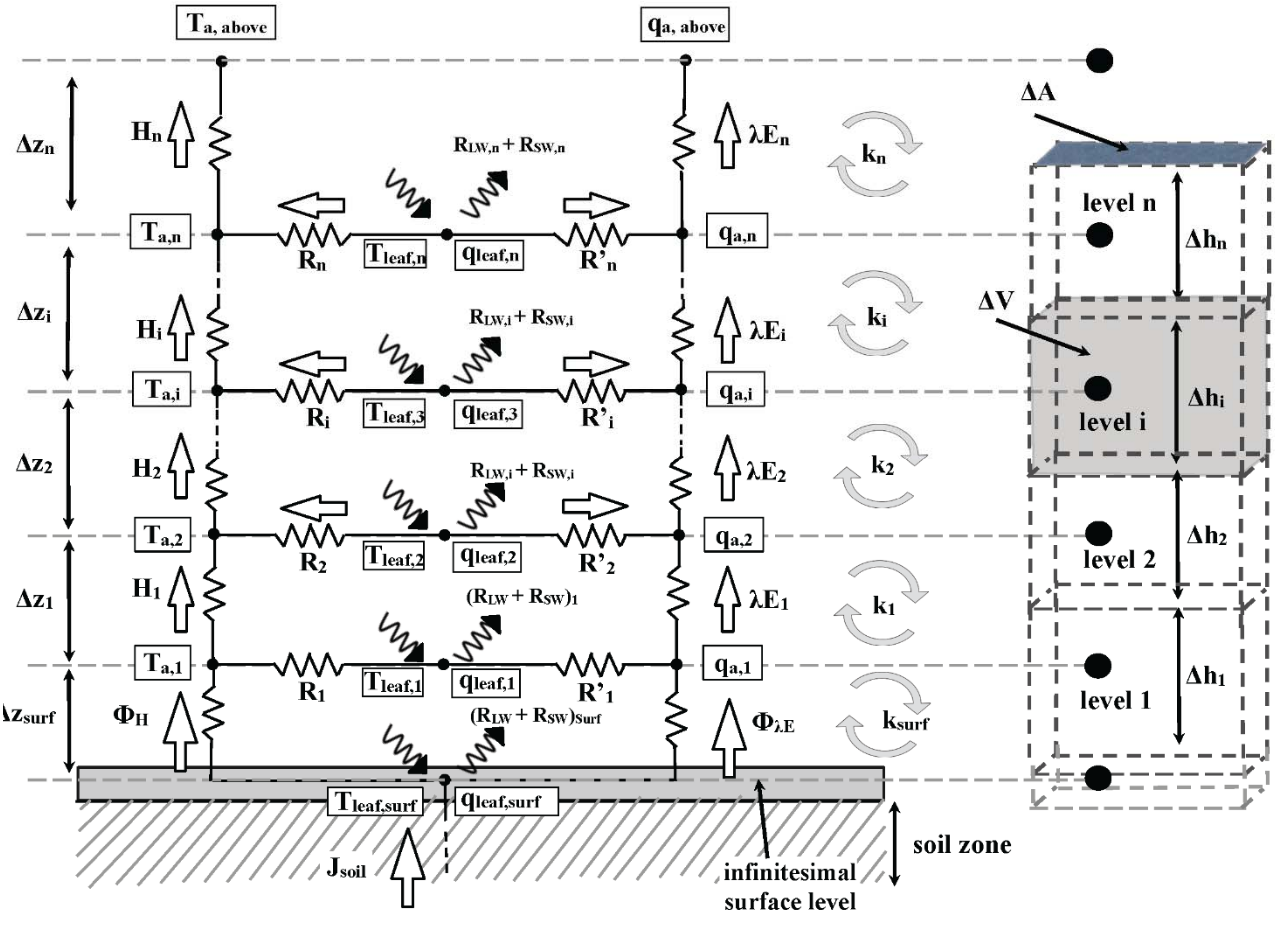
Referee#1

We would like to thank both reviewers for their insightful comments. Below we discuss how we will address their concerns in the revised manuscript.

#1 Equation 7 uses the threshold of 298.15 K. What is the physical basis for this threshold - or is is an empirical value?

We would like to thank the reviewer for pointing out the aforementioned issue, i.e. "The threshold of 298.15 K may be only suitable for sites in the temperate climate zone (with temperate grass species)". Indeed, this threshold temperature should reflect the geographical variation for different sites or locations. To the extent of the current approach to global applications, the generic temperature of 298.15 K will need to be replaced by a localized threshold.

Equation 7 describes the seasonality of the soil-atmosphere interface, which we believe is driven by the under-story and its phenology (Launiainen et al., 2015). Currently, the model does not simulate the production nor the phenology of the under-story. As a substitute for this rather complex process, we made use of a weighting coefficient for the conductance of the soil-atmosphere interface (K_{surf}) or, in other words, the calculation of the water vapor exchange between the soil layer and the first air column ($\Phi_{\lambda E}$) (see the $\Phi_{\lambda E}$ and K_{surf} in the figure below and the formal description of using K_{surf}, which is given in the supplementary material of Ryder et al. (2016), in Equation S4.30 and S4.31).



In Equation 7, we used 298.15 K as a threshold to simulate over-story phenology. Above this threshold, we use the sum of the canopy gaps as a proxy for the under-story phenology. In other words, the current approach assumes that when the long-term (21 days) mean t2m temperature exceeds 15°C (298.15 K), shading from the over-storey will become the main driver over the under-story phenology. Given the spatial distribution of our study sites, this is a crude but defendable assumption.

As an intermediate solution between this validation exercise and the global application in the next study, we will search for a more general parameterization of this threshold temperature and we will try to modify the reference temperature in Equation 7 by using a global soil temperature map instead. This, implies that we will have to rerun the model optimization work for the tuning coefficients a_8 to a_{10}.

#2 Equation 11 describes the calculation of stomata resistance dependent on photosynthesis activity of the plant (Farquhar model). This leaf photosynthesis model does not consider interaction between stomata resistance and soil water availability (stomata regulation by trees in case of disturbed water supply from soil).

The reviewer expressed concern for the absence of soil water availability in the calculation of stomatal resistance in Equation 11. After re-reading the text we understand where this concern originates, but our model formulation accounts for soil water stress in the calculation of actual transpiration and in turn in stomatal conductance and photosynthesis. ORCHIDEE-CAN calculates the supply of the water available for transpiration (F_{Trs}) as the pressure difference between the soil and the leaves (p_{delta}) divided by the sum of hydraulic resistances of fine roots (R_r), sapwood (R_{sap}) and leaves (R_l), i.e., $F_{Trs}=p_{delta}/(R_r+R_{sap}+R_l)$ (see Equation 20 in Naudts et al., 2015). The atmospheric demand of water for transpiration is calculated as the vapor pressure difference between the leaves and atmosphere divided by the sum of boundary layer resistance (R_b) and stomatal resistance (R_s) (see Equations 9, 14 and 15 in Ryder et al., 2016). When the supply can satisfy the demand, there is no water stress and photosynthesis (A) is calculated. When the demand is limited by the supply term, A and R_s are recalculated such that they satisfy the supply. Water stress thus enters Equation 11 in the value of A. Through Equation 11, we add a weighting factor (W_{sr}) to the original calculation of stomatal resistance (R_s) to tune the final calculation of the transpiration demand term (this tuning factor represents the coupling of the canopy to the atmosphere). Following the above reasoning, we will improve the description of equation 11 to eliminate the misunderstanding concerning how ORCHIDEE-CAN accounts for soil water stress.

#3 The authors should explain how they want to tackle the mismatch between rough resolution of driving data (reanalysis 0.5 degree) and high vertically resolved vegetation layer. Is it necessary in this case to leave the bigleaf concept?

Using forcing data of a rough spatial resolution to drive the model may contain information derived from several different land cover types, thus this comment touches upon an interesting issue: how to account for the average surface fluxes from the contribution of different subgrid scale land cover types? The present ORCHIDEE single-layer model calculates a weighted average of different PFTs across a grid square to calculate a total representative flux. An alternative approach, and one that we are investigating using this multi-layer model, is to calculate the heat fluxes of each vegetation type separately (sub-grid scale modeling) so that the mixing occurs above the canopy. We will add this point to the discussion.

#4 Apart from that, it is doubtful whether reanalysis data with a resolution of 0.5x0.5 degree give a realistic information for soil water pool.

For the spin-up of the initial state of the soil water pool, 20 years of climate data are required. We had a choice between using local high resolution climate observations for a usually very limited time period or using low resolution regional re-analysis for a much longer time period. Using the local high resolution data would have the advantage that local information is used, but due to the fact that some time series are only 2 to 4 years long (**Table 3** Period IV in Chen et al.), the spin-up would have to cycle 5 to 10 times over the same data. Although local data could then still have been used, cycling gives a lot of weight to the climatic events in the time series and may as such result in a biased spin-up. The alternative is to use 20 years of a climate re-analysis, these data represent the inter-annual variability better than cycling over the same 2 or 4 years of data but has the disadvantage that the data are less likely to represent the local conditions (especially in mountainous regions). Given the fact that we did not have access to soil water content data, we could not evaluate which method is better to spin-up the soil water content in the model. For this reason, we performed a sensitivity analysis of the parameterization of the initial soil water content at one of the driest sites used in this study (In the section 3.1 Model parameterization: Page 12 Line 23-25 and **Fig. S7** in the supplementary information from Chen et al.). Note that the model calibration and validation were based on the site level observations because that part of the study did not require cycling of the same data. In short, in the absence of a rigorous validation of both approaches to the spin-up of the soil water content, it is not possible to rank one method above the other. In the revised text we will clarify the strengths and weaknesses of the two present different approaches.

#5 The model performance strongly depend on the model tuning. There are a couple of tuning parameters without plausible natural background. This fact makes a transferability of the results to other sites difficult. Could the authors discuss this problem?

This comment refers to a long-standing issue in model development and model validation which is very well discussed by Oreskes et al. (1994). Despite the direction of the land surface model community towards the development of more mechanistic models, all large-scale land surface models contain an important level of empiricism. When the model is carefully developed and validated the empirical parameters mimic an overly complex (for the purpose of the model) or poorly understood process. As we tried to follow this philosophy we believe that our parameters have a plausible natural background but this does not overcome the issue of equifinality of the model. Ideally, future developments should aim at replacing such parameters by a more mechanistic approach if the empirical module represents a process that is at the core of the objectives of the model.

Tuning parameter names used in this study	**Physical parameter**	**Empirical representation of**
a_1 to a_5	effective surface drag	Bending of tree branches to increase the contact surface
a_6 to a_7	eddy diffusivity	Inner canopy turbulent mixing induced by canopy structure
a_8 to a_{10}	surface-atmosphere conductance	Sub-canopy phenology

W_{br}	layer boundary resistance	Upscaling the atmospheric coupling for the heat transfer from a single leaf to the entire canopy
W_{sr}	layer stomatal resistance	Upscaling the atmospheric coupling for the water vapor transfer from a single leaf to the entire canopy

In Ryder et al. 2016, the model was developed and tested for a single site. In the current manuscript we aim to test the model for more diverse environmental conditions in order to demonstrate that the numerics can deal with the variation that can be found in global ecosystems. For this we granted ourselves the freedom to derive a separate parameter set for each site. By doing so we learned about the strengths and weaknesses of the model and its parameters. Next, we will have to derive a single parameter set for each PFT and test how well the model reproduces global patterns in, for example, evapotranspiration. This is the point of the development and validation chain, where we will learn about the transferability of the parameters. We will address this issue in the manuscript by rephrasing parts of the introduction and adding a paragraph to the discussion.

#6 The multi-layer approach shows an improvement especially in soil heat flux. Is it relevant for climate? Apart from that, for inter-annual cycle soil heat flux must be about zero (not fulfilled in Fig. 4)!

Comparing the observed magnitude of soil heat flux with other components of the surface energy budget shows that at forest sites the soil heat flux is almost one order of magnitude smaller than the other components. The reported result - that the multi-layer simulation shows a better model prediction skill is interesting (as discussed), but is unlikely to be sufficient to justify the added complexity of a multi-layer model. However, the soil heat flux is an essential aspect in simulating the snow phenology (Wang et al., 2015). Therefore, improved simulations of the soil heat fluxes could have important indirect effects on climate simulations of regions with a pronounced snow season.

The reviewer remarks that the inter-annual cycle of soil heat flux should be zero. This is indeed to be expected for graphs showing the absolute soil heat flux. **Fig. 4**, however, shows the model skill for different components in the energy budget – the annual sum of the model skill should not be zero. We will prepare new figures showing the absolute values for both the observations and simulations at the diurnal and inter-annual scale.

References:
Naudts et al., 2015: A vertically discretised canopy description for ORCHIDEE (SVN r2290) and the modifications to the energy, water and carbon fluxes, *Geoscientific Model Development*, 8, 2035–
Ryder et al., 2016: A multi-layer land surface energy budget model for implicit coupling with global atmospheric simulations, *Geoscientific Model Development*, 9, 223–245, doi:10.5194/gmd-9-223-2016, 2016.
Oreskes et al., 1994: Verification, validation, and confirmation of numerical models in the Earth sciences, *Science*, 263, 641-646, 1994.
Launiainen et al., 2015: Coupling boreal forest CO_2, H_2O and energy flows by a vertically structured forest canopy–soil model with separate bryophyte layer, *Ecological Modelling*, 312 (24), 385–405,

2015.
Chen et al., 2016: Evaluating the performance of the land surface model ORCHIDEE-CAN on water and energy flux estimation with a single- and a multi- layer energy budget scheme, *Geosci. Model Dev. Discuss.*, doi:10.5194/gmd-2016-26, 2016
Wang et al., 2015: Impacts of Satellite-Based Snow Albedo Assimilation on Offline and Coupled Land Surface Model Simulations, *PLoS ONE* 10(9):e0137275 · September, 2015.
2065, doi:10.5194/gmd-8-2035-2015, 2015.