A sampling method for improving the representation of spatially varying precipitation and soil moisture using the Simple Biosphere Model

Isaac D. Medina,[1] A. Scott Denning,[1] Ian T. Baker,[1] Jorge A. Ramirez,[2] and David A. Randall[1]

Received 18 July 2013; revised 15 November 2013; accepted 10 December 2013.

[1] Representing spatially varying precipitation for current grid length scales used in General Circulation Models (GCMs) is a continuing challenge. Furthermore, to fully capture the hydrologic effects of nonuniform precipitation, a representation of soil moisture heterogeneity and distribution of spatially varying precipitation must exist within the same framework. For this study, the explicit and sampling methods of *Sellers et al*. (2007) are tested off-line using the Simple Biosphere Model (SiB3) in an arid, semiarid, and wet site, and are numerically compared to the bulk method, which is currently used in GCMs. To carry out the numerical experiments, an arbitrary grid area was defined by (1) a single instance of SiB3 (bulk method), (2) 100 instances of SiB3 (explicit method), and (3) less than 100 instances of SiB3 (sampling method). Precipitation was randomly distributed over fractions of the grid area for the explicit and sampling methods, while the standard SiB3 exponential distribution relating precipitation intensity to the grid area wet fraction was used in the bulk method. Comparing the sampling and bulk method to the explicit method indicates that 10 instances of SiB3 in the sampling method better captures the spatial variability in soil moisture and grid area flux calculations produced by the explicit method, and deals realistically with spatially varying precipitation at little additional computational cost to the bulk method.

Citation: Medina, I. D., A. S. Denning, I. T. Baker, J. A. Ramirez, and D. A. Randall (2014), A sampling method for improving the representation of spatially varying precipitation and soil moisture using the Simple Biosphere Model, *J. Adv. Model. Earth Syst.*, *5*, doi:10.1002/2013MS000251.

1. Introduction

[2] It has been shown that the representation of soil moisture heterogeneity in General Circulation Models (GCMs) is very important for modeling earth's climate system [e.g., *Ronda et al*., 2002; *Ryu and Famiglietti*, 2006]. Aside from the oceans, soil moisture is another slowly varying component that influences weather and climate through its impact on evaporation and other surface energy fluxes [*Schär et al*., 1999; *Koster et al*., 2004]. Furthermore, studies have shown that in continental midlatitude summers, oceanic impacts on precipitation are small relative to soil moisture [*Koster et al*., 2000, 2002, 2003, 2004, 2006; *Guo et al*., 2006]. With the lack of soil moisture heterogeneity in early climate models, the use of a single soil moisture value for the entire grid area (bulk approach) along with the nonlinear relationship between soil moisture and evapotranspiration has caused inaccuracies in the calculation of the grid area evapotranspiration rate [*Sellers et al*., 2007]. From equation (1), the grid average of the

$$\overline{f(x)} \neq f(\overline{x}) \tag{1}$$

effects of the heterogeneous variable x on the nonlinear function f is not equivalent to the grid average of the heterogeneous variable x applied to the nonlinear function f [*Giorgi and Avissar*, 1997]. To illustrate this problem, when the value of the water stress factor relating the area-averaged soil moisture to evapotranspiration (e.g., Figure 1) lies on the most nonlinear part of the curve, this can lead to inaccuracies in the calculation of the grid area evapotranspiration rate during a precipitation event (e.g., small-scale or convective storm) as a result of a significant jump in the value of the water stress factor. To mitigate this problem [see *Avissar*, 1992; *Giorgi and Avissar*, 1997; *Essery et al*., 2003; *Sellers et al*., 2007], tiling (patches) and binning approaches have been introduced to represent land surface heterogeneity by dividing the grid area into separate fractional

[1]Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado, USA.
[2]Department of Civil and Environmental Engineering, Colorado State University, Fort Collins, Colorado, USA.

Corresponding author: I. D. Medina, Department of Atmospheric Science, Colorado State University, 1371 Campus Delivery, Fort Collins, CO 80523-1371, USA. (imedina@atmos.colostate.edu)

1942-2466/14/10.1002/2013MS000251

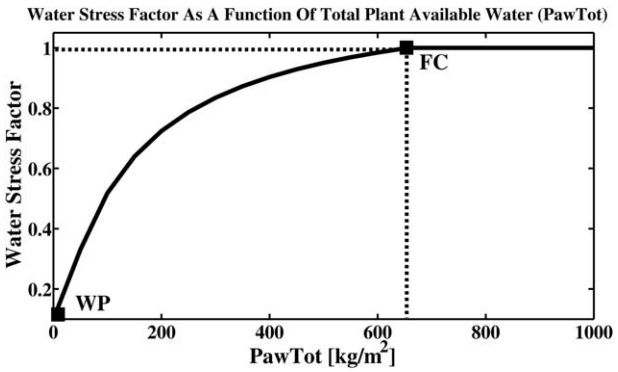

Figure 1. Water stress factor plotted as a function of total plant available water (PawTot, see section 3.2) for the semiarid site. This figure was made using PawTot values ranging from *wp* to total saturation (see section 3.2, equations (13–16)), and soil parameters such as *fc*, *wp*, etc. are derived from literature values and standard SiB3 parameters for a short vegetation/C4 grassland (see section 3.1).

areas (each area can be represented by a different vegetation type, soil moisture value, etc.), and since energy fluxes are calculated separately for each fractional area before calculating a grid average, this allows for an increase in sampling along the nonlinear curve relating soil moisture and evapotranspiration. Furthermore, other approaches have been introduced and represent subgrid-scale heterogeneity through the integration of processes over analytical or empirical probability density functions (PDFs). Regardless of the method used to represent land surface heterogeneity, it is important to note that using a spatially varying land surface with evenly distributed precipitation does not completely describe the spatial variability (and vice versa), and for the case of soil moisture heterogeneity, representing the distribution of spatially varying precipitation is very important, but is poorly characterized in GCMs.

[3] In a study done by *Dickinson and Henderson-Sellers* [1988], the Biosphere-Atmosphere Transfer Scheme (BATS) was used to evaluate the climatic impacts of the deforestation of the Amazon Basin, where simulated interception loss for the entire basin, when compared to interception loss measured at a single basin site, showed overestimations of about 150%. *Dickinson* [1989] explained that the overestimation of surface net radiation and a large canopy storage capacity were the sources for the high interception loss. However, in another study, *Shuttleworth and Dickinson* [1989] suggested that a much more serious source of error was the neglect of spatial variability in precipitation. Currently in GCMs, large-scale (frontal or stratiform) and small-scale (convective) precipitation is predicted as a grid area average or a single grid area precipitation value for every time step. The even distribution of the area-averaged large-scale precipitation rate is sufficient, but evenly distributing the area-averaged small-scale precipitation rate is problematic and decreases the physical realism of small-scale events that alter the balance between evapotranspiration and runoff [*Pitman et al.*, 1990]. A number of studies have explored the disaggregation of precipitation in climate models [*Sato et al.*, 1989b; *Pitman et al.*, 1990; *Gao and Sorooshian*, 1994; *White et al.*, 1997; *Onof et al.*, 1998; *Arora et al.*, 2001; *Hahmann*, 2003; *Tang et al.*, 2007], and many have reported improvements in simulated hydrology and changes in the partitioning of energy fluxes.

[4] For the case of the Simple Biosphere Model (SiB3), the precipitation disaggregation method of *Sato et al.* [1989b] has been implemented for the distribution of small-scale precipitation. The precipitation intensity determines what fraction of the grid area receives precipitation based on an exponential distribution. Unfortunately, due to the lack of land surface heterogeneity, precipitation reaching the land surface from a fraction of the grid is evenly distributed within the entire soil column, and does not completely capture the hydrological effects of nonuniform precipitation. To capture the full effect of precipitation variability in SiB3, it is clear that a representation of land surface heterogeneity and an improved distribution of precipitation must exist within the same framework. Therefore, in this study, we focus on improving the representation of soil moisture heterogeneity and spatially varying precipitation at little additional computational cost to the bulk method (see section 3.4) by applying the methods of *Sellers et al.* [2007], to SiB3. Section 2 gives a brief description of SiB3, methods are described in section 3, results are discussed in section 4, and section 5 has concluding remarks as well as future work related to this research.

2. The Simple Biosphere Model (SiB3)

[5] The Simple Biosphere Model (SiB3) is based on a land-surface parameterization scheme originally used to compute biophysical exchanges of energy, water, and momentum in climate models [*Sellers et al.*, 1986], and later adapted to include ecosystem metabolism [*Sellers et al.*, 1996a; *Denning et al.*, 1996]. The parameterization of photosynthetic carbon assimilation is based on enzyme kinetics originally developed by *Farquhar et al.* [1980], and is linked to stomatal conductance and thence to the surface energy budget and atmospheric climate [*Collatz et al.*, 1991, 1992; *Sellers et al.*, 1996a; *Randall et al.*, 1996]. The model has been updated to include prognostic calculation of temperature, moisture, and trace gases in the canopy air space [*Vidale and Stöckli*, 2005], and the model has been evaluated against eddy covariance measurements at a number of sites [*Baker et al.*, 2003; *Hanan et al.*, 2004; *Baker et al.*, 2008]. Direct-beam and diffuse solar radiation are treated separately for calculations of photosynthesis and transpiration of sunlit and shaded canopy fractions, using algorithms similar to those of *De Pury and Farquhar* [1997]. Other recent improvements include biogeochemical fractionation and recycling of stable carbon isotopes [*Suits et al.*, 2005], improved treatment of soil hydrology and thermodynamics, and the introduction of a multilayer snow model based on the Community Land Model [*Dai et al.*, 2003; *Stöckli et al.*,

2008a], a prognostic phenology algorithm that assimilates vegetation imagery [*Stöckli et al.*, 2008b], the biogeochemical cycling of carbon among decomposing organic pools [*Schaefer et al.*, 2008], and the ecophysiology of corn, soy, and wheat crops [*Lokupitiya et al.*, 2009]. The model is now referred to as SiB3.

3. Methods

3.1. Description of Sites

[6] Derived from an integration of satellite data (e.g., vegetation parameters), literature values and standard SiB3 parameters (e.g., soil parameters) [*Sellers et al.*, 1996b], numerical simulations for an arid, semiarid, and wet climate region (hereafter called sites) were used in this study for a 2 month period. Each site was forced with meteorology from a US city that best described the site climatology (sites do not explicitly represent those cities), and was compiled from a combination of observations and stochastically produced from a weather generator. The arid site is a broadleaf shrub and bare soil mix forced with meteorology from Phoenix, AZ with a total of 50.8 mm of precipitation, the semiarid site is a short vegetation/C4 grassland forced with meteorology from Oklahoma City, OK with a total of 177.9 mm of precipitation, and the wet site is a broadleaf and needleleaf tree mix forced with meteorology from Baton Rouge, LA with a total of 302.77 mm of precipitation.

[7] Since meteorological data used to force SiB3 (air temperature (K), vapor pressure (Pa), atmospheric surface pressure (Pa), wind speed (m/s), short wave radiation (W/m^2), long wave radiation (W/m^2), and small-scale and large-scale precipitation (mm/s)) were not readily available at the necessary temporal and spatial resolution (30 min temporal resolution and a spatial resolution of a point or single weather station observation) for each location, 30 min point observations (or single weather station observations) from the Atmospheric Radiation Measurements (ARM) Program flux tower site (data available from Ameriflux Website: ameriflux.lbl.gov/), located in north-central Oklahoma, were adjusted to create six instances of July meteorology for each site (6 instances $\times$ 3 sites = 18 total July meteorological sets). This was done for the purpose of simulating spatially varying meteorology (i.e., at each site, all six sets of meteorology occurred at the same time, but over different fractions of the grid area), and to complete the 2 month (62 days) duration of the study each set of meteorology was used to force both months. Adjustments were made to ARM site observations from July 2003 (arbitrarily chosen), using the Weather Generator (WGEN) program [*Richardson*, 1981; *Richardson and Wright*, 1984]. Therefore, from the single ARM site observation set from July 2003, 18 sets of meteorology were created where each site had six meteorological sets derived from a combination of meteorology from the ARM site observation set and stochastically produced for a US city that best described the site climatology using the WGEN program.

[8] The WGEN program [*Richardson*, 1981; *Richardson and Wright*, 1984] is a stochastic weather generator that calculates point (or weather station) values of daily precipitation (inches), maximum temperature (°F), minimum temperature (°F), and daily averaged short wave radiation (ly) with two user options. With the first option, daily values of the four variables are produced for a specified number of years with program supplied statistical data (statistical data are derived from actual weather station data corresponding to the specified site), and with option two, the program reads user supplied values of precipitation and the three other variables are calculated based on the user and program supplied data. Using option one for this study, the model first generates precipitation as an independent variable, while the other three variables are calculated based on the wet or dry status of the day. The wet or dry status of each day is determined using a first-order Markov chain model and the precipitation amount on a wet day is generated using a two-parameter gamma distribution. The Markov chain model only depends on the precipitation status of the previous day (wet or dry), where a wet day is defined as a day with a rainfall measurement greater than or equal to 0.01 in. [*Haan*, 1977]. Finally, the procedure for generating daily values of short wave radiation, maximum, and minimum temperatures is based on the weakly stationary generating process [*Matalas*, 1967]. Generated variable values will be close to monthly means obtained from actual data, however, due to temporal and spatial smoothing from the model or topography and other factors, a correction procedure is offered for temperature and precipitation output.

[9] For the meteorological adjustments to ARM site observations for July 2003, the observed minimum and maximum daily temperatures were adjusted to WGEN output, and the remainder of the temperatures for that day, were adjusted to fit between the new minimum and maximum temperature. Short wave radiation was adjusted by multiplying the observed short wave radiation values by the ratio of the WGEN daily averaged short wave radiation to the observed daily averaged short wave radiation. Precipitation had three possible adjustments, the first was for the case of observed precipitation and precipitation generated by WGEN, the second was for the case of observed precipitation and no precipitation generated by WGEN, and the third was for the case of no observed precipitation and precipitation generated by WGEN. For the first case, adjustments were made by multiplying the observed precipitation values by the ratio of daily WGEN precipitation to the daily sum of observed precipitation, case two required setting the values of observed precipitation to zero, and the third case divided daily WGEN precipitation by four, creating a 2 h precipitation event that occurred at a randomly chosen time of the day. Despite the different precipitation rates and times of occurrence, all six 1 month meteorological sets received the same amount of monthly precipitation. Fractions of the grid area received the same meteorology (i.e., at each site, all six sets of meteorology occurred at the same time, but over different fractions of the grid area), and to avoid subgrid-scale variability within subgrid-scale variability within SiB3, only the large-scale precipitation distribution option was used for the explicit (see section 3.3) and

sampling (see section 3.5) methods, while in the bulk method (see section 3.4) the standard exponential distribution (see sections 1 and 3.2), where rainfall intensity determines the wet fraction of the grid area, was used.

3.2. Diagnostic Variable

[10] The diagnostic variable used for applying the methods of *Sellers et al*. [2007] is now presented. Total plant available water (PawTot, kg/m^2) is a SiB3 variable that is dependent on the saturation of the soil column and is used to diagnose the water stress factor for the vegetation canopy. As the soil column saturation changes over time, this is reflected in PawTot, thus, it is the diagnostic variable used in this study. For the calculation of PawTot in SiB3, we begin with the surface hydrology. To account for the hydrologic effects of small-scale and large-scale precipitation, two precipitation distributions are used in SiB3. Rainfall is distributed according to equation (2), where $I_{c,p(X)}$ is the relative amount of small-scale (c) or large-scale (p)

$$I_{c,p(X)} = a_{c,p} e^{-b_{c,p} X} + c_{c,p} \tag{2}$$

precipitation as a function of the fractional area (x) and the small-scale or large-scale constants a_c, b_c, c_c, a_p, b_p, and c_p [*Sato et al*., 1989b]. Since both types of precipitation can occur together, the total amount of precipitation is given in equation (3), where P is the total

$$PI_{(X)} = (P_c a_c + P_p a_p) e^{-bX} + (P_c c_c + P_p c_p) \tag{3}$$

precipitation during a time step and P_c and P_p are small-scale and large-scale precipitation rates (mm). Therefore, with this distribution scheme, the large-scale precipitation rate is distributed evenly across the grid area, but the intensity of the small-scale precipitation rate determines what fraction of the grid area receives precipitation (area gets smaller as the rainfall rate intensity increases). As mentioned in section 1, once precipitation reaches the land surface, it is evenly distributed within the entire soil column, making this scheme unrealistic for the average small-scale precipitation rate, but satisfactory for the average large-scale precipitation rate. The dynamics of precipitation intercepted by the vegetation canopy is described in equation (4),

$$\frac{\partial M_c}{\partial t} = PI_{(x)} - D \tag{4}$$

where M_c is the amount of canopy interception, $PI_{(x)}$ is the precipitation and D is the canopy drainage rate [*Sato et al*., 1989b; *Sellers et al*., 1996a]. Defining evaporative fluxes [see *Sellers et al*., 1996a], equation (5) describes the evaporation (λE_{ci}) rate from the wet

$$\lambda E_{ci} = \frac{(e^*(T_c) - e_a)}{r_b} \frac{(\rho H2O) cp}{\gamma} \tag{5}$$

portion of the vegetation canopy, where $e^*(T_c)$ is the saturation vapor pressure at the canopy temperature, e_a is the canopy air space (CAS) vapor pressure, r_b is the bulk canopy boundary layer resistance, ρH_2O is the density of water, cp is the specific heat of air, and γ is the psychrometric constant. The surface interception storage is calculated as the sum of the inputs (D) and the outputs (runoff, infiltration, evaporation). Evaporation from surface interception (λE_{gi}) is calculated using equation (6), where T_g is the ground temperature, r_d is

$$\lambda E_{gi} = \frac{e^*(T_g) - e_a}{r_b} \frac{(\rho H2O) cp}{\gamma} w_g \tag{6}$$

the aerodynamic resistance between the ground and CAS, and w_g is the ground wetness fraction. Overland flow is generated for the fraction of the grid where the residual rainfall rate exceeds the locally derived soil hydraulic conductivity. Evaporation from the topsoil layer (λE_{gs}) is calculated using equation (7), where h_{soil} is the relative humidity of the soil pore space and r_{soil} is the soil surface resistance.

$$\lambda E_{gs} = \frac{h_{soil} e^*(T_g) - e_a}{r_{soil} + r_d} \frac{(\rho H2O) cp}{\gamma} (1 - w_g) \tag{7}$$

[11] Moving into the subsurface, plant available water (Paw, volumetric) is calculated for soil layers ($i = 1, I$).

$$Paw_i = \left[\frac{W_i}{(dz_i)(\rho H2O)} \right] - wp \tag{8}$$

containing roots using equation wp is the wilting point (volumetric), W_i is the soil water mass per unit area (kg/m^2) for the ith soil layer, and dz_i is the thickness of the ith soil layer (m). Water is removed from each soil layer based on the root fraction in that layer and vegetation stress increases rapidly as soil moisture drops to the wp. This response is realistic in plants, and with this approach in SiB3, we see a gradual reaction in soil moisture stress as the soil moisture decreases [*Baker et al*., 2008; *Harper et al*., 2010]. Equation (9) is used to govern the evolution of soil water mass per unit area

$$\frac{\partial W_i}{\partial t} = -\frac{\partial q}{\partial z} - (f_{root} \lambda E_{ct}) + M_{il} \tag{9}$$

[*Sellers et al*., 1996a; *Dai et al*., 2003]. For the ith soil layer, W_i is the soil water mass (kg/m^2), f_{root} is the root fraction, M_{il} is the mass rate of melting (+) or freezing (−) of soil ice, and q is the water flow. Transpiration (λE_{ct}) is calculated using equation (10)

$$\lambda E_{ct} = \frac{e^*(T_c) - e_a}{(1/g_c) + 2r_b} \frac{(\rho H2O) cp}{\gamma} (1 - w_c) \tag{10}$$

where $e^*(T_c)$ is the saturation vapor pressure at the canopy temperature, e_a is the canopy air space (CAS) vapor pressure, cp is the specific heat of air, w_c is the canopy wet fraction, g_c is the canopy conductance, r_b is the bulk canopy boundary layer resistance, and γ is the

psychrometric constant. Darcy's law expresses the vertical flow within soil layers (equation (11)). K is the hydraulic

$$q = -K\left(\frac{\partial \psi}{\partial z} - 1\right) \tag{11}$$

conductivity, ψ is the soil negative potential, and the $+1$ term accounts for gravitational drainage of water out of the bottom of the soil column to create base flow is simply the hydraulic conductivity of layer 10 (equation (12)), and any excess water from saturated soil

$$q_{10} = K_{10} \tag{12}$$

layers is also added to base flow. There is no horizontal exchange of hydrologic fluxes between grid cells in SiB3, and once a grid cell soil column has saturated, it is assumed that runoff flows directly into the ocean. PawTot is calculated using equation (13). The fraction of PawTot (PawFrac, unitless) that is available to plants

$$PawTot = \sum_{i=1}^{I} [(Paw_i)(dz_i)(\rho H2O)] \tag{13}$$

is calculated as the ratio of PawTot to the maximum amount of PawTot (PawMax, kg/m^2) using equations (14) and (15), where fc is the field capacity (volumetric).

$$PawMax = \sum_{i=1}^{I} [(fc - wp)(dz_i)(\rho H2O)] \tag{14}$$

$$PawFrac = \frac{PawTot}{PawMax} \tag{15}$$

Finally, soil moisture deficit below fc for each layer is aggregated and total column water stress or the water stress factor is calculated using equation (16), where S is the shape parameter, which currently has a value of

$$Water\_Stress\_Factor = \left[\frac{(1+S)(PawFrac)}{(S+PawFrac)}\right] \tag{16}$$

0.2 [*Baker et al.*, 2008]. For a completely dry soil column up to wp, the water stress factor has a value of 0.1, and from fc to a completely saturated soil column, the value is 1.0. Water stress is maximized when the value approaches 0.1 and there is no stress at and beyond fc. The idea for this study is to increase the sampling along the highly nonlinear curve of the water stress factor plotted as a function of PawTot (see Figure 1) when calculating grid area evapotranspiration rates and soil wetness to avoid unrealistic jumps during precipitation events at little additional computational cost to the bulk method (see section 3.4).

3.3. Explicit Method

[12] Following the methods of *Sellers et al.* [2007], an arbitrary grid area normalized to an area of unity (1.0) was divided into 100 cells of equal area and randomly initialized with a Gaussian PawTot distribution (Figure 2). In the initial distribution, with all values lying on the most nonlinear part of the curve representing the water stress factor as a function of PawTot (see Figure 1), each cell was represented by an instance of SiB3 and soil column saturation representative of the PawTot value assigned to that cell (it is important to note that soil parameters such as fc, wp, etc. were different at each site, see section 3.1). At each location, each of the six sets of July meteorology (see section 3.1) was randomly assigned a fractional area. Daily precipitation events from all six sets were forced to occur on randomly selected cells following *Hahmann* [2003], until the preassigned fractional area for every set was met (recall, the conventional exponential distribution relating the precipitation rate to the grid area wet fraction was not used with this method to avoid subgrid-scale variability within subgrid-scale variability). Every fractional area had its own "atmosphere" following the approach of *White et al.* [1997], but in this study, all meteorological forcings (not just precipitation as in *White et al.* [1997]) were different. To illustrate this, let us say that on day two, precipitation occurred with three of the six sets of meteorology. The three wet instances would first be distributed randomly across the grid area until the preassigned fractional areas for each set were met, and then the remaining three instances of meteorology (dry) would be randomly assigned among the unoccupied cells until those preassigned fractional areas were met.

[13] Every instance of SiB3 was forced with its assigned meteorology until the next daily precipitation event was observed and the random distribution was repeated. To complete the 2 month (62 days) duration of the study each set of meteorology was used to force 2 months (preassigned fractional areas remained the same for the second month, but each set was still randomly distributed across the grid area) and for every time step (10 min time step with hourly output) 100 instances of SiB3 took place. Grid averages were calculated by the integration of fluxes across all instances of SiB3, and individual cell fluxes and runoff did not interact with neighboring cells and were assumed to go directly into the atmosphere and ocean, respectively. As in *Sellers*

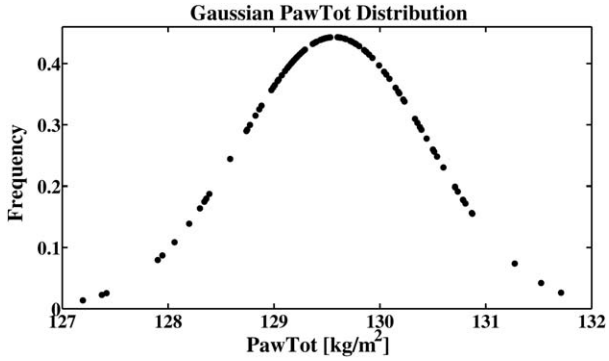

Figure 2. Initial Gaussian distribution of PawTot centered on a mean value of 129.57 kg/m^2 for the semiarid site.

et al. [2007], the explicit method was used as the standard for judging the other methods as it explicitly resolved subgrid-scale variability. Clearly, this method is computationally expensive and is not practical for global GCM simulations.

3.4. Bulk Method

[14] The initial PawTot distribution used in the explicit method was area averaged to give a single PawTot value for the entire grid area (equation (17)).

$$\langle PawTot \rangle = \int_A (PawTot)\,da \tag{17}$$

The grid area was defined by a single instance of SiB3 with a soil column saturation representative of the area-averaged PawTot value, and was forced with a single meteorological data set compiled from the area-averaged meteorology from the six meteorological data sets used in the explicit method. The conventional exponential distribution relating the precipitation rate to the wet fraction of the grid area was used with this method. As mentioned in section 1, once precipitation reaches the land surface, it is evenly distributed within the entire soil column, making this scheme unrealistic for the average small-scale precipitation rate, but satisfactory for the average large-scale precipitation rate. This method only required a single model run for every time step and is computationally inexpensive.

3.5. Sampling Method

[15] The alternative binned method of *Sellers et al*. [2007] or sampling method is a modified version of the explicit method. The initial PawTot distribution was sorted and averaged based on the number of bins (j) that were used. Using 10 bins, every 10 values of the sorted (ascending order) PawTot distribution were averaged giving 10 PawTot values representative of the driest bin up to the wettest bin. The grid area was then divided into 10 equal fractional areas (a_j) and each fractional area was represented by an instance of SiB3 and soil column saturation representative of its assigned PawTot value. Daily meteorological forcings were distributed based on the percentage of the grid area that received precipitation in the explicit method, where the number of bins being used multiplied that percentage and that value rounded to the nearest integer equaled the number of bins that were randomly chosen to receive precipitation. Precipitation rates were based on the total fractional area occupied by the wet bins and the area average precipitation, and all other meteorological forcings were an average of the wet

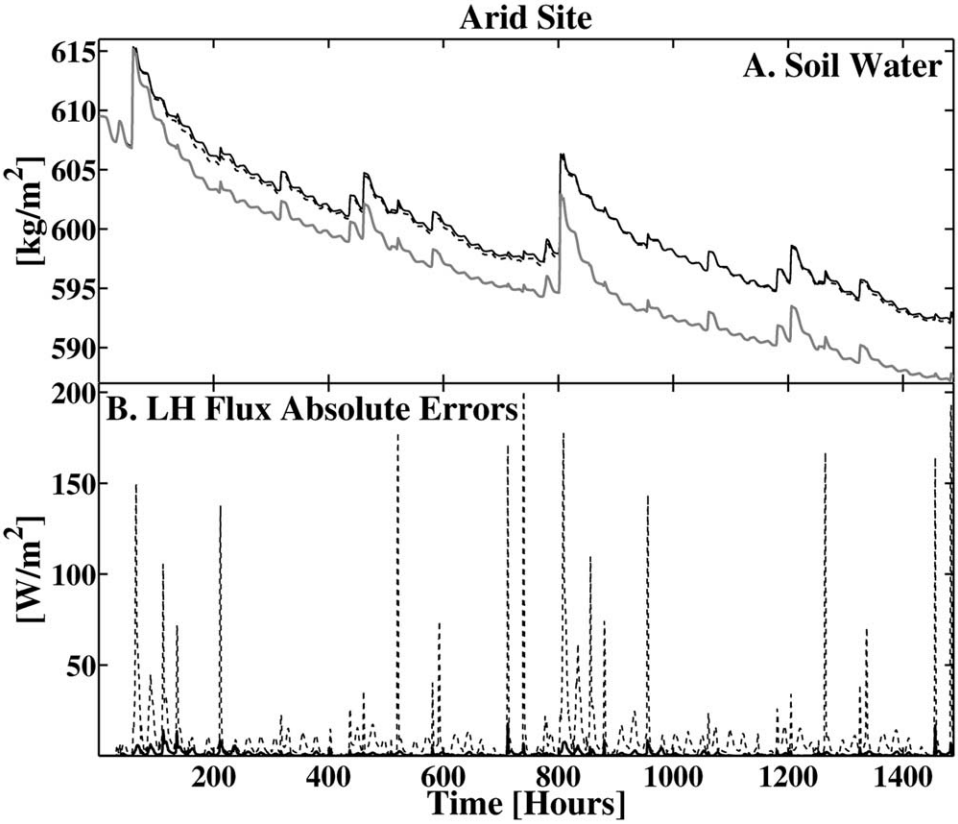

Figure 3. (a) A time series of column integrated soil water mass for the explicit (solid black line), bulk (solid gray line), and sampling (10 bins; dashed black line) methods. Peaks correspond to precipitation events and meandering peaks and troughs represent the diurnal cycle. (b) A time series of absolute errors for the grid area LH flux for the bulk (dashed gray line) and sampling (10 bins; solid black line) methods.

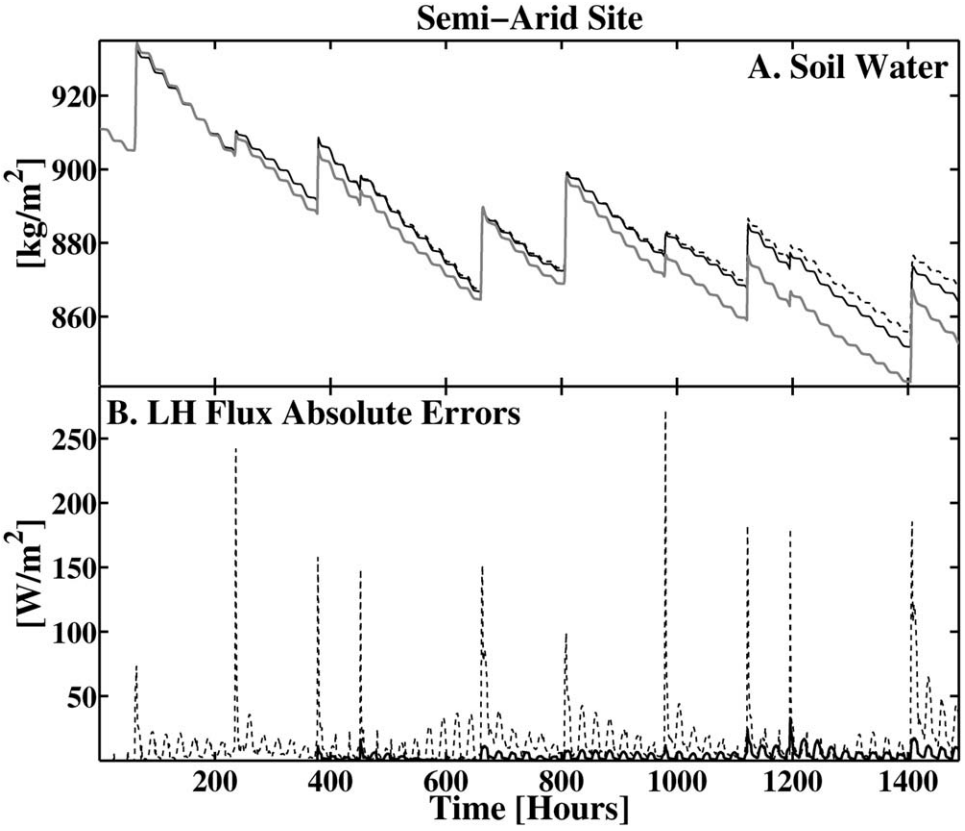

Figure 4. (a) A time series of column integrated soil water mass for the explicit (solid black line), bulk (solid gray line), and sampling (10 bins; dashed black line) methods. Peaks correspond to precipitation events and meandering peaks and troughs represent the diurnal cycle. (b) A time series of absolute errors for the grid area LH flux for the bulk (dashed gray line) and sampling (10 bins; solid black line) methods.

meteorological sets used for the wet areas in the explicit method. The remaining bins received meteorological forcings that were an average of the dry meteorological sets used for the dry areas in the explicit method. Grid values were calculated using equation (18).

$$Grid\_Value = \sum_{j=1}^{J} \left[(SiB)_j a_j \right] \tag{18}$$

4. Results

4.1. Column Integrated Soil Water and Stress

[16] In this study, PawTot was the diagnostic variable used for applying the different methods (see section 3.2), and was chosen because it is a good representation of the saturation of the entire soil column. Therefore, here, we present time series of the grid-averaged column integrated soil water mass to show that applying these methods to PawTot is appropriate. Furthermore, time series of the grid-averaged PawTot are similar to the column integrated soil water mass. Plot a of Figures 3–5 is time series of grid-averaged column integrated soil water mass for the arid, semiarid, and wet site, respectively. The sampling method performed the best at all sites, and with an increasing bin count (10, 20, and 50), time series converged to the explicit method. It was concluded that with the use of 10 bins, the spatial heterogeneity similar to the one simulated with the explicit method was sufficiently represented. For all sites, the bulk method showed a drier time series relative to the explicit method as a result of area-averaged meteorology. Even though water was evenly distributed over the grid area soil column in this method (the other methods were comprised of multiple columns), very light rainfall rates and very high canopy interception rates dominated and resulted in a drier time series. For the calculation of grid area water stress factor (time series not shown here because the plots are too erratic) the single PawTot value in the bulk method calculation at every time step caused erratic behavior in the calculation of the water stress factor as changing PawTot values moved up and down the highly nonlinear curve in Figure 1. For the other methods, increased sampling along the nonlinear curve with multiple instances of SiB3 eliminated that behavior, and it was concluded that the dry bins contributed more to the grid average water stress factor than the wet bins. From further analysis, the change to the initial Gaussian soil moisture distribution as documented in *Sellers et al.* [2007], was also documented here. During dry periods, the initial

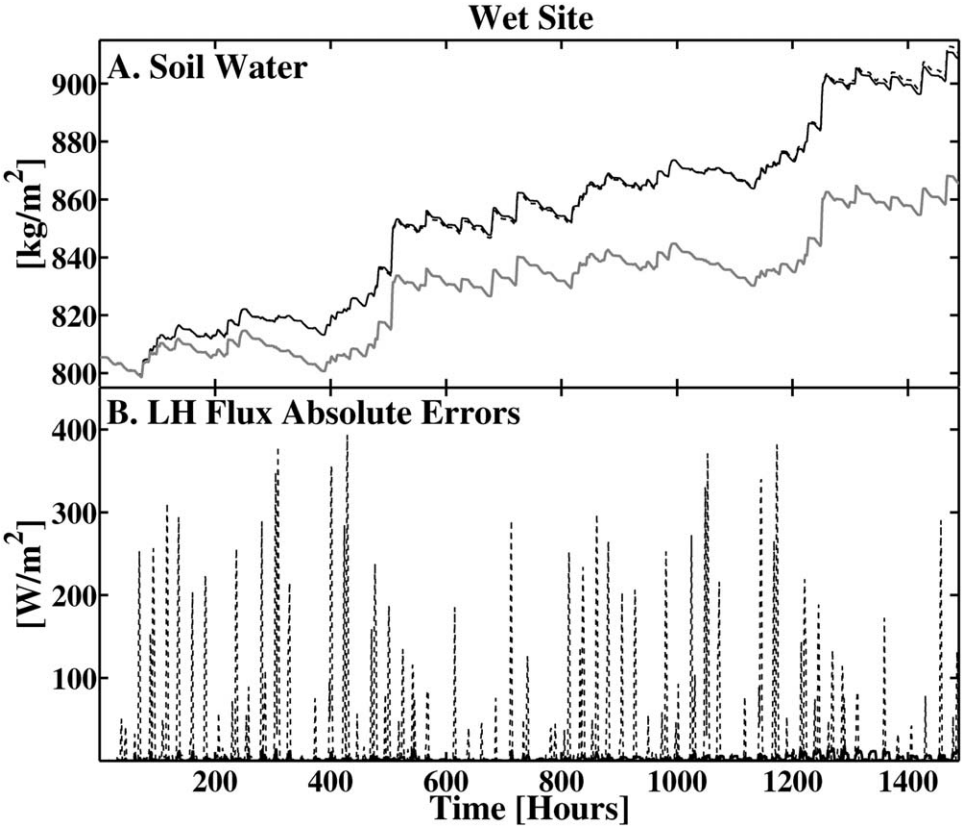

Figure 5. (a) A time series of column integrated soil water mass for the explicit (solid black line), bulk (solid gray line), and sampling (10 bins; dashed black line) methods. Peaks correspond to precipitation events and meandering peaks and troughs represent the diurnal cycle. (b) A time series of absolute errors for the grid area LH flux for the bulk (dashed gray line) and sampling (10 bins; solid black line) methods.

PawTot distribution for the explicit method dried down uniformly (distribution moved from right to left, see Figure 2) and during precipitation events, the distribution developed wet peaks (wet fractions of the grid area) to the right of the remainder of the distribution that stayed dry, and as the wet peaks dried down, they moved to the left and eventually converged with the rest of the distribution. Relative to the bulk method, the sampling method better captured this progression.

4.2. Latent Heat (LH) and Sensible Heat (SH) Fluxes

[17] The bulk method grid-averaged LH flux was always the highest relative to all methods, with the greatest contributions coming from the canopy intercepted LH flux as more precipitation was intercepted with this method, and the ground LH flux, where water that did reach the surface contributed more to the total LH flux as water was evenly distributed over the entire grid rather than in a fraction of the grid as in the other methods. The grid-averaged LH flux for the sampling method was very close to the explicit method for all bin counts, and this was not surprising because during precipitation events the total area that received precipitation was much closer to the explicit method. Absolute errors (defined as the absolute value of the difference between an explicit method grid value and the equivalent grid value from the other methods [*Sellers et al.*, 2007]) for the bulk and sampling methods for the grid-

Table 1. Time Integrated Hydrologic Flux Percentage Errors[a]

	Arid Site Bulk Method (%)	Arid Site Sampling Method (%)	Semiarid Site Bulk Method (%)	Semiarid Site Sampling Method (%)	Wet Site Bulk Method (%)	Wet Site Sampling Method (%)
Total LH flux	15	-	12	−3	28	−2
Canopy intercepted LH flux	179	10	215	−2	273	5
Ground LH flux	6	-	45	−3	4	−4
Transpiration	2	-	−8	−3	−7	–
Runoff	-	-	−99	23	−98	24

[a]Time integrated hydrologic flux percentage errors for all methods compared to the explicit method. Errors below 1% are indicated by a dash, and negative and positive values represent an underestimation and overestimation, respectively.

averaged LH flux for all sites are presented in plot b of Figures 3–5. Table 1 shows time integrated percentage errors for all hydrologic fluxes, where all methods are compared to the explicit method. Errors below 1% are indicated by a dash, and negative and positive values represent an underestimation and overestimation, respectively. For the SH flux, the bulk method always had the lowest values relative to all methods, while the grid-averaged SH flux for the sampling method was very close to the explicit method for all bin counts.

4.3. Runoff, Canopy Air Space (CAS) Temperature, and Relative Humidity (RH)

[18] In the bulk method, time integrated total runoff was lower than the explicit method with the exception of the arid site, where runoff was about the same for the bulk, explicit, and sampling method due to very light rainfall rates in the experiment, which resulted in very little total runoff. Relative to the other methods, the bulk method had the lowest CAS temperature as well as the highest CAS relative humidity as a result of having the largest LH flux. During precipitation events, in the bulk method all components that contributed to the grid area total LH flux came from the entire grid area rather than a fraction of the grid as in the other methods, where the total LH flux was a grid average of the wet and dry fractions. Additionally, in the sampling method the CAS temperature and CAS relative humidity were not too different when compared to the explicit method, as a result of similar meteorological and soil moisture distributions.

5. Concluding Remarks and Future Work

5.1. Concluding Remarks

[19] To improve the representation of soil moisture heterogeneity and spatially varying precipitation, the methods of *Sellers et al*. [2007] were applied to the land surface model, SiB3. By applying the methods to PawTot, a SiB3 variable representative of the soil column saturation and used to diagnose the water stress factor for the vegetation canopy, the representation of the spatially varying soil column saturation as well as surface fluxes for the bulk and sampling methods were compared to the explicit method for a 2 month period. Precipitation was randomly distributed over the grid area for the explicit and sampling method, while the exponential distribution currently used in SiB3 was used for the bulk method.

[20] When compared to the explicit method, the bulk method was dominated by canopy interception due to lighter rainfall rates and had a much higher LH flux. The greatest contributions to the LH flux came from the canopy interception as well as the ground. Even though the time series for the column integrated soil water and PawTot were drier than in the explicit method, the water that did reach the soil surface during precipitation events allowed for a higher ground LH flux since the flux came from the entire grid area rather than a fraction of the grid as in the explicit method. Additionally, as a result of the high LH flux, the CAS temperature was lower and the CAS relative humidity was higher than in the explicit method.

[21] For the duration of the study in the sampling method, the total fractional area occupied by the wet bins was very close to the area occupied in the explicit method. This resulted in similar rainfall rates, total runoff, canopy interception, and surface fluxes, and suggests that the sampling method better captures the spatial heterogeneity in the distribution of PawTot and the total column saturation produced by the explicit method. Choosing the sampling method over the bulk method, the grid area transitions from being dominated by canopy interception, and having a cool and humid CAS to a grid area dominated by a warmer and less humid CAS. Of course, using a large bin count will produce results that are in closer agreement with the explicit method, but it was concluded that using a bin count of 10 is an improvement from the bulk method and better captures subgrid-scale heterogeneity at little additional computational cost. To illustrate this, the wall clock time for each method was averaged across all sites and was normalized to the bulk method giving values of 1, 6, and 51. This means that on average, the explicit method was about 50 times more expensive computationally while the sampling method (using 10 bins) was only about 5. It is important to note that simulations were performed using an iMac system with a single processor and values may vary across different computing systems.

5.2. Future Work

[22] We plan on applying the methods from this research to other sites to evaluate the performance there, and use a number of different distributions of PawTot for initialization. Other experiments have been suggested and involve saturating certain soil layers and perhaps mixing biomes within the grid area. Finally, we plan on performing online runs using a single column model (SCM) as well as running these methods within a GCM.

[23] **Acknowledgments.** This work is supported by the Integrated Water Atmosphere Ecosystems Education and Research (IWATER) NSF IGERT Program, managed by Colorado State University grant 0966346. AmeriFlux observations in Oklahoma (US-ARM) are supported by the Office of Biological and Environmental Research of the US Department of Energy under contract DE-AC02–05CH11231 as part of the Atmospheric Radiation Measurement Program and the Atmospheric System Research Program. We are also very grateful to the contributions made by anonymous reviewers that assisted in making this paper clearer.

References

Arora, V. K., F. H. Chiew, and R. B. Grayson (2001), Effect of subgrid-scale variability of soil moisture and precipitation intensity on surface runoff and streamflow, *J. Geophys. Res.*, *106*(D15), 17,073–17,091, doi:10.1029/2001JD900037.

Avissar, R. (1992), Conceptual aspects of a statistical-dynamical approach to represent landscape subgrid-scale heterogeneities in atmospheric models, *J. Geophys. Res.*, *97*(D3), 2729–2742, doi:10.1029/91JD01751.

Baker, I. T., A. S. Denning, N. Hanan, L. Prihodko, P.-L. Vidale, K. Davis, and P. Bakwin (2003), Simulated and observed fluxes of sensible and latent heat and CO_2 at the WLEF TV tower using SiB2.5, *Global Change Biol.*, *9*, 1262–1277, doi:10.1046/j.13652486.2003.00671.x.

Baker, I. T., L. Prihodko, A. S. Denning, M. Goulden, S. Miller, and H. R. da Rocha (2008), Seasonal drought stress in the Amazon:

Reconciling models and observations, *J. Geophys. Res.*, *113*, G00B01, doi:10.1029/2007JG000644.

Collatz, G. J., J. T. Ball, C. Grivet, and J. A. Berry (1991), Physiological and environmental regulation of stomatal conductance, photosynthesis, and transpiration: A model that includes a laminar boundary layer, *Agric. For. Meteorol.*, *54*, 107–136, doi:10.1016/0168-1923(91)90002-8.

Collatz, G. J., M. Ribas-Carbo, and J. A. Berry (1992), Coupled photosynthesis-stomatal conductance model for leaves of C4 plants, *Aust. J. Plant Physiol.*, *19*, 519–538, doi:10.1071/PP9920519.

Dai, Y., et al. (2003), The common land model, *Bull. Am. Meteorol. Soc.*, *84*, 1013–1023, doi:10.1175/BAMS-84-8-1013.

De Pury, D. G. G., and G. D. Farquhar (1997), Simple scaling of photosynthesis from leaves to canopies without the errors of big-leaf models, *Plant Cell Environ.*, *20*, 537–557, doi:10.1111/j.1365–3040.1997.00094.x.

Denning, A. S., J. G. Collatz, C. Zhang, D. A. Randall, J. A. Berry, P. J. Sellers, G. D. Colello, and D. A. Dazlich (1996), Simulations of terrestrial carbon metabolism and atmospheric CO_2 in a general circulation model. Part 1: Surface carbon fluxes, *Tellus, Ser. B*, *48*, 521–542, doi:10.1034/j.1600-0889.1996.t01-1-00010.x.

Dickinson, R. E. (1989), Implications of tropical deforestation for climate: A comparison of model and observational descriptions of surface energy and hydrological balance, *Philos. Trans. R. Soc. London B*, *324*, 423–430, doi:10.1098/rstb.1989.0056.

Dickinson, R. E., and A. Henderson-Sellers (1988), Modeling tropical deforestation: A study of GCM land surface parameterizations, *Q. J. R. Meteorol. Soc.*, *144*, 439–462, doi:10.1002/qj.49711448009.

Essery, R. L. H., M. J. Best, R. A. Betts, P. M. Cox, and C. M. Taylor (2003), Explicit representation of sub-grid heterogeneity in a GCM land-surface scheme, *J. Hydrometeorol.*, *4*, 530–543, doi:10.1175/1525-7541(2003)004<0530:EROSHI>2.0.CO;2.

Farquhar, G. D., S. V. Caemmerer, and J. A. Berry (1980), A biochemical model of photosynthetic CO_2 assimilation in leaves of C3 species, *Planta*, *149*, 78–90, doi:10.1007/BF00386231.

Gao, X., and S. Sorooshian (1994), A stochastic precipitation disaggregation scheme for GCM Applications, *J. Clim.*, *7*, 238–247, doi:10.1175/1520-442(1994)007<0238:ASPDSF>2.0.CO;2.

Giorgi, F., and R. Avissar (1997), Representation of heterogeneity effects in Earth system modeling: Experience from land surface modeling, *Rev. Geophys.*, *35*(4), 413–437, doi:10.1029/97RG01754.

Guo, Z., et al. (2006), GLACE: The global land-atmosphere coupling experiment. Part II: Analysis, *J. Hydrometeorol.*, *7*, 611–625, doi:10.1175/JHM511.1.

Haan, C. T. (1977), Statistical Methods in Hydrology, Iowa State Univ. Press, Ames, Iowa.

Hahmann, A. N. (2003), Representing spatial subgrid-scale precipitation variability in a GCM, *J. Hydrometeorol.*, *4*, 891–900, doi:10.1175/1525–7541(2003)004<0891:RSSPVI>2.0.CO;2.

Hanan, N. P., J. A. Berry, S. B. Verma, E. A. Walter-Shea, A. E. Suyker, G. G. Burba, and A. S. Denning (2004), Model analyses of biosphere-atmosphere exchanges of CO_2, water and energy in Great Plains tall grass prairie and wheat ecosystems, *Agric. For. Meteorol.*, *131*, 162–179.

Harper, A. B., A. S. Denning, I. T. Baker, M. D. Branson, L. Prihodko, and D. A. Randall (2010), Role of deep soil moisture in modulating climate in the Amazon rainforest, *Geophys. Res. Lett.*, *37*, L05802, doi:10.1029/2009GL042302.

Koster, R. D., M. J. Suarez, and M. Heiser (2000), Variance and predictability of precipitation at seasonal-to-interannual timescales, *J. Hydrometeorol.*, *1*, 26–46, doi:10.1175/1525–7541(2000)001<0026:VAPOPA>2.0.CO;2.

Koster, R. D., P. A. Dirmeyer, A. N. Hahmann, R. Ijpelaar, L. Tyahla, P. Cox, and M. J. Suarez (2002), Comparing the degree of land-atmosphere interaction in four atmospheric general circulation models, *J. Hydrometeorol.*, *3*, 363–375, doi:10.1175/1525–7541(2002)003<0363:CTDOLA>2.0.CO;2.

Koster, R. D., M. J. Suarez, R. W. Higgins, and H. M. Van den Dool (2003), Observational evidence that soil moisture variations affect precipitation, *Geophys. Res. Lett.*, *30*(5), 1241, doi:10.1029/2002GL016571.

Koster, R. D., et al. (2004), Regions of strong coupling between soil moisture and precipitation, *Science*, *305*, 1138–1140, doi:10.1126/science.1100217.

Koster, R. D., et al. (2006), GLACE: The global land-atmosphere coupling experiment. Part I: Overview, *J. Hydrometeorol.*, *7*, 590–610, doi:10.1175/JHM510.1.

Lokupitiya, E., A. S. Denning, K. Paustian, I. T. Baker, K. Schaefer, S. Verma, T. Meyers, C. Bernacchi, A. Suyker, and M. Fischer (2009), Incorporation of crop phenology in Simple Biosphere Model (SiBcrop) to improve land-atmosphere carbon exchanges from croplands, *Biogeosciences*, *6*, 969–986, doi:10.5194/bg-6-969-2009.

Matalas, N. C. (1967), Mathematical assessment of synthetic hydrology, *Water Resour. Res.*, *3*(4), 937–945, doi:10.1029/WR003i004p00937.

Onof, C., N. G. Mackay, L. Oh, and H. S. Wheater (1998), An improved rainfall disaggregation technique for GCMs, *J. Geophys. Res.*, *103*(D16), 19,577–19,586, doi:10.1029/98JD01147.

Pitman, A., A. Henderson-Sellers, and Z. L. Yang (1990), Sensitivity of regional climates to localized precipitation in global models, *Nature*, *346*, 734–737, doi:10.1038/346734a0.

Randall, D. A., et al. (1996), A revised land surface parameterization (SiB2) for GCMs. Part III: The greening of the Colorado State University general circulation model, *J. Clim.*, *9*, 738–763, doi:10.1175/1520-0442(1996)009<0738:ARLSPF>2.0.CO;2.

Richardson, C. W. (1981), Stochastic simulation of daily precipitation, temperature and solar radiation, *Water Resour. Res.*, *17*(1), 182–190, doi:10.1029/WR017i001p00182.

Richardson, C. W., and D. A. Wright (1984), WGEN: A Model for Generating Daily Weather Variables (ARS-8), U.S. Dep. of Agric, copies of this publication are available from the National Technical Information Service (NTIS), 5285 Port Royal Road, Springfield, VA 22161.

Ronda, R. J., B. J. J. M. van den Hurk, and A. A. M. Holtslag (2002), Spatial heterogeneity of the soil moisture content and its impact on surface flux densities and near-surface meteorology, *J. Hydrometeor*, *3*, 556–570, doi:10.1175/1525-7541(2002)003<0556:SHOTSM>2.0.CO;2.

Ryu, D., and J. S. Famiglietti (2006), Multi-scale spatial correlation and scaling behavior of surface soil moisture, *Geophys. Res. Lett.*, *33*, L08404, doi:10.1029/2006GL025831.

Sato, N., P. J. Sellers, D. A. Randall, E. K. Schneider, J. Shukla, J. L. Kinter III, Y.-T. Hou, and E. Albertazzi (1989b), Implementing the Simple Biosphere Model (SiB) in a general circulation model: Methodologies and results, NASA Contractor Report 185509, 70 pp., NASA HQ, Washington, D. C.

Schaefer, K., P. Tans, A. S. Denning, I. Baker, J. Berry, L. Prihodko, N. Suits, and A. Philpott, (2008), The combined Simple Biosphere/Carnegie-Ames-Stanford Approach (SiBCASA) model, *J. Geophys. Res.*, *113*, G03034, doi:10.1029/2007JG000603.

Schär, C., D. Lüthi, U. Beyerle, and E. Heise (1999), The soil-precipitation feedback: A process study with a regional climate model, *J. Clim.*, *12*, 722–741, doi:10.1175/1520-0442(1999)012<0722:TSPFAP>2.0.CO;2.

Sellers, P. J., Y. Mintz, Y. C. Sud, and A. Dalcher (1986), A Simple Biosphere Model (SiB) for use within general circulation models, *J. Atmos. Sci.*, *43*, 505–531, doi:10.1175/1520-0469(1986)043<0505:ASBMFU>2.0.CO;2.

Sellers, P. J., D. A. Randall, G. J. Collatz, J. A. Berry, C. B, Field, D. A. Dazlich, C. Zhang, G. D. Collelo, and L. Bounoua (1996a), A revised land surface parameterization (SiB2) for atmospheric GCMs. Part I: Model formulation, *J. Clim.*, *9*, 676–705, doi:10.1175/1520-0442(1996)009<0676:ARLSPF>2.0.CO;2.

Sellers, P. J., S. O. Los, C. J. Tucker, C. O. Justice, D. A. Dazlich, G. J. Collatz, and D. A. Randall (1996b), A revised land surface parameterization (SiB2) for atmospheric GCMs. Part II: The generation of global fields of terrestrial biophysical parameters from satellite data, *J. Clim.*, *9*, 706–737, doi:10.1175/1520-0442(1996)009<0706:ARLSPF>2.0.CO;2.

Sellers, P. J., M. J. Fennessy, and R. E. Dickinson (2007), A numerical approach to calculating soil wetness and evapotranspiration over large grid areas, *J. Geophys. Res.*, *112*, D18106, doi:10.1029/2007JD008781.

Shuttleworth, W. J., and R. E. Dickinson (1989), Comments on modeling tropical deforestation: A study of GCM land-surface parameterizations, *Q. J. R. Meteorol. Soc.*, *115*, 1117–1179, doi:10.1002/qj.49711548910.

Stöckli, R., D. M. Lawrence, G.-Y. Niu, K. W. Oleson, P. E. Thornton, Z.-L. Yang, G. B. Bonan, A. S. Denning, and S. W. Running (2008a), The use of fluxnet in the community land model development, *J. Geophys. Res.*, *113*, G01025, doi:10.1029/2007JG000562.

Stöckli, R., T. Rutishauser, D. Dragoni, P. E. Thornton, L. Lu, and A. S. Denning (2008b), Remote sensing data assimilation for a prognostic phenology model, *J. Geophys. Res.*, *113*, G04021, doi:10.1029/2008JG000781.

Suits, N. S., A. S. Denning, J. A. Berry, C. J. Still, J. Kaduk, and J. B. Miller (2005), Simulation of carbon isotope discrimination of the terrestrial biosphere, *Global Biogeochem. Cycles*, *19*, GB1017, doi:10.1029/2003GB002141.

Tang, Q., T. Oki, S. Kanae, and H. Hu (2007), The influence of precipitation variability and partial irrigation within grid cells on a hydrological simulation, *J. Hydrometeorol.*, *8*, 499–512, doi:10.1175/JHM589.1.

Vidale, P.-L and R. Stöckli (2005), Prognostic canopy air space solutions for land surface exchanges. *Theor. And Appl. Climatol.*, *80*, 245–257, doi:10.1007/s00704-004-0103-2.

White, C. B., P. R. Houser, A. M. Arain, Z. L. Yang, K. Syed, and W. J. Shuttleworth (1997), The aggregate description of semi-arid vegetation with precipitation-generated soil moisture heterogeneity, *Hydrol. Earth Syst. Sci. Discuss.*, *1*, 205–212, doi:10.5194/hess-1-205-1997.