

Dear Dr Andréassian,

Thank you for providing us with the review of our submitted manuscript to the Hydrology and Earth System Sciences Journal entitled "Technical Note: Reducing the spin-up period of integrated surface water-groundwater models", hess-2014-169. We appreciate the reviews from Referee#1 and Referee#2 and believe that their comments and suggestions have significantly improved our manuscript. In the following, we address specific reviewer comments.

## **Anonymous Referee #1**

### **General comments**

*1. This paper focuses on the description and evaluation of a procedure for reducing the time of model spin-up, which is a commonly adopted strategy to initialize integrated/ coupled hydrological models such as ParFlow.CLM. In my opinion, there is a core issue within this paper related to its very basic idea, i.e., the assumption that an equilibrium state (achieved over no matter how many years of forcing data) can represent a correct (or even reasonable) initial catchment state. Although I acknowledge that this is a common assumption, I believe that not only it is not true in general, but the number of cases where this could be reasonable is limited, in theory, only to catchments where i) the land use do not change over time and ii), most importantly, the inter-annual variability of the weather forcing is very small. The latter point is equivalent to the assumption that a single year (or two, three) of forcing data can be considered representative of the whole climatic regime of the catchment, an hypothesis that is never realistic in practice. Unfortunately for hydrologists, catchments are always dynamic systems and never in a state of equilibrium; therefore, I am afraid that the whole procedure proposed in the paper is not worth the effort from the very beginning. Instead, the only way to achieve a correct or reasonable initial state is to use a "warm-up" procedure, where the model must be run using a long enough time series of forcing data before the period of interest; the necessary warm-up duration will be obviously catchment-specific and can be evaluated by starting the model with two or more different initial guesses and checking that after the warm-up all the simulations converged to the same final (dynamic) state.*

We agree with the reviewer comments regarding the short comings of equilibrium based initialization for initializing coupled/integrated hydrologic models. Despite these shortcomings this initialization method is commonly used. This technical note provides a method for improving the efficiency of this commonly used initialization technique. While in the land surface modelling community various experiments have been performed across multiple sites and models to assess the impact of initialization approaches and spin-up criteria (Yang et al., 1995, Rodell et al., 2005) on simulated response, the issue of model initialization has not been fully explored for the coupled or integrated hydrologic models. As we stated in our objectives, here the goal was to reduce the spin-up period for equilibrium based initializations. The equilibrium based initializations have been used previously for exploring land surface-groundwater coupling (Kollet and Maxwell, 2008) and assessing the impact of climate change on groundwater-land surface interactions using an integrated hydrologic model (Ferguson and Maxwell, 2010). Here, similar to the Project for Intercomparison of Land Surface Parameterization Schemes (PILPS), only one year of forcing data is used for ParFlow.CLM spin-up. The adjustment method introduced can be examined when multiple years of forcing data is used for the warm-up period. We could not

examine the impact of multiple years of forcing on the spin-up time due to the intensive computational demand. In Ajami et al. (2014), the impact of different initializations approaches as examined in the literature was briefly discussed and we refer the reader to that summary. The issue of initialization is very important particularly in coupled or integrated hydrologic models and coordinated efforts to perform such experiments across multiple models and sites are required in hydrologic modelling community.

The following point is added to the revised Introduction (Page 2, Lines 30-31 and Page 3, Lines 1-2):

The two most common initialization approaches in coupled or integrated distributed hydrologic models are: (1) initial depth to water table is specified at a certain uniform depth below the land surface (Kollet and Maxwell, 2008) and the impact of initialization is reduced through recursive simulations over either a single or multiple years of forcing data, until equilibrium conditions are reached, which are usually related to spin-up criteria based on changes in groundwater heads (Refsgaard, 1997) or changes in water and energy balances (Kollet and Maxwell, 2008); or (2) the model is initialized from a fully saturated condition and simulations are continued until modelled baseflow matches the observations (Jones et al., 2008). Equilibrium based initializations have been utilized previously for exploring land surface-groundwater coupling (Kollet and Maxwell, 2008) and assessing the impact of climate change on groundwater-land surface interactions using an integrated hydrologic model (Ferguson and Maxwell, 2010).

Further, we added application of equilibrium based initialization in the objective section of the revised manuscript (Page 3, Lines 19-28):

The objective of the current study is to develop a *hybrid* spin-up approach that significantly reduces the number of years of spin-up required for model state equilibrium. The equilibrium based initialization represents a correct initial state for catchments in which the land use does not change over time and the inter-annual variability of atmospheric forcing is very small: assumptions that are common to most simulation frameworks. This technical note provides a method for improving the efficiency of this commonly used initialization technique. The performance of the proposed approach in reducing the spin-up period for a catchment scale application of the ParFlow.CLM model is evaluated against the standard continuous recursive simulation approach that is commonly applied for land surface model spin-up, and referred to here as the baseline spin-up approach.

*2. Another issue of this paper regards the lack of important details, such as (at least) a brief description of ParFlow.CLM, and some steps of the procedure that are not described with sufficient clarity. See below in the list of specific comments.*

A brief description of ParFlow.CLM is added to the revised manuscript. Please see Section 2.1. Section 2.2 is revised to include further details about the approach.

### **Specific comments:**

*1. Page 6971, line 20: please define “service unit”.*

Definition of a service unit is added in the revised manuscript (Page 3, Lines 8-9).  
(a service unit is equivalent to 1 hour of time used by one processor)

*2. Page 6972, section 2.1: despite the title, no description whatsoever of the model is provided, but only a description of the two catchments.*

A brief description of ParFlow.CLM is added at the beginning of section 2.1 as follows: ParFlow is a 3D variably saturated groundwater flow model that solves the mixed form of the three-dimensional Richards equation for the subsurface (Ashby and Falgout, 1996; Jones and Woodward, 2001; Maxwell et al., 2014). ParFlow has a fully integrated overland flow simulator (Kollet and Maxwell, 2006) and performs routing of the ponded water on the land surface via the kinematic wave equation. The Common Land Model (CLM 3.0) (Dai *et al.*, 2003) is integrated into ParFlow to simulate water and energy fluxes at the land surface (Maxwell and Miller, 2005; Kollet and Maxwell, 2008). ParFlow.CLM versions 605 and 653 were used for the Skjern River and Baldry simulations respectively, which are described below. The terrain following grid of Maxwell (2013) is not implemented in these modelling set-ups.

*3. Page 6975: lines 7-19: this section is rather difficult to follow. Is the DTWT function used to re-initialize the model spatially variable or uniform? And the resulting DTWT distribution after re-initialization? Also, it is not clear how the “best performing” DTWT functions were chosen: what objective function was used to evaluate the best performance, root mean square difference, mean absolute error or the semi-variogram?*

The DTWT function produces spatially distributed DTWT for re-initialization. The procedure for generating spatially distributed DTWT is as follows: 1) develop the DTWT function based on percent changes in mean annual DTWT values across the domain for six cycles of ParFlow.CLM simulations (global DTWT function), and 2) Implement the DTWT function at every grid cell to re-initialize the ParFlow.CLM model. The updated DTWT at every grid cell depends on its initial value. To clarify this point, the manuscript is revised as follows (Page 7, Lines 28-32):

The empirical DTWT functions calculated above estimate percentage changes in mean annual DTWT as a function of simulation year. To predict spatially distributed mean annual DTWT from a global DTWT function, the mean annual DTWT from the final cycle of the ParFlow.CLM spin-up simulation for every grid cell is used as the initial value to successively estimate DTWT distributions as a function of simulation year. These DTWT distributions are based on the predicted percent change values from the global DTWT function.

Here we are comparing the performance of multiple DTWT functions against the baseline simulation using multiple objective functions. As presented in our results a single objective function does not constantly perform best for all the cases and each of the objective functions provides a summary statistic regarding a certain aspect of the model performance. While mean absolute error (MAE) and root mean square difference (RMSD) provide an overall average of model error, RMSD is more sensitive to extreme values. Percent bias gives information about the average tendency of the model prediction to be larger or smaller than the baseline simulation. We revised the manuscript as follows for the Skjern River sub-catchment section (Section 3.1, Page 10-11):

Optimum parameter values for single and double exponential DTWT functions were obtained using nonlinear least squares method. Performance of the single and double exponential DTWT functions in predicting 14 years of DTWT were compared against ParFlow.CLM baseline spin-up simulations (years 7 through 20) of Ajami et al. (2014) to find the optimum

empirical DTWT function for the Skjern River sub-catchment. Post-simulation analysis indicates that global DTWT functions based on domain or catchment averaged percentage change values are better predictors of DTWT response compared to local DTWT functions developed for every grid cell. Instability of local DTWT functions occurs in grid cells where percent changes in DTWT oscillate between positive and negative values through initial spin-up simulations. Spatial distribution of these grid cells are shown in Fig S1.

Calculated RMSD and percent bias relative to the baseline spin-up simulations indicate that global double exponential functions using ParFlow.CLM spin-up simulations 2 to 6 provide a better fit compared to various single exponential functions obtained from different spin-up simulation years (e.g. 2 to 3, 2 to 4, etc.). Because the first six cycles of ParFlow.CLM simulations were the same between the baseline spin-up simulations and DTWT distributions from DTWT functions presented in Fig 4, comparisons were made with simulations 7 to 20 of the baseline spin-up approach of Ajami et al. (2014).

As can be seen from Fig. 4a, the mean annual DTWT over the domain derived from the single exponential functions (fitted to percentage change data from simulations 2 to 6) under-predict the baseline spin-up simulations, due to their consistent small underestimates in comparison to double exponential functions fitted to the same data points. Only for the mean absolute error (MAE) calculated at each pixel do single exponential functions based on simulations 2 to 6 perform slightly better and produce smaller errors on average than the double exponential functions (Fig. 4b). It should be noted that the percent bias in mean annual DTWT for simulation cycle 20 is -1.6% for the domain based double exponential function and -6.2% for the single exponential function, with both functions derived from simulation cycles 2 to 6. Therefore, single exponential functions are not further examined in re-initializations of the DTWT. In terms of mean DTWT across the domain (Fig. 4a), the catchment delineated double exponential DTWT function provides a better prediction and the smallest mean bias when compared to the function based on the entire model domain. However, Figure 4b indicates that the mean absolute error values are slightly smaller for the domain based double exponential function. The higher MAE of the catchment based double exponential function is a result of slightly more regions with over and underestimated DTWT values that contribute to a good overall mean DTWT (Fig. 4a), but contains more errors spatially compared to the domain based double exponential function.

The section is followed by an overall conclusions in the revised manuscript (Page 11, Line 32):

In summary, double exponential functions are chosen as they have less bias compared to single exponential functions and there is very little difference in terms of MAE amongst predictions. The choice is further supported by the RMSD and semi-variograms.

*4. Page 6976, lines 2-3: why is the pressure head profile in the UZ adjusted with a (basically) instantaneous distribution, taken from the last day of the sixth cycle, while the re-initialized DTWT is assumed as an annual mean? I see a possible lack of consistency that should be discussed.*

To clarify this point, the manuscript is revised as follows:

Section 2.2. Page 6, Lines 25-28:

Sensitivity of spin-up functions across multiple criteria and variables showed that the estimated spin-up period based on mean annual DTWT were more stable when compared to other spin-up criteria, such as changes in the mean DTWT for the last day of recursive simulations (Ajami et al., 2014)....

Section 2.2. Page 9, Lines 8-19:

In the adjusted pressure head approach, the hydrostatic equilibrium assumption is used in regions between the new DTWT and the initial DTWT. The ParFlow.CLM pressure head distribution is adjusted to begin at the new pressure head from the initial WT such that the vertical profile is maintained (Fig. 3). This adjustment may represent a lack of consistency in the proposed approach as the DTWT function estimates mean annual DTWT, while pressure head adjustments in the unsaturated zone are taken from the last day of the sixth cycle of ParFlow.CLM. While it is possible to use DTWT values from the last day of simulations to develop a DTWT function, estimated DTWT values from such a function exhibit larger variability and result in a larger bias. For the Skjern River sub-catchment, percent bias between the estimated DTWT values from the DTWT functions and the baseline simulation of Ajami et al. (2014) were -4% and -1.6% for the DTWT functions based on the last day and mean annual DTWT values respectively.

*5. Page 6977, lines 1-17 and Fig. 3: I am quite puzzled by these results. From Fig. 3a, I would expect that i) the MAE of the Exp2-Catchment curve decreased with time, not the contrary, especially after year 14, and ii) the MAE of the two Exp1 curves was larger, not smaller, than the Exp2 curves. Have the authors any explanation for this?*

Here the MAE is calculated on a pixel basis by computing the average absolute difference between the estimated DTWT from the DTWT functions and the baseline spin-up simulation. We included the formulas for calculating the objective functions in the revised manuscript to clarify this point. Figure 4a shows the mean annual DTWT over the domain and it indicates that the Exp2-Catchment function results in the smallest mean bias and the single exponential functions have the worst overall mean. Figure 4b shows the MAE calculated at each pixel and it indicates that the Exp2-Catchment function has slightly more regions with over and underestimated DTWT values resulting in a good overall mean, but contain more errors spatially. Meanwhile, the single exponential function estimates result in consistent small underestimates which produce slightly smaller errors when averaged spatially but has a worse overall mean.

Therefore, i) Figure 4b of the revised manuscript indicates that the performance of DTWT functions deteriorate for later time period (simulations 7 through 20) and the MAE increases in later simulations. ii) Similarly, average model error on a pixel basis (MAE) is smaller for the two single exponential functions than the Exp2 curves as shown in Figure 4b.

To clarify this point, the manuscript is revised. Please see the response to specific comment #3.

6. Page 6977, lines 18-29: *it is not clear how the semi-variograms were calculated. Was the mean annual DTWT used?*

Mean annual DTWT at every grid cell was used to calculate semi-variograms. The manuscript is revised to describe the procedure. See Page 11, Lines 14-23:

To investigate this result further, three empirical semi-variograms were generated. As the impact of an east-west spatial trend in the mean annual DTWT values was evident in the semi-variograms, the trend should first be removed from the mean annual DTWT values. To remove the trend, a plane was fitted to the observed mean annual DTWT values, with an equation of the form:

$$z = ax + by + c \quad (6)$$

where  $a$ ,  $b$  and  $c$  are fitted coefficients,  $x$  and  $y$  are the coordinates of every grid cell, and  $z$  is the mean annual DTWT. Residuals are computed by subtracting the estimated mean annual DTWT from Equation 6 from the observed mean annual DTWT values. Finally, the semi-variogram of the residuals as a function of distance is calculated.

7. Page 6978, lines 20-22: *this sentence is not clear, please rephrase.*

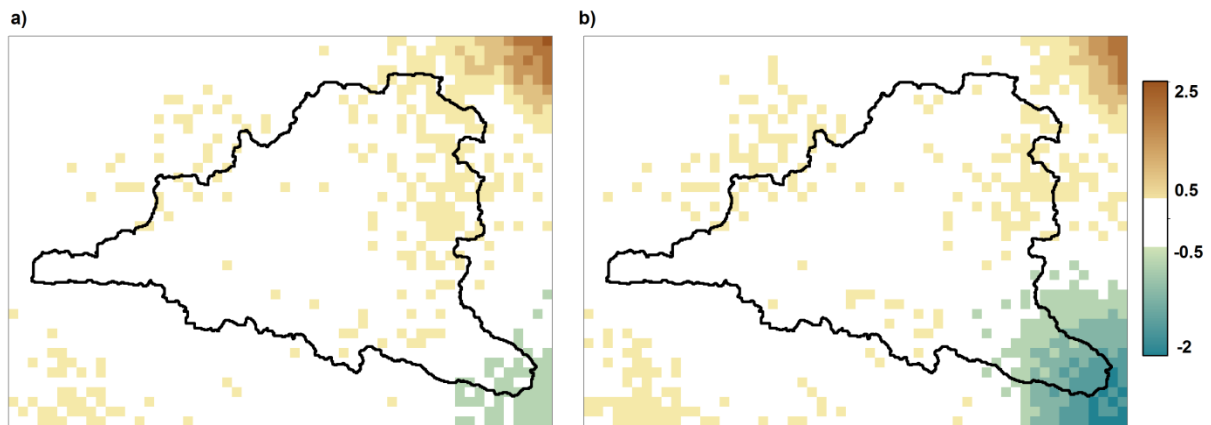
The sentences are rephrased in the revised manuscript (Page 12, Lines 23-31 and Page 13, Lines 1-6):

While in both re-initializations, DTWT and subsequently groundwater storage volume were the same at the start of the simulations, unsaturated zone storage of the hydrostatic equilibrium option was drier than the adjusted pressure head option. Additional ParFlow.CLM simulations after re-initialization ensured equilibrium of groundwater storage. As can be seen from Fig. 6a, hydrostatic re-initialization results in a deeper WT at equilibrium (simulation 12) relative to the baseline equilibrium year (simulation 20). Higher DTWT values of the hydrostatic option at equilibrium correspond to smaller groundwater storage and subsequently larger unsaturated zone storage compared to the baseline spin-up (Fig. 5). It should be noted that in ParFlow.CLM, groundwater and unsaturated zone storages are not explicitly determined by fixed size compartments and the extent of an unsaturated zone is determined by the location of the water table. Percent changes in mean annual unsaturated zone storage between the last two years of recursive simulations were 0.1% for the hydrostatic equilibrium and 0.3% for the adjusted pressure head re-initializations, indicating unsaturated zone equilibrium at different threshold levels.

8. Page 6979, lines 5-9 and Fig. 6: *from the figure I cannot see how the adjusted vertical pressure distribution produces better results than the hydrostatic profiles, nor I can see the bias with the latter. Perhaps, would be a good idea to show the experimental pdf (artícula) of the differences along with their spatial distribution.*

Figure 6a shows that the difference between DTWT distributions from the hydrostatic equilibrium option and the baseline simulation is mostly positive, while for the adjusted pressure head option the differences in DTWT values are negative in the upper part of the catchment. We generated the kernel density plots of the differences; however, the figure was not informative especially when the bandwidth was set the same for both density plots. Therefore, we only included the spatial distribution of the differences in the revised manuscript and changed the colour scheme of the Figure to present this bias.





**Figure 6.** Differences in equilibrium DTWT between ParFlow.CLM simulations after re-initializations and ParFlow.CLM after 20 years of baseline spin-up simulations in (m), where a) is based on hydrostatic pressure distribution above the water table for the initial condition, while b) is based on adjusted pressure head distribution above the water table for the Skjern River sub-catchment. White regions correspond to grid cells where the differences in equilibrium DTWT are less than 0.5.

9. Page 6979, line 16: *why does the smaller Baldry catchment require more service units than the larger Skjern catchment? Is it because the former has a larger grid size due to a better DEM resolution?*

A major factor in the Baldry catchment requiring more service units is that it required 40% longer simulations to reach equilibrium. The total number of service units for one year of simulation is larger for the Baldry sub-catchment than the Skjern River sub-catchment despite its lower number of computational nodes (467712 nodes in Baldry compared to 909440 in the Skjern River sub-catchment). It is difficult to solely attribute the increases in computational time to the DEM cell size. Based on our experience, increases in the number of vertical nodes result in longer computational time. Here, the increases in computational time in Baldry are due to multiple factors. As indicated in the revised manuscript, two different versions of ParFlow have been used for these catchments and we used different options for storing the CLM output files (silo versus PFB). For the Skjern River sub-catchment, CLM output files were saved as distributed silo files and after every model restart (every 15 days), a one processor job was submitted to un-distribute the silo files and save them as PFB files. For the Baldry sub-catchment, the PFB option was used and the un-distribution of PFB files was performed with the main ParFlow TCL script that uses 64 processors. Therefore, this set-up has led to unrealistic increases in service units for the Baldry. Based on the ParFlow user forum, it seems that the issue with the un-distributed PFB files can be resolved by using different setting when compiling the ParFlow.CLM code.

We should note that the number of processors for every catchment scale simulation was determined by performing parallel efficiency tests.

10. Page 6981, lines 1-2: *I do not agree that the proposed procedure “has the potential to assist in parameter calibration”. Due to equifinality, if a wrong initial state is used, such as the one likely to achieve by assuming equilibrium, a calibration procedure could lead to strongly biased parameters.*

We acknowledge the reviewer concerns here. We are not attempting to address equifinality here, we simply present a technique that can make previously used calibration approaches more efficient. We have removed this statement in the revised manuscript. It should be noted that a period of spin-up has been often implemented during calibration. We revised the Summary section as follows (Page 15, Lines 12-16):

Previous efforts in calibrating coupled or integrated hydrologic models required a spin-up process after every parameter update (Stisen et al., 2011; Weill et al., 2013). Development of a computationally efficient spin-up approach will enable this type of systematic calibration of integrated or coupled hydrologic models.

*11. Page 6990, Fig. 5: I am surprised that the hydrostatic equilibrium procedure underestimates the groundwater storage even from the start of the simulation. If the DTWT at re-initialization is the same as for the adjusted pressure profile, how can the Authors explain that large bias?*

To address reviewer comment, the manuscript is modified as follows (Page 12, Lines 23-31 and Page 13, Lines 1-6):

While in both re-initializations, DTWT and subsequently groundwater storage volume were the same at the start of the simulations, unsaturated zone storage of the hydrostatic equilibrium option was drier than the adjusted pressure head option. Additional ParFlow.CLM simulations after re-initialization ensured equilibrium of groundwater storage. As can be seen from Fig. 6a, hydrostatic re-initialization results in a deeper WT at equilibrium (simulation 12) relative to the baseline equilibrium year (simulation 20). Higher DTWT values of the hydrostatic option at equilibrium correspond to smaller groundwater storage and subsequently larger unsaturated zone storage compared to the baseline spin-up (Fig. 5). It should be noted that in ParFlow.CLM, groundwater and unsaturated zone storages are not explicitly determined by fixed size compartments and the extent of an unsaturated zone is determined by the location of the water table. Percent changes in mean annual unsaturated zone storage between the last two years of recursive simulations were 0.1% for the hydrostatic equilibrium and 0.3% for the adjusted pressure head re-initializations, indicating unsaturated zone equilibrium at different threshold levels.

Figure 5b shows time series of groundwater storage for the equilibrium year for three cases, baseline simulation, adjusted pressure head and hydrostatic options. The equilibrium year corresponds to simulation cycles of 10, 12 and 20 for the adjusted pressure head, hydrostatic equilibrium, and baseline simulations, respectively. Figure 5b caption is revised to clarify this:

**Figure 5.** Comparison of a) unsaturated and b) groundwater storages of ParFlow.CLM equilibrium year using the hybrid and baseline spin-up approaches (Ajami et al., 2014). The equilibrium year corresponds to simulation cycles of 10, 12 and 20 for the adjusted pressure head, hydrostatic equilibrium, and baseline simulations, respectively. The dynamics of groundwater and unsaturated zone storages are closely reproduced by the adjusted pressure head distribution approach relative to the baseline spin-up approach for the Skjern River sub-catchment.



12. Page 6992, Fig. 7: *there seems to be a spatial pattern, with streaks of DTWT overestimation in the south of the catchment. How can this be explained?*

The contours of DTWT overestimation occurred along the direction of flow lines from high elevation areas in the catchment toward the catchment outlet. As indicated in the manuscript, the performance of global DTWT functions were deteriorated in high elevation areas. The Figure caption is revised as follows:

**Figure 7.** Differences in equilibrium DTWT of Baldry ParFlow.CLM simulations after re-initialization with the adjusted pressure head distribution above the water table and ParFlow.CLM after 28 years of baseline spin-up simulations in (m). The contours of DTWT overestimation are along the direction of flow lines from high elevation areas toward the catchment outlet.

### **Technical corrections**

1. Page 6970, lines 18-19: *change the sentence to “The issue of model initialization is important for hydrologic predictions as the initial state has a major impact on the catchment’s model response”.*

It is revised.

2. Page 6979, line 21: *correct “particular”.*

It is corrected.

3. Page 6974, line 17: *DTWT was 3 m only for the Skjern catchment.*

Initial DTWT for the Baldry sub-catchment is included in the revised manuscript.

4. Page 6977, line 20: *change “semi-variance” to “semi-variogram”.*

It is modified.

### **References:**

Rodell, M., P. R. Houser, A. A. Berg, and J. S. Famiglietti (2005), Evaluation of 10 methods for initializing a land surface model, *J. Hydrometeorol.*, 6(2), 146-155.

Yang, Z. L., R. E. Dickinson, A. Henderson-Sellers, and A. J. Pitman (1995), Preliminary study of spin-up processes in land surface models with the first stage data of Project for Intercomparison of Land Surface Parameterization Schemes Phase 1(a), *J. Geophys. Res.*, 100(D8), 16553-16578.

## **Anonymous Referee #2**

*The technical note presents a methodology to nudge the predicted groundwater table depth, thereby reducing the number of years required for spin-up of integrated surfacewater-groundwater model Parflow.CLM, based on subsurface storage spin-up criteria. The methodology however does not reduce the real computation time of the model itself, but only reduces the number of years of recursive runs required to initialize the model based on the spin-up criteria. Also, it does not distinguish between the computation time required for each year of spin-up on whether it decreases or it is constant. Although the problem size used in this study cannot be considered to be computationally intensive, which is also a relative term, the idea presented does show some potential to reduce the spin-up period to generate initial soil moisture data. In general, the manuscript is very well written, but at the same time, there are some shortcomings in the paper that needs to be addressed. There are several instances where the content of the paper is intangible, and inadequacy in experiment designs for the proposed methodology. In addition, the figure quality are very poor in terms of the size of figure, fonts and scale, rendering them unreadable.*

We agree with the reviewer that the methodology does not decrease the computational time and it only reduces the spin-up period. As the reviewer suggested we changed spin-up time to spin-up period to clarify this point in the revised manuscript. We do not have the exact record of the computational time for each year of simulation during the spin-up, but our observations show no significant decrease in computational time as system equilibrates. We have significantly revised the manuscript to clearly describe the methodology and highlight the limitations of our approach.

We checked the quality of all the figures and improved them where possible.

### **Specific comments:**

*1. Benchmark the term “computationally intensive”, which is loosely used throughout the manuscript. Eg., in comparisons to : : .*

We revised the manuscript as follows (Page 3, Lines 10-17):

The challenge lies in designing methodologies to reduce spin-up period in computationally intensive integrated hydrologic models such as ParFlow.CLM (Ashby and Falgout, 1996; Jones and Woodward, 2001; Kollet and Maxwell, 2006) when initialization from equilibrium states is required for transient simulations. In integrated hydrologic models like ParFlow, numerical solution of the Richards equation in 3D increases computational time (Kim et al., 1997; Maxwell et al., 2014) in comparison to approaches that use a 1D Richards equation for the vadoze zone and a 2D groundwater flow formulation for simulating subsurface flow.

*2. The title “Reducing the spin-up time” appears to be misleading in the sense, whether it is reducing the computation time itself or the number of iterative years required, it needs to be cleared. Eg. “Reducing the spin-up period”...*

In the revised manuscript, we changed spin-up time to spin-up period throughout the text.

*3. Model description is absent. Which version of the model is being used here, is it the terrain following co-ordinate system or the older version? Are the catchments delineated*

*for the simulation or a box domain is used? This needs to be all clarified. If the terrain following co-ordinate is used, the number of vertical levels can be reduced. In addition, the real computation time can also be reduced using delineated catchments.*

We included a brief description of the ParFlow.CLM as well as the version numbers in the revised manuscript. The terrain following coordinate system was not used in our simulations. We used the box domain in our simulations to reduce the impact of boundary conditions on catchment scale fluxes.

A brief description of ParFlow.CLM is added at the beginning of section 2.1 as follows:

ParFlow is a 3D variably saturated groundwater flow model that solves the mixed form of the three-dimensional Richards equation for the subsurface (Ashby and Falgout, 1996; Jones and Woodward, 2001; Maxwell et al., 2014). ParFlow has a fully integrated overland flow simulator (Kollet and Maxwell, 2006) and performs routing of the ponded water on the land surface via the kinematic wave equation. The Common Land Model (CLM 3.0) (Dai *et al.*, 2003) is integrated into ParFlow to simulate water and energy fluxes at the land surface (Maxwell and Miller, 2005; Kollet and Maxwell, 2008). ParFlow.CLM versions 605 and 653 were used for the Skjern River and Baldry simulations respectively, which are described below. The terrain following grid of Maxwell (2013) is not implemented in these modelling set-ups.

Section 2.1.1 is revised as follows (Page 4, Lines 28-29 and Page 5, Lines 1-2):

To reduce the impact of boundary conditions on catchment scale fluxes, the computational domain is extended beyond the delineated catchment boundary. As such, the ParFlow.CLM model domain covered a 28 km by 20 km area that encompasses the Skjern River sub-catchment (Fig. 2).

Section 2.1.2 is revised as follows (Page 5, Lines 24-25):

The ParFlow.CLM model of the site was set up over a 2.9 by 2.9 km area encompassing the Baldry sub-catchment (Fig. 2) in order to reduce the impact of boundary conditions on catchment scale fluxes.

*4. In both studies, spatially uniform atmospheric forcing is used, could this be possibly one of the reason why the domain mean DTWT function performs well for the relatively flat topography used in this study. How will it effect the empirical DTWT functions, if spatially varying forcing is used? A case study with relatively larger extent, and spatially varying forcing should be presented to prove the presented methodology for its suitability in other applications.*

We agree with the reviewer comments regarding examining the performance of DTWT functions across sites with steep topography, larger extent and spatially varying forcing. However, performing additional experiments requires huge computational demand and it is beyond the scope of this Technical Note. We revised the summary section to include this point (Page 15, Lines 20-21):

In addition, the role of topography and spatially distributed forcing should be further examined.

*5. Pg. 6977, Ln 1-17, This paragraph is very confusing, show the formulation of calculation of MAE and RMSD in terms of the grid points, and then proceed to discussion, else the figure says otherwise. What does the mean DTWT in y-axis refer to, is it the*

domain mean or catchment mean? Fig. 3B is addressed before discussion about Fig. 3 itself.

We included the formula for calculating the objective functions in the methodology section (Page \*, Lines 6-15):

Root mean square difference (RMSD), mean absolute error (MAE) and bias were computed to find the best performing DTWT function. These objective functions are calculated as follows:

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (B_i - M_i)^2} \quad (3)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |B_i - M_i| \quad (4)$$

$$\%bias = \frac{\sum_{i=1}^N (B_i - M_i)}{\sum_{i=1}^N (B_i)} \times 100 \quad (5)$$

where N is the number of grid cells in the domain, B is the mean annual DTWT from the baseline simulation of Ajami et al. (2014) for every grid cell, and M is the estimated mean annual DTWT at a grid cell obtained from a DTWT function.

We revised the Results section. Y-axis refers to the domain mean. We revised section 3.1 as follows:

Optimum parameter values for single and double exponential DTWT functions were obtained using nonlinear least squares method. Performance of the single and double exponential DTWT functions in predicting 14 years of DTWT were compared against ParFlow.CLM baseline spin-up simulations (years 7 through 20) of Ajami et al. (2014) to find the optimum empirical DTWT function for the Skjern River sub-catchment. Post-simulation analysis indicates that global DTWT functions based on domain or catchment averaged percentage change values are better predictors of DTWT response compared to local DTWT functions developed for every grid cell. Instability of local DTWT functions occurs in grid cells where percent changes in DTWT oscillate between positive and negative values through initial spin-up simulations. Spatial distribution of these grid cells are shown in Fig S1.

Calculated RMSD and percent bias relative to the baseline spin-up simulations indicate that global double exponential functions using ParFlow.CLM spin-up simulations 2 to 6 provide a better fit compared to various single exponential functions obtained from different spin-up simulation years (e.g. 2 to 3, 2 to 4, etc.). Because the first six cycles of ParFlow.CLM simulations were the same between the baseline spin-up simulations and DTWT distributions

from DTWT functions presented in Fig 4, comparisons were made with simulations 7 to 20 of the baseline spin-up approach of Ajami et al. (2014).

As can be seen from Fig. 4a, the mean annual DTWT over the domain derived from the single exponential functions (fitted to percentage change data from simulations 2 to 6) underpredict the baseline spin-up simulations, due to their consistent small underestimates in comparison to double exponential functions fitted to the same data points. Only for the mean absolute error (MAE) calculated at each pixel do single exponential functions based on simulations 2 to 6 perform slightly better and produce smaller errors on average than the double exponential functions (Fig. 4b). It should be noted that the percent bias in mean annual DTWT for simulation cycle 20 is -1.6% for the domain based double exponential function and -6.2% for the single exponential function, with both functions derived from simulation cycles 2 to 6. Therefore, single exponential functions are not further examined in re-initializations of the DTWT. In terms of mean DTWT across the domain (Fig. 4a), the catchment delineated double exponential DTWT function provides a better prediction and the smallest mean bias when compared to the function based on the entire model domain. However, Figure 4b indicates that the mean absolute error values are slightly smaller for the domain based double exponential function. The higher MAE of the catchment based double exponential function is a result of slightly more regions with over and underestimated DTWT values that contribute to a good overall mean DTWT (Fig. 4a), but contains more errors spatially compared to the domain based double exponential function.

#### **Minor Comments:**

*1.Ln 6970, Ln. 24 : Rephrase.*

We rephrased the sentence in the revised manuscript (Page 2, Lines 12-14):

Since information on the spatial pattern of water table and soil moisture distributions is generally unavailable, various approaches have been developed to determine the initial DTWT variation.

*2.Pg. 6971, Ln 23: spin-up period*

It is revised.

*3.Pg. 6971, Ln 27: number of years of spin-up required for*

The sentence is revised (Page 3, Lines 19-20):

The objective of the current study is to develop a *hybrid* spin-up approach that significantly reduces the number of years of spin-up required for model state equilibrium.

*4.Pg. 6972, Ln 4: Confusing statement, Fig. 1 mentions 3 stages, but the paragraph begins with two stages.*

The sentence is modified (Page 4, Lines 2-3):

The hybrid approach consists of three main stages: a two-stage model simulation step and an intermediate state-updating step using the DTWT function.

*5.Pg. 6972, Ln 7: “against the equilibrated initial condition for the subcatchment of : : ..., using the ParFlow.CLM model.”*

It is revised. See Page 4, Lines 4-7:

The utility of the proposed scheme is compared against the equilibrated initial condition for a sub-catchment of the Skjern River basin in Denmark, using the ParFlow.CLM model as developed by Ajami et al. (2014) that employed a traditional baseline spin-up approach.

*6.Pg. 6972, Ln 20: Mention grid point numbers. Also mention the annual precipitation received and min-max annual temperatures in the text.*

This section is revised as follows:

The modelling grid had a horizontal resolution of 500 m and a vertical discretization of 0.5 m. Catchment topography was determined via a 500 m digital elevation model (DEM) and the bottom elevation of the domain was a uniform -75 m, resulting in a  $56 \times 40 \times 406$  dimension grid...

Page 5, Lines 11-12:

In 2003, annual precipitation was 801.6 mm and minimum and maximum daily air temperature were 261.2 K and 295.2 K respectively.

*7.Pg. 6973, Ln 24: Also mention the annual precipitation received and min-max annual temperatures in the text. Why 400m deep layer here?*

The annual precipitation and min-max annual temperature is added for the Baldry sub-catchment in the revised manuscript (Page 6, Lines 7-8):

In 2004, annual precipitation was 674.8 mm, and minimum and maximum daily air temperature were 277 K and 305.5 K respectively.

The subsurface thickness was not 400 m. The bottom elevation of the domain was 400m. The manuscript is revised as follows (Page 5, Lines 27-28):

The bottom elevation of the modelling grid was a uniform 400 m resulting in a subsurface thickness of 43 to 101 m across the computational domain.

*8.Pg. 6974, Ln 4: Does these function depend on the initial condition of prescribed groundwater table depth ?*

The coefficients and shape of these functions depends on the initial condition of prescribed groundwater table depth. The sentence is revised as follows (Page 6, Lines 15-21):

Analysis of ParFlow.CLM spin-up behavior using the baseline spin-up approach for the sub-catchment of the Skjern River identified that percentage changes in subsurface storages and DTWT had the form of an exponential decay for a model initialized from a uniform 3 m DTWT (Ajami et al., 2014). Due to spatial adjustment of the water table during the spin-up, groundwater levels declined near the catchment divide and reached the land surface along the channel network causing an overall decline in mean annual DTWT relative to the initial condition.

*9.Pg. 6975, Ln7-14: Paragraph not comprehensible. Rephrase.*

The paragraph is revised (Page 7, Lines 28-32, and Page 8, Lines 1-6):

The empirical DTWT functions calculated above estimate percentage changes in mean annual DTWT as a function of simulation year. To predict spatially distributed mean annual DTWT from a global DTWT function, the mean annual DTWT from the final cycle of the ParFlow.CLM spin-up simulation for every grid cell is used as the initial value to



successively estimate DTWT distributions as a function of simulation year. These DTWT distributions are based on the predicted percent change values from the global DTWT function. Sensitivity of DTWT functions to the number of ParFlow.CLM cycles was also examined by developing a number of DTWT functions using data from 2 to 6 cycles of ParFlow.CLM. To assess the performance of these DTWT functions, estimated mean annual DTWT from the DTWT functions were compared against mean annual DTWT from the ParFlow.CLM model of Ajami et al. (2014) that had been spun-up for 20 years.

10.Pg.6975, Ln 19: *predicted for : : :...*

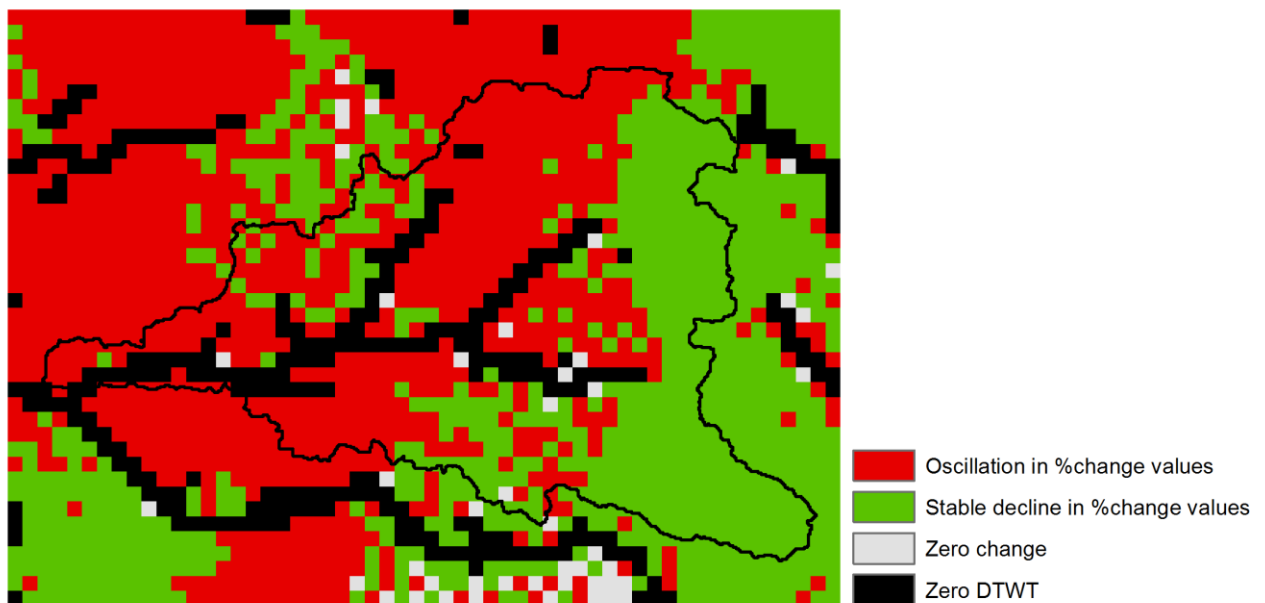
It is revised.

11.Pg. 6976, Ln 20: *It has be to discussed clearly, whether the computation domain consists of delineated catchment or a rectangular domain in the experiment description itself.*

This point is clarified in the revised manuscript. The computational domain consists of the rectangular domain which includes the delineated catchment.

12.Pg. 6976. Ln 24: *A plot showing these oscillations will be illustrative.*

A Figure is added to the supplementary information that illustrates regions in the Skjern River sub-catchment were changes in DTWT oscillate.



**Fig. S1.** Delineating four regions in the modelling domain according to percent changes in mean annual DTWT values from six cycles of ParFlow.CLM simulations. Red region represents grid cells where percent changes in DTWT values oscillate between negative and positive values, while the green region corresponds to grid cells with stable decline in DTWT. In grey regions, percent changes in DTWT have reached zero and black region corresponds to the channel network, where DTWT is zero.

13.Pg. 6977, Ln 19: *Show the formulation of semi-variograms calculations.*

The formulation of semi-variograms are included in the revised manuscript. See page 11- Lines 14-23:

To investigate this result further, three empirical semi-variograms were generated. As the impact of an east-west spatial trend in the mean annual DTWT values was evident in the semi-variograms, the trend should first be removed from the mean annual DTWT values. To remove the trend, a plane was fitted to the observed mean annual DTWT values, with an equation of the form:

$$z = ax + by + c \quad (6)$$

where  $a$ ,  $b$  and  $c$  are fitted coefficients,  $x$  and  $y$  are the coordinates of every grid cell, and  $z$  is the mean annual DTWT. Residuals are computed by subtracting the estimated mean annual DTWT from Equation 6 from the observed mean annual DTWT values. Finally, the semi-variogram of the residuals as a function of distance is calculated.

14. Pg. 6979, Ln 2: *Is this the result from the simulation using the initial condition from the different methods ?*

yes. To clarify this point, the manuscript is revised as follows (Page 13, Lines 14-18):

At equilibrium, differences in simulated DTWT from the last day of the ParFlow.CLM simulations after re-initializations (hydrostatic equilibrium and adjusted pressure head distribution) and the baseline spin-up approach varied by up to 2 m inside the catchment boundary (Fig. 6), although most areas were within 0.5m.

15.Pg. 6980, Ln 11: *Is it the case in reality?*

In our simulations, we assumed that the Baldry sub-catchment is covered by the plantation forest as indicated in the methodology section. In reality half of the catchment is covered by pasture and stage recordings at the catchment outlet indicate 335 days of no flow in 2004.

16.Pg. 6980, Ln 17: *“simulation year”. Is the computation time same for each year of simulation. Does is also exhibit some pattern?*

We do not have the exact computation time for each year of simulation. In general, the computational time slightly decreases as the system equilibrates but the reduction in computational time is not significant.

17.Pg. 6981: Ln 1-7: *Far-fetching conclusions. Please remove it.*

We removed this statement and revised the manuscript as follows:

Previous efforts in calibrating coupled or integrated hydrologic models required a spin-up process after every parameter update (Stisen et al., 2011; Weill et al., 2013). Development of a computationally efficient spin-up approach will enable this type of systematic calibration of integrated or coupled hydrologic models.

18.Pg. 6981: Ln 9: *“reducing number of years to ...”*

It is revised.

19.Pg. 6981: Ln 10: *“spin-up years”*

It is revised.

# Technical Note: Reducing the spin-up ~~time period~~ of integrated surface water-groundwater models

H. Ajami<sup>1,2</sup>, J. P. Evans<sup>3,4</sup>, M. F. McCabe<sup>5</sup>, S. Stisen<sup>6</sup>

[1]{School of Civil and Environmental Engineering, University of New South Wales, Sydney, Australia}

[2]{Connected Waters Initiative Research Centre, University of New South Wales, Sydney, Australia}

[3]{Climate Change Research Centre, University of New South Wales, Sydney, Australia}

[4]{ARC Centre of Excellence for Climate System Science, University of New South Wales, Sydney, Australia}

[5]{Water Desalination and Reuse Center, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia}

[6]{Geological Survey of Denmark and Greenland, Copenhagen, Denmark}

Correspondence to: H. Ajami (h.ajami@unsw.edu.au)

## Abstract

One of the main challenges in ~~catchment-scale~~the application of coupled ~~or~~ /integrated hydrologic models, is specifying a catchment's initial conditions in terms of soil moisture and depth to water table (DTWT) distributions. One approach to reduce uncertainty in model initialization is to run the model recursively using ~~either~~ a single or multiple years of forcing data until the system equilibrates with respect to state and diagnostic variables. However, such "spin-up" approaches often require many years of simulations, making them computationally intensive. In this study, a new hybrid approach was developed to reduce the computational burden of ~~the spin-up procedure time for an integrated groundwater surface water-land surface model (ParFlow-CLM)~~ by using a combination of ~~ParFlow-CLM~~model simulations and an empirical DTWT function. The methodology is examined ~~in~~-~~across~~ two ~~distinct~~ catchments located in ~~the a~~ temperate ~~region of Denmark~~ and ~~a~~ semi-arid regions of ~~Denmark and~~ Australia ~~respectively~~. Our results illustrate that the hybrid approach reduced

1 the spin-up ~~period time~~ required for an integrated groundwater-surface water-land surface  
2 model (ParFlow-CLM) by ParFlow-CLM by up to 50%; To generalize results to different  
3 climate and catchment conditions, and we outline a methodology that is applicable to other  
4 coupled / or integrated modelling frameworks when initialization from an equilibrium state is  
5 required.

## 7 **1 Introduction**

8 The issue of model initialization is important for hydrologic simulation and predictions<sub>s</sub>, as the  
9 ~~model~~ initial state has a major impact on a catchment's modelled response (Berthet et al.,  
10 2009). In coupled / or integrated surface-subsurface models<sub>s</sub>, uncertainty in a catchment  
11 antecedent condition is of particular importance ~~because as~~ both the soil moisture distribution  
12 and depth to water table (DTWT) need to be specified at the start of a simulation (Ivanov et  
13 al., 2004; Noto et al., 2008).

15 Since ~~there is often no a priori~~ information on the spatial pattern of water table and soil  
16 moisture distributions is generally unavailable, various approaches have been developed to  
17 determine the initial DTWT variation. Sivapalan et al. (1987) used a topography-soil index to  
18 map the spatial distribution of initial DTWT. In another approach, Troch et al. (1993) used  
19 recession flow analysis to estimate the effective water table height of a catchment. Regardless  
20 of the choice of initial DTWT, the uncertainty involved is such that a period of spin-up is  
21 always required (Cloke et al., 2003)<sub>s</sub>, as the applied atmospheric forcing is often inconsistent  
22 with the hydrodynamic initialization of the catchment inferred from limited observations  
23 (Ajami et al., 2014).

25 The two most common initialization approaches in coupled / or integrated distributed  
26 hydrologic models are ~~as follows~~: (1) initial depth to water table is specified at a certain  
27 uniform depth below the land surface (Kollet and Maxwell, 2008)<sub>s</sub>, ~~with and~~ the impact of  
28 initialization is reduced through recursive simulations over either a single or multiple years of  
29 forcing data, until equilibrium conditions are reached, which are: usually related to spin-up  
30 criteria based on changes in groundwater heads (Refsgaard, 1997) or changes in water and  
31 energy balances (Kollet and Maxwell, 2008); or (2) the model is initialized from a fully

1 saturated condition and simulations are continued until modelled baseflow matches the  
2 observations (Jones et al., 2008). [Equilibrium based initializations have been utilized](#)  
3 [previously for exploring land surface-groundwater coupling \(Kollet and Maxwell, 2008\) and](#)  
4 [assessing the impact of climate change on groundwater-land surface interactions using an](#)  
5 [integrated hydrologic model \(Ferguson and Maxwell, 2010\).](#)

6  
7 Results of a ParFlow.CLM spin-up study for a catchment in Denmark showed that at least 20  
8 years of recursive simulations were required to reach equilibrium in subsurface storages,  
9 defined as [occurring](#) when percent changes in monthly unsaturated and saturated zone  
10 storages were less than 0.1% and 0.01% respectively (Ajami et al., 2014). For reference, 20  
11 years of spin-up simulations required 20,000 service units ([a service unit is equivalent to 1](#)  
12 [hour of time used by one processor](#)) on a high performance parallel computing cluster:  
13 equivalent to over 26 days of computation using 32 processors. The challenge lies in  
14 designing methodologies to reduce spin-up [period time](#)—in computationally intensive  
15 integrated [hydrologic](#) models such as ParFlow.CLM (Ashby and Falgout, 1996; Jones and  
16 Woodward, 2001; Kollet and Maxwell, 2006) when initialization from equilibrium states is  
17 required for transient simulations. [In integrated hydrologic models like ParFlow, numerical](#)  
18 [solution of the Richards equation in 3D increases computational time \(Kim et al., 1997;](#)  
19 [Maxwell et al., 2014\) in comparison to approaches that use a 1D Richards equation for the](#)  
20 [vadoze zone and a 2D groundwater flow formulation for simulating subsurface flow.](#)

21  
22 The objective of the current study is to develop a *hybrid* spin-up approach that significantly  
23 reduces the [number of years of spin-up time](#) required for model state equilibrium. [The](#)  
24 [equilibrium based initialization represents a correct initial state for catchments in which the](#)  
25 [land use does not change over time and the inter-annual variability of atmospheric forcing is](#)  
26 [very small: assumptions that are common to most simulation frameworks. This technical note](#)  
27 [provides a method for improving the efficiency of this commonly used initialization](#)  
28 [technique.](#) The performance of the proposed approach in reducing the spin-up [period time](#)—for  
29 a catchment scale application of the ParFlow.CLM model is evaluated against ~~a~~-[the](#) standard  
30 continuous recursive simulation approach [that is](#) commonly applied for land surface model  
31 spin-up, and referred to here as ~~a~~-[the](#) baseline spin-up approach.

## 2 Data and methodology

The hybrid approach consists of three main stages: a two-stage model simulation step and an intermediate state-updating step using the DTWT function. Figure 1 illustrates this hybrid spin-up approach. The ~~applicability-utility~~ of the proposed scheme is ~~examined-compared~~ against the equilibrated initial condition for a sub-catchment of the Skjern River basin in Denmark, using the ParFlow.CLM model ~~as of the sub-catchment of the Skjern River basin in Denmark,~~ developed by Ajami et al. (2014) ~~using that employed~~ a traditional baseline spin-up approach. ~~Further~~An additional assessment of ~~performance of~~ the hybrid approach in reducing the spin-up period time is ~~evaluated-undertaken~~ by developing ~~a~~ ParFlow.CLM ~~model~~ for a semi-arid catchment in Australia.

### 2.1 Overview of the ParFlow.CLM models

ParFlow is a 3D variably saturated groundwater flow model that solves the mixed form of the three-dimensional Richards equation for the subsurface (Ashby and Falgout, 1996; Jones and Woodward, 2001; Maxwell et al., 2014). ParFlow has a fully integrated overland flow simulator (Kollet and Maxwell, 2006) and performs routing of the ponded water on the land surface via the kinematic wave equation. The Common Land Model (CLM 3.0) (Dai et al., 2003) is integrated into ParFlow to simulate water and energy fluxes at the land surface (Maxwell and Miller, 2005; Kollet and Maxwell, 2008). ParFlow.CLM versions 605 and 653 were used for the Skjern River and Baldry simulations respectively, which are described below. The terrain following grid of Maxwell (2013) is not implemented in these modelling set-ups.

#### 2.1.1 Sub-catchment of the Skjern River Basin, Denmark

The sub-catchment of the Skjern River basin in western Denmark has an area of 208 km<sup>2</sup> (Fig. 2) ~~that and~~ is characterized by mild topography and a temperate climate (Jensen and Illangasekare, 2011). Agricultural land is the dominant cover type (78%), with the remainder of the catchment area covered by evergreen needle leaf forest. To reduce the impact of boundary conditions on catchment scale fluxes, the computational domain is extended beyond the delineated catchment boundary. As such, ~~T~~the catchment's ParFlow.CLM model domain



1 covered a 28 km by 20 km area [that encompasses the Skjern River sub-catchment \(Fig. 2\)](#).  
2 The modelling grid had a horizontal resolution of 500 m and a vertical discretization of 0.5 m.  
3 Catchment topography was determined via a 500 m digital elevation model (DEM) and the  
4 bottom elevation of the domain was a uniform -75 m, [resulting in a 56 × 40 × 406 dimension](#)  
5 [grid](#).

6  
7 At the land surface, the ParFlow free-surface overland flow boundary condition was assigned.  
8 A no-flow boundary condition was specified for the sides and bottom boundary. Spatially  
9 uniform hourly atmospheric forcing (air temperature, wind speed, specific humidity, air  
10 pressure, precipitation, incoming shortwave and downward longwave radiation) for the year  
11 2003 ~~were~~ [was](#) used for spin-up. [In 2003, annual precipitation was 801.6 mm and minimum](#)  
12 [and maximum daily air temperature were 261.2 K and 295.2 K respectively](#). Initial DTWT  
13 was assigned uniformly at 3m below the land surface. Ground surface temperature was set to  
14 the mean annual air temperature (281 K) at the start of a simulation. Prescribed subsurface  
15 hydraulic parameters include the saturated hydraulic conductivity ( $0.3 \text{ m h}^{-1}$ ), porosity (0.39),  
16 van Genuchten parameters ( $\alpha=1.5 \text{ m}^{-1}$  and  $n=2$ ), and relative residual saturation (0.1).

### 18 **2.1.2 Baldry sub-catchment, Australia**

19 The Baldry sub-catchment, located in central west New South Wales of Australia, has an area  
20 of  $1.9 \text{ km}^2$  with an elevation range from 443 m to 500 m [inside the catchment boundary](#) (Fig.  
21 2). For the spin-up experiment, the catchment land cover was assumed to be evergreen  
22 broadleaf forest representing eucalyptus plantation.

23  
24 The ParFlow.CLM model of the site was set up over a 2.9 by 2.9 km area encompassing the  
25 Baldry [sub-catchment \(Fig. 2\) in order to reduce the impact of boundary conditions on](#)  
26 [catchment scale fluxes](#). Catchment topography was represented using a 60 m pre-processed  
27 DEM. The bottom elevation of the modelling grid was a uniform 400 m [resulting in a](#)  
28 [subsurface thickness of 43 to 101 m across the computational domain](#). The modelling grid had  
29 a 60 m resolution in the x and y directions and ~~its a~~ vertical discretization ~~was of~~ 0.5 m,  
30 resulting in a  $48 \times 48 \times 203$  [dimension](#) grid.

1 As for the Skjern River implementation, At the land surface, the ParFlow free-surface  
2 overland flow boundary condition was assigned at the land surface. A no-flow boundary  
3 condition was specified for the lateral and bottom boundaries of the computational domain  
4 (gray domain in Fig. 2). Hourly forcing data for the year 2004 were obtained from a weather  
5 station at the site. For the hourly downward longwave radiation, the Modern Era  
6 Retrospective Analysis for Research and Applications (MERRA) reanalyses data interpolated  
7 to  $0.25^\circ \times 0.25^\circ$  resolution ~~was~~ were used (Decker et al., 2012). In 2004, annual precipitation  
8 was 674.8 mm, and minimum and maximum daily air temperature were 277 K and 305.5 K  
9 respectively. Prescribed subsurface hydraulic parameters include the saturated hydraulic  
10 conductivity ( $0.18 \text{ m h}^{-1}$ ), porosity (0.25), van Genuchten parameters ( $\alpha=1.5 \text{ m}^{-1}$  and  $n=2$ ),  
11 and relative residual saturation (0.1). The model was initialized with a uniform DTWT of 2m  
12 below the land surface. Ground surface temperature was set to mean annual air temperature  
13 288.1 [K].

## 15 **2.2 Development of empirical DTWT functions for model re-initialization**

16 Analysis of ParFlow.CLM spin-up behavior ~~via~~ using the baseline spin-up approach for the  
17 sub-catchment of the Skjern River identified that percentage changes in subsurface storages  
18 and DTWT had the form of an exponential decay for a model initialized from a uniform 3 m  
19 DTWT (Ajami et al., 2014). Due to spatial adjustment of the water table during the spin-up,  
20 groundwater levels declined near the catchment divide and reached the land surface along the  
21 channel network causing an overall decline in mean annual DTWT relative to the initial  
22 condition. Using the functional relationships between the number of simulation years and the  
23 percentage change of a variable, Ajami et al. (2014) ~~developed~~ produced a series of spin-up  
24 functions based on 16 years of initial ParFlow.CLM simulations. These spin-up functions  
25 were used to predict the number of years required until the model equilibrated, based on a  
26 predefined threshold i.e. 0.1% or 0.01% change for a given variable. Sensitivity of spin-up  
27 functions across multiple criteria and variables showed that the estimated spin-up period  
28 based on mean annual DTWT were more stable when compared to other spin-up criteria, such  
29 as changes in the mean DTWT for the last day of recursive simulations (Ajami et al., 2014).

1 ~~Conversely,~~†The inverse of a spin-up function for DTWT predicts percent changes in DTWT  
2 as a function of simulation years, which hereinafter is referred to as the empirical DTWT  
3 function. In this study, we examined the capabilities of empirical DTWT functions as a means  
4 for updating DTWT and hence groundwater storage after just a few initial ParFlow.CLM  
5 spin-up simulations. The expectation is that this state updating should reduce the total number  
6 of spin-up years of simulation, substantially reducing the computational burden. To do this, a  
7 series of spin-up simulations were performed based on an arbitrary initial state (DTWT ~~was of~~  
8 3 m below the land surface ~~for the Skjern River sub-catchment sub-catchments,~~ –as in Ajami  
9 et al. (2014)), in order to identify the minimum number of data points required to develop an  
10 empirical DTWT function (stage 1 of model simulation).

11  
12 Due to the anticipated large changes in mean annual DTWT values between the first and  
13 second year of the spin-up simulation, the first year of data is removed from the analysis. As a  
14 minimum of four data points (i.e. 6 cycles of ParFlow.CLM simulations) are required to fit a  
15 double exponential function, six years of spin-up simulations are performed using a single  
16 year of forcing data. To assess the sensitivity of DTWT functions to the number of  
17 ParFlow.CLM cycles, various exponential functions with single (Eq. 1) or double exponential  
18 (Eq. 2) terms are fit to the ParFlow.CLM cycles 2 to 6:

19 
$$y = a \exp(bx) \tag{1}$$

20 
$$y = a \exp(bx) + c \exp(dx) \tag{2}$$

21 where  $y$  is the percentage change in DTWT,  $x$  is the number of simulation years, and  $a, b, c, d$   
22 are the fitting parameters. Coefficient of determination and root mean square error are used as  
23 goodness of fit measures. This means that a minimum of four data points (i.e. 6 cycles of  
24 ParFlow.CLM simulations) are required to fit a double exponential function to percentage  
25 change values. Therefore, six years of spin-up simulations were performed using forcing data  
26 for the year 2003. Further, the performances of global versus local scales DTWT functions are  
27 evaluated. In this analysis, a domain based global DTWT function is based on percent  
28 changes in mean annual DTWT values from all the grid cells inside the computational  
29 domain, while a catchment based DTWT function is based on the grid cells inside the  
30 catchment boundary. Local DTWT functions are developed for every grid cell based on  
31 percent changes in mean annual DTWT values at that grid cell. Percent changes in catchment  
32 and domain averaged annual DTWT values, as well as changes in DTWT at every grid cell,

1 were used to fit exponential functions at the global and local scales respectively, using single  
2 (Eq. 1) or double exponential (Eq. 2) terms:

$$3 \quad y = a \exp(bx) \quad (1)$$

$$4 \quad y = a \exp(bx) + c \exp(dx) \quad (2)$$

5 where  $y$  is the percentage change in DTWT,  $x$  is the number of simulation years, and  $a, b, c, d$  are the  
6 fitting parameters.

7  
8 ~~These~~The empirical DTWT functions calculated above estimate percentage changes in mean  
9 annual DTWT as a function of simulation years. ~~Depending on the number of ParFlow.CLM~~  
10 ~~cycles used to fit the DTWT functions (i.e. 2 to 6, cycles),~~ To predict spatially distributed  
11 mean annual DTWT estimate from a global DTWT function, the mean annual DTWT from  
12 the ~~last~~ final cycle of the ParFlow.CLM spin-up simulation for every grid cell ~~was~~ is used as  
13 the initial value to successively estimate DTWT distributions as a function of simulation year.  
14 These DTWT distributions are based on the predicted percent change values from the global  
15 DTWT function. Sensitivity of DTWT functions to the number of ParFlow.CLM cycles was  
16 also examined by developing a number of DTWT functions using data from 2 to 6 cycles of  
17 ParFlow.CLM. To assess the performance of these DTWT functions, estimated mean annual  
18 DTWT from the DTWT functions were compared against mean annual DTWT from the  
19 ParFlow.CLM model of Ajami et al. (2014) that had been spun-up for 20 years. - Root mean  
20 square difference (RMSD), mean absolute error (MAE) and bias were computed to find the  
21 best performing DTWT function. These objective functions are calculated as follows:

$$23 \quad RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (B_i - M_i)^2} \quad (3)$$

$$24 \quad MAE = \frac{1}{N} \sum_{i=1}^N |B_i - M_i| \quad (4)$$

$$25 \quad \%bias = \frac{\sum_{i=1}^N (B_i - M_i)}{\sum_{i=1}^N (B_i)} \times 100 \quad (5)$$

1 [where N is the number of grid cells in the domain, B is the mean annual DTWT from the](#)  
2 [baseline simulation of Ajami et al. \(2014\) for every grid cell, and M is the estimated mean](#)  
3 [annual DTWT at a grid cell obtained from a DTWT function.](#)

4  
5 In the state updating stage, the best performing empirical DTWT function (a double  
6 exponential DTWT function [as discussed in section 3.1](#)) was used to estimate percentage  
7 changes in DTWT as a function of simulation years, until percentage changes reached the  
8 0.01% threshold. Using the percent change values and mean annual DTWT distribution from  
9 the sixth cycle of the ParFlow.CLM spin-up simulation, spatially distributed DTWT was  
10 predicted [for the entire computational domain](#).

11  
12 In the second stage of model simulations, the ParFlow.CLM was re-initialized using newly  
13 estimated DTWT values from a double exponential DTWT function, and spin-up simulations  
14 were continued until equilibration based on subsurface storage spin-up criteria. The second  
15 stage of spin-up simulations was necessary to ensure equilibrium after re-initialization,  
16 especially for the unsaturated zone storage.

17  
18 One issue with the re-initialization of DTWT using the DTWT function is that the distribution  
19 of soil moisture above the water table cannot be estimated. Here, we considered two  
20 approaches to define pressure head distribution above the water table: (1) implementing the  
21 commonly used hydrostatic equilibrium assumption, where pressure head at the water table  
22 was linearly decreased as a function of elevation head towards the land surface; and (2)  
23 adjusting the pressure head distribution of the unsaturated zone from the last day of the sixth  
24 cycle of ParFlow.CLM spin-up simulations based on new DTWT values [from the DTWT](#)  
25 [function. In the adjusted pressure head approach, the hydrostatic equilibrium assumption is](#)  
26 [used in regions between the new DTWT and the initial DTWT. The ParFlow.CLM pressure](#)  
27 [head distribution is adjusted to begin at the new pressure head from the initial WT such that](#)  
28 [the vertical profile is maintained \(Fig. 3\). This adjustment may represent a lack of consistency](#)  
29 [in the proposed approach as the DTWT function estimates mean annual DTWT, while](#)  
30 [pressure head adjustments in the unsaturated zone are taken from the last day of the sixth](#)  
31 [cycle of ParFlow.CLM. While it is possible to use DTWT values from the last day of](#)  
32 [simulations to develop a DTWT function, estimated DTWT values from such a function](#)  
33 [exhibit larger variability and result in a larger bias.](#)

1 [For the Skjern River sub-catchment, percent bias- between the estimated DTWT values from](#)  
2 [the DTWT functions and the baseline simulation of Ajami et al. \(2014\) were -4% and -1.6%](#)  
3 [for the DTWT functions based on the last day and mean annual DTWT values respectively.](#)

### 5 **2.3. ~~Evaluation~~ Evaluating of the hybrid spin-up approach**

6 [The P](#)performance of the hybrid spin-up approach in reducing the spin-up ~~timeperiod~~ [-is](#)  
7 evaluated by developing a ParFlow.CLM model for the Baldry sub-catchment. The baseline  
8 spin-up simulations were performed using spatially uniform, hourly forcing data for the year  
9 2004 [and an arbitrary initial state \(DTWT of 2 m below the land surface\)](#). The equilibrium  
10 condition was achieved when percent changes in catchment averaged monthly groundwater  
11 storages were below 0.1% threshold level. [Similar to the Skjern River sub-catchment,](#)  
12 [sensitivity of empirical DTWT functions to the number of ParFlow.CLM cycles \(i.e. cycles 2](#)  
13 [to 6\) are explored and assessed against the baseline spin-up simulations of Baldry sub-](#)  
14 [catchment.](#) In the next step, the hybrid spin-up approach outlined in section 2.2 was  
15 implemented to re-initialize the ParFlow.CLM using [a domain based double exponential](#)  
16 [DTWT function to estimate spatially distributed DTWT and](#) the adjusted pressure head  
17 distribution approach above the water table. Recursive simulations after re-initialization  
18 continued until equilibrium condition was achieved.

## 20 **3 Results**

### 21 **3.1 Performance of empirical DTWT functions in predicting DTWT in the** 22 **Skjern River sub-catchment**

23 [Optimum parameter values for single and double exponential DTWT functions were obtained](#)  
24 [using nonlinear least squares method.](#) Performance of the single and double exponential  
25 DTWT functions in predicting 14 years of DTWT were compared against ParFlow.CLM  
26 baseline spin-up simulations (years 7 through 20) of Ajami et al. (2014) to find the optimum  
27 empirical DTWT function for the Skjern River sub-catchment. Post-simulation analysis  
28 indicates that global DTWT functions based on domain or catchment averaged percentage  
29 change values are better predictors of DTWT response compared to local DTWT functions



1 developed for every grid cell. Instability of local DTWT functions occurs in grid cells where  
2 percent changes in DTWT oscillate between positive and negative values through initial spin-  
3 up simulations. Spatial distribution of these grid cells are shown in Fig S1.

4  
5 Calculated ~~root mean square difference (RMSD)~~ and percent bias relative to the baseline spin-  
6 up simulations indicate that global double exponential functions using ParFlow.CLM spin-up  
7 simulations 2 to 6 provide a better fit compared to various single exponential functions  
8 obtained from different spin-up simulation years (e.g. 2 to 3, 2 to 4, etc.). Because the first six  
9 cycles of ParFlow.CLM simulations were the same between the baseline spin-up simulations  
10 and DTWT distributions from DTWT functions presented in Fig 4, comparisons were made  
11 with simulations 7 to 20 of the baseline spin-up approach of Ajami et al. (2014).

12  
13 ~~Only for the mean absolute error (MAE) do single exponential functions based on simulations~~  
14 ~~2 to 6 perform better than the double exponential functions (Fig. 3b). Because the first six~~  
15 ~~cycles of ParFlow.CLM simulations were the same between the baseline spin-up simulations~~  
16 ~~and DTWT distributions presented in Fig 3, comparisons were made with simulations 7 to 20~~  
17 ~~of the baseline spin-up approach of Ajami et al. (2014).~~ As can be seen from Fig. 34a, the  
18 mean annual DTWT over the domain derived from the single exponential functions (fitted to  
19 percentage change data from simulations 2 to 6) under-predict the baseline spin-up  
20 simulations, due to their consistent small underestimates in comparison to double exponential  
21 functions fitted to the same data points. ~~Only for the mean absolute error (MAE) calculated at~~  
22 each pixel do single exponential functions based on simulations 2 to 6 perform slightly better  
23 and produce smaller errors on average than the double exponential functions (Fig. 34b). It  
24 should be noted that the percent bias in mean annual DTWT for simulation cycle 20 is -1.6%  
25 for the domain based double exponential function and -6.2% for the single exponential  
26 function, with both functions derived from simulation cycles 2 to 6. Therefore, single  
27 exponential functions are not further examined in re-initializations of the DTWT. In terms of  
28 mean DTWT across the domain (Fig. 34a), the catchment delineated double exponential  
29 DTWT function provides a better prediction and the smallest mean bias when compared to the  
30 function based on the entire model domain. However, Figure 34b indicates that the mean  
31 absolute error values are slightly smaller for the domain based double exponential function.  
32 The higher MAE of the catchment based double exponential function is a result of slightly  
33 more regions with over and underestimated DTWT values that contribute to a good overall

1 mean DTWT (Fig. 4a), but contain more errors spatially compared to the domain based  
2 double exponential function.

3  
4 To investigate this result further, three empirical semi-variograms were generated. As the  
5 impact of an east-west spatial trend in the mean annual DTWT values was evident in the  
6 semi-variograms, the trend should first be removed ~~ing the trend~~ from the mean annual  
7 DTWT values, using a polynomial function and calculating the semi-variance of the residuals  
8 as a function of distance. To remove the trend, a plane was fitted to the observed mean annual  
9 DTWT values, with an equation of the form:

$$10 \quad z = ax + by + c \quad (6)$$

11 where  $a$ ,  $b$  and  $c$  are fitted coefficients,  $x$  and  $y$  are the coordinates of every grid cell, and  $z$  is  
12 the mean annual DTWT. Residuals are computed by subtracting the estimated mean annual  
13 DTWT from Equation 6 from the observed mean annual DTWT values. Finally, the semi-  
14 variogram of the residuals as a function of distance is calculated. The semi-variance is a  
15 measure ~~for of~~ spatial variance ~~of a variable~~ and presents average (dis-)similarity between  
16 data pairs at a given distance. Investigating the empirical semi-variograms of mean annual  
17 DTWT values (Fig. 3e4c) indicates that the domain based double DTWT function is a better  
18 predictor of mean annual DTWT, because the spatial structure of DTWT is sufficiently  
19 reproduced by the domain based function, and the catchment based function ~~had~~ has a higher  
20 variance compared to the baseline simulations. Therefore, it is recommended to use the  
21 domain based DTWT function, as it contains data from high elevation regions on the eastern  
22 side of the domain that contribute to topographically driven flow and equilibrates slower  
23 than in other regions (Ajami et al., 2014). In summary, double exponential functions are  
24 chosen as they have less bias compared to single exponential functions and there is very little  
25 difference in terms of MAE amongst predictions. The choice is further supported by the  
26 RMSD and semi-variograms.

### 3.2 Impact of unsaturated zone re-initialization on ParFlow.CLM spin-up

Impacts of re-initializing the unsaturated zone using the hydrostatic equilibrium versus adjusted vertical pressure distribution on the spin-up time-period were also explored using the ParFlow.CLM simulations of the Skjern River sub-catchment. As can be seen from Fig. 43, the difference between the two initialization methods is more pronounced in areas of deep water table, where hydrostatic pressure head distribution results in a drier unsaturated zone compared to adjusted pressure head distribution. Results indicate that after re-initialization, the system equilibrated after 6 additional years of spin-up simulation when using the hydrostatic equilibrium option. With the adjusted pressure head distribution option, only 4 additional years of spin-up simulation were required. Therefore, depending on the pressure head distribution above the water table, either 10 or 12 years of ParFlow.CLM simulations were sufficient to ensure subsurface storage equilibrium, reducing the spin-up time by 40% or 50%, compared to the baseline spin-up approach.

The ~~Improved~~-improved performance of the adjusted pressure distribution is related to the fact that information about soil moisture distribution from stage 1 of spin-up simulations is preserved in this approach. In both initialization approaches, the groundwater storage was equilibrated at the 0.01% threshold level, based on changes in mean monthly values. In comparison to the baseline spin-up approach, both groundwater and unsaturated zone storages of the equilibrium year are closely reproduced by the adjusted pressure head distribution option (Fig. 5). While in both re-initializations, DTWT and subsequently groundwater storage volume were the same at the start of the simulations, unsaturated zone storage of the hydrostatic equilibrium option was drier than the adjusted pressure head option. Additional ParFlow.CLM simulations after re-initialization ensured equilibrium of groundwater storage. As can be seen from Fig. 6a, hydrostatic re-initialization results in a deeper WT at equilibrium (simulation 12) relative to the baseline equilibrium year (simulation 20). For the hydrostatic equilibrium—Higher DTWT values of the hydrostatic option at equilibrium correspond to smaller groundwater storage and subsequently larger, —increases in—unsaturated zone storage compared to the baseline spin-up (Fig. 5). It should be noted that in ParFlow.CLM, groundwater and unsaturated zone storages are not explicitly determined by fixed size compartments and the extent of an unsaturated zone is determined by the location of the water table. Percent changes in mean annual unsaturated zone storage between the last two years of

1 [recursive simulations were 0.1% for the hydrostatic equilibrium and 0.3% for the adjusted](#)  
2 [pressure head re-initializations, indicating ~~resulted in~~ unsaturated zone equilibrium at](#)  
3 [different threshold levels.](#)

4  
5 Changes in annual water balance after re-initialization were also compared against the  
6 baseline spin-up approach of Ajami et al. (2014). While changes in annual evapotranspiration  
7 were approximately 1 mm between the two spin-up approaches (annual baseline  
8 evapotranspiration of 447.3 mm), percent bias in annual discharge against observations  
9 decreased by about 2% compared to the baseline approach (Table 1). In the hybrid approach,  
10 changes in groundwater storage were positive, because after re-initialization, DTWT  
11 decreased as simulations proceed and the system reached equilibrium (Fig. 5). [At equilibrium,](#)  
12 [differences in simulated DTWT from the last day of the ParFlow.CLM simulations after re-](#)  
13 [initializations \(hydrostatic equilibrium and adjusted pressure head distribution\)](#) and the  
14 baseline spin-up approach varied by up to 2 m inside the catchment boundary (Fig. 6),  
15 although most areas were within 0.5m. Differences were more pronounced in areas of higher  
16 elevation in the catchment. Figure 6 shows that the hydrostatic equilibrium pressure head  
17 adjustment leads to a clear bias with consistent over estimation of the DTWT, while the  
18 adjusted vertical pressure distribution produces a distribution of pressure head errors centred  
19 on the expected value.

### 21 **3.3 Evaluation of the hybrid spin-up approach [over the Baldry sub-catchment](#)**

22 Similar to the Skjern River sub-catchment, percent changes in monthly groundwater storages  
23 were used to assess [the](#) equilibrium condition. However, for the Baldry sub-catchment, a  
24 threshold level of 0.1% was chosen as the convergence criterion. Results indicated that 28  
25 years of recursive simulations were required until the model equilibrated based on monthly  
26 groundwater storage changes. For reference, 28 years of baseline spin-up simulations for  
27 Baldry required 37000 service units, [which is](#) equivalent to 24 days of computation using 64  
28 processors of a high performance computing cluster.

1 [To obtain optimum parameter values for single and double exponential DTWT functions](#)  
2 [nonlinear least squares method is used.](#) Similar to the Skjern River sub-catchment, a double  
3 exponential DTWT function using simulations 2 to 6 resulted in WT distributions with the  
4 smallest RMSD and percent bias, relative to the baseline spin-up simulations. For Baldry, a  
5 domain based double exponential function had the closest mean [over the domain](#) (Fig.  
6 [S1aS2a](#)) and the smallest mean absolute error relative to the baseline simulation (Fig.  
7 [S1bS2b](#)). However, DTWT semi-variograms showed higher variances in the domain based  
8 double exponential function relative to the catchment based function ([Fig. S2c](#)). Despite the  
9 slight differences in the predictive power of DTWT functions between the two catchments, a  
10 double exponential function seems to perform best based on most of the criteria.

11  
12 To re-initialize the ParFlow.CLM model of the Baldry sub-catchment, the new DTWT  
13 distribution obtained from the domain based double exponential function was used. After re-  
14 initialization with the adjusted pressure head distribution, only 8 additional simulation [years](#)  
15 were required until percent changes in monthly groundwater storages reached below the 0.1%  
16 level. This result indicates a 50% reduction in the spin-up [period time](#) of a semi-arid  
17 catchment when the hybrid spin-up approach is used.

18  
19 Comparison of WT distributions from the last day of equilibrium simulations (baseline  
20 simulation and the hybrid approach) illustrated differences of up to 1 m (Fig. 7). However, for  
21 the majority of cells inside the catchment, differences were [up to less than](#) 0.5 m. Similar to  
22 the Skjern River sub-catchment, the largest differences in WT distribution were observed in  
23 higher elevation areas [in the southern part of the catchment](#). Lower WT levels in the hybrid  
24 spin-up approach resulted in larger unsaturated zone storage compared to the baseline spin-up  
25 (Fig. [S2S3](#)). In this semi-arid catchment, no stream flow was generated at the catchment's  
26 outlet for the equilibrium year. The difference in annual evapotranspiration was only 0.2 mm  
27 between the two equilibrium simulations (Table 2).

#### 28 29 **4 Summary**

30 We present a hybrid approach for reducing the [number-period](#) of spin-up [simulations](#)  
31 [required to reach equilibrium with](#) the integrated hydrological model ParFlow.CLM. In the

1 | case of the Skjern River and the Baldry sub-catchments, ~~the simulation period~~ ~~time year~~  
2 | decreased by 50% compared to the baseline spin-up approach when an adjusted pressure head  
3 | distribution was specified above the water table. Although, ParFlow.CLM was used as a  
4 | modelling platform, the developed methodology is applicable to other coupled ~~or~~ /integrated  
5 | hydrologic models.

6 |  
7 | ~~Therefore, a~~ general approach to spin-up should include the following steps: (1) perform six  
8 | years of hydrologic model simulations with DTWT initialized via ~~an expert~~  
9 | ~~knowledge~~/informed guess (here 2m and 3m below the land surface for the Baldry and Skjern  
10 | River sub-catchments respectively); (2) calculate a global double exponential DTWT function  
11 | using domain wide data and estimate the new DTWT for the desired equilibration level; (3)  
12 | implement the adjusted pressure head approach for the unsaturated zone initialization; and (4)  
13 | continue spin-up simulations until desired equilibration level is reached. ~~Beyond being used~~  
14 | ~~to define initial states of the model, this process has the potential to assist in parameter~~  
15 | ~~calibration.~~ Previous efforts in calibrating coupled ~~or~~ /integrated hydrologic models required a  
16 | spin-up process after every parameter update (Stisen et al., 2011; Weill et al., 2013).  
17 | Development of ~~a computationally efficient spin-up the hybrid spin-up~~ approach ~~could be one~~  
18 | ~~step towards is required for will~~ enabling ~~this type of~~ systematic calibration of integrated ~~or~~  
19 | ~~coupled~~ hydrologic models.

20 |  
21 | ~~However, further refinement is required to facilitate automatic calibration approaches.~~  
22 | Additional experiments across multiple catchments with different climate and subsurface  
23 | heterogeneity and DTWT initializations are ~~also~~ required to assess the efficiency of the  
24 | proposed approach in reducing ~~number of years~~ ~~the simulation period~~ ~~time~~ to equilibration in a  
25 | variety of settings. ~~In addition, the role of topography and spatially distributed forcing should~~  
26 | ~~be further examined.~~ Reducing the required spin-up ~~time-years~~ of ~~coupled or~~ integrated/  
27 | ~~coupled~~ hydrologic models will expand their application for hydrological investigations and  
28 | facilitate the use of these models to investigate both real world and theoretical system  
29 | behavior.



## 1 **Acknowledgements**

2 This research was funded by the Australian Research Council and the National Water  
3 Commission through support of the National Centre for Groundwater Research and Training  
4 (NCGRT). [Jason Evans was supported by the Australian Research Council Future Fellowship](#)  
5 [FT110100576](#). We acknowledge the support provided by the National Computational  
6 Infrastructure at the Australian National University through the National Computational Merit  
7 Allocation Scheme, Intersect partner share and University of New South Wales LIEF grant  
8 (LE120100181) for high performance computing. The use of high performance computing  
9 facilities at the University of New South Wales is also appreciated. We acknowledge the  
10 HOBE hydrological observatory and the New South Wales Department of Primary Industries  
11 for providing the Skjern River and Baldry data respectively. [We would also like to thank the](#)  
12 [two anonymous reviewers for their valuable and insightful comments.](#)  
13

## 1 **References**

- 2 Ajami, H., McCabe, M. F., Evans, J. P., and Stisen, S.: Assessing the impact of model spin-up  
3 on surface water-groundwater interactions using an integrated hydrologic model, *Water*  
4 *Resour. Res.*, 50, 2636-2656, doi:10.1002/2013WR014258, 2014.
- 5 Ashby, S. F., and Falgout, R. D.: A parallel multigrid preconditioned conjugate gradient  
6 algorithm for groundwater flow simulations, *Nuclear Science and Engineering*, 124, 145-159,  
7 1996.
- 8 Berthet, L., Andréassian, V., Perrin, C. and Javelle, P.: How crucial is it to account for the  
9 antecedent moisture conditions in flood forecasting? Comparison of event-based and  
10 continuous approaches on 178 catchments, *Hydrol. Earth Syst. Sci.*, 13(6), 819-831, 2009
- 11 Cloke, H. L., Renaud, J. P., Claxton, A. J., McDonnell, J. J., Anderson, M. G., Blake, J. R.,  
12 and Bates, P. D.: The effect of model configuration on modelled hillslope-riparian  
13 interactions, *J. Hydrol.*, 279, 167-181, doi:10.1016/S0022-1694(03)00177-X, 2003.
- 14 [Dai, Y., Zeng, X., Dickinson, R. E., Baker, I., Bonan, G.B., Bosilovich, M. G., Denning, A.](#)  
15 [S., Dirmeyer, P. A. Houser, P. R. Niu, G., Oleson, K. W., Schlosser, C. A., and Yang, Z.L.:](#)  
16 [The Common Land Model, \*Bulletin of the American Meteorological Society\*, 84\(8\), 1013-](#)  
17 [1023, 2003.](#)
- 18 Decker, M., Pitman, A. J., and Evans, J. P.: Groundwater constraints on simulated  
19 transpiration variability over southeastern Australian forests, *Journal of Hydrometeorology*,  
20 14(2), 543-559, 2012.
- 21 [Ferguson, I. M., and Maxwell, R. M.: Role of groundwater in watershed response and land](#)  
22 [surface feedbacks under climate change, \*Water Resour. Res.\*, 46\(10\), W00F02,](#)  
23 [doi:10.1029/2009WR008616, 2010.](#)
- 24 Ivanov, V. Y., Vivoni, E. R., Bras, R. L., and Entekhabi, D.: Catchment hydrologic response  
25 with a fully distributed triangulated irregular network model, *Water Resour. Res.*, 40,  
26 W11102, doi:10.1029/2004wr003218, 2004.
- 27 Jensen, K. H., and Illangasekare, T. H.: HOBE: A Hydrological Observatory, *Vadose Zone*  
28 *Journal*, 10, 1-7, doi:10.2136/vzj2011.0006, 2011.

1 Jones, J. E., and Woodward, C. S.: Newton–Krylov-multigrid solvers for large-scale, highly  
2 heterogeneous, variably saturated flow problems, *Adv. Water Resour.*, 24, 763-774, doi:  
3 10.1016/s0309-1708(00)00075-0, 2001.

4 Jones, J. P., Sudicky, E. A., and McLaren, R. G.: Application of a fully-integrated surface-  
5 subsurface flow model at the watershed-scale: A case study, *Water Resour. Res.*, 44,  
6 W03407, doi:10.1029/2006wr005603, 2008.

7 [Kim, C.P., Stricker, J. N.M. and Feddes, R. A.: Impact of soil heterogeneity on the water  
8 budget of the unsaturated zone, \*Water Resour. Res.\*, 33,991-999, doi: 10.1029/97WR00364,  
9 1997.](#)

10 Kollet, S. J., and Maxwell, R. M.: Integrated surface–groundwater flow modeling: A free-  
11 surface overland flow boundary condition in a parallel groundwater flow model, *Adv. Water  
12 Resour.*, 29, 945-958, doi:10.1016/j.advwatres.2005.08.006, 2006.

13 Kollet, S. J., and Maxwell, R. M.: Capturing the influence of groundwater dynamics on land  
14 surface processes using an integrated, distributed watershed model, *Water Resour. Res.*, 44,  
15 W02402, doi:10.1029/2007wr006004, 2008.

16 [Maxwell, R.M.: A terrain-following grid transform and preconditioner for parallel, large-  
17 scale, integrated hydrologic modeling, \*Adv. Water Resour.\*, 53, 109-117, doi:  
18 10.1016/j.advwatres.2012.10.001, 2013.](#)

19 [Maxwell, R.M, and Miller, N.L.: Development of a coupled land surface and groundwater  
20 model. \*Journal of Hydrometeorology\*, 6: 233-247, doi:10.1175/JHM422.1, 2005](#)

21 [Maxwell, R. M., Putti, M., Meyerhoff, S., Delfs, J., Ferguson, I.M., Ivanov, V., Kim, J.,  
22 Kolditz, O., Kollet, S.J., Kumar, M., Lopez, S., Niu, J., Paniconi, C., Park, Y., Phanikumar,  
23 M.S., Shen, C., Sudicky, E.A., and Sulis, M.: Surface-subsurface model intercomparison: A  
24 first set of benchmark results to diagnose integrated hydrology and feedbacks, \*Water Resour.  
25 Res.\*, 50, doi:10.1002/2013WR013725, 2014.](#)

26 Noto, L. V., Ivanov, V. Y., Bras, R. L., and Vivoni, E. R.: Effects of initialization on response  
27 of a fully-distributed hydrologic model, *J. Hydrol.*, 352, 107-125, doi:  
28 10.1016/j.jhydrol.2007.12.031, 2008.

29 Refsgaard, J. C.: Parameterisation, calibration and validation of distributed hydrological  
30 models, *J. Hydrol.*, 198, 69-97, doi:10.1016/s0022-1694(96)03329-x, 1997.

- 1 Sivapalan, M., Beven, K., and Wood, E. F.: On hydrologic similarity: 2. A scaled model of  
2 storm runoff production, *Water Resour. Res.*, 23, 2266-2278,  
3 doi:10.1029/WR023i012p02266, 1987.
- 4 Stisen, S., Sonnenborg, T. O., Hojberg, A. L., Troldborg, L., and Refsgaard, J. C.: Evaluation  
5 of climate input biases and water balance issues using a coupled surface-subsurface model,  
6 *Vadose Zone Journal*, 10, 37-53, doi:10.2136/vzj2010.0001, 2011.
- 7 Troch, P. A., Mancini, M., Paniconi, C., and Wood, E. F.: Evaluation of a distributed  
8 catchment scale water balance model, *Water Resour. Res.*, 29, 1805-1817,  
9 doi:10.1029/93wr00398, 1993.
- 10 Weill, S., Altissimo, M., Cassiani, G., Deiana, R., Marani, M., and Putti, M.: Saturated area  
11 dynamics and streamflow generation from coupled surface–subsurface simulations and field  
12 observations, *Adv. Water Resour.*, 59, 196-208, doi:10.1016/j.advwatres.2013.06.007, 2013.  
13

1 **Table 1.** Skjern River sub-catchment annual water balance for the equilibrium year after the  
 2 baseline spin-up approach and the hybrid approach using the hydrostatic equilibrium and  
 3 adjusted pressure head distribution options above the water table. Annual precipitation is  
 4 801.6 mm.

Simulation Name	Number of simulations	%bias <sup>1</sup> Q	ET [mm/y]	dS <sup>2</sup> GW <sup>3</sup> [mm]	dS UZ <sup>4</sup> [mm]
ParFlow baseline simulation	20	20.3	447.3	-3.3	-0.2
ParFlow+DTWT function (Hydrostatic equilibrium)	12	18.1	446.8	3.3	-0.7
ParFlow+DTWT function (Adjusted Pressure)	10	18.5	446.3	3	-1.6

5 <sup>1</sup>Percent bias is based on observed discharge at the gauge shown in Fig. 2:-

6 <sup>2</sup>changes in storage; <sup>3</sup>groundwater storage; <sup>4</sup>unsaturated zone storage

7

8

9

10

11

12

13

14

15

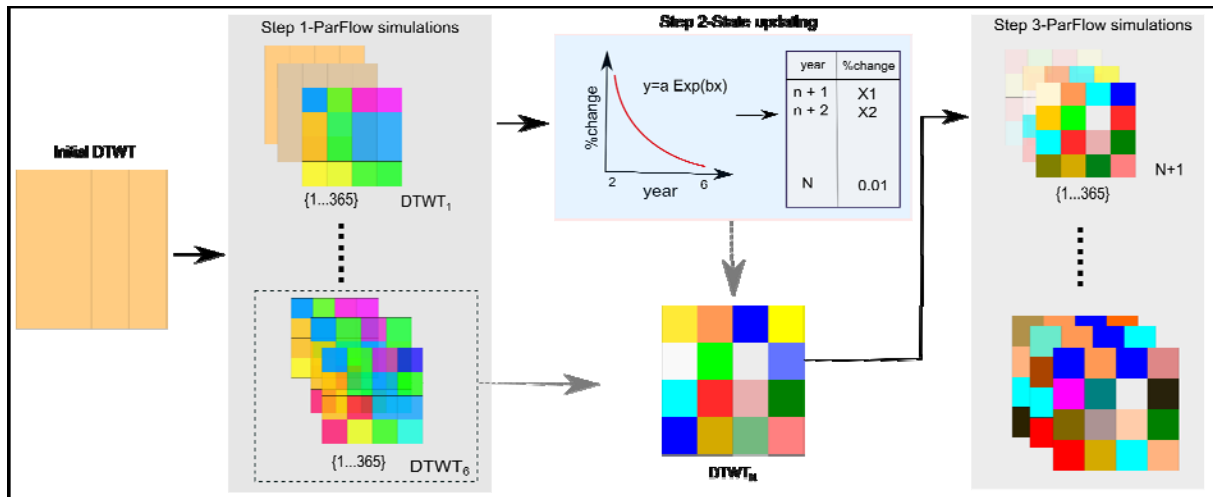
1 **Table 2.** Baldry sub-catchment annual water balance for the equilibrium year after the  
 2 baseline spin-up approach and the hybrid approach with the adjusted pressure head  
 3 distribution option above the water table. Annual precipitation is 674.8 mm.

Simulation Name	Number of Simulations	ET [mm/y]	dS <sup>1</sup> GW <sup>2</sup> [mm]	dS UZ <sup>3</sup> [mm]
ParFlow baseline simulation	28	519.2	-12	4.4
ParFlow+DTWT function (Adjusted Pressure)	14	519	-0.2	1.0

4 <sup>1</sup>changes in storage; <sup>2</sup>groundwater storage; <sup>3</sup>unsaturated zone storage

5

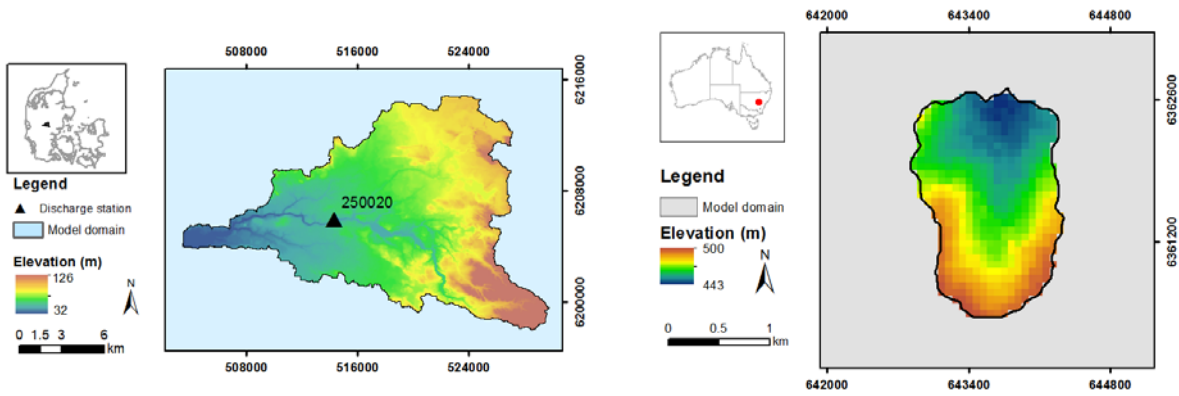
6



1  
 2 **Figure 1.** The hybrid spin-up approach consists of three main steps: (1) initial ParFlow.CLM  
 3 spin-up simulations based an arbitrary DTWT distribution, (2) state updating step by  
 4 developing a DTWT function based on percent changes in mean annual DTWT in initial spin-  
 5 up simulations, and (3) stage 2 of ParFlow.CLM spin-up simulations until desired  
 6 equilibration level is reached.  
 7



1



2

3 **Figure 2.** Sub-catchment of the Skjern river basin located in western Denmark (reproduced  
4 from Ajami et al. 2014) (left), and Baldry sub-catchment in Australia (right). Modelling  
5 domains –are extended beyond the catchment boundary to remove the impact of boundary  
6 conditions on catchment fluxes.

7

8

9

10

11

12

13

14

15

16

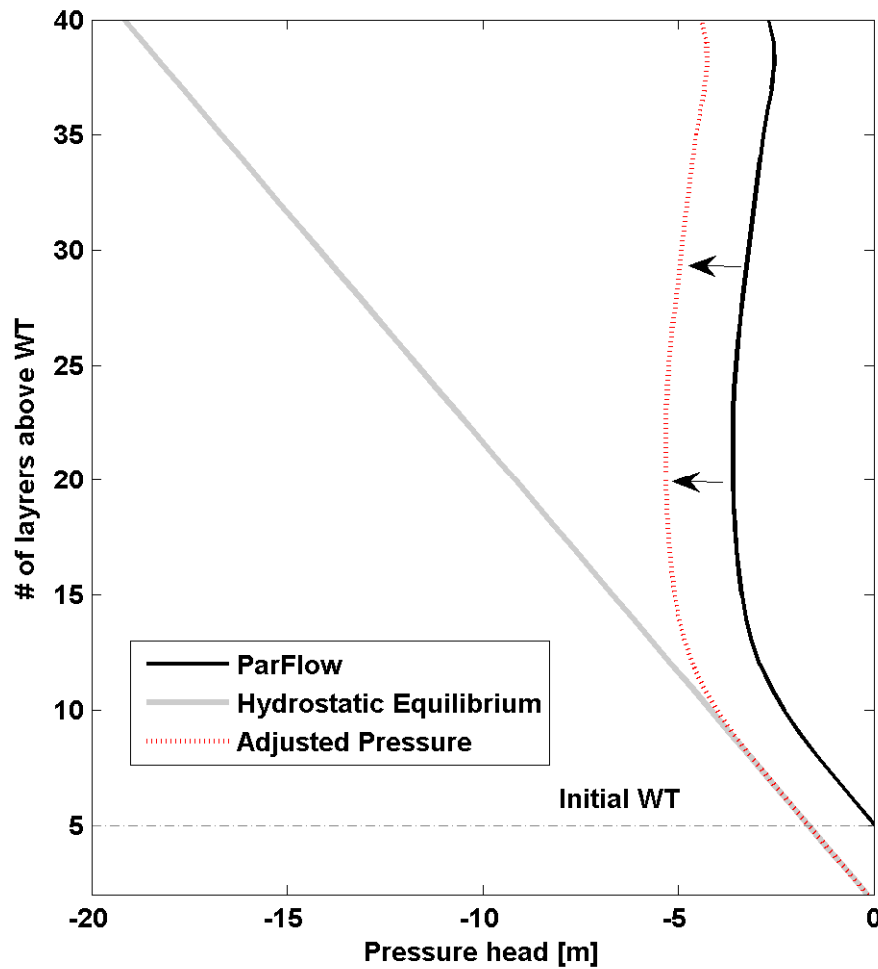
17

18

19

20

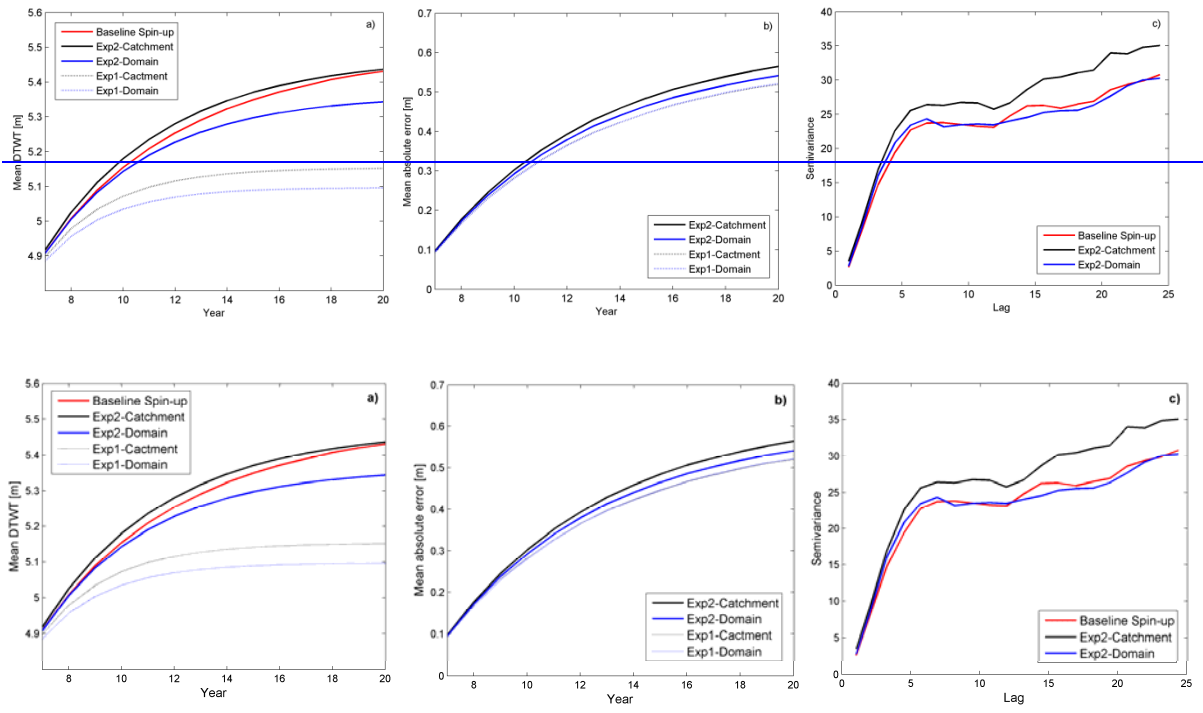
1



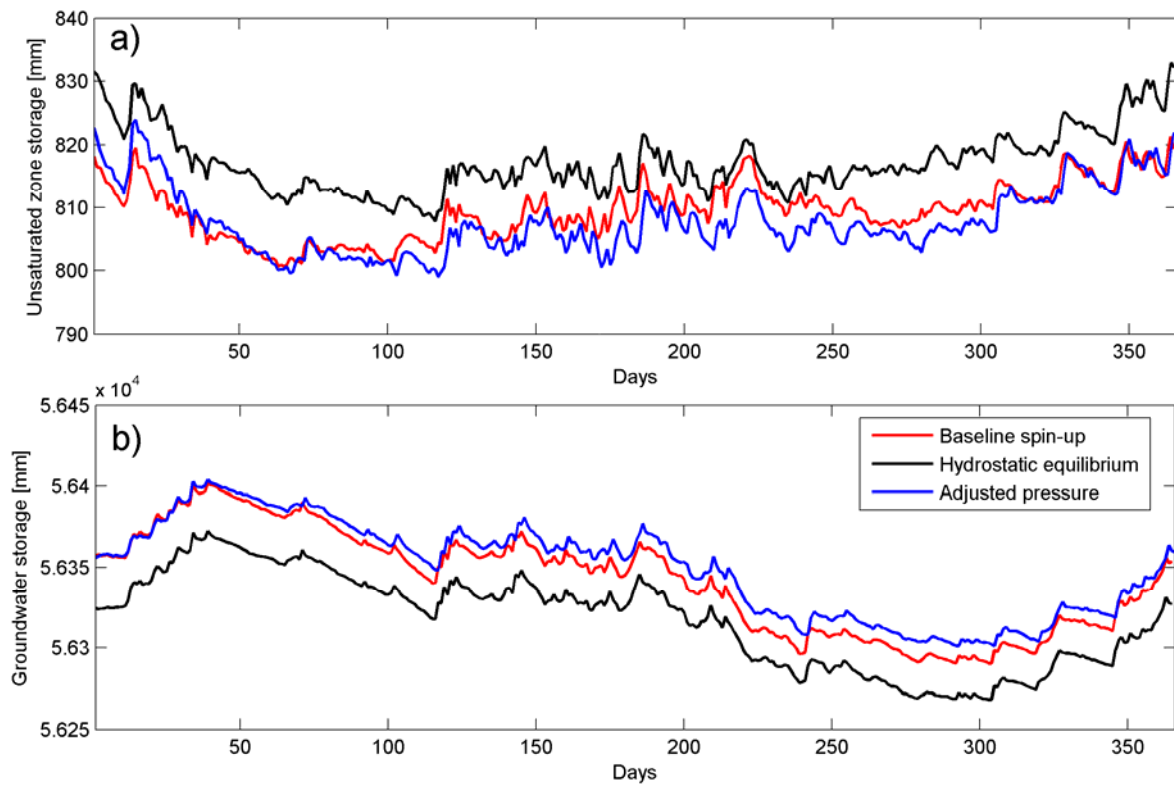
2

3 **Figure 3.** Adjusted pressure head distribution above the estimated DTWT from the DTWT  
4 function. Pressure head distribution of the last day of ParFlow.CLM spin-up simulation 6 was  
5 adjusted at every grid cell based on the position of DTWT estimated from the DTWT  
6 function. In this approach, the hydrostatic equilibrium assumption is used in regions between  
7 the new DTWT and the initial DTWT. The ParFlow.CLM pressure head distribution is  
8 adjusted to begin at the new pressure head from the initial WT such that the vertical profile is  
9 maintained. Hydrostatic pressure distribution is shown as a reference for the new DTWT,  
10 which is lower than the WT in simulation 6.

11

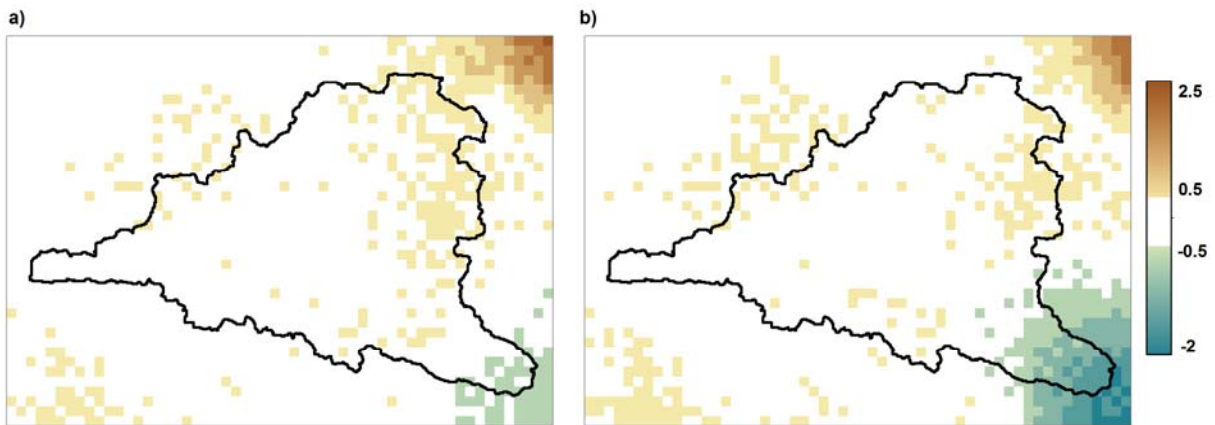
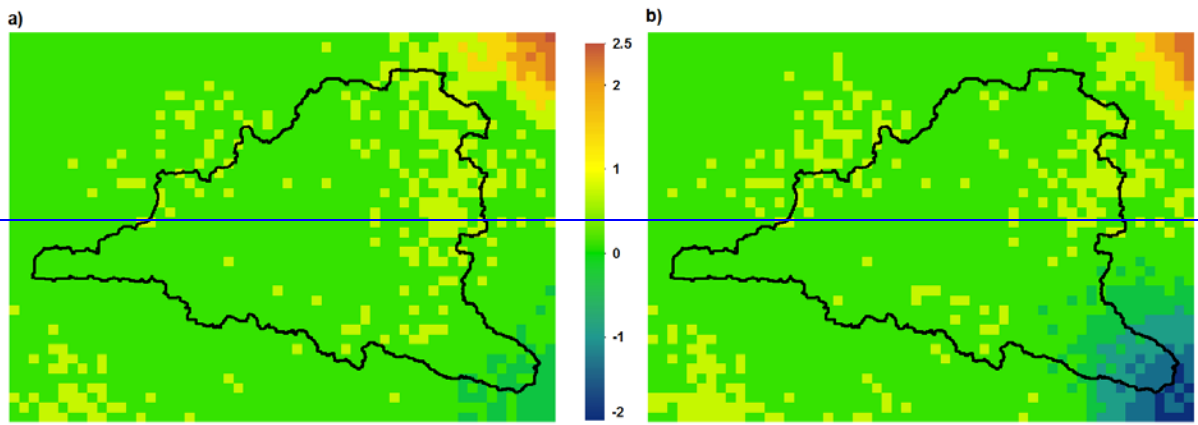


**Figure 34.** a) Comparison between the simulated mean annual DTWT obtained from the baseline spin-up approach of ParFlow.CLM of the Skjern River sub-catchment and empirical DTWT functions. The single exponential model was formulated using the domain and catchment averaged data from spin-up simulations 2 to 6; b) estimated mean absolute error based on simulated DTWT from the baseline spin-up together with both catchment and domain averaged single and double exponential functions; and c) experimental semi-variograms of mean annual DTWT from ParFlow.CLM equilibrium year (after 20 years of simulations) and DTWT from catchment and domain averaged double exponential functions, showing that catchment based semi-variances are higher than the baseline simulation. Exp1 and Exp2 refer to single and double exponential functions respectively.

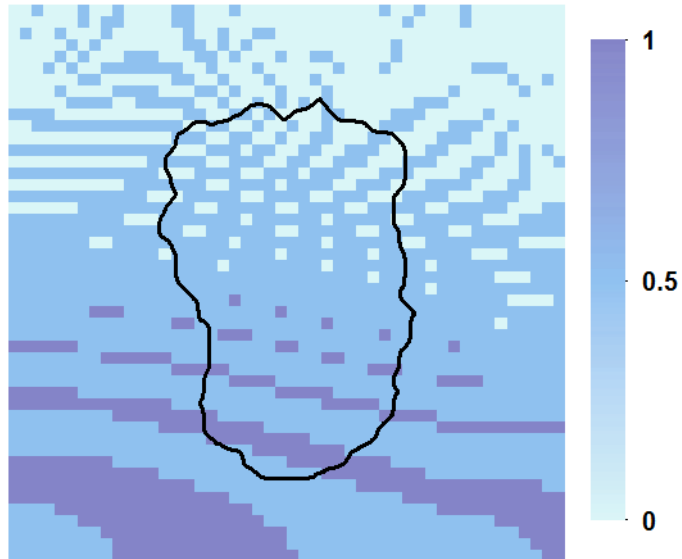


1  
2  
3  
4  
5  
6  
7  
8  
9

**Figure 5.** Comparison of a) unsaturated and b) groundwater storages of ParFlow.CLM equilibrium year using the hybrid and baseline spin-up approaches (Ajami et al., 2014). [The equilibrium year corresponds to simulation cycles of 10, 12 and 20 for the adjusted pressure head, hydrostatic equilibrium, and baseline simulations, respectively.](#) The dynamics of groundwater and unsaturated zone storages are closely reproduced by the adjusted pressure head distribution approach relative to the baseline spin-up approach for the Skjern River sub-catchment.



**Figure 6.** Differences in equilibrium DTWT between ParFlow.CLM simulations after re-initializations and ParFlow.CLM after 20 years of baseline spin-up simulations in (m), where a) is based on hydrostatic pressure distribution above the water table for the initial condition, while b) is based on adjusted pressure head distribution above the water table for the Skjern River sub-catchment. White regions correspond to grid cells where the differences in equilibrium DTWT are less than 0.5.



1

2 **Figure 7.** Differences in equilibrium DTWT of Baldry ParFlow.CLM simulations after re-  
3 initialization with the adjusted pressure head distribution above the water table and  
4 ParFlow.CLM after 28 years of baseline spin-up simulations in (m). The contours of DTWT  
5 overestimation are along the direction of flow lines from high elevation areas toward the  
6 catchment outlet.

7

8

9

10

11

12

13

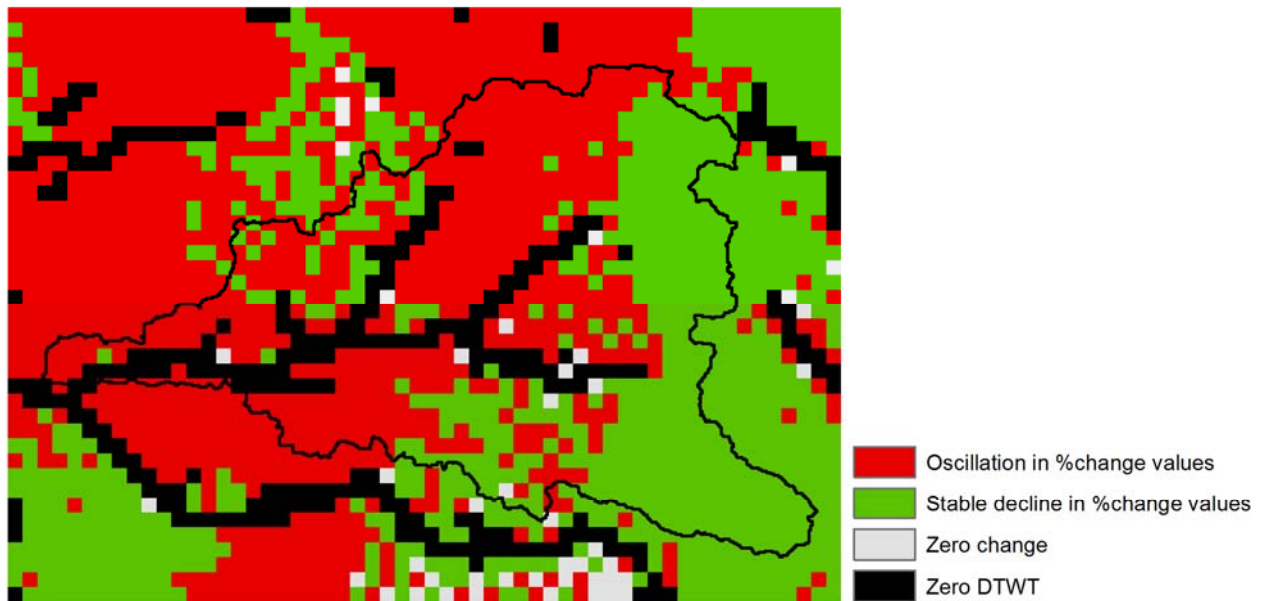
14

15

16

17

18



2

3 [Fig. S1. Delineating four regions in the modelling domain according to percent changes in](#)  
4 [mean annual DTWT values from six cycles of ParFlow.CLM simulations. Red region](#)  
5 [represents grid cells where percent changes in DTWT values oscillate between negative and](#)  
6 [positive values, while the green region corresponds to grid cells with stable decline in DTWT.](#)  
7 [In grey regions, percent changes in DTWT have reached zero and black region corresponds to](#)  
8 [the channel network, where DTWT is zero.](#)

9

10

11

12

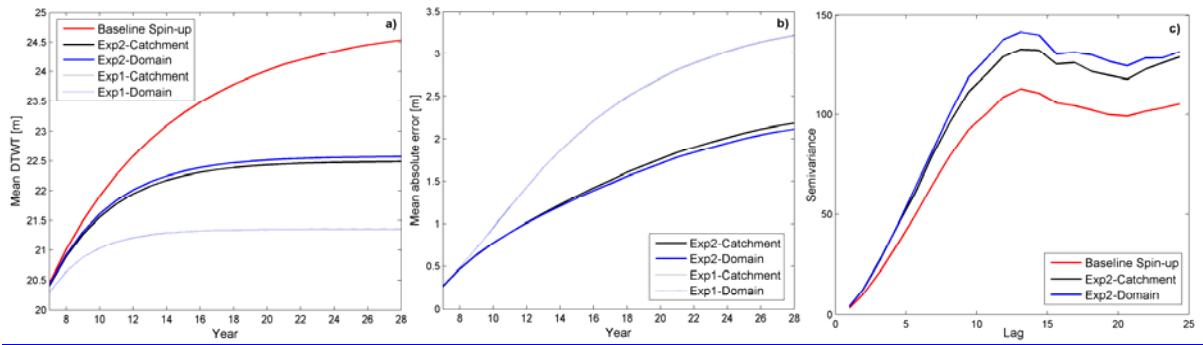
13

14

15

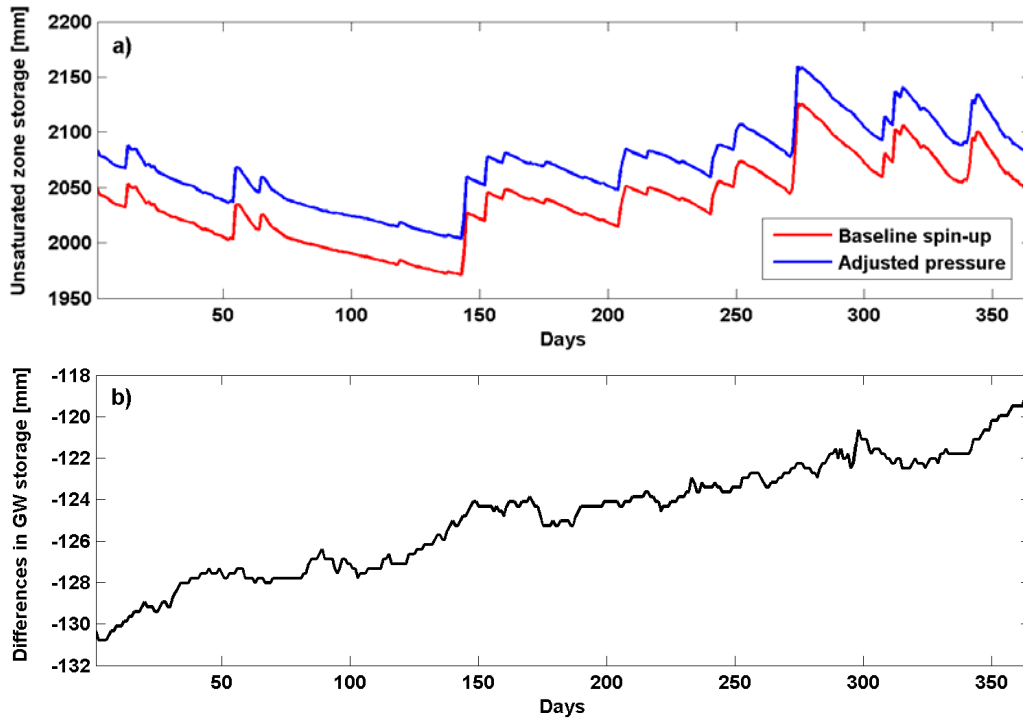
16





1  
2 **Fig. S2.** a) Comparison between the simulated mean annual DTWT obtained from the  
3 baseline spin-up approach of ParFlow.CLM and empirical DTWT functions for the Baldry  
4 sub-catchment. The single exponential model was formulated using the domain and  
5 catchment averaged data from spin-up simulations 2 to 6; b) estimated mean absolute error  
6 based on simulated DTWT from the baseline spin-up together with both catchment and  
7 domain averaged single and double exponential functions; and c) experimental semi-  
8 variograms of mean annual DTWT from ParFlow.CLM equilibrium year (after 28 years of  
9 simulations) and DTWT from catchment and domain averaged double exponential functions.  
10 DTWT distributions from the DTWT functions had higher variances relative to the baseline  
11 simulation. Exp1 and Exp2 refer to single and double exponential functions respectively.

12  
13



1  
 2 **Fig. S3.** a) Comparison of unsaturated zone storage dynamics of the equilibrium year  
 3 obtained from the hybrid spin-up approach and the baseline spin-up simulations in Baldry  
 4 sub-catchment. b) Differences in groundwater storages of the equilibrium year obtained from  
 5 the baseline spin-up simulation and the hybrid approach.

6