Assessing the impact of model spin-up on surface water-groundwater interactions using an integrated hydrologic model

Hoori Ajami1,2, Matthew F. McCabe3, Jason P. Evans4,5, and Simon Stisen6

1School of Civil and Environmental Engineering, University of New South Wales, Sydney, New South Wales, Australia, 2Connected Waters Initiative Research Centre, University of New South Wales, Sydney, New South Wales, Australia, 3Water Desalination and Reuse Center, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, 4Climate Change Research Centre, University of New South Wales, Sydney, New South Wales, Australia, 5ARC Centre of Excellence for Climate System Science, University of New South Wales, Sydney, New South Wales, Australia, 6Geological Survey of Denmark and Greenland, Copenhagen, Denmark

Abstract
Integrated land surface-groundwater models are valuable tools in simulating the terrestrial hydrologic cycle as a continuous system and exploring the extent of land surface-subsurface interactions from catchment to regional scales. However, the fidelity of model simulations is impacted not only by the vegetation and subsurface parameterizations, but also by the antecedent condition of model state variables, such as the initial soil moisture, depth to groundwater, and ground temperature. In land surface modeling, a given model is often run repeatedly over a single year of forcing data until it reaches an equilibrium state: the point at which there is minimal artificial drift in the model state or prognostic variables (most often the soil moisture). For more complex coupled and integrated systems, where there is an increased computational cost of simulation and the number of variables sensitive to initialization is greater than in traditional uncoupled land surface modeling schemes, the challenge is to minimize the impact of initialization while using the smallest spin-up time possible. In this study, multicriteria analysis was performed to assess the spin-up behavior of the ParFlow.CLM integrated groundwater-surface water-land surface model over a 208 km² subcatchment of the Ringkobing Fjord catchment in Denmark. Various measures of spin-up performance were computed for model state variables such as the soil moisture and groundwater storage, as well as for diagnostic variables such as the latent and sensible heat fluxes. The impacts of initial conditions on surface water-groundwater interactions were then explored. Our analysis illustrates that the determination of an equilibrium state depends strongly on the variable and performance measure used. Choosing an improper initialization of the model can generate simulations that lead to a misinterpretation of land surface-subsurface feedback processes and result in large biases in simulated discharge. Estimated spin-up time from a series of spin-up functions revealed that 20 (or 21) years of simulation were sufficient for the catchment to equilibrate according to at least one criterion at the 0.1% (0.01%) threshold level. Amongst a range of convergence metrics examined, percentage changes in monthly values of groundwater and unsaturated zone storages produced a slow system convergence to equilibrium, whereas criteria based on ground temperature allowed a more rapid spin-up. Slow convergence of unsaturated and saturated zone storages is a result of the dynamic adjustment of the water table in response to a physically arbitrary or inconsistent initialization of a spatially uniform water table. Achieving equilibrium in subsurface storage ensured equilibrium across a spectrum of other variables, hence providing a good measure of system-wide equilibrium. Overall, results highlight the importance of correctly identifying the key variable affecting model equilibrium and also the need to use a multicriteria approach to achieve a rapid and stable model spin-up.

1. Introduction
Partitioning of water and energy budgets at the land surface is controlled by complex interactions between climate, soil, geology, and vegetation types. Integrated groundwater-surface water-land surface models that simulate terrestrial hydrologic processes are valuable tools to understand and characterize catchment functions such as partitioning, storage, and release [Wagener et al., 2007]. These models also enable an
understanding of feedback processes between the land surface and atmosphere and how these can impact upon catchment behavior, particularly in response to changes in climate variability and land cover.

The successful operation of an integrated hydrologic model is impacted by conceptual model design elements that characterize a catchment’s land surface and geological structure, define model parameters, and develop model evaluation measures [Yilmaz et al., 2008]. Likewise, the fidelity of model simulations is impacted not only by the vegetation and subsurface parameterizations, but also by antecedent condition of model state variables, such as the initial soil moisture and ground temperature [Levis et al., 1996; Shrestha and Houser, 2010]. Biases in model states of soil moisture, temperature, and snow can also lead to incorrect partitioning of water and energy balances [Rodell et al., 2004].

In most cases, spatially distributed information describing catchment states (i.e., spatial distribution of the water table and soil moisture) is unavailable. Likewise, the temporal and spatial variability of applied atmospheric forcing is often inconsistent with the hydrodynamic initialization of the catchment, which is usually inferred from limited observations or an initial guess. In order to reduce the impact of initial conditions on consequent simulations, land surface models (LSMs) that simulate the effects of soil and vegetation on exchanges of heat, moisture, and momentum between the land and the atmosphere, are often run repeatedly using a single year of forcing data until they reach an equilibrium state in the water and energy balances [Wood et al., 1998]. The purpose of this “spin-up” process is to reach a model equilibrium state, broadly understood to be the condition at which there is minimal artificial drift in the model state or prognostic variables [Cosgrove et al., 2003]. Although a number of studies have investigated the spin-up behavior of LSMs [Cosgrove et al., 2003; Rodell et al., 2005; Yang et al., 1995], there is neither clear consensus on the definition of an equilibrium state [Yang et al., 1995], the specification of spin-up evaluation criteria, and variables used to define an equilibrium state nor an optimal method for spinning up a LSM [Shrestha and Houser, 2010].

Yang et al. [1995] defined equilibrium state as the state at which a “model’s state at year n + 1 is identical to that at year n.” As it is difficult to obtain identical states between two recursive runs, a number of approaches for arriving at this point have been proposed. For instance, the required spin-up can be determined by identifying the simulation time at which the difference between mean annual latent and sensible heat fluxes (between two recursive runs) is less than 0.1 W m⁻² [Yang et al., 1995]. Alternatively, the Simplified Simple Biosphere (SSiB) modeling group used a 0.1% threshold for changes in mean annual energy fluxes, while the Goddard Institute for Space Studies (GISS) modeling group suggested a threshold value of 0.01% for all the monthly means of skin temperature, latent and sensible heat fluxes, total runoff, and snow water equivalent (SWE) to define equilibrium state [Yang et al., 1995]. For the Variable Infiltration Capacity (VIC) model [Liang et al., 1994], an equilibrium state was assumed to be reached when differences in mean, minimum, and maximum of land surface fluxes in recursive simulations were identical to the 5th significant digit [Yang et al., 1995].

In Phase 2a of the Project for Intercomparison of Land Surface Parameterization Schemes (PILPS) experiment, Chen et al. [1997] defined the equilibrium state to be the time at which the difference in recursive runs of January mean values of surface temperature, latent and sensible heat fluxes and root zone soil moisture were less than 0.01 K, 0.1 W m⁻², and 0.1 mm, respectively. Cosgrove et al. [2003] used percent cutoff-based (PC) time, the e-folding time, and anomaly values to characterize spin-up using variables including the total column soil moisture, root zone soil moisture, deep soil temperature, and evaporation. The PC time indicates the amount of time required until the percentage change in the average monthly values between two recursive runs decreases to a specific threshold. The e-folding time is obtained by fitting an exponential relationship to the autocorrelation function of monthly soil moisture anomalies and also represents temporal characteristics of the LSM spin-up process. Anomalies are computed based on the difference in monthly means from the equilibrium state. Therefore, the accuracy of the e-folding time depends on the definition of the equilibrium state [Cosgrove et al., 2003]. Although total column soil moisture or root zone moisture is often used to define the equilibrium state, Yang et al. [2011] used vertical soil moisture and temperature profiles for point-scale LSM modeling. The equilibrium state was reached when changes in soil moisture and temperature of every soil layer at the end of each spin-up cycle were less than 0.5 mm and 0.5 K, respectively.

Recently, a number of distributed, physically based coupled and integrated hydrologic models have been developed to simulate saturated and unsaturated zone flow processes in combination with land surface
processes. The major functional difference between the two approaches is that in integrated hydrologic models, surface and subsurface flow equations are solved simultaneously. In coupled hydrologic models, groundwater flow, unsaturated zone flow, and overland flow equations are solved separately, and through an iterative scheme, continuity along the interfaces of flow domains are ensured [de Rooij et al., 2013; Kollet and Maxwell, 2006]. Examples of these different schemes include MIKE SHE [Abbott et al., 1986], the Integrated Hydrology Model (InHM) [Vanderkwaak, 1999], HydroGeoSphere [Therrien et al., 2008], and ParFlow [Kollet and Maxwell, 2008].

In order to determine initial conditions in coupled/integrated hydrologic models, various approaches have been implemented based on the particular modeling objectives. In the case of an application of InHM for an event-based rainfall-runoff simulations with no evapotranspiration over a 0.1 km² catchment, initial water table position was determined by performing a drainage experiment from a fully saturated state, where the water table was at the land surface and initial condition was achieved when the best fit between the observed and simulated discharge was obtained [VanderKwaak and Loague, 2001]. To simulate a transient surface and subsurface hydrologic response at a catchment scale using the InHM model, Jones et al. [2008] applied a net rainfall rate to a fully saturated system until the best fit between the observed and simulated discharge was obtained [Jones et al., 2008]. On the other hand, Ebel et al. [2008] used a 1 year warm up period to initialize the InHM model for continuous simulations [Ebel et al., 2008]. Other examples include Goderniaux et al. [2009] who used calibrated steady state initial conditions to perform transient simulation with the HydroGeoSphere model at the catchment scale and Refsgaard [1997] who ran the MIKE SHE model for a period of 9 years (1969–1978) using informed estimates of initial conditions for 1969 to remove the impact of initialization. Recursive simulations were continued until simulated 1978 matched the simulated 1978 head from the previous simulation [Refsgaard, 1997]. In order to initialize the ParFlow hydrologic model when fully coupled to the Common Land Model (CLM; hereafter referred to ParFlow.CLM), Kollet and Maxwell [2008] used 1 year of forcing data for recursive simulations until the difference between starting and ending day of water and energy storages for that year, dropped below a certain threshold. On the other hand, Rihani et al. [2010] used normalized changes in annual storage of below 1% as a spin-up measure in their ParFlow.CLM simulations.

As can be seen, a variety of approaches have been utilized to address this critical modeling element. Despite the importance of spin-up in reducing uncertainty in model state initialization, the spin-up behavior of coupled/integrated hydrologic models has not been fully explored. In the nearest analogous case of LSMs, these typically require multiple years of spin-up cycles until hydraulic and thermal equilibrium is reached.
Recursive simulations are required to reduce bias in integrated states such as soil moisture, temperature, and snow, and improve accuracy in future predictions [Kumar et al., 2006]. In most studies, often only 1 year of forcing data are used in recursive simulations to spin-up a LSM. In the present study, the spin-up scheme of Phase 1 of the PILPS experiment [Pitman et al., 1999] was employed to evaluate the spin-up behavior of ParFlow.CLM at the catchment scale. ParFlow.CLM model simulations were repeated on a single year of forcing until an equilibrium state was obtained. Using these simulations as a basis of investigation, the objectives of this study were to: (1) perform multiple criteria analysis using different spin-up measures and thresholds to define equilibrium state in ParFlow.CLM at the catchment scale; (2) investigate differences in spin-up time across multiple energy and water balance variables; (3) evaluate model performance; and (4) quantify changes in surface water-groundwater interactions as a function of spin-up time.

2. Data and Methodology

2.1. Study Site
The study site is one of the subcatchments of the Ringkobing Fjord catchment, located in western Denmark (Figure 1). The study catchment domain is approximately 208 km$^2$, with agricultural lands covering 78% of the area, mixed with evergreen needle leaf forest on the eastern side. The climate of the region is classified as temperate and characterized by mild winters and relatively cold summers [Jensen and Illangasekare, 2011]. The average annual temperature is 8.2°C and August and January are the warmest and coldest months of the year, respectively [Stisen et al., 2011]. Catchment surface geology is mainly glacial material from the Quaternary period, with a highly permeable coarse sandy soil type [Stisen et al., 2011].

2.2. The ParFlow.CLM Model
ParFlow is a 3-D variably saturated groundwater flow model that solves the three-dimensional Richards’ equation using a cell-centered finite difference scheme in space and an implicit backward Euler scheme in time [Ashby and Falgout, 1996; Jones and Woodward, 2001; Kollet and Maxwell, 2006]. ParFlow has a fully integrated overland flow simulator which solves the kinematic wave equation [Kollet and Maxwell, 2006] and is fully coupled with the Common Land Model (CLM) 3.0 [Dai et al., 2003] to simulate water and energy balances at the land surface [Kollet and Maxwell, 2008; Maxwell and Miller, 2005]. In this coupling, CLM soil column/root zone formulation and runoff generation processes are replaced by ParFlow, and CLM computes soil and canopy evaporation, plant transpiration, ground and sensible heat fluxes, and freeze-thaw
processes. ParFlow.CLM is designed to run on massively parallel computer systems using distributed forcing and land cover types [Kollet et al., 2010].

### 2.3. ParFlow.CLM Model Setup

The ParFlow.CLM model domain was established over a 28 km by 20 km area encompassing the study catchment. Land surface topography was defined using a preprocessed 500 m digital elevation model (DEM) of the catchment and the bottom elevation of the model domain was set to −275 m. The model was configured with a horizontal grid resolution of 500 m and a vertical discretization of 0.5 m, resulting in 56,840 cells in the x, y, and z dimensions, respectively. Overland flow boundary condition was specified at the land surface and no-flow boundary conditions were specified for the bottom boundary and along all lateral boundaries. Subsurface geology was considered to be an isotropic medium with a uniform saturated hydraulic conductivity of 0.3 (m h⁻¹). The remaining subsurface hydraulic parameters were prescribed as follows: porosity (0.39), van Genuchten parameters (α = 1.5 m⁻¹ and n = 2), and relative residual saturation (0.1). Gridded land cover data were used to specify land cover types for every model cell based upon the International Geosphere-Biosphere Programme (IGBP) classification scheme. Hourly forcing data of downward shortwave radiation, temperature, wind speed, atmospheric pressure, and specific humidity were obtained from the Danish Meteorological Institute for 2003. Hourly downward longwave radiation was obtained from the Modern Era Retrospective-Analysis for Research and Applications (MERRA) reanalyses product developed by NASA (https://gmao.gsfc.nasa.gov/merra/), with spatial resolution of 0.5°/0.666° longitude. Hourly precipitation values were estimated by evenly distributing daily precipitation data to hourly intervals, with the catchment average daily precipitation data derived from 2 km krigged precipitation fields produced via rain gauges within and around the catchment. This data have been corrected for undercatch based on wind and temperature data [Allerup et al., 1997; Stisen et al., 2012]. In 2003, the annual total precipitation was 801.6 mm and the annual mean air temperature was 281.3 K. Figure 2 presents meteorological details of the forcing data for the year 2003.

The simulation time step was hourly for both the land surface and subsurface. The model initialization was as follows: depth to water table at 3 m below the land surface; initial ground temperature at 281 K (equal to the mean daily temperature); and initial snow cover was zero. The subsurface initialization of ParFlow.CLM with a predefined depth to water table uses hydrostatic pressure head distribution above the water table to

<table>
<thead>
<tr>
<th>Criteria</th>
<th>GW Storage</th>
<th>Unsaturated Storage</th>
<th>S⁺ CLM</th>
<th>S⁺ Root Zone</th>
<th>Depth to Water Table</th>
<th>Discharge</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1%</td>
<td>0.10%</td>
<td>0.01%</td>
<td>1%</td>
<td>0.10%</td>
<td>0.01%</td>
</tr>
<tr>
<td>Start</td>
<td>2</td>
<td>2⁰⁺</td>
<td>17</td>
<td>10</td>
<td>19</td>
<td>7</td>
</tr>
<tr>
<td>End</td>
<td>2</td>
<td>6</td>
<td>16</td>
<td>9</td>
<td>20</td>
<td>6</td>
</tr>
<tr>
<td>Start-end</td>
<td>2</td>
<td>2⁰⁺</td>
<td>16</td>
<td>11</td>
<td>15⁺</td>
<td>6</td>
</tr>
<tr>
<td>Minimum</td>
<td>2</td>
<td>6</td>
<td>16</td>
<td>10⁻²⁺</td>
<td>19</td>
<td>7</td>
</tr>
<tr>
<td>Maximum</td>
<td>2</td>
<td>7</td>
<td>17</td>
<td>9⁻²⁺</td>
<td>19</td>
<td>2⁻²⁺</td>
</tr>
<tr>
<td>PC time</td>
<td>2</td>
<td>7</td>
<td>19</td>
<td>12</td>
<td>19</td>
<td>7</td>
</tr>
<tr>
<td>Annual</td>
<td>2</td>
<td>7</td>
<td>17</td>
<td>10</td>
<td>10</td>
<td>7</td>
</tr>
</tbody>
</table>

⁻² Relative saturation.
⁻²⁺ This is the first instance that the percentage change falls below a certain threshold. However, percentage change has risen a few times during the simulations.
determine soil moisture distribution in the unsaturated zone. This means that in our setup, all the land surface nodes had a pressure head of $-3$ m at the start of the simulation. Simulations were performed on a high-performance parallel computing cluster with 32 processors. One year of simulation required 1000 service units, with a service unit equivalent to 1 h of time used by one processor (i.e., 1000 physical hours required for 1 year of model simulation).

### 2.4. Spin-Up Evaluation Measures

Several criteria were used to assess the ParFlow.CLM spin-up at daily, monthly, and annual time scales. These criteria were based on changes in catchment averaged values of the state and diagnostic variables.
and integrated measures of catchment response in recursive simulations (Table 1). The integrated measures included discharge, runoff coefficient (ratio of discharge to precipitation), and relationships between energy fluxes and depth to water table. All the spin-up criteria were calculated based on grid cells inside the catchment boundary (Figure 1) to minimize the impact of boundary conditions on catchment-scale fluxes. Since all criteria were not necessarily examined at every time scale, the sections below detail the evaluation metrics employed at each specific simulation resolution.

2.4.1. Daily Time Scale
Five criteria were used to assess the ParFlow.CLM spin-up at the daily time scale. Differences in water and energy variables were calculated between the first days of recursive simulations (start day criterion) and last days of recursive simulations (end day criterion), and also the beginning and end of a given simulation year (start-end criterion). The latter criterion has been used previously to define equilibrium state in ParFlow.CLM [Kollet and Maxwell, 2008]. Changes in daily minimum and maximum values of the variables listed in Table 1 were also computed in recursive simulations to estimate spin-up time.

2.4.2. Monthly and Annual Time Scales
Three criteria based on changes in monthly water and energy fluxes were used to assess spin-up at the monthly time scale including: (1) percent cutoff-based time [Cosgrove et al., 2003]; (2) differences in mean January energy fluxes and ground surface temperature over recursive runs [Chen et al., 1997]; and (3) mean monthly anomalies of state and diagnostic variables presented in Table 1 [Cosgrove et al., 2003]. Percent cutoff-based time determines the amount of time required for yearly changes in monthly averaged model output to decrease to a certain threshold. The percentage change is calculated as follows:

\[
PC = 100 \frac{M_1 - M_2}{M_2}
\]

where \(M_1\) is the monthly mean of a variable from the previous year and \(M_2\) is the monthly average for the current year [de Goncalves et al., 2006]. Monthly anomalies were estimated by assuming the last year of recursive simulations represents the spin-up year [Cosgrove et al., 2003].

To determine the equilibrium state at the annual time scale, percentage changes and absolute differences in mean annual model outputs in recursive simulations were calculated. Based on the absolute differences, the spin-up time is defined as the time at which the difference in mean annual latent and sensible heat fluxes (between two recursive runs) was less than 0.1 W m\(^{-2}\) [Yang et al., 1995].

3. Results

3.1. Multicriteria Analysis of Spin-Up Time Based on State and Diagnostic Variables
To assess if the model has reached equilibrium state according to the spin-up criteria, three predefined threshold levels (1%, 0.1%, and 0.01%) were specified except for the criteria based on the absolute
differences in model outputs [e.g., Yang et al., 1995; Chen et al., 1997]. Based on these thresholds, spin-up time was defined as the number of years that it took until the percentage changes in a given criterion reached the predefined thresholds (Table 2). The results in this section are based on the 20 years of simulations. In some instances, the model failed to reach equilibrium at high threshold levels even after 16 years of recursive simulations, so spin-up functions were developed (section 3.2) for estimating the number of simulations required to ensure equilibrium.

### 3.1.1. Groundwater Storage

Spatially distributed adjustment of the water table in ParFlow.CLM resulted in large changes in depth to water table distribution in the catchment as the subsurface was initialized from a uniform water table distribution (3 m below the land surface). Subsequently, these large adjustments in water table distribution impacted subsurface storages during the spin-up. As simulation time increased, percentage changes in various spin-up measures decreased for all subsurface variables (groundwater storage, unsaturated storage, saturation in the root zone (5 soil layers), and saturation in CLM soil layers (10 layers)). As can be seen from Table 2, changes in variables related to the subsurface storage, with the exception of groundwater storage and start-end criterion for relative saturation in CLM layers, did not reach the 0.01% threshold even after 20 years of simulations for any criteria. Groundwater storage decreased throughout simulations (Figure 3) and equilibrated at the 0.01% level after 17 years of simulations based on daily and annual criteria and after 19 years of simulations based on the monthly PC criterion (Table 2). Increases in storage changes in year 2 caused the largest difference in groundwater storage to occur between the start and end days of year 2 and first days of year 2 and 3 (Figure 3). This variability in groundwater storage was not captured by the start and start-end daily criteria, because percent changes in storage for the first two simulations were below 0.1%.

### 3.1.2. Unsaturated Storage and Depth to Water Table (DTWT)

Unsaturated zone storage and depth to water table were only able to achieve equilibrium at the 1% level, when evaluated against all of the spin-up criteria (Table 2). Unsaturated zone storage reached equilibrium at the 0.1% threshold level based on the start days, end days, and maximum daily criteria. Overall changes in unsaturated zone storage became positive through recursive simulations, causing expansion of the unsaturated zone (Figure 3) and increases in depth to water table in all years except the first, where changes in storage were negative. Negative changes in storage in year 1 resulted in lower storage values in the first day of year 2 compared to year 1 and impacted the start-end and start daily criteria (Figure 3). Similar to the case of groundwater storage, more simulations were required for the model to equilibrate based on the monthly PC values compared to other criteria (Table 2). The slow convergence of unsaturated zone storage to equilibrium was likely a direct result of the spatially distributed adjustment of the water table from a physically arbitrary and nonphysical initialization of a spatially uniform DTWT. As the system equilibrated, the water table responded by approaching the land surface along the channel network and declining along the catchment divides. Spin-up time of the unsaturated zone was further impacted by the dynamic response of the water table to climate forcing and topographically driven flow.

---

**Table 3. Number of Years Required Until the System Reached an Equilibrium State According to Predefined Thresholds Based on Changes in Daily, Monthly, and Annual Criteria of Energy Balance Fluxes and Surface and Bottom Layer Soil Temperature**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1%</td>
<td>0.10%</td>
<td>0.01%</td>
<td>1%</td>
<td>0.10%</td>
<td>0.01%</td>
</tr>
<tr>
<td>Start</td>
<td>3</td>
<td>10</td>
<td>12*</td>
<td>3</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>End</td>
<td>2</td>
<td>6</td>
<td>12*</td>
<td>2</td>
<td>5</td>
<td>10*</td>
</tr>
<tr>
<td>Start-end</td>
<td>2</td>
<td>5</td>
<td>10*</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Minimum</td>
<td>2</td>
<td>2</td>
<td>8*</td>
<td>2</td>
<td>2</td>
<td>4*</td>
</tr>
<tr>
<td>Maximum</td>
<td>2</td>
<td>2</td>
<td>8*</td>
<td>2</td>
<td>2</td>
<td>4*</td>
</tr>
<tr>
<td>PC</td>
<td>3</td>
<td>12*</td>
<td>3*</td>
<td>8</td>
<td>3</td>
<td>4*</td>
</tr>
<tr>
<td>Chen et al. [1997]</td>
<td>3</td>
<td>8*</td>
<td>3*</td>
<td>8</td>
<td>3</td>
<td>4*</td>
</tr>
<tr>
<td>Annual</td>
<td>2</td>
<td>9</td>
<td>16*</td>
<td>6</td>
<td>3*</td>
<td>5*</td>
</tr>
<tr>
<td>Yang et al. [1995]</td>
<td>7</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

*This is the first instance that the percentage change falls below a certain threshold. However, percentage change has risen a few times during the simulations.

bFor these criteria, spin-up time is defined based on differences in monthly or annual mean of respective variables.
3.1.3. Root Zone Storage
The initial root zone relative saturation is also impacted by the subsurface initialization approach. Here ini-
tial DTWT was below the root zone (rooting depth was 2.5 m below the land surface) and hydrostatic pres-
sure distribution above the water table caused initial relative saturation of 0.46 in the root zone. As
simulations in year 1 proceed, precipitation and redistribution of the water table caused higher relative sat-
uration in the first day of year 2 compared to year 1 (Figure 3). In subsequent simulations, saturation in the
root zone decreased due to a decrease in recharge. Similar to results for the unsaturated zone, changes in
root zone saturation did not reach the 0.01% threshold by simulation 20. However, by simulation 15, the
system had equilibrated at the 0.1% level based on all of the criteria (Table 2). Root zone saturation in the
first day of a simulation was higher than the last day of the simulation up to simulation 8. After this point,
the trend reversed. Simulation 8 had the smallest difference between the start and last day of a simulation
(0.02%). Starting from simulation 16, differences between mean annual values of root zone saturation were
zero up to three decimal places, but the percentage change in annual means did not reach the 0.01% threshold. Relative saturation in the CLM zone required the same or higher number of simulations com-
pared to the root zone to equilibrate (Table 2).

3.1.4. Discharge
A similar pattern to that for groundwater storage was observed for evaluation measures based on daily dis-
charge, except that the magnitudes of change were higher due to small discharge values. Unsaturated con-
dition at the gauge due to initialization of DTWT at 3 m resulted in no discharge for 47 days, and
subsequently impacted initial values of start-end and start day criteria (Figure 3). For daily discharge, differ-
ces between the beginning and end of a year reached its minimum after 10 years (0.02%) and then
increased. Equilibrium at the 0.01% level was not reached for any criteria based on discharge; however,
absolute differences were zero up to two decimal places by 18 years of simulations for all the criteria.

3.1.5. Snow Water Equivalent (SWE)
The zeroed initialization of SWE caused the largest difference between the first and last day of first year of
simulations due to an accumulation of SWE as the simulation proceeded (Figure 4). By simulation 7, differ-
cences in start and end days stabilized at 1% due to precipitation and frost, with the smallest difference
observed for simulation 2 (Figure 4). At least 12 years of simulations were required until equilibrium at the
0.01% level were obtained for the remaining daily and annual criteria based on changes in SWE. Including
leap years in the recursive simulations (CLM uses a standard Gregorian calendar) impacted percentage
change values of SWE computed between recursive simulations and caused a cyclic pattern in percentage
changes (Figure 4). A similar pattern was observed for ground surface temperature and energy fluxes.

3.1.6. Ground Surface and Bottom Soil Layer Temperature
In terms of ground surface and bottom layer soil temperature, the impact of initialization using uniform ground
surface temperature disappeared after 2–4 years of recursive simulations, depending on the criteria and
threshold level (Table 3). By the forth simulation, thermal equilibrium was reached at the 0.01% level for all the
criteria (Table 3). Differences in January means fell below 0.01 K after 3 years of simulations based on the Chen
et al. [1997] criterion for the equilibrium state. As expected, the magnitudes of change were much higher for
the soil surface layer compared with the bottom layer in earlier simulations for the start, end day and maximum
daily criteria. Decreases in mean annual values of ground temperature were cyclical and were impacted by
including leap years in the simulations. Adding an extra day in February slightly decreased the mean annual
ground surface temperature compared to nonleap years and subsequently influenced energy fluxes.

3.1.7. Energy Fluxes
For all of the energy flux evaluations, differences between the first and last day of a simulation were higher in
the first year due to initial conditions. As the simulations progressed, differences stabilized to constant values
due to differences in applied forcing between the first and last days of a year (Figure 5). Similar behavior for
the start-end criterion was observed for the ground surface temperature and SWE (Figure 5). Likewise, differ-
cences in start days and last days of recursive simulations were largest between years 1 and 2. Cyclical oscilla-
tions in the start and end criteria were caused by including leap years in simulations, resulting in higher
ground surface temperature, ground heat flux and upward longwave radiation, and lower sensible and latent
heat fluxes for the last day of simulation in leap years compared with nonleap years. For the ground heat flux,
it took larger numbers of simulations until the model equilibrated at the 0.01% threshold level compared to
the other energy fluxes at the daily time scale, except for the maximum daily criterion (Table 3). Based on the
Chen et al. [1997] criterion for energy fluxes, equilibrium was reached after 3 years of recursive simulations when differences in January energy fluxes fell below 0.1 W m\(^{-2}\). Except for the upward longwave radiation, none of the energy fluxes equilibrated at the 0.01% level based on monthly PC values. However, absolute differences in monthly values were less than 0.01 W m\(^{-2}\) after 17 years of simulation for the latent heat and sensible heat fluxes and after 12 years of simulations for the ground heat flux. Similar to the monthly PC values, percentage changes in annual values were impacted by small differences in energy fluxes in recursive simulations. For sensible heat flux, differences reached 0.004 W m\(^{-2}\) between simulation year 19 and 20.

For energy balance variables overall, but particularly for the sensible and ground heat fluxes, it took longer for the model to equilibrate at the monthly PC time scale compared to the annual time scale due to variations in the spin-up time of monthly energy fluxes based on monthly PC criterion. According to all the criteria, the model equilibrated fastest (after 6 years) based on changes in upward longwave radiation.

Figure 5. Estimated percent change between (left) the first and last day of a simulation in a given year, (middle) the first days, and (right) the last days of recursive simulations. Variables examined include (from top-to-bottom) latent heat flux (LH), sensible heat flux (SH), ground heat flux (G), and upward longwave radiation (LW up).
3.2. Development of Spin-Up Functions

As illustrated in Tables 2 and 3, even after 16 years of recursive simulations the model failed to reach an equilibrium state at either the 0.1% or 0.01% threshold for many of the criteria. This was especially true for variables related to subsurface storage and discharge, which failed to achieve predefined thresholds. In order to define the spin-up time at the finer resolution (0.01%) for different variables and criteria, we developed a series of spin-up functions for each criterion based on the relationship between the percentage change of a given variable and the number of simulation years until the respective change was obtained, using 16 years of simulations. As can be seen in Figure 3, the relationship between the percentage change values and the number of simulation years has the form of an exponential decay function. To fit the exponential function, the first year of data were removed from the analysis. Our analysis shows that an exponential function with two terms provided a better fit than a single exponential, having higher adjusted $R^2$ and lower root mean squared error (RMSE) across the multiple criteria and variables examined here (see Tables S1–S6 of the Supporting Information).

Spin-up functions were developed for the following criteria: five daily criteria (section 2.4.1), mean annual monthly PC based on average of monthly PC values for a given year, and percentage change in annual mean values. The spin-up functions estimated the number of simulation years required until a given threshold was reached, based on each criterion (Table 4). Overall, 23 years was required until the system equilibrated according to every criterion at the 0.01% level. However, 20 (or 21) years of simulation were sufficient for everything to equilibrate according to at least one criterion at the 0.1% (0.01%), respectively. Based on these results, further simulations were performed until groundwater storage equilibrated at 0.01%.

Results presented in Tables S1–S6 suggest that empirical spin-up functions were able to capture the overall behavior of changes in subsurface storages and discharge across multiple criteria during spin-up and predict spin-up time. However, their coefficients remain site specific and are expected to change as a function of topography, subsurface characteristics and climate. Sensitivity of double exponential spin-up functions to the number of recursive simulations (6–16 simulations) across multiple criteria and variables was also explored. The minimum number of simulations required to develop double exponential functions is 6, as the percentage change value between simulation 1 and 2 is large and at least four data points are required. Estimated spin-up times based on changes in mean annual depth to water table was more stable compared to other variables and criteria. Estimated spin-up times for 0.01% threshold level from the water table spin-up function ranged between 13 and 22 years, corresponding to 6 and 16 recursive simulations, respectively. Therefore, an exponential function based on changes in mean annual DTWT can be a useful tool in predicting water table response as a function of time and reduce spin-up time (H. Ajami et al., unpublished data, 2014).

3.3. Monthly Anomalies

Examining the monthly anomalies across a range of studied variables show just how far the initial conditions are from the equilibrium state. Assuming that simulation 20 was the equilibrated year, monthly anomalies were computed for a number of variables (Figure 6). Results show that the initial depth to water table of 3.0 m was far from the equilibrium state (catchment average of 5.2 m), and as the simulations proceeded through time, water table fluctuated as a result of topographically driven subsurface flow, discharge due to streamflow, and evapotranspiration and recharge as a result of precipitation through recursive simulations. Monthly ground surface temperature, which was initialized based on mean monthly air temperature (281 K), quickly adjusted itself after 3 years of simulations, with mean monthly values ranging between 272.3 and 290.8 K. For the first year, ground heat flux was far from the equilibrium state (January flux was $-5.1 \, \text{W m}^{-2}$), although it reached equilibrium at a faster rate than either the latent heat or sensible heat fluxes at the monthly time scale (after 4 years, anomalies were less than 0.25 $\text{W m}^{-2}$). In terms of the upward longwave radiation, the model equilibrated faster than other energy fluxes (after 2 years, anomalies

<table>
<thead>
<tr>
<th>Variables</th>
<th>0.10%</th>
<th>0.01%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Groundwater storage</td>
<td>7–8</td>
<td>18–19</td>
</tr>
<tr>
<td>Unsaturated zone storage</td>
<td>17–20</td>
<td>17–21</td>
</tr>
<tr>
<td>Relative saturation (CLM zone)</td>
<td>13–16</td>
<td>16–20</td>
</tr>
<tr>
<td>Relative saturation (root zone)</td>
<td>14–15</td>
<td>18–20</td>
</tr>
<tr>
<td>Depth to water table</td>
<td>20–21</td>
<td>21–22</td>
</tr>
<tr>
<td>Discharge</td>
<td>16–21</td>
<td>16–23</td>
</tr>
</tbody>
</table>

Table 4. Estimated Number of Simulations Required Until a Predefined Threshold Is Reached According to Five Daily Criteria, Monthly, and Annual Criteria Using Spin-Up Functions
were below 0.25 W m\(^{-2}\), as the upward longwave radiation is primarily controlled by ground surface temperature.

Standard deviations of the differences in monthly means through recursive simulations were also estimated in order to examine the roles of climate forcing and initialization on reaching equilibrium. Variability of the differences in monthly values was highest in January for the unsaturated zone and May through July for groundwater storage, indicating larger changes in storage for these months relative to the rest of the year (Figure 7). The decline of the unsaturated zone storage in the first year of simulation caused large differences between January storages of simulation 1 and 2. For groundwater storage, changes in recharge direction in February through April of the second year lowered the water table and caused the largest variability of monthly differences in May through July, relative to the rest of the year. For a number of variables including discharge, root zone saturation, snow water equivalent, and ground surface and bottom layer soil temperature, January had the highest variability in monthly differences as a result of initialization. In the case of discharge, initializing the water table at 3 m below the land surface resulted in pressure head of \(-3\) m at the gauge, and it took 47 days until streamflow was first generated at the gauge in the first year of simulation. In subsequent simulations, due to topography-driven flow and distribution of groundwater levels, grid cells along the channel network became saturated and only the ponding depth changed at the gauge.

The largest variability in difference in monthly means was in January for most energy fluxes except for the latent heat flux and upward longwave radiation (Figure 7). These two flux terms displayed a peak variability

![Figure 6. Monthly anomaly from equilibrium condition for: (a) depth to water table; (b) groundwater (saturated) and unsaturated zone storages; (c) ground surface and bottom layer soil temperature; and (d) latent heat, sensible heat, and ground heat fluxes, and upward longwave radiation. The 20th simulation was assumed to be the equilibrated year.](image-url)
in August and February respectively because of large variations in August evapotranspiration and variability of ground surface temperature in February as a result of including leap years in the simulations.

The variability of monthly differences highlights the role of climate forcing and seasonality, in conjunction with dynamic changes in catchment storage, in reaching equilibrium. For example, transitioning to a drier period in August contributes to variability of monthly evapotranspiration differences in this month.
Similarly, increased variability of sensible and ground heat fluxes in March is related to a transition to a warmer period in March relative to February. These variations in monthly differences impact estimated spin-up time using the PC time, as percent changes for all months should fall below a specific threshold.

3.4. Dynamics of the Water and Energy Balances During Spin-Up

As illustrated in section 3.1, the number of years required until the system equilibrates depends upon the spin-up measure used. To investigate the influence of this behavior further, the dynamics of the catchment-scale water balance during the spin-up process were examined. Percent biases in runoff ratios relative to observations were high, especially for the second year of simulations (62.3%), with percent bias decreasing as simulations proceed through time (20.3% after simulation 20; Figure 8). Unsaturated condition at the gauge for the first 47 days of year 1 caused smaller annual runoff and lower bias in runoff ratio of year 1 compared to year 2. Similarly, percent bias in monthly simulated discharge decreased as simulations progressed. The largest percent bias by simulation 20 was observed for November and December discharge (64%), corresponding to higher precipitation compared to previous months, while the smallest bias in simulated discharge occurred in March (1%). With the exception of a few dry days in July, percent bias in daily discharge decreased as the system reached equilibrium.

Mean annual evapotranspiration decreased from 459.7 mm/year to 447.3 mm/year after 20 years of simulations (Figure 8). Higher evapotranspiration rates in earlier simulations were due to the shallower water table.
As the system equilibrated, redistribution of ground water levels impacted annual evapotranspiration, and variability in mean annual evapotranspiration declined after 13 years of simulations. This decline is related to an earlier equilibrium of the root zone (Table 2) and changes in the degree of connectivity between groundwater and latent heat flux through recursive simulations [Maxwell and Kollet, 2008]. The extent of the groundwater coupling zone was identified based on the relationship between mean annual depth to water table and mean annual evapotranspiration for every pixel in the catchment (Figure 9). According to this relationship, three distinct zones were identified: a temperature controlled zone (DTWT < 0.6 m), a precipitation controlled zone (DTWT > 5 m) where evapotranspiration is controlled by precipitation, and a groundwater controlled zone (0.6 < DTWT < 5). In the groundwater controlled zone, changes in the water table impact the magnitude of land surface fluxes. As can be seen in Figure 10, in earlier simulations the groundwater controlled zone constituted up to 70% of the catchment. Due to spatially distributed adjustments of the water table, the extent of the groundwater controlled zone stabilized at around 44% after 13 years of simulations. The extent of the temperature controlled zone quickly stabilized at around 20% after 6 years of simulations, as changes in land surface fluxes in this zone do not depend on changes in water table depth.

Annual changes in unsaturated and saturated zone storages were larger in the earlier simulations compared to latter simulations and reached to −3.3 and −0.2 mm by simulation 20, indicating that the system had nearly achieved a steady state condition (Figure 8). Changes in the saturated zone storage showed an overall decrease in groundwater storage as simulations proceed due to increases in the catchment’s mean DTWT compared to initial condition. For unsaturated zone storage, the amount of water entering the system was higher than discharge, causing increases in unsaturated zone storage through recursive simulations. The magnitude and direction of monthly changes in unsaturated zone storage changed through the spin-up simulations. In the case of groundwater storage, only the magnitude of changes in storage was altered, with the exception of May, where changes in storage became positive as the system reached equilibrium.

Relationships between energy flux variables and depth to water table were explored during the spin-up simulations for the forest and agricultural regions within the catchment. For both land cover types, simulation 1 had the largest difference in the degree of coupling between ground heat flux and depth to water table. After year 1, relationships were quite similar, except in areas of shallow water tables. Here water levels were higher in earlier simulations, especially in the forested sites. Mean annual latent heat flux was larger for the forested site compared to the agricultural lands. Ground heat flux was larger in the agricultural site as a response to the average ground surface temperature being slightly higher (by approximately 1 K), compared to the forested site. While partitioning of energy fluxes depends on the temperature and moisture condition of the system, if moisture is limited, energy will generally be used to increase sensible heat flux.

From Figure 9 it can be seen that in agricultural areas with deeper ground water levels, higher outgoing longwave radiation is observed than in areas with shallow water table. Higher plant transpiration in areas of shallow water table results in a cooler surface temperature and decreases in outgoing longwave radiation. Despite the decrease in mean annual ground surface temperature in water table depths between 2 and 10 m, the outgoing longwave radiation at the forest site first slightly increased when depths were between 2 and 6 m, and then declined for deeper water table levels (6–10 m). The reason for this fluctuation is that in
forest sites, where the depth to water table was between 2 and 6 m, latent heat fluxes decreased and became moisture limited. In this situation, energy was likely used to increase sensible and longwave radiation fluxes, reflecting simulation responses. At depths between 6 and 10 m, the latent heat flux decoupled from the water table and outgoing longwave radiation decreased as a result of a decrease in ground surface temperature.

4. Discussion

4.1. Are the Spin-Up Times Consistent Across Multiple Variables and Criteria Used to Define Equilibrium State?

Multicriteria analysis of ParFlow.CLM spin-up through recursive simulations illustrated that spin-up time is a function of variable and the criteria used to define spin-up. Thermal equilibrium was achieved faster than saturated and unsaturated zone equilibrium, similar to results obtained from LSM based spin-up studies [Cosgrove et al., 2003]. Faster thermal equilibrium is attributed to smaller variational inertia and interannual variability of soil temperature relative to soil moisture [Rodell et al., 2005]. Results for this particular catchment and the modeling setup showed that more simulations were required for the unsaturated zone storage to equilibrate at predefined threshold levels compared with the groundwater storage. Although a uniform initial depth to water table and the spatially distributed adjustment of the water table contributed to this issue, the particular structure of the ParFlow.CLM in simulating two-way feedback processes between the land surface and subsurface may contribute as well. In ParFlow.CLM, saturated and unsaturated zones are not explicitly specified by fixed compartments. Saturation in the unsaturated zone is controlled by infiltration and the rising and falling of the water table.

Our results illustrate that the lower limit (minimum number) of simulations required for the model to equilibrate is governed by ground surface temperature response and that the upper limit for equilibrium to be achieved is based on subsurface storage response. Spin-up functions can provide a guide in estimating the spin-up time required when a finer equilibrium threshold is desired. But the question remains: which spin-up criteria should be considered for equilibrating coupled/integrated hydrologic models?

Many of the spin-up criteria introduced for LSM-based studies have focused on energy fluxes, and in particular, the latent heat flux. Because LSMs constitute the lower boundary of climate models, accurate estimate of evapotranspiration is important, since it is the main energy sink from the land surface impacting sensible heat flux and outward longwave radiation and describes the key mechanism for water vapor transport into the atmosphere [McCabe et al., 2005; Lofgren and Gronewold, 2013]. The important role of soil moisture in LSM spin-up studies has also been highlighted [Cosgrove et al., 2003; Shrestha and Houser, 2010], as it controls the partitioning of water and energy balances at the land surface and can significantly impact precipitation in numerical weather prediction models [Vivoni et al., 2009]. As can be seen from our analyses, when using an integrated hydrologic model like ParFlow.CLM, selecting spin-up criteria based on groundwater storage is more important in defining the equilibrium state as it ensures equilibrium across other criteria as well. In particular, equilibrium of subsurface storages is very important for this catchment because groundwater plays a major role in stream discharge (the base flow index, which is the ratio of base flow to total streamflow, is greater than 0.7), and equilibrium to a finer threshold is required to reduce bias in stream discharge. Our results also showed that the dependency of evapotranspiration on groundwater equilibrium was reduced as the extent of groundwater controlled zone stabilized in the catchment after 13 years of recursive simulations.

The choice of spin-up criteria in determining equilibrium state depends on a catchments’ climatic condition as well. For example, in catchments where snowmelt constitutes a major component of discharge in spring or in areas with permanent permafrost, equilibrium in snow water equivalent and ice content should be considered in defining the equilibrium state.

4.2. How Does Spatial Variability of Water and Energy Fluxes Change According to Various Equilibrium States?

To assess the sensitivity of surface water-groundwater exchanges to various equilibrium conditions, three equilibrium states were identified based on the minimum time required for subsurface storages and discharge to equilibrate at the 0.1% threshold level (Table 2). Minimum spin-up times were identified as
follows: groundwater storage (2 years); root zone storage and discharge (10 years); and unsaturated zone storage (20 years). As can be seen from Figure 11, changes in local groundwater storage were up to 3.5 m on the eastern side of the domain after 2 years of simulations, and the threshold level of 0.1% was not sufficient for the groundwater system to equilibrate in this catchment (changes between the first and second year of simulations were less than 0.1%). Changes in unsaturated zone storage revealed different patterns of surface water-groundwater exchange across multiple equilibrium states. For instance, high elevation recharge zones became discharge zones after root zone equilibrium (simulation 20). Also, spatial variability of changes in storage showed that during spin-up, larger changes in storage occurred in higher elevation zones and along the ridgeline, where changes in DTWT relative to simulation 20 were the largest. These storage changes are caused by the spatial distribution of DTWT from the uniform initialization and the impact of climate forcing. In terms of land cover, differences in mean annual evapotranspiration from the equilibrium simulation (simulation 20) were highest in earlier simulations in the forested areas due to deeper rooting depth of evergreen needle leaf forest relative to croplands and distribution of trees in higher elevations. Therefore, equilibrium of groundwater to higher spin-up thresholds was important in this catchment to represent recharge and discharge patterns as a function of a given climate and reducing variability in DTWT in high elevation areas.

The spatial variability of differences in subsurface storages was explored by computing pixel-based root mean square difference (RMSPD) across successive simulations, and estimating catchment-scale coefficient of variation (CV) from local RMSDs. Results showed decreases of catchment-scale mean and standard deviation of RMSPD and increases in CVs of subsurface storages as the system equilibrated (Figure 12). Despite the decline in mean and standard deviation of RMSPD, CVs of saturated and unsaturated zone storages increased consistently toward equilibrium (standard deviation of RMSPD did not decline at the same rate as the mean RMSPD). This increase in CV is indicative of the spatial heterogeneity of local RMSPD and the nonuniform

Figure 11. Annual changes in (a–c) groundwater and (d–f) unsaturated zone storages (in m) for three equilibrium states based on groundwater storage, root zone storage and discharge, and unsaturated zone storage criteria. (g–i) Differences in mean annual DTWT of corresponding simulations from simulation 20.
decline of RMSD across the catchment during spin-up. It is also consistent with the increased spatial variability of groundwater levels at the equilibrium state compared to initial DTWT. For the root zone storage, CV of RMSD values remained at about 2.1 after three spin-up cycles, indicating a more uniform decline of RMSD as the majority of the catchment is covered by croplands.

Despite the role of topography in the variability of spin-up times across the catchment, the global spin-up criteria used here were adequate to ensure catchment-scale equilibrium. By simulation 20, changes in land surface fluxes and DTWT became very small. It is expected that at the regional scale and in catchments with large variations in topography and land cover types, differences in spin-up time may become larger. Further research is required to examine the application of spatial analysis techniques in defining the equilibrium state for regional-scale application of coupled/integrated hydrologic models.

### 4.3. What Is the Role of Predefined Thresholds in Estimating Spin-Up Time?

Generally, spin-up times are longer for higher threshold levels across all the spin-up criteria. As has been highlighted in our analysis, setting up predefined thresholds can sometimes produce misleading results. For example, groundwater storage was equilibrated even after two recursive simulations at the 1% level, while the largest changes in storage occurred for year 2, which was not captured by the spin-up criteria at the 1% threshold. In the case of DTWT, the model did not equilibrate based on the 0.1% and 0.01% thresholds even after 20 years of simulations. However, differences in the mean annual depth to water table of select equilibrium states from simulation 20 reached 0.1 m for the majority of pixels inside the catchment (Figure 11). For the energy fluxes, as simulations proceed small differences in energy fluxes inflated percentage change values. Therefore, it is recommended to examine the differences in variables as well as the percentage change values during the spin-up assessment process.

Based on the start-end criterion, the system did not converge to a set of predefined thresholds for energy fluxes, ground temperature and SWE. Instead, this criterion stabilized around a constant value due to inherent differences in the input forcing between the first and last days of 2003. For state variables like SWE, soil moisture, and groundwater storage, the value of this criterion is further impacted by the history of system
states for a certain time span (memory). Therefore, differences between these 2 days can become very small depending on the system response, but they cannot reach zero. As evident in Figures 3–5, convergence for this criterion should be considered in terms of stabilization of percent change values rather than via predefined thresholds.

4.4. What Is the Impact of Different Initialization Approaches on Catchment-Scale Simulations and Spin-Up Time?

Although, specifying the initial conditions is amongst the first-order decisions that need to be made by modelers [Cloke et al., 2003], this element of uncertainty is often overlooked within the hydrologic modeling community. Rushton and Wedderburn [1973] assessed the impact of four initialization approaches for transient groundwater simulations using a series of one-dimensional simulations. They concluded that for most aquifers, initial heads should be in the state of dynamic balance (i.e., aquifer inputs and outputs are in overall equilibrium for some time period before the start of a transient simulation [Miles and Rushton, 1983]) to minimize error in simulated heads and the impacts of stress, such as pumping on groundwater systems. An initial hydraulic head of zero or hydraulic head obtained from a steady state solution given either a particular or average values of inflows and outflows into or from the aquifer, led to larger error in simulated head. Wood and Calver [1992] argue that it would not be possible to reach the true steady state when both saturated and unsaturated flows were simulated, like modeling subsurface flow in hillslopes. Cloke et al. [2003] simulated the impact of three commonly used initial conditions for water table in an idealized setting, including: (1) a hydrostatic curvilinear water table; (2) a hydrostatic inclined water table; and (3) a hydrostatic pressure below linear water table with nonhydrostatic conditions above the water table. Their results showed that even by applying high infiltration rates at the land surface, the impact of various initializations can last for up to several hundred days over subsequent simulations. Therefore, implementing a spin-up approach would be required to reduce bias in simulated hydrologic fluxes.

Here we assessed the impact of the choice of spin-up measure on the time to equilibrium, when equilibrium-based simulations are used to initialize the model. Thus, this has to be explored by calculating the spin-up measures on a set of spin-up simulations that start from the same initial state. Using a different method to choose this initial state may change the total time it takes to reach equilibrium, but the relative consequence of choosing the spin-up measure would not change. Needless to say, the choice of climate forcing will impact model states even for equilibrium-based simulations. If the wettest year on record was used for spin-up simulations, the model will reach equilibrium at a very wet state and it will impact changes in storage and hydrologic fluxes in subsequent simulations.

Spin-up times were not only impacted by the meteorological forcing. Factors such as boundary conditions, subsurface parameters like hydraulic conductivity, initial depth to water table, and catchment topography also have a considerable influence. Our results suggest that subsurface components and groundwater fed discharge were considerably slower at equilibrating. Therefore, special attention is required to provide a better initial guess of groundwater level in the catchment. This could perhaps be based on a steady state run with average climate forcing or application of a combined topography-soil index [Sivapalan et al., 1987; Troch et al., 1993] to produce better initial groundwater levels. Another challenge is the distribution of soil moisture above the water table, especially in areas of deep groundwater levels. This issue can be overcome by performing additional spin-up simulations to ensure unsaturated zone equilibrium. Further research is required to assess the impact of different initialization approaches on spin-up time in integrated hydrologic models.

5. Conclusions

The performance of a number of spin-up measures to assess the equilibrium state for the ParFlow.CLM, with reference to a real-world application at the catchment scale, has been examined. Results illustrate that for all of the spin-up measures employed here it takes longer for the subsurface storage and discharge to equilibrate than for the ground surface temperature and energy fluxes. Although thermal equilibrium is quickly reached after 2–4 years, depending on the spin-up measure, it takes longer for energy fluxes other than the upward longwave radiation to equilibrate based on these same measures. Our analyses not only highlights the advantages of using a multicriteria approach for the subsurface storage in assessing the spin-up, but also that information obtained from the partitioning of the water balance is quite useful in examining the equilibrium state and how catchment fluxes are evolving as the system equilibrates. For example, in a groundwater fed stream
such as the Ringkobing Fjord catchment, groundwater equilibrium is important in reducing bias in the simulated discharge. In terms of evapotranspiration, the dependency of evapotranspiration on groundwater equilib-rium is reduced as the extent of the groundwater controlled zone stabilized in the catchment after 13 years of recursive simulations. Our results illustrate that changes in storage and DTWT occurred in areas of higher elevation in the catchment, and highlighted the role of topography in controlling spin-up time.

Sensitivity analysis of spin-up functions revealed that a spin-up function based on changes in mean annual depth to water table produced more stable spin-up times compared to other variables and criteria examined here. Therefore, the inverse of a spin-up function which predicts changes in mean annual DTWT as a function of time constituted the basis for development of a new approach to reduce spin-up time in integrated/coupled hydrologic models (H. Ajami et al., unpublished data, 2014). In the case of Ringkobing Fjord catchment, the spin-up time reduced by 50% (H. Ajami et al., unpublished data, 2014).

The study also sheds light on the importance of spin-up when using an integrated hydrologic model. It should be recognized that the spin-up time obtained from these simulations remain site specific. Spin-up time depends on the location and forcing data, and is impacted by precipitation intensity, solar radiation, soil depth, soil and vegetation type and initial moisture content [Rodell et al., 2005; Yang et al., 2011; Yang et al., 1995]. Similar to activities within the LSM community, coordinated efforts are required to perform catchment-scale experiments across multiple sites. Such an effort would include examining a range of coupled/integrated hydrologic models to assess spin-up time in areas with different climate, land surface and geologic structure, and to design methodologies to reduce spin-up time.

Generally, the increased computational cost of implementing catchment-scale integrated models has limited the evaluation of the impact of different initialization approaches. Likewise, the influence of incorporating multiple years of forcing instead of 1 year of forcing to reaching equilibrium state remains underinvestigated. Models that fail to adequately examine their sensitivity to initial condition and also simulate from a nonequilibrium state run the risk of developing large biases in water and energy fluxes. In such cases, the model implications would be exacerbated for studies of limited temporal duration, for those investigating streamflow forecasting, or when integrated hydrologic models are coupled to numerical weather forecasting systems. Although issues related to the uncertainty in state initialization are commonly discussed in the streamflow forecasting literature [Li et al., 2009] and in LSM spin-up studies, understanding catchment behavior at the equilibrium condition is certainly of importance when impacts of changes in climate [Ferguson and Maxwell, 2010] or in landcover on the partitioning of the water and energy balance are being investigated.

References


Chen, T. H., et al. (1997), Cabauw experimental results from the project for intercomparison of land-surface parameterization schemes, J. Clim., 10(6), 1194–1215.


