Reviewer 2

General comments

The manuscript describes how the incorporation of crop specific phenology data improves ET and soil erosion estimates for large-scale simulations using SWAT+ throughout the Nile basin as compared to the default phenology implemented in the model. The simulated LAI and ET values agree much better with validation data obtained from remote sensing. The estimated erosion rates are substantially lower as compared to the default model.

The topic of the manuscript fits the scope of HESS. The results are relevant because they demonstrate the how important it may be to account in an adequate manner for regional differences of crop-specific phenologies.

Unfortunately, the method section is not very well written and is often rather confusing. The key element for the improved model set-up is the use of a global data set on plant and harvest dates for specific crops (Global Gridded Crop Model Intercomparison (GGCMI), see Tab. 1, L. 164 - 168) instead of using the default heat unit approach implemented in SWAT+ by default. It is pointed out in the Abstract and the Introduction that this default approach often fails in tropical regions because crop development is strongly affected by precipitation (e.g., L. 17, 47) while temperature is well suited for temperate regions. However, it remains obscure how the GGCMI data account for this deficiency. It is not explained whether these phenology data are based on observational reference data or on model simulations. If they were model based, one should know how the model accounts for precipitation and temperature compared to the SWAT+ concept. Irrespective of whether the dataset is observation or model based, one should know whether the data represent long-term averages or account for yearly variations. It remains also obscure what the spatial resolution of the dataset is.

These temporal aspects are also neglected in the analysis of the results. The authors used seven years of data for model validation (L. 173). However, they only present results averaged across the entire study period (2009 – 2015). They don’t present any data on inter-annual variability (e.g. of precipitation) that might have affected the results. At least in some regions, rainfall varied affecting also the crops simulated in the manuscript (e.g., Epule, Dhiba et al. 2021). Such inter-annual differences can also be expected for erosion, which is very much triggered by few events.

Recommendation

The manuscript requires substantial improvements regarding

1. the presentation of the methods
2. the temporal aspects of the data series that have been analysed.
3. the issues listed in the detailed comments

References:

Epule, T. E., D. Dhiba, D. Etongo, C. Peng and L. Lepage (2021). "Identifying maize yield and precipitation gaps in Uganda." SN Applied Sciences **3**(5): 537.

Response: Thank you for a positive evaluation of our study and your time to provide critical comments to improve the manuscript.

The GGCMI crop calendar (Jägermeyr et al., in revision) is an observation-based product, combining first-hand data sources from various agricultural ministries. These data do not include modeling results and therefore overcome the missing temperature-precipitation seasonality in the old model version. In the new model version we schedule planting dates and cultivar selection based on real-world observational planting and harvest data. Planting thus happens at the prescribed day per crop in each 0.5° grid cell. The harvest day varies in each year as physiological maturity depends on accumulated phenological heat units. In a warm growing season maturity is reached earlier than in a cold season. On average, cultivars are selected to match the observational harvest day.

With regards to the temporal aspects of the analysis, we focused on the seasonal representation of the LAI and its subsequent impact on erosion (for illustration of the concept) because as shown in the example Figures 1 and 2 (below), the seasonal pattern is similar in all years. Even though there is some interannual variability in the magnitude of the LAI signal and precipitation, the seasonal pattern is consistent. Consequently, for clarity of the message, we focused on showing the variability in the months (season) within which the LAI and erosion estimates peak since the season is the same throughout the year regardless of the interannual variability in the magnitudes of the signals. Of course, when focusing on the crop outputs such as yield or water productivity, analysis of the interannual variability is very important with the spatial scale of analysis put into consideration. The interannual variability discussion will be highlighted in the revised manuscript.

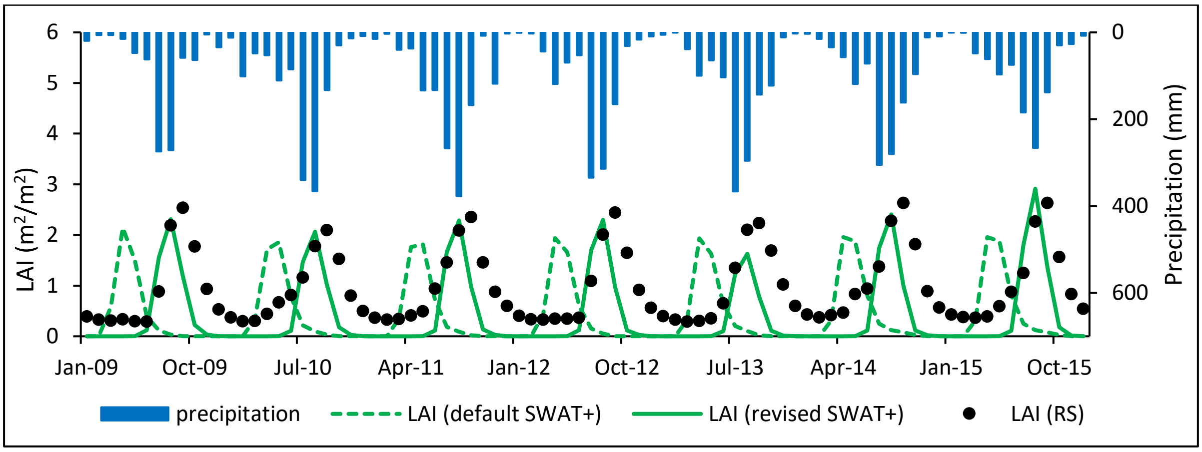

Figure 1: LAI comparison for rainfed maize in the Blue Nile basin

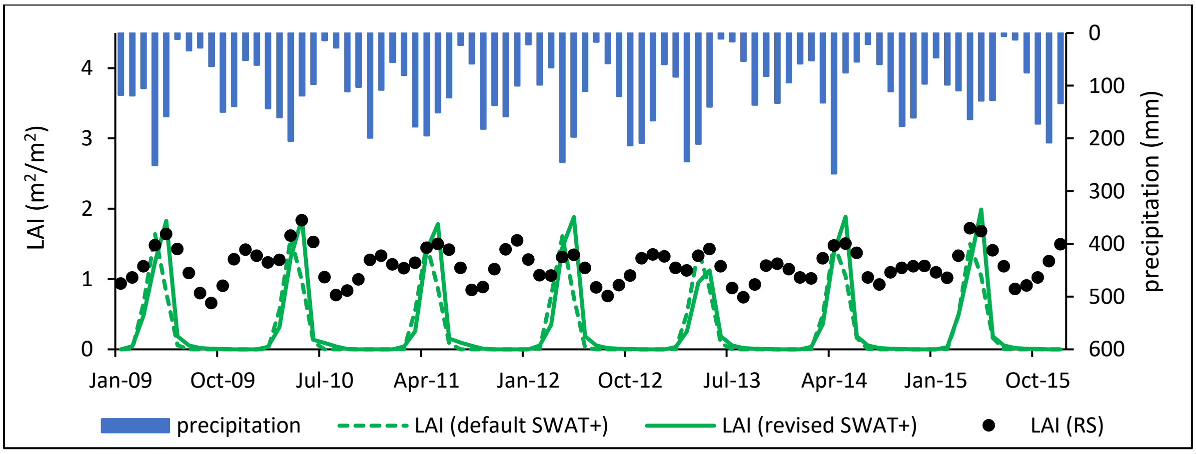

Figure 2: LAI comparison for rainfed wheat in the Victoria basin

Detailed comments:

- L. 87 – 88: Please describe more precisely (in the Method section) what these tables and datasets provide.

Response: Decision tables are already described in the 'Method section" from Lines 125 to Line 129. Global datasets of rainfed and irrigated cropland, associated management practices of Fertilization (Nitrogen and Phosphorus) are specified in Table 1, Line 152. However, additional descriptions of the datasets will be added in the revised manuscript for clarity.

- L. 97: Study area: Please describe the study period as well.

Response: The study period from 2009 to 2015 is to be further described in the revised manuscript.

- L. 146: "approached suggested by Chawanda": Approach for doing what?

Response: Chawanda et al., (2020) proposed an approach for setting up a SWAT+ model using the harmonized land use product (LUH2; Hurtt et al., 2020)) that was utilized in this study. LUH2 product is formatted as NetCDF with each layer containing a percentage of land use instead of a single-layer raster file. The SWAT+ code has to be adapted to include routines that read the LUH2 NetCDF data as suggested by Chawanda et al., (2020). A brief description of the approach will be added in the revised manuscript.

- L. 152 (Table 1): Confusing: what is used for model set-up and what for comparison between the default and the revised model. Also linguistically. the sentence is strange (data sets are not used for crop management). GGCMI: please provide more details to address the questions mentioned above (general comments).

Response: Table 1 only includes datasets used for model setup. Datasets used for model comparison (validation) are described in "section 2.6 Validation of model results – Line 171". A brief description about GCCMI dataset together with other datasets is to be added in the revised manuscript. The caption of Table 1, Line 152 is to be rectified and a separate table showing datasets used in model validation will be added in the revised manuscript.

- L. 153: Section 2.5: It is not clear whether this section presents the revised SWAT version only or the land use for both the default and the revised version. If it's only about the revised version describe how land use was established for the default version. Otherwise, clearly indicate which part only refers to the revised version.

Response: The default agricultural land use is precisely stated in the "section 2.4 Default Model set up". By default, the cropland use was represented in a generic way using heat units to trigger the cropping season [Line 148 – line Line149]. This is how SWAT+ represents agricultural land use by default (Arnold et al., 2013). Section 2.5 gives the proposed scheduling using the plant and harvest dates extracted from the global phenology dataset (GCCMI data) and the fertilization extracted from (Hurtt et al., 2020) and (Lu and Tian, 2017) as described in the section. This will be clearly indicated in the revised manuscript.

- L. 164: crop phenology data: please provide more information (see above).

Response: More information about the global phenology dataset (GGCMI data) is to be added in the revised manuscript.

- **L. 191: What is a scientific validation? As it reads later in the manuscript it seems to be a plausibility check.**

Response: Biondi et al., (2012) draws a distinction between performance validation and scientific validation. Performance validation is a typical approach adopted to evaluate model performance that requires the comparison between simulated outputs on a set of observations that were not initially used for calibration. This involves the use of graphical techniques or performance metrics. A scientific validation aims at evaluating the consistency of the model thought as input-state-output system. This concept was derived from the idea that verifying the model performance by simply comparing outputs and observations does not assure that the model is correct from a scientific point of view. This validation may include and extend the performance validation and is specifically required in cases when the quality and quantity of observation data is not sufficient to allow an adequate validation. Studies such as (Haregeweyn et al., 2017) have used the term 'scientific validation' in validating their results. Just like our research, they did the validation by comparing the soil erosion model outputs with previous studies and the scant observations.

However, since the term 'scientific validation' seems to be confusing in this context, it is to be replaced with 'plausibility check' in the revised manuscript.

- **L. 194: What's the meaning of "moreover" at that point?**

Response: The intended meaning was "in addition to the previous statement". However, "moreover" is to be omitted and the statement rephrased in the revised manuscript.

- **L. 197 – 200: These two sentences are not clear.**

Response: The statements are to be rephrased but what they intend to mean is that in order to isolate the uncertainty in the default and revised model setups due to agricultural land use representation, the models were compared in default parameters considering that calibrations change with different catchments. So only the water balance was checked to ensure that we have consistency of the model thought as an input-state-output system. This means that the differences seen in the model setups originate primarily from the crop representation and management practices. However, the two statements will be rephrased for clarity in the revised manuscript.

L. 204 – 207: How has the model be parameterized? No information is provided. How well did the model perform e.g., when compared to discharge? The water balance information provided are not very conclusive since L. 254 demonstrates that the model can be quite off for one component (ET) while still closing the water balance well (compensating errors).

Response: A default parameterization of the model was used, (this will be added in the revised manuscript). The study aims at improving the default model simulation by better representing the physical land processes (crop growth and evapotranspiration), prior to any calibration. A flow calibration would not much affect these processes but rather aims at improving flow simulations, which was not the objective of our study. We only focused on the ET component of the water balance to show how representation of crop phenology can impact the ET estimates which should ideally precede any calibration efforts [Line 200 – Line 203]. Besides, SWAT was developed with the objective of predicting the impact of management on water, sediment and agricultural yields in large 'ungauged' basins where usually no data is available for calibrations (Arnold et al., 1998; Srinivasan et al., 2010). Additionally, matching flow observations at the outlet/gauging station does not necessarily mean that the internal processes e.g ET are realistically simulated.

- **L. 208: Results: Please provide a short description of the hydro-climatic characterization of the study period (2009 – 2015) including metrics of temporal variability.**

Response: The hydro-climatic description of the study period is to be added in the "Results" section of the revised manuscript.

- **L. 210 – 220: Provide information about variability and model performance across years as well.**

Response: We originally focused on the seasonal variations of the LAI and erosion estimates as we believe the annual cycle information is relevant for highlighting the value of this representation. For example other studies such as Levis et al., (2012) that incorporated agriculture in a Community Earth System Model (CESM1), evaluated vegetation traits e.g. LAI, on a seasonal/annual basis. However, the interannual variability will also be discussed in the revised manuscript.

- **L. 224: Fig. 4: Indicate in the caption whether the data represent average values. If they are provide standard deviations in the figures.**

Response: The standard deviation is to be added in the revised manuscript for data representing averages and captions will be adjusted.

- **L. 226: Fig. 5: Same comment as above. Additionally: the LAI of both models are rather similar. How does it come that the erosion rates differed so strongly? Explain.**

Response: We noticed that a slight increase in the LAI magnitude had a strong impact on the erosion peaks. Even though the cropping season in both the default and revised model setups captures only one cropping season, there is still a reduction in the HRU erosion estimates because the revised SWAT+ LAI, representative of an actual crop is slightly greater than the default LAI representative of a generic crop. Additionally, with a slightly higher LAI magnitude in the revised SWAT+ model, more biomass is generated which results in more residue that could be more effective in reducing soil erosion even after the cropping season. Residue intercepts rain droplets near the soil surface that drops regain no fall velocity. Thus, a given percentage of residue is more effective than the same percentage of canopy cover (Neitsch et al., 2011).(lines 274-279). This will be explicitly explained in the revised manuscript.

L. 230: The situations are probably denoted by Cases 1 and 2 in Fig. A2. However, this is not explained. How could one differentiate between the two cases in the spatial data?

Response: Cases 1 refer to growing season (irrigated) occurring in the dry season of the season while case 2 refers to growing season (irrigated) occurring during the first rainy season. So case 2 is mainly supplementary irrigation. This information is to be added to text in the revised manuscript.

- **L. 240: Same comment as for Fig. 4; Additionally: Why is LAI for the default so low? Was there an underestimation of irrigated wheat (acreage)? If yes, this would imply that not only phenology but also land use differed between model versions. This was not made explicit so far. Clarify.**

Response: By default, no management practices (i.e. irrigation and fertilization) are implemented. Hence, being in the Nile delta that predominantly relies on irrigation for plant growth, the plant growth is constrained by default causing the low LAI. However, with the implemented crop phenology and associated management practices, we see an improvement in the LAI simulation. This information will be added to text in the revised manuscript.

- **L. 260 – 263: Would one not expect an overestimation of irrigation and therefore ET from an ideal unlimited water source for irrigation?**

Response: Water is applied with limits (50mm every application for this case study) which avoids the continuous over irrigation. The water is irrigated when the water stress is below a specified threshold of 0.7. By unlimited source (which is the deep aquifer in the model), we mean that the water is always available for irrigation when the water stress in the field during the growing season goes below the specific threshold. However in real life scenarios, the farmer may not always have water for irrigation when needed on a specific day due to different management schedules of irrigation schemes. . Additionally, not all irrigation water comes from the deep aquifer on all irrigation sites. The statement will be rephrased and explained better in the revised manuscript.

- **L. 266: Fig. 7: A difference map between model predictions and remote sensing observations would be more instructive. How can one distinguish between agricultural and non-agricultural ET? Units: correct to mm y^{-1}.**

Response: Fig.7 in the manuscript masks out non-agricultural areas and presents only ET for agricultural areas to highlight the impact of the agricultural areas since this study focused on the ET response to agricultural land use representation. Presenting agricultural ET further highlights the impact of human activities (planting, harvesting, irrigation, fertilization) which is significant for developing agricultural water resources management strategies (Wang et al., 2008). However, a difference map on model ET output will be included too in the revised manuscript. The units will be corrected as well in the revised manuscript.

- **L280 – 283: Below, a specific comparison is presented for the Blue Nile region. Please provide also the relative change for this area to help the reader linking the two aspects.**

Response: The relative change in the Upper Blue Nile is to be provided in the revised manuscript.

- **L. 297: On L. 295 a max of 20500 t km^{-2} y^{-1} is mentioned. Can you explain?**

Response: The simulated annual average soil erosion ranged from 0 to 20500 t km^{-2} y^{-1} in the whole region. The standard deviation is to be added for this value in the revised manuscript. On Line 297, we mention an annual average soil erosion range of 0 to 13000 t km^{-2} y^{-1} in the Upper Blue Nile basin. The statements will be rephrased in the revised manuscript.

References

Arnold, J. G., Srinivasan, R., Muttiah, R. S., and Williams, J. R.: Large Area Hydrologic Modeling and Assessment Part I: Model Development1, 34, 73–89, https://doi.org/doi:10.1111/j.1752-1688.1998.tb05961.x, 1998.

Arnold, J. G., Kiniry, J. R., Srinivasan, R., Williams, J. R., Haney, E. B., and Neitsch, S. L.: SWAT 2012 input/output documentation, Texas Water Resources Institute, 2013.

Biondi, D., Freni, G., Iacobellis, V., Mascaro, G., and Montanari, A.: Validation of hydrological models: Conceptual basis, methodological approaches and a proposal for a code of practice, Physics and Chemistry of the Earth, Parts A/B/C, 42–44, 70–76, https://doi.org/10.1016/j.pce.2011.07.037, 2012.

Chawanda, C. J., Arnold, J., Thiery, W., and Griensven, A. van: Mass balance calibration and reservoir representations for large-scale hydrological impact studies using SWAT+, Climatic Change, 1–21, https://doi.org/10.1007/s10584-020-02924-x, 2020.

Haregeweyn, N., Tsunekawa, A., Poesen, J., Tsubo, M., Meshesha, D. T., Fenta, A. A., Nyssen, J., and Adgo, E.: Comprehensive assessment of soil erosion risk for better land use planning in river basins: Case study of the Upper Blue Nile River, Science of The Total Environment, 574, 95–108, https://doi.org/10.1016/j.scitotenv.2016.09.019, 2017.

Hurtt, G. C., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B. L., Calvin, K., Doelman, J. C., Fisk, J., Fujimori, S., Goldewijk, K. K., Hasegawa, T., Havlik, P., Heinimann, A., Humpenöder, F., Jungclaus, J., Kaplan, J., Kennedy, J., Kristzin, T., Lawrence, D., Lawrence, P., Ma, L., Mertz, O., Pongratz, J., Popp, A., Poulter, B., Riahi, K., Shevliakova, E., Stehfest, E., Thornton, P., Tubiello, F. N., van Vuuren, D. P., and Zhang, X.: Harmonization of Global Land-Use Change and Management for the Period 850–2100 (LUH2) for CMIP6, 1–65, https://doi.org/10.5194/gmd-2019-360, 2020.

Jägermeyr et al.: Climate change signal in global agriculture emerges earlier in new generation of climate and crop models, in revision.

Levis, S., Bonan, G. B., Kluzek, E., Thornton, P. E., Jones, A., Sacks, W. J., and Kucharik, C. J.: Interactive Crop Management in the Community Earth System Model (CESM1): Seasonal Influences on Land–Atmosphere Fluxes, 25, 4839–4859, https://doi.org/10.1175/JCLI-D-11-00446.1, 2012.

Lu, C. C. and Tian, H.: Global nitrogen and phosphorus fertilizer use for agriculture production in the past half century: shifted hot spots and nutrient imbalance, 9, 181, 2017.

Neitsch, S. L., Arnold, J. G., Kiniry, J. R., and Williams, J. R.: Soil and water assessment tool theoretical documentation version 2009, Texas Water Resources Institute, 2011.

Srinivasan, R., Zhang, X., and Arnold, J.: SWAT ungauged: hydrological budget and crop yield predictions in the Upper Mississippi River Basin, 53, 1533–1546, 2010.

Wang, S., Kang, S., Zhang, L., and Li, F.: Modelling hydrological response to different land-use and climate change scenarios in the Zamu River basin of northwest China, 22, 2502–2510, 2008.