Thank you for your kind response and your stimulating questions. It is always a pleasure to hear that your work is causing interest and is providing new insights. Here, we have tried to respond as short and clear as possible each question formulated by you.

1- Are the model parameters different from cell to cell? If yes, which parameters are identical and which parameters are different?

To answer this question is important to better understand the concept of split-structure for the effective parameter value at each cell. This calibration strategy consists on an application of a scalar multiplier to each prior parameter field (specified from data describing watershed characteristics: soils, vegetation, topography, land use, etc.) and to estimate a "best" value for this multiplier via calibration. This so-called "multiplier" approach makes the assumption that the prior parameter field properly describes the spatial pattern of parameter variation (the pattern of relative magnitudes from cell to cell), but that the magnitudes of all the parameter values must be adjusted to achieve a better simulation of the model response.

Hence, the effective parameter at each cell (i.e. the parameter value used when running the model) is compounded by two parts: (1) a common correction factor for each type of parameter that takes into account the model, information and input errors and the temporal and spatial scale effects; and (2) the estimated parameter value at each cell.

Hence, the parameters are different from cell to cell while the correction factor is common from map to map. The estimated parameter values were extracted from the field work done and presented in the doctoral thesis by Franz (2007) and following the recommendations provided by the TETIS model's support team. Two of the authors are actually active members of this team and we also used our own experience.

2- I did not understand how the model calculate the LAI which then is used to calculate the transpiration?

The LAI is calculated by the dynamic vegetation sub-model called LUE-Model. The LUE-Model computes the leaf biomass (B_l) according to the following equation:

$$\frac{dB_l}{dt} = (LUE * \varepsilon * PAR * fPAR - Re) * \varphi_l(B_l) - k_l * B_l$$

where LUE is the Ligth Use Efficiency, ε takes into account the reduction in LUE due to stress sources, Re is the respiration, $\varphi_l(B_l)$ is the fractional leaf allocation and k_l is the leaf natural decay factor to reproduce the senescence.

Once B_l is computed it can be transformed into LAI by using the specific leaf area (SLA) and the vegetation fractional cover (f_c) according to the next equation:

$$LAI = B_l * SLA * f_c$$

In the current version of the manuscript only the references about this model were mentioned and they should be specified. In this way, readers will only have to check the references if they are interested in specific details. These two equations together with the explanation will be provided in the next version of the manuscript. More detailed description can be found in Pasquato *et al.* (2015) and Ruiz-Pérez *et al.* (2016) (references embedded in the manuscript).

3- Maybe I missed, but what is the resolution of the implemented model?

You did not miss, we forgot to give that information. The temporal resolution is already specified and it is daily while the spatial resolution was 90X90 meters. It will be included in following versions of the manuscript.

4- How did the manual calibration help to find the best parameters? How the parameters' ranges have been constrained? In table 1, LUE tree and shrub is out of specified range (Shrub is misspelled).

In this case, the manual calibration was considered mandatory as long as the model had never been used at catchment scale and, therefore, we had not clues about its suitability. Although non-statistical indicators were reported, the manual calibration helped to find the best parameters and constrain the searching boundaries in this following three senses:

1. The best set of parameters obtained after the manual calibration was used as seed for the automatic calibration. We think this fact reduced the computational time required by the automatic calibration as long as this starting point or seed is supposed to be closer to the best global solution than a random starting point.
2. We were allowed to double-ckeck the values of the parameters after the manual calibration with those ones recommended in literature. In this way, we assured that the searching boundaries to be used during the automatic calibration process were consistent and wide enough. The manual calibration pointed out that wider ranges were not required and, in this sense, it constrained this searching boundaries.
3. A manual calibration always gives clues about the potential inter-relationships between parameters. These clues can be used to guide the automatic calibration process (this research was not the case) and to be critic with the results obtained after the automatic calibration (it was the case here) since a sense of relative values was provided by the manual calibration. In that sense, the manual calibration can be extremely helpful to find the best and with physical consistency parameters.

Finally, thanks for the observation about Table three. The boundary for all three cases was 1.12 instead of 1.2 and 'Srhub' will be corrected in the whole table.

5- A clearer explanation regarding EOF_i would be appreciated. What does different i exactly mean?

If we apply the EOF decomposition (also called Principal Component Analysis) to a simple matrix, the EOFi is the i eigenvector. We always assume that the eigenvectors are ordered according to their corresponding eigenvalues (i.e. the amount of variance explained by them). Hence, EOF_1 is the first eigenvector associated with the first eigenvalue and, therefore, which explained more amount of variance. Therefore, i means the position of the eigenvalue when is sorted according the explained variance.

In our research, however, we wanted to apply this methodology to analyse spatio-temporal data. That's why the first step was to transform this data into a matrix. Basically, we construct a matrix (F) in which each column is the temporal variation of the data in a particular cell while each row represents the cells values during a particular time step. Once the matrix was constructed, we applied then the EOF analysis as usual. Therefore, we obtained the eigenvectors as usual. However, these eigenvectors can be regarded as maps by considering the same ordering criterion as used in F construction. In this way, the i-eigenvector becomes to the i-main/principal pattern/map. Hence, EOF_i is the principal pattern associated with the i eigenvalue.

Having reviewed the current manuscript, we found inconsistencies in line 5 and equation 5. We should have kept the same sub-index i instead of j. Otherwise, it might be confusing. We will improve this section and we will check the mathematical consistency within the equations.

6- How would be the model performance with and without calibration on observed satellite data? Any gains or losses there? This would be great to be addressed.

We completely agree with your suggestions. In fact, we are working on it in new on-going projects. In this new applications, we want to use different sources of information (field observations, remote sensing data, etc.) with different resolutions (point measurements, spatio-temporal data, etc.) in order to determine whether models performance improve. However, the study area of this manuscript was discarded for this analysis because this Kenyan catchment can be considered as scarce-data catchment. In fact, the available data is really poor and for this reason, it was precisely selected for this experience. We wanted to face the issue of no having available observations. The calibration was completely 'blind' in terms of observed discharge, i.e. observed discharge was not even known at the beginning of this research. In this way, we assured that the calibration relied only on the satellite data. The main reason to do so was because we did not want to analyse the potential performance improvement by including satellite data, but how far we can arrive by using ONLY satellite data when this data is used as properly as possible. This latter goal leaded and defined all the strategy followed in this research.

Anyway, as mentioned, we also are interested in your suggestion but we would recommend to achieve this goal in study areas with good quality of field data. Hopefully, we can discuss in-depth this topic in following applications.