation drought responses, have been identified as critical for capturing drought processes and shown to improve CABLE performance during droughts but were not explicitly explored here.

The simulation of the effects of soil moisture limitation on photosynthesis and stomatal conductance remains a key uncertainty for drought responses in LSMs (Zhou et al., 2013). Models rely on differing assumptions about the effects of water stress on photosynthesis and stomatal conductance (Egea et al., 2011; Keenan et al., 2009) but generally assume similar drought responses across different PFTs (including CABLE

as employed here) despite experimental evidence pointing to systematic differences in plant adaptations to drought (De Kauwe et al., 2015b; Zhou et al., 2013). In common with many other LSMs (Verhoef and Egea, 2014), CABLE limits gas exchange during low soil moisture using the dimensionless scalar β following Eq. (11). The function is strongly linked to soil properties (through wilting point and field capacity parameters) and does not directly consider vegetation characteristics beyond rooting depth (which varies little by PFT). De Kauwe et al. (2015b) evaluated CABLE against flux site observations during the 2003 European drought using an alternative drought model with experimentally derived drought sensitivities. They showed significant underestimations of Q_E using the default CABLE but these were improved using different vegetation sensitivities to drought (varying from low sensitivity in xeric environments to high in mesic environments in line with experimental evidence) and a dynamic weighting of water uptake across soil layers. Experimental data to inform the parameterisation of PFT-specific drought responses, however, remains limited (De Kauwe et al., 2015b), complicating the implementation of such responses into LSMs. Li et al. (2012) showed the underestimation of CABLE-simulated Q_E under water-stressed conditions could be improved by employing an alternative root water uptake scheme. The default root water uptake function in CABLE employed here (Wang et al., 2011) assumes a constant efficiency of water uptake per unit root length (Li et al., 2012). CABLE with the alternative scheme, combining a function allowing variable root-density distribution (Lai and Katul, 2000) with a hydraulic redistribution scheme (which allows roots to move water from wetter to drier soil layers), was shown to correctly capture the magnitude of seasonal-scale droughts across three flux tower sites. The implementation of more realistic vegetation responses and adaptations to droughts should further refine the performance of the new hydrological scheme during dry-down periods.

Furthermore, in the simulations described here prescribed monthly MODIS LAI was used. Whilst CABLE and many other LSMs are capable of simulating LAI dynamically, it is common practice, particularly in coupled online simulations, to rely on prescribed monthly climatology instead of time-varying LAI. This limits the realistic simulation of reductions in LAI during severe droughts and consequent feedbacks with radiative and evaporative processes such as interception losses. Canopy defoliation may, for example, decrease transpiration and interception but also

1 increase radiation reaching the soil surface, potentially increasing soil evaporation in
2 the presence of available moisture. As these feedbacks were not considered in this
3 study, the rate of dry-down may have been overestimated at sites which experienced
4 LAI reductions during rainfall deficits, but which may not have been captured in the
5 MODIS LAI inputs. However, as only the magnitude of LAI was varied in this study,
6 it is not possible to quantify the effects of temporal errors in LAI on simulated Q_E.
7 Since both hydrological models were forced with identical LAI, it is unlikely
8 uncertainties in the prescribed LAI explain the excessive dry-down in the default
9 hydrological scheme.
10
11 We have limited our analysis to short-term, seasonal-scale rainfall deficits. Multi-
12 annual droughts, such as the Millennium drought in eastern Australia (van Dijk et al.,
13 2013), are likely to exhibit different dynamics in terms of vegetation responses and
14 consequent feedbacks with land surface fluxes, soil moisture states and albedo.
15 Prudhomme et al. (2011), for example, showed the JULES LSM to more successfully
16 reproduce long-term hydrological droughts than short-term events in terms of
17 duration and severity. Realistic representations of plant adaptations to drought and
18 dynamically varying LAI are likely to be increasingly important for representing
19 vegetation resilience and coupled land surface processes during long-term droughts.
20 We therefore suggest future studies of LSM performance under water-stressed
21 conditions should evaluate models against drought events at different temporal scales.
22
23
24 **5 Conclusions**
25
26 This study evaluated the CABLE land surface model for seasonal-scale precipitation
27 deficits using 20 flux tower sites distributed globally. We varied the soil hydrological
28 and stomatal conductance parameterisations, and the inputs for LAI and soil
29 properties. Our goal was to determine whether CABLE could capture dry-down
30 associated with rainfall deficits as these components of the model are varied, or
31 whether the model lacks the physical parameterisations to simulate this phenomenon.
32
33 On long time scales (annual and above), compensating biases mean that the two
34 versions of CABLE performed similarly. However, as our analysis focused more on

periods of rainfall deficit, a new hydrological parameterisation based on Decker (2015) clearly improved the capability of CABLE to simulate Q_E. However, neither version of CABLE, and no reasonable choice of soil parameter, LAI or stomatal conductance resolved systematic seasonal-scale biases in excessive spring soil evaporation. The reasons for these biases cannot be determined in isolation and we will next pursue these model limitations using the PLUMBER multi-model benchmarking framework (Best et al., 2015).

Our study highlights some opportunities for land modellers. First, our study again demonstrates the value in freely-available flux tower data for identifying systematic biases in LSMs. The value of these data extends well beyond their common use in evaluating means or seasonal cycles. Second, a major role for LSMs is to simulate feedbacks to the atmosphere associated with rainfall deficits. We have demonstrated that there is skill in CABLE in simulating these feedbacks as a landscape dries but clearly more work needs to be invested in capturing all the elements of a drying soil and its impacts on Q_E. While the parameterisation of hydrology has been explored over the years, we remind the community that there are on-going challenges in modelling soil moisture and links between soil moisture and evaporation that are not yet resolved. Third, we note that CABLE performs reasonably relative to other LSMs (Abramowitz et al., 2007; Best et al., 2015) and yet when we interrogate the model's performance at timescales when compensating biases are limited, CABLE displays some concerning behaviour. It is inevitable that other LSMs, if examined using these periods of precipitation deficit, will also exhibit problems. Clearly, formally testing LSMs against more extreme conditions, and in the context of specific phenomenon (e.g. drought or heatwave) is a necessary step to build confidence in the projections from climate models that utilise LSMs.

Acknowledgements

References

Abramowitz, G.: Towards a public, standardized, diagnostic benchmarking system for land surface models, Geosci. Model Dev., 5, 819–827, doi:10.5194/gmd-5-819-2012, 2012.

Abramowitz, G., Pitman, A., Gupta, H., Kowalczyk, E. and Wang, Y.: Systematic Bias in Land Surface Models, J. Hydrometeorol., 8, 989–1001, doi:10.1175/JHM628.1, 2007.

Abramowitz, G., Leuning, R., Clark, M. and Pitman, A.: Evaluating the performance of land surface Models, J. Clim., 21, 5468–5481, doi:10.1175/2008JCLI2378.1, 2008.

Allen, C. D., Macalady, A. K., Chenchouni, H., Bachelet, D., McDowell, N., Vennetier, M., Kitzberger, T., Rigling, A., Breshears, D. D., Hogg, E. H. (Ted), Gonzalez, P., Fensham, R., Zhang, Z., Castro, J., Demidova, N., Lim, J. H., Allard, G., Running, S. W., Semerci, A. and Cobb, N.: A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests, For. Ecol. Manage., 259, 660–684, doi:10.1016/j.foreco.2009.09.001, 2010.

Ball, T., Woodrow, I. and Berry, J.: A model predicting stomatal conductance and its contribution to the control of photosynthesis under different environmental

conditions, in Progress in Photosynthesis Research, edited by J. Biggins, pp. 221–224, Martinus Nijhoff, Leiden, the Netherlands., 1987.

Best, M. J., Abramowitz, G., Johnson, H. R., Pitman, A. J., Balsamo, G., Boone, A., Cuntz, M., Decharme, B., Dirmeyer, P. A., Dong, J., Ek, M., Guo, Z., Haverd, V., van den Hurk, B. J., Nearing, G. S., Pak, B., Peters-Lidard, C., Santanello, J. A., Stevens, L. and Vuichard, N.: The plumbing of land surface models: benchmarking model performance, J. Hydrometeorol., 16, 1425–1442, doi:10.1175/JHM-D-14-0158.1, 2015.

Bi, D., Dix, M., Marsland, S. J., O'Farrell, S., Rashid, H. a, Uotila, P., Hirst, A. C., Kowalczyk, E., Golebiewski, M., Sullivan, A., Yan, H., Hannah, N., Franklin, C., Sun, Z., Vohralik, P., Watterson, I., Zhou, X., Fiedler, R., Collier, M., Ma, Y., Noonan, J., Stevens, L., Uhe, P., Zhu, H., Griffies, S. M., Hill, R., Harris, C., Puri, K. and Farrell, S. O.: The ACCESS coupled model: description, control climate and evaluation, Aust. Meteorol. Oceanogr. J., 63, 41–64, 2013.

Blyth, E., Clark, D. B., Ellis, R., Huntingford, C., Los, S., Pryor, M., Best, M. and Sitch, S.: A comprehensive set of benchmark tests for a land surface model of simultaneous fluxes of water and carbon at both the global and seasonal scale, Geosci. Model Dev., 4, 255–269, doi:10.5194/gmd-4-255-2011, 2011.

Clapp, R. B. and Hornberger, G. M.: Empirical equations for some soil hydraulic properties, Water Resour. Res., 14, 601–604, 1978.

Collins, M., Knutti, R., Arblaster, J., Dufresne, J. L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W. J., Johns, T., Krinner, G., Shongwe, M., Tebaldi, C., Weaver, A. J. and Wehner, M.: Long-term Climate Change: Projections, Commitments and Irreversibility, in Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley, pp. 1029–1136, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA., 2013.

Cruz, F. T., Pitman, A. J., McGregor, J. L. and Evans, J. P.: Contrasting Regional Responses to Increasing Leaf-Level Atmospheric Carbon Dioxide over Australia, J. Hydrometeorol., 11, 296–314, doi:10.1175/2009jhm1175.1, 2010.

Dai, A.: Increasing drought under global warming in observations and models, Nat. Clim. Chang., 3, 52–58, doi:10.1038/nclimate1633, 2012.

De Kauwe, M. G., Disney, M. I., Quaife, T., Lewis, P. and Williams, M.: An assessment of the MODIS collection 5 leaf area index product for a region of mixed coniferous forest, Remote Sens. Environ., 115, 767–780, doi:10.1016/j.rse.2010.11.004, 2011.

De Kauwe, M. G., Kala, J., Lin, Y.-S., Pitman, A. J., Medlyn, B. E., Duursma, R. A., Abramowitz, G., Wang, Y.-P. and Miralles, D. G.: A test of an optimal stomatal conductance scheme within the CABLE land surface model, Geosci. Model Dev., 8, 431–452, doi:10.5194/gmd-8-431-2015, 2015a.

De Kauwe, M. G., Zhou, S.-X., Medlyn, B. E., Pitman, A. J., Wang, Y. P., Duursma, R. A. and Prentice, I. C.: Do land surface models need to include differential plant species responses to drought? Examining model predictions across a latitudinal gradient in Europe, Biogeosciences, 12, 12349–12393, doi:10.5194/bgd-12-12349-2015, 2015b.

Decker, M.: Development and evaluation of a new soil moisture and runoff parameterization for the CABLE LSM including subgrid-scale processes, J. Adv. Model. Earth Syst., 7, doi:10.1002/2013MS000282.Received, 2015.

Dirmeyer, P. A.: A history and review of the Global Soil Wetness Project (GSWP), J. Hydrometeorol., 12, 729–749, doi:10.1175/JHM-D-10-05010.1, 2011.

Egea, G., Verhoef, A. and Vidale, P. L.: Towards an improved and more flexible representation of water stress in coupled photosynthesis-stomatal conductance models, Agric. For. Meteorol., 151, 1370–1384, doi:10.1016/j.agrformet.2011.05.019, 2011.

Fisher, J. I. and Mustard, J. F.: Cross-scalar satellite phenology from ground, Landsat, and MODIS data, Remote Sens. Environ., 109, 261–273, doi:10.1016/j.rse.2007.01.004, 2007.

Global Soil Data Task Group: Global gridded surfaces of selected soil characteristics (IGBP-IDS), [online] Available from: http://www.daac.ornl.gov, 2000.

Gordon, H. B., Rotstayn, L. D., McGregor, J. L., Dix, J. L., Kowalczyk, E. A., O'Farrell, S. P., Waterman, L. J., Hirst, A. C., Wilson, S. G., Collier, M. A.,

Watterson, I. G. and Elliott, T. I.: The CSIRO Mk3 climate system model, Aspendale, Victoria, Australia. [online] Available from: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:The+CSIRO+Mk3 +Climate+System+Model#0, 2002.

Haverd, V. and Cuntz, M.: Soil-Litter-Iso: A one-dimensional model for coupled transport of heat, water and stable isotopes in soil with a litter layer and root extraction, J. Hydrol., 388, 438–455, doi:10.1016/j.jhydrol.2010.05.029, 2010.

Haverd, V., Raupach, M. R., Briggs, P. R., Canadell, J. G., Isaac, P., Pickett-Heaps, C., Roxburgh, S. H., van Gorsel, E., Viscarra Rossel, R. A. and Wang, Z.: Multiple observation types reduce uncertainty in Australia's terrestrial carbon and water cycles, Biogeosciences, 10, 2011–2040, doi:10.5194/bg-10-2011-2013, 2013.

Huang, M., Piao, S., Sun, Y., Ciais, P., Cheng, L., Mao, J., Poulter, B., Shi, X., Zeng, Z. and Wang, Y.: Change in terrestrial ecosystem water-use efficiency over the last three decades, Glob. Chang. Biol., 21, 2366–2378, doi:10.1111/gcb.12873, 2015.

IPCC: Climate change 2013: The physical science basis. Contribution of working group I to the fifth assessment report of the Intergovernmental Panel on Climate Change, edited by T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA., 2013.

Kala, J., Decker, M., Exbrayat, J.-F., Pitman, A. J., Carouge, C., Evans, J. P., Abramowitz, G. and Mocko, D.: Influence of leaf area index prescriptions on simulations of heat, moisture, and carbon fluxes, J. Hydrometeorol., 15, 489–503, doi:10.1175/JHM-D-13-063.1, 2014.

Keenan, T., García, R., Friend, A. D., Zaehle, S., Gracia, C. and Sabate, S.: Improved understanding of drought controls on seasonal variation in Mediterranean forest canopy CO_2 and water fluxes through combined in situ measurements and ecosystem modelling, Biogeosciences, 6, 1423–1444, doi:10.5194/bgd-6-2285-2009, 2009.

Koster, R. D., Guo, Z., Yang, R., Dirmeyer, P. A., Mitchell, K. and Puma, M. J.: On the nature of soil moisture in land surface models, J. Clim., 22, 4322–4335, doi:10.1175/2009JCLI2832.1, 2009.

Kowalczyk, E. A., Garratt, J. R. and Krummel, P. B.: Implementation of a soil-canopy

scheme into the CSIRO GCM: Regional aspects of the model response. [online] Available from: http://www.csa.com/partners/viewrecord.php?requester=gs&collection=TRD&recid=N9524192AH, 1994.

Kowalczyk, E. A., Wang, Y. P., Law, R. M., Davies, H. L., Mcgregor, J. L. and Abramowitz, G.: The CSIRO Atmosphere Biosphere Land Exchange (CABLE) model for use in climate models and as an offline model, Aspendale, Victoria, Australia., 2006.

Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S. and Prentice, I. C.: A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, Global Biogeochem. Cycles, 19, GB1015, doi:10.1029/2003GB002199, 2005.

Lai, C.-T. and Katul, G.: The dynamic role of root-water uptake in coupling potential to actual transpiration, Adv. Water Resour., 23, 427–439, doi:10.1016/S0309-1708(99)00023-8, 2000.

Lawrence, P. J. and Chase, T. N.: Representing a new MODIS consistent land surface in the Community Land Model (CLM 3.0), J. Geophys. Res. Biogeosciences, 112, G01023, doi:10.1029/2006JG000168, 2007.

Leuning, R.: A critical appraisal of a combined stomatal-photosynthesis model for C_3 plants, Plant. Cell Environ., 18, 339–355, 1995.

Leuning, R., van Gorsel, E., Massman, W. J. and Isaac, P. R.: Reflections on the surface energy imbalance problem, Agric. For. Meteorol., 156, 65–74, doi:10.1016/j.agrformet.2011.12.002, 2012.

Li, L., Wang, Y.-P., Yu, Q., Pak, B., Eamus, D., Yan, J., van Gorsel, E. and Baker, I. T.: Improving the responses of the Australian community land surface model (CABLE) to seasonal drought, J. Geophys. Res., 117, G04002, doi:10.1029/2012JG002038, 2012.

Lin, Y.-S., Medlyn, B. E., Duursma, R. A., Prentice, I. C., Wang, H., Baig, S., Eamus, D., de Dios, V. R., Mitchell, P., Ellsworth, D. S., de Beeck, M. O., Wallin, G., Uddling, J., Tarvainen, L., Linderson, M.-L., Cernusak, L. a., Nippert, J. B., Ocheltree, T. W., Tissue, D. T., Martin-StPaul, N. K., Rogers, A., Warren, J. M., De

1 Angelis, P., Hikosaka, K., Han, Q., Onoda, Y., Gimeno, T. E., Barton, C. V. M.,

2 Bennie, J., Bonal, D., Bosc, A., Löw, M., Macinins-Ng, C., Rey, A., Rowland, L.,

3 Setterfield, S. A., Tausz-Posch, S., Zaragoza-Castells, J., Broadmeadow, M. S. J.,

4 Drake, J. E., Freeman, M., Ghannoum, O., Hutley, L. B., Kelly, J. W., Kikuzawa, K.,

5 Kolari, P., Koyama, K., Limousin, J.-M., Meir, P., Lola da Costa, A. C., Mikkelsen,

6 T. N., Salinas, N., Sun, W. and Wingate, L.: Optimal stomatal behaviour around the

7 world, Nat. Clim. Chang., 5, 459–464, doi:10.1038/nclimate2550, 2015.

8 Lorenz, R. and Pitman, A. J.: Effect of land-atmosphere coupling strength on impacts

9 from Amazonian deforestation, Geophys. Res. Lett., 41, 5987–5995,

10 doi:10.1002/2014GL061017, 2014.

11 Lorenz, R., Pitman, A. J., Donat, M. G., Hirsch, A. L., Kala, J., Kowalczyk, E. A.,

12 Law, R. M. and Srbinovsky, J.: Representation of climate extreme indices in the

13 ACCESS1.3b coupled atmosphere–land surface model, Geosci. Model Dev., 7, 545–

14 567, doi:10.5194/gmd-7-545-2014, 2014.

15 Medlyn, B. E., Duursma, R. A., Eamus, D., Ellsworth, D. S., Prentice, I. C., Barton,

16 C. V. M., Crous, K. Y., De Angelis, P., Freeman, M. and Wingate, L.: Reconciling

17 the optimal and empirical approaches to modelling stomatal conductance, Glob.

18 Chang. Biol., 17, 2134–2144, doi:10.1111/j.1365-2486.2010.02375.x, 2011.

19 Oleson, K. W., Niu, G. Y., Yang, Z. L., Lawrence, D. M., Thornton, P. E., Lawrence,

20 P. J., Stöckli, R., Dickinson, R. E., Bonan, G. B., Levis, S., Dai, A. and Qian, T.:

21 Improvements to the community land model and their impact on the hydrological

22 cycle, J. Geophys. Res. Biogeosciences, 113, doi:10.1029/2007JG000563, 2008.

23 Oleson, K. W., Lawrence, D. M., Bonan, G. B., Drewniak, B., Huang, M., Koven, C.

24 D., Levis, S., Li, F., Riley, W. J., Subin, Z. M., Swenson, S. C., Thornton, P. E.,

25 Bozbiyik, A., Fisher, R., Heald, C. L., Kluzek, E., Lamarque, J.-F., Lawrence, P. J.,

26 Leung, L. R., Lipscomb, W., Muszala, S., Ricciuto, D. M., Sacks, W., Sun, Y., Tang,

27 J. and Yang, Z.-L.: Technical description of version 4.5 of the Community Land

28 Model (CLM), Boulder, Colorado., 2013.

29 Pitman, A. J.: The evolution of, and revolution in, land surface schemes designed for

30 climate models, Int. J. Climatol., 23, 479–510, doi:10.1002/joc.893, 2003.

31 Powell, T. L., Galbraith, D. R., Christoffersen, B. O., Harper, A., Imbuzeiro, H. M. a,

32 Rowland, L., Almeida, S., Brando, P. M., da Costa, A. C. L., Costa, M. H., Levine, N.

M., Malhi, Y., Saleska, S. R., Sotta, E., Williams, M., Meir, P. and Moorcroft, P. R.: Confronting model predictions of carbon fluxes with measurements of Amazon forests subjected to experimental drought, New Phytol., 200, 350–365, doi:10.1111/nph.12390, 2013.

Prudhomme, C., Parry, S., Hannaford, J., Clark, D. B., Hagemann, S. and Voss, F.: How Well Do Large-Scale Models Reproduce Regional Hydrological Extremes in Europe?, J. Hydrometeorol., 12, 1181–1204, doi:10.1175/2011JHM1387.1, 2011.

Prudhomme, C., Giuntoli, I., Robinson, E. L., Clark, D. B., Arnell, N. W., Dankers, R., Fekete, B. M., Franssen, W., Gerten, D., Gosling, S. N., Hagemann, S., Hannah, D. M., Kim, H., Masaki, Y., Satoh, Y., Stacke, T., Wada, Y. and Wisser, D.: Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment., Proc. Natl. Acad. Sci. U. S. A., 111, 3262–7, doi:10.1073/pnas.1222473110, 2014.

Puri, K., Dietachmayer, G., Steinle, P., Dix, M., Rikus, L., Logan, L., Naughton, M., Tingwell, C., Xiao, Y., Barras, V., Bermous, I., Bowen, R., Deschamps, L., Franklin, C., Fraser, J., Glowacki, T., Harris, B., Lee, J., Le, T., Roff, G., Sulaiman, a, Sims, H., Sun, X., Sun, Z., Zhu, H., Chattopadhyay, M. and Engel, C.: Implementation of the initial ACCESS numerical weather prediction system, Aust. Meteorol. Oceanogr. J., 63, 265–284, 2013.

Sakaguchi, K. and Zeng, X.: Effects of soil wetness, plant litter, and under-canopy atmospheric stability on ground evaporation in the Community Land Model (CLM3.5), J. Geophys. Res. Atmos., 114, 1–14, doi:10.1029/2008JD010834, 2009.

Samanta, A., Ganguly, S., Hashimoto, H., Devadiga, S., Vermote, E., Knyazikhin, Y., Nemani, R. R. and Myneni, R. B.: Amazon forests did not green-up during the 2005 drought, Geophys. Res. Lett., 37, 1–5, doi:10.1029/2009GL042154, 2010.

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B. and Teuling, A. J.: Investigating soil moisture–climate interactions in a changing climate: A review, Earth-Science Rev., 99, 125–161, doi:10.1016/j.earscirev.2010.02.004, 2010.

Sheffield, J. and Wood, E.: Drought: past problems and future scenarios, Earthscan, London, United Kingdom., 2011.

Tallaksen, L. M. and Stahl, K.: Spatial and temporal patterns of large-scale droughts in Europe: model dispersion and performance, Geophys. Res. Lett., 41, 429–434, doi:10.1002/2013GL058954.Received, 2014.

Trenberth, K. E., Dai, A., Schrier, G. Van Der, Jones, P. D., Barichivich, J., Briffa, K. R. and Sheffield, J.: Global warming and changes in drought, Nat. Clim. Chang., 4, 17–22, doi:10.1038/NCLIMATE2067, 2014.

van Dijk, A. I. J. M., Beck, H. E., Crosbie, R. S., de Jeu, R. A. M., Liu, Y. Y., Podger, G. M., Timbal, B. and Viney, N. R.: The Millennium Drought in southeast Australia (2001-2009): Natural and human causes and implications for water resources, ecosystems, economy, and society, Water Resour. Res., 49, 1–18, doi:10.1002/wrcr.20123, 2013.

Verhoef, A. and Egea, G.: Modeling plant transpiration under limited soil water: Comparison of different plant and soil hydraulic parameterizations and preliminary implications for their use in land surface models, Agric. For. Meteorol., 191, 22–32, doi:10.1016/j.agrformet.2014.02.009, 2014.

Wang, Y. and Leuning, R.: A two-leaf model for canopy conductance , photosynthesis and partitioning of available energy I : Model description and comparison with a multi-layered model, Agric. For. Meteorol., 91, 89–111, 1998.

Wang, Y. P., Kowalczyk, E., Leuning, R., Abramowitz, G., Raupach, M. R., Pak, B., van Gorsel, E. and Luhar, A.: Diagnosing errors in a land surface model (CABLE) in the time and frequency domains, J. Geophys. Res., 116, G01034, doi:10.1029/2010JG001385, 2011.

Williams, M., Richardson, A. D., Reichstein, M., Stoy, P. C., Peylin, P., Verbeeck, H., Carvalhais, N., Jung, M., Hollinger, D. Y., Kattge, J., Leuning, R., Luo, Y., Tomelleri, E., Trudinger, C. M. and Wang, Y.-P.: Improving land surface models with FLUXNET data, Biogeosciences, 6, 1341–1359, doi:10.5194/bg-6-1341-2009, 2009.

Yuan, H., Dai, Y., Xiao, Z., Ji, D. and Shangguan, W.: Reprocessing the MODIS Leaf Area Index products for land surface and climate modelling, Remote Sens. Environ., 115, 1171–1187, doi:10.1016/j.rse.2011.01.001, 2011.

Zeng, X. and Decker, M.: Improving the numerical solution of soil moisture–based Richards equation for land models with a deep or shallow water table, J.

Hydrometeorol., 10, 308–319, doi:10.1175/2008JHM1011.1, 2009.

Zhang, H., Pak, B., Wang, Y. P., Zhou, X., Zhang, Y. and Zhang, L.: Evaluating surface water cycle simulated by the Australian Community Land Surface Model (CABLE) across different spatial and temporal domains, J. Hydrometeorol., 14, 1119–1138, doi:10.1175/JHM-D-12-0123.1, 2013.

Zhou, S., Duursma, R. A., Medlyn, B. E., Kelly, J. W. G. and Prentice, I. C.: How should we model plant responses to drought? An analysis of stomatal and non-stomatal responses to water stress, Agric. For. Meteorol., 182-183, 204–214, doi:10.1016/j.agrformet.2013.05.009, 2013.

Zhou, X., Zhang, Y., Wang, Y., Zhang, H., Vaze, J., Zhang, L., Yang, Y. and Zhou, Y.: Benchmarking global land surface models against the observed mean annual runoff from 150 large basins, J. Hydrol., 470-471, 269–279, doi:10.1016/j.jhydrol.2012.09.002, 2012.

Zobler, L.: Global soil types, 1-degree grid, , doi:10.3334/ ORNLDAAC/418, 1999.

Figures

1

2
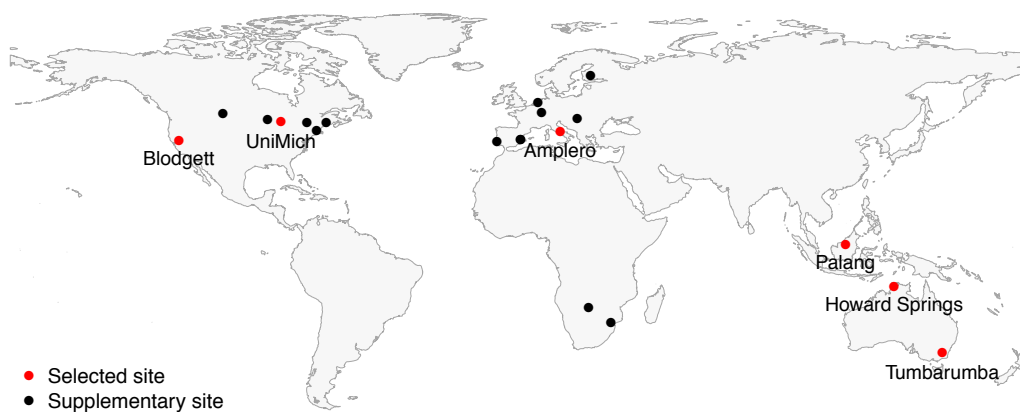
3

4 **Fig. 1.** Location of selected and supplementary flux tower sites.

5

6

7

8

9

10

11

12

13

14

15

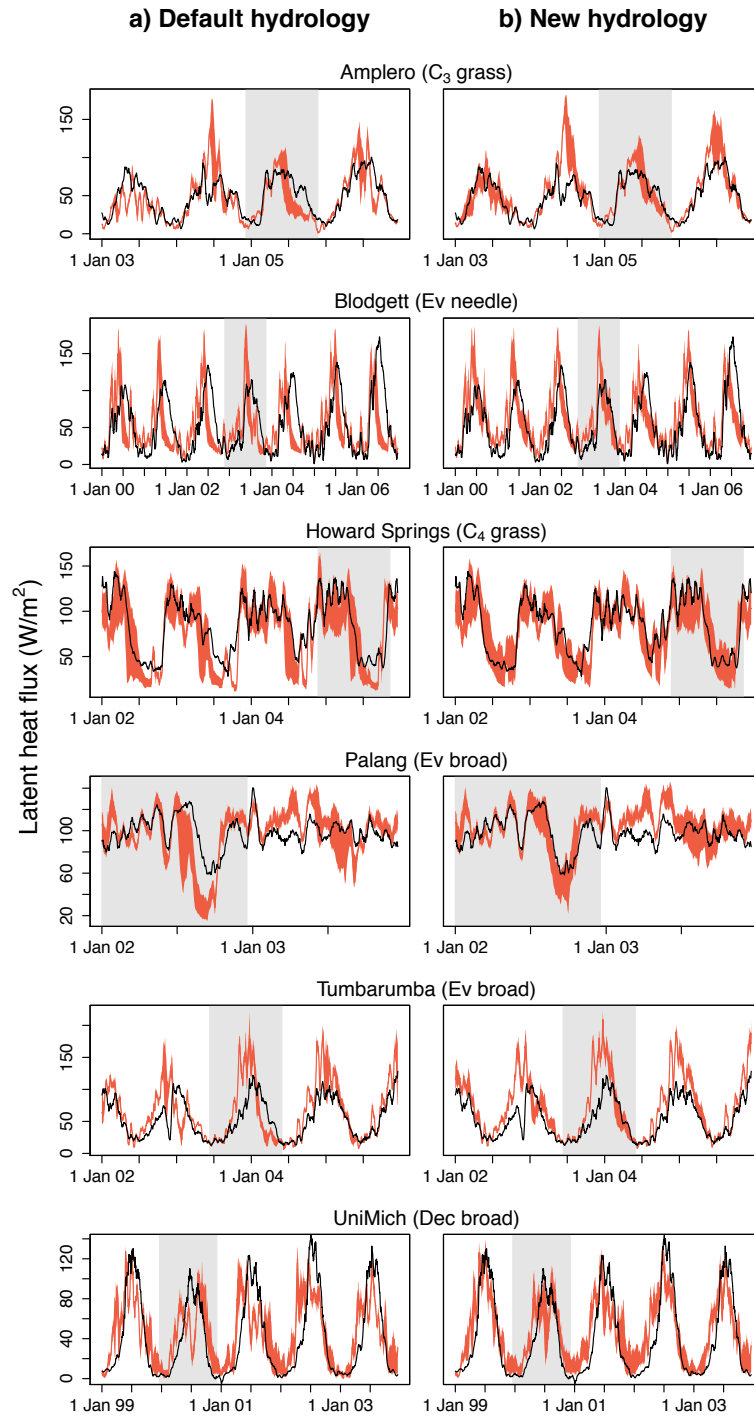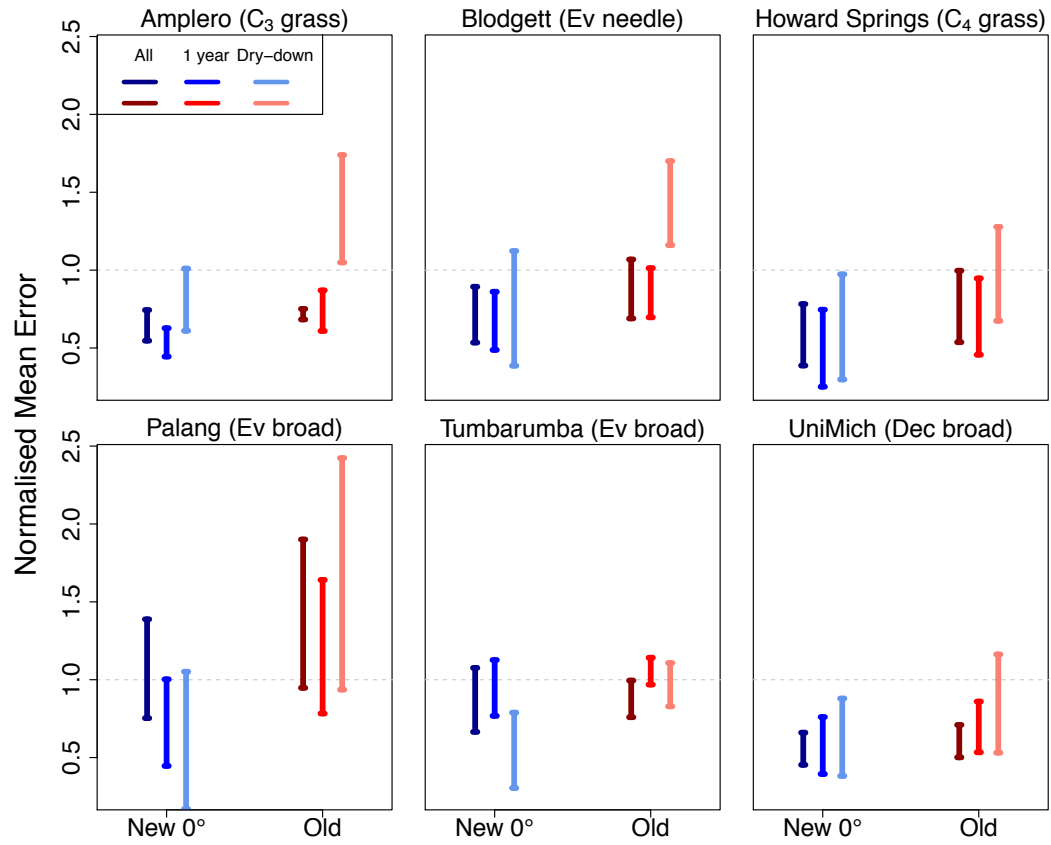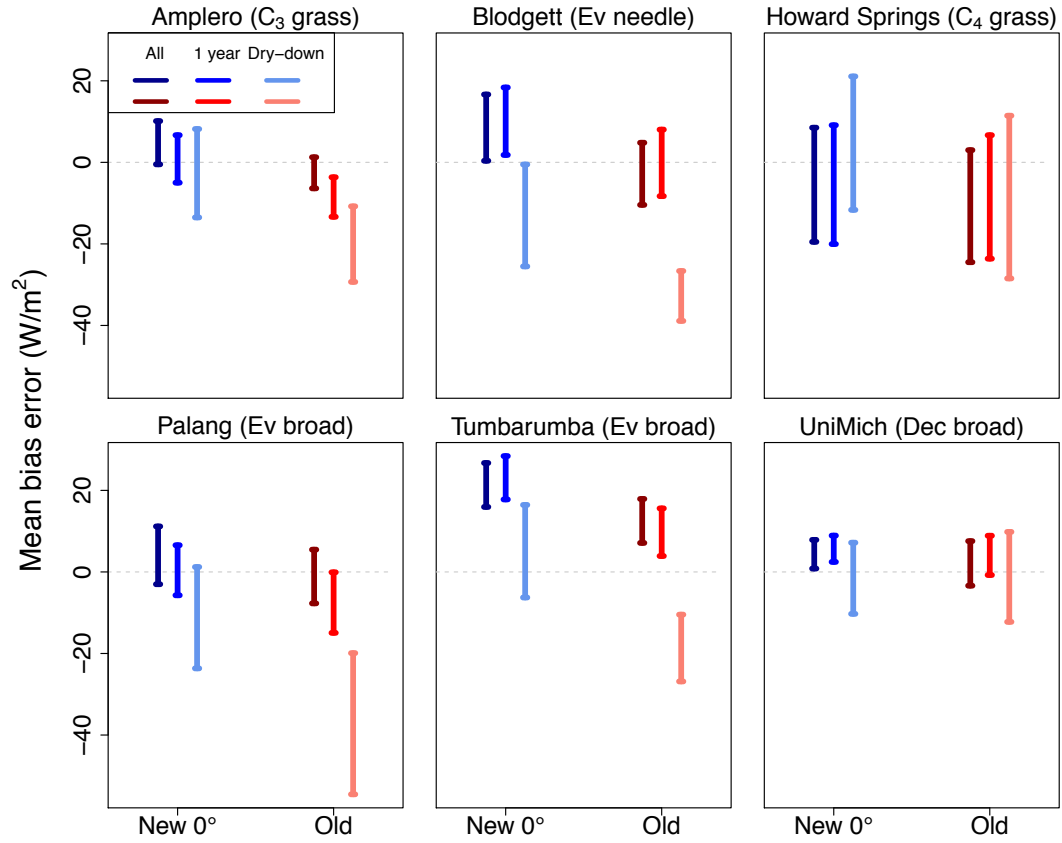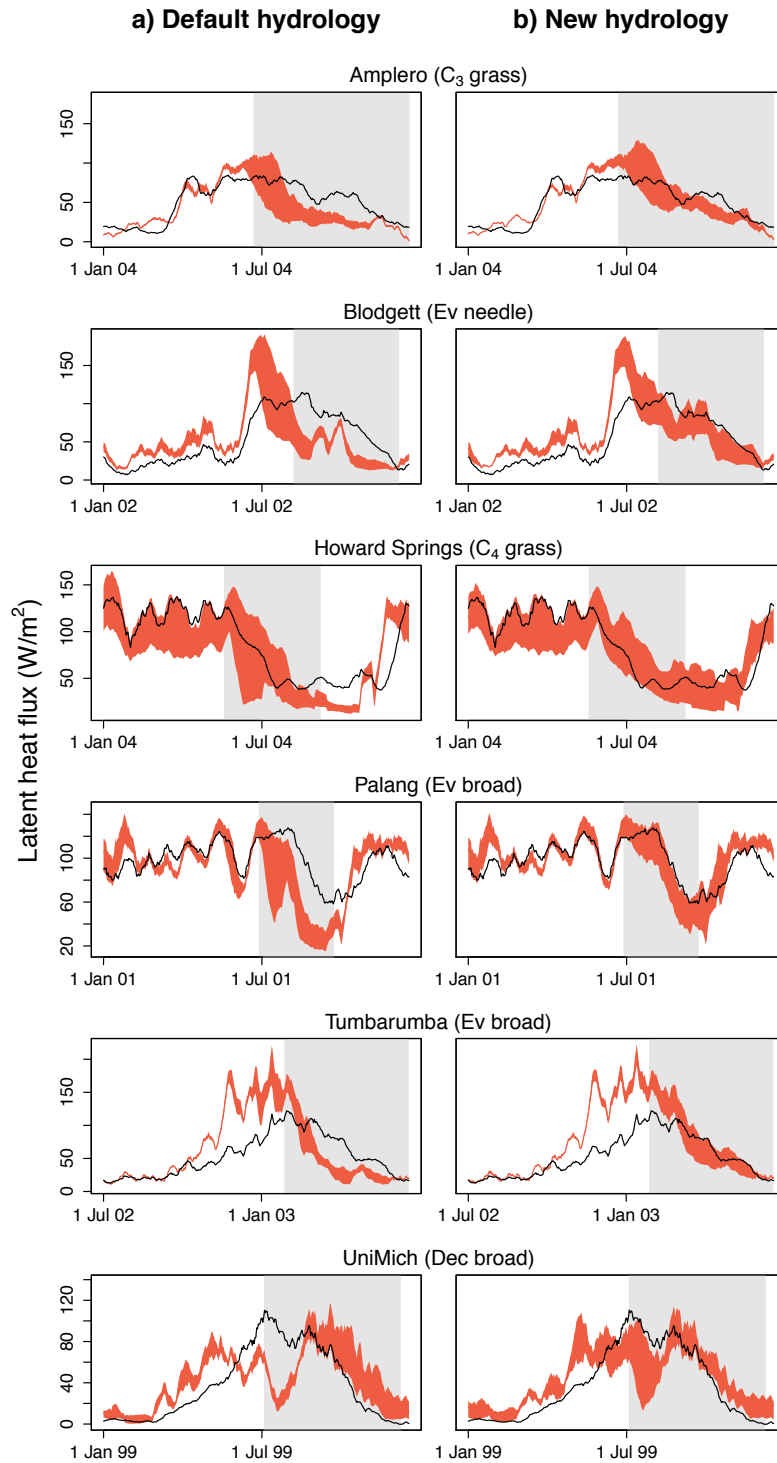16

17

18

19

20

21

22

23

24

25

26

1

2 **Fig. 2**. The range in simulated latent heat (red) during the whole observational data

3 period using the default (left panel) and new (right panel) hydrological schemes with

4 alternative LAI, g_s and soil parameterisations. Observed latent heat is shown in black.

5 The grey shading denotes the selected one-year period.

1

2 **Fig. 3**. The range in normalised mean error metrics of latent heat simulations using

3 the default (red) and new (blue) hydrological schemes with alternative LAI, g_s and

4 soil parameterisations during the whole, annual and dry-down periods. Values closer

5 to 0.0 indicate better model performance.

6

7

8

9

10

11

12

13

14

15

16

17

18

1

2 **Fig. 4.** The range in mean bias error metrics of latent heat simulations using the
3 default (red) and new (blue) hydrological schemes with alternative LAI, g_s and soil
4 parameterisations during the whole, annual and dry-down periods. Values closer to
5 0.0 indicate better model performance.

6

7

8

9

10

11

12

13

14

15

16

17

18

a) Default hydrology   **b) New hydrology**

Amplero (C$_3$ grass)

Blodgett (Ev needle)

Howard Springs (C$_4$ grass)

Latent heat flux (W/m^2)

Palang (Ev broad)

Tumbarumba (Ev broad)

UniMich (Dec broad)

1

Fig. 5. The range in simulated latent heat (red) during the one-year period using the default (left panel) and new (right panel) hydrological schemes with alternative LAI, g_s and soil parameterisations. Observed latent heat is shown in black. The grey shading denotes the selected dry-down period. All time series run from January to December, except Tumbarumba which run from July to the following June.
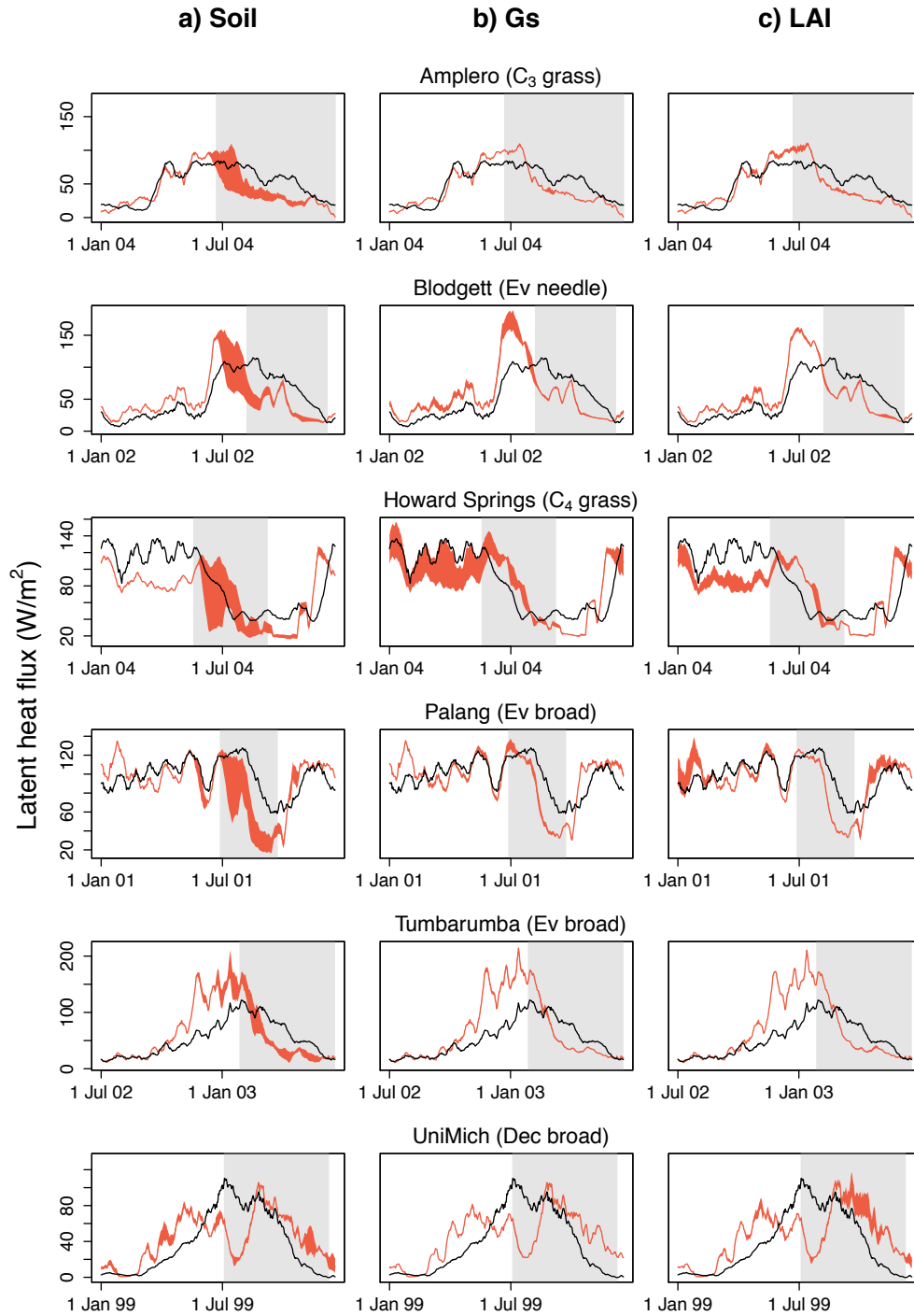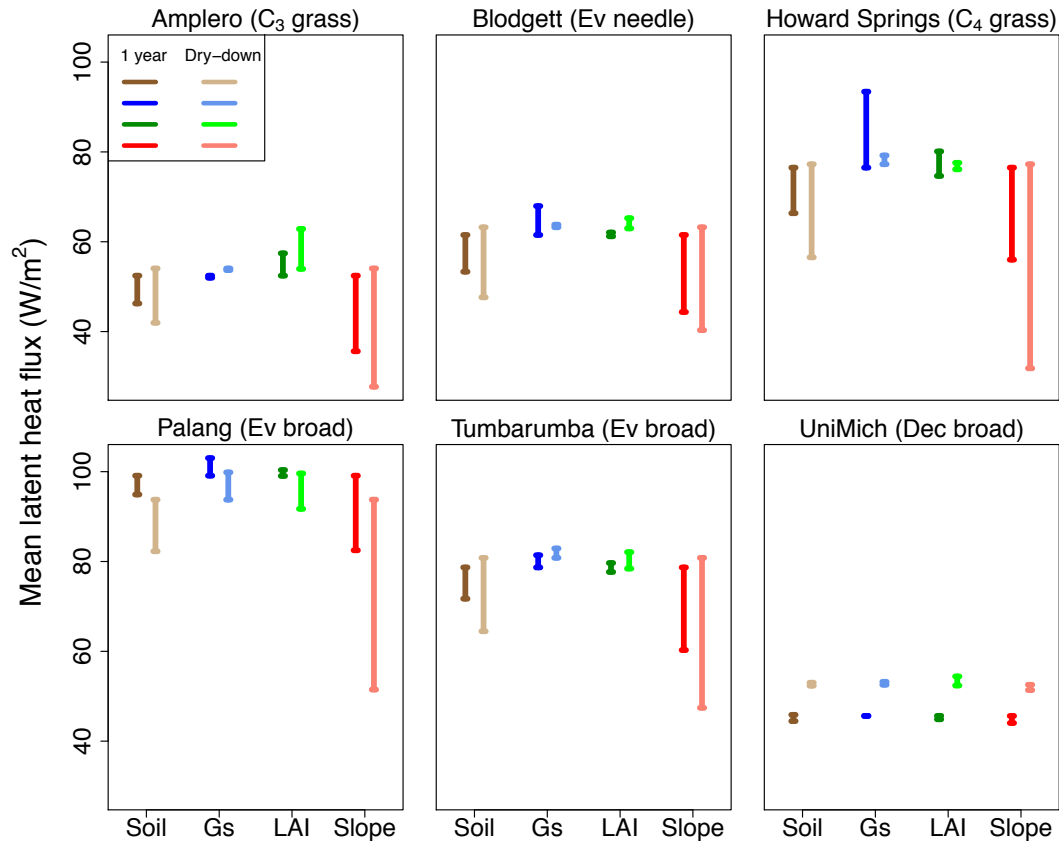
Fig. 6. The range in simulated latent heat (red) arising from the individual effects of soil parameters (left panel), g_s (centre panel) and LAI (right panel) using the default hydrological scheme during the one-year period. Observed latent heat is shown in black. The grey shading denotes the selected dry-down period. The individual effects were determined by fixing the other parameterisations at their default values (medium soil, Medlyn g_s and centre LAI).
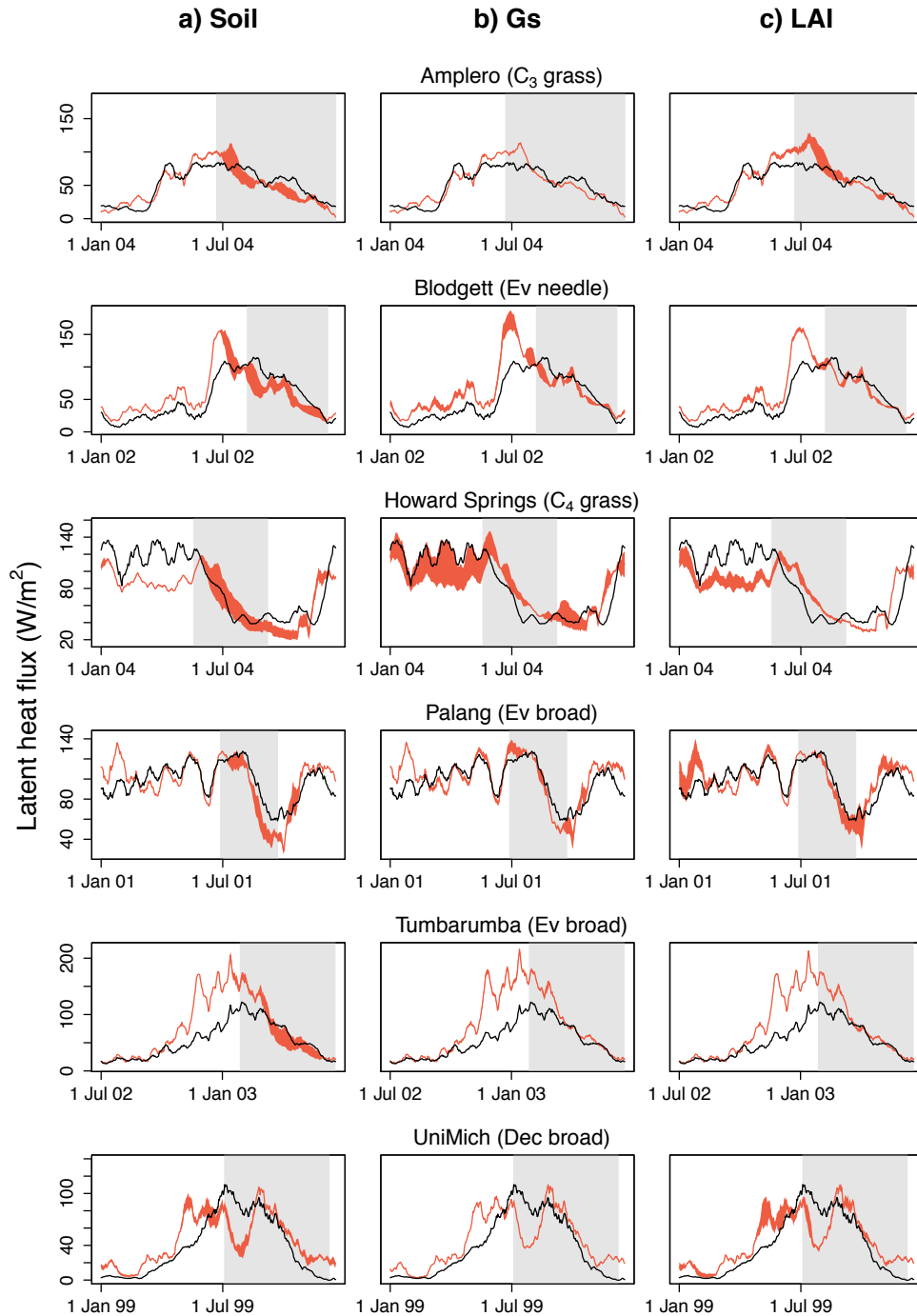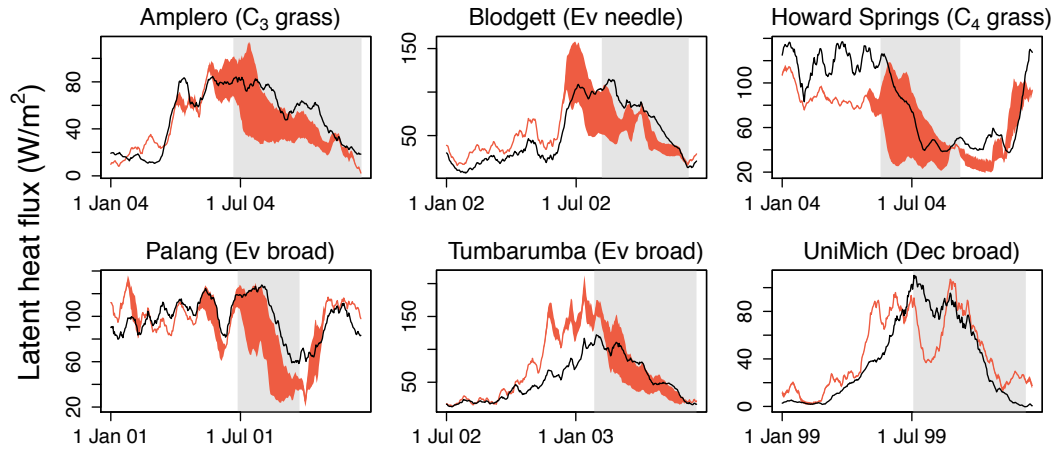
Fig. 7. The range in simulated latent heat (red) arising from the individual effects of soil parameters (left panel), g_s (centre panel) and LAI (right panel) using the new hydrological scheme during the one-year period. Observed latent heat is shown in black. The grey shading denotes the selected dry-down period. The individual effects were determined by fixing the other parameterisations at their default values (medium soil, Medlyn g_s, centre LAI and 0° slope).

Fig. 8. The range in simulated mean latent heat arising from the individual effects of soil (brown), g_s (blue), LAI (green) and slope (red) parameterisations using the new hydrological scheme during the one-year and dry-down periods. The individual effects were determined by fixing the other parameterisations at their default values (medium soil, Medlyn g_s, centre LAI and 0° slope).

1

2 **Fig. 9.** The range in simulated latent heat (red) arising from the individual effects of

3 the slope parameter using the new hydrological scheme during the one-year period.

4 Observed latent heat is shown in black. The grey shading denotes the selected dry-

5 down period. The individual effects were determined by fixing the other

6 parameterisations at their default values (medium soil, Medlyn g_s and centre LAI).