Author’s response to Referee #3

First of all, we would like to thank the Referee for valuable comments and suggestions. We will provide additional analysis to address the comments, and make necessary changes in the revised manuscript. We have provided direct responses to the comments here as well.

Major Comments

1) **This paper mainly focus on investigating the impact of “space and time varying vegetation parameters” on defining infiltration, root water uptake and transpiration processes and decomposing the TWS. It sounds strange to use the words such as “importance of vegetation”, “including vegetation”, “including vegetation characteristics”, “including vegetation data” and “contribution of vegetation”. These words cannot reflect well the research objective, and I think the impact of vegetation/vegetation characteristics/ vegetation data is also embedded within the baseline experiment using the static and globally uniform parameter values.**

AC: We thank the Referee for this constructive criticism and agree that the baseline experiment implicitly accounts for the role of vegetation in its globally uniform parameters. We clarified the statements mentioned by the Referee by rephrasing into “including vegetation explicitly”, “including varying vegetation characteristics” or “importance of variation of vegetation properties” at the respective occasions. We further clarify this aspect by introducing the following sentences in *2.1 Overview*:

> *[... In the VEG experiment, we describe vegetation related parameters as the linear product of a calibrated scalar and spatio-temporal varying vegetation variables. By calibrating the scalar, we include the continuous pattern from the data, but weight it to best fit with observational constraints.]*
>
> Even though the optimized parameters of the baseline experiment implicitly account for the effect of vegetation, its parameters are global constants and do not vary spatially. In the **VEG** experiment vegetation related parameters vary explicitly spatially and partly temporally.
>
> *[...]*

2) **The authors calibrate and validate the model performance almost for the same period (01/2002-12/2014 vs. 03/2000-12/2014). From the view of traditional calibration and validation procedure, it will be better to use part of the observations (e.g. period of 2002-2008) to calibrate the model, and then using the remaining observations (e.g. period of 2008-2014) to validate the model.**

AC: This is a good point. Due to the rather short time period of observational constraints (01/2002-12/2014, which are 144 monthly values), we decided to not further reduce the temporal information. Instead of splitting the time period, we divided the calibration and validation data spatially. This is similar to the proxy basin test in traditional catchment hydrology, when a model is calibrated for one basin and evaluated for another. We will add a sentence to the revised manuscript to clarify this.

3) **For the regional analysis of model performance, the authors derive 5 hydroclimatic regions by performing a cluster analysis, but it sounds strange to treat almost the whole China and Europe as the same group, i.e., the moderate mid latitudes (Temperate). This is contrast to the common sense. The authors are thus suggested to use a better classification such as the Köppen–Geiger climate classification.**

AC: Our regional classification was based on clusters of seasonal dynamics of the observation data used in this study (ET, Q, and TWS). It additionally includes the latitude along which the main gradient of the regional climate exists. Note that the classification was agnostic to geographical regions or boundaries, and the regions showing similar hydro-climatic features were grouped into a cluster. We are aware that it is common to use the Koeppen-Geiger classification for regionalization. However, one of the major shortcomings of Koeppen-Geiger classes is the inclusion of grid cells from Southern and Northern latitudes into one class. These regions with opposing seasonal cycles potentially distort the seasonality of hydrological variables at the regional scale. We will therefore keep the regionalization based on hydroclimatic clusters as the basis of our analysis and results.

Nevertheless, following the Referee's suggestion, we performed a regional analysis for Koeppen-Geiger climate zones instead of hydroclimatic clusters. To do so, we aggregated Koeppen-Geiger subgroups considering the main climate group and distinguishing between humid and semi-arid conditions. The resulting zones are shown in Fig. 1. Fig. 2 evaluates model performance for the Koeppen-Geiger regions and Fig. 3 shows the composition of seasonal TWS variations therein. The results presented below will be included in the supplement of the revised manuscript.

Note that most parts of the *Polar* and *Boreal* Koeppen-Geiger (KG) zone are included in the *Cold* region (R1) of the hydroclimatic cluster classification. We find that the regional averages are very similar for both classification schemes in terms of model performance and composition of seasonal TWS variations.

The Northern Hemisphere *Temp* and *Boreal-sa* KG zones are both included in the *Temperate* hydroclimatic region (R2). *Temp* KG and the *Temperate* region (R2) agree well regarding model performance and seasonal cycles, although we see a slightly better performance for the *Temp* KG regarding wTWS and Q. In the *Boreal-sa* KG, **B** and **VEG** don't reproduce the spring peak of Q and precede the observed wTWS significantly, decreasing model performance slightly when combining the *Temp* and *Boreal-sa* KG zones in one hydroclimatic region. Therefore, it would make sense to further distinguish the *Temperate* hydroclimatic cluster region. However, *Boreal-sa* KG spans Northern China, where poorer model performance is also evident from the performance maps in Fig. 4 of the manuscript.

However, as mentioned in the manuscript, the advantage of the hydroclimatic cluster regionalization becomes obvious when interpreting results of the *Arid* and *Temp-sa* KG zones. This is because these climate zones are distributed across the Southern and Northern Hemisphere, causing 2 peaks in the regional seasonal cycles for wTWS, ET and Q, due to opposing seasonal dynamics. The *Arid* KG zone includes the *Semi-arid* cluster regions (R5) in the Southern Hemisphere, as well as parts of the *Temperate* region (R2) (mainly in North America). The *Temp-sa* KG zone covers a rather small fraction of the study area, that is spread over the *Temperate* region (R2) in the Northern Hemisphere and the *Semi-arid* region (R5) of the Southern Hemisphere.

Likewise, the effect of opposing seasonal cycles is apparent in the *Tropic* KG zone, although less apparent due to the proximity to the equator where the climate is more homogeneous and seasonality is low. The *Tropic* KG corresponds to the *Humid* cluster region (R3) on the Southern Hemisphere, and parts of the *Sub-humid* region

(R4) on the Northern Hemisphere. Compared to the hydroclimatic cluster regions, the *Tropic* KG has less seasonal variation (a smaller amplitude) of wTWS, ET and Q, due to its larger area North and South of the equator. Both, **B** and **VEG** underestimate the ongoing depletion of wTWS from September to December in *Tropic* KG, which is likely related to the opposing seasonal cycles of wTWS in the *Humid* (R3) and the *Sub-humid* (R4) cluster regions. In the *Tropic* KG, Q peaks in March (as in *Humid* (R3)) and has a second, smaller peak in September (when Q peaks in the *Sub-humid* region (R4)). However, model performance is very similar for *Tropic* KG and the *Humid* and *Sub-humid* cluster regions.

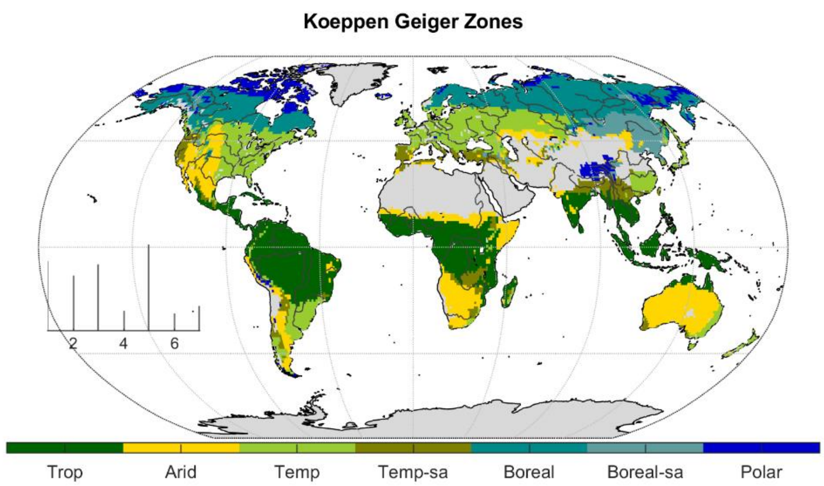

Figure 1: Regions based on Koeppen-Geiger climate zones (Trop = Af, Am, As, Aw; Arid = BSh, BSk, BWh, BWk; Temp = Cfa, Cfb, Cfc, Dfa, Dfb; Temp-sa = Csa, Csb, Csc, Cwa, Cwb, Cwc; Boreal = Dfc, Dfd; Boreal-sa = Dsa, Dsb, Dsc, Dwa, Dwb, Dwc, Dwd; Polar = EF, ET).

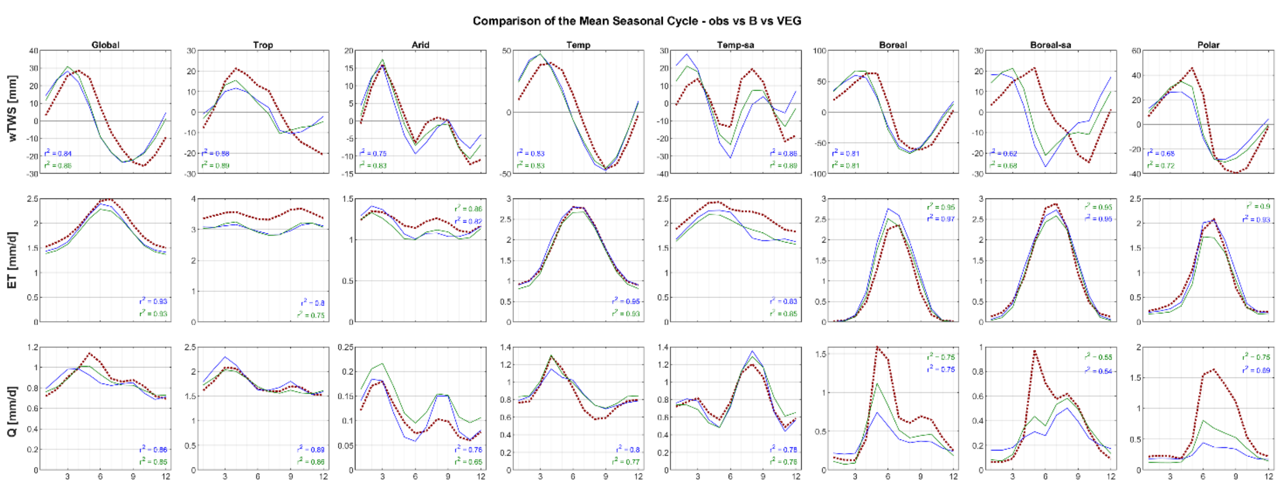

Figure 2: Global and regional mean seasonal cycles of total water storage (wTWS), evapotranspiration (ET) and runoff (Q) for the B and VEG experiments compared to the observational constraints by GRACE (wTWS), FLUXCOM (ET) and GRUN (Q).

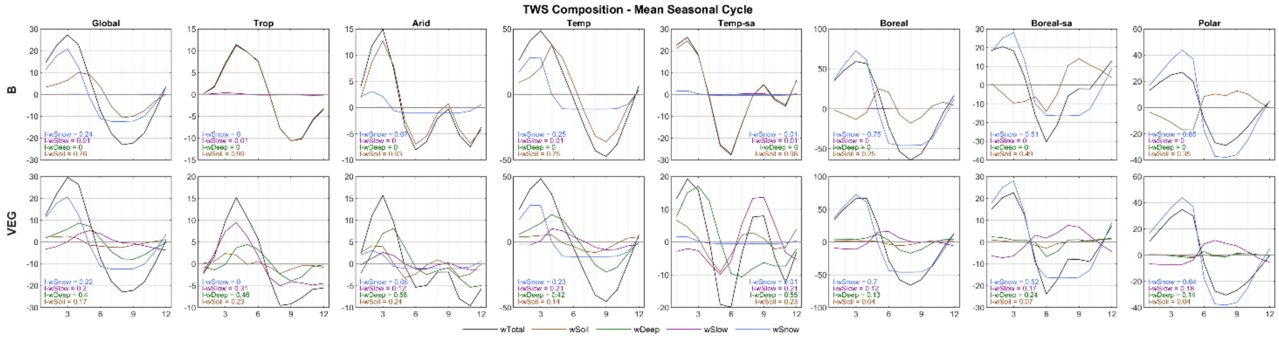

Figure 3: Global and regional mean seasonal cycles of simulated total water storage and its components for B and VEG, including the regional Impact Index *I* for each storage.

Minor Comments

1) The title of Section 2 is wrong and the structure of this section can be improved. E.g., "1.1 Methods" should be "2 Methods", "1.2 Overview" should be "2.1 Overview". In addition, it's suggested to modifying "2.2.1" to "2.3".

AC: We thank the Referee for pointing out the mixed-up section numbering. The suggestion also aligns with Referee #1's minor comments 4) and 6), and we will follow the suggestions in the revised manuscript.

2) It's suggested to modifying all the tables to the standard table format.

AC: Thanks for the suggestion! We included grey shaded rows as 'subsections' of the table to improve readability. However, we will adapt the layout of the tables and figures when and if requested by HESS.

3) Line 410: "Fig. 2" should be "Fig. 3"

AC: Thanks for pointing out the reference to the wrong figure. We will change accordingly.

4) The number in Fig.3 is difficult to read

AC: We will increase the font sizes of the numbers in Fig. 3.