

# Comments from the Reviewers and the corresponding revision

## UNCERTAINTY REDUCTION AND PARAMETERS ESTIMATION OF A DISTRIBUTED HYDROLOGICAL MODEL WITH GROUND AND REMOTE SENSING DATA

by F. Silvestro, S. Gabellani, R. Rudari, F. Delogu, P. Laiolo and G. Boni

*Dear Editor and Reviewers,*

*Please find below the replies to reviewers' comments and the main changes done to the manuscript or that we can do, if you agree, to modify the manuscript.*

*The comments are reported in red, while the replies are in blue. The text references are reported in Italic black.*

*If the Editor agrees we can produce a new version of the manuscript that includes the corrections. In the meanwhile we report a part of these modifications in the following answers.*

*Best regards.*

### **Reviewer 1**

We thank the reviewer for her/his valuable suggestions that helped us to clarify and improve the manuscript. A detailed answer to each comment is reported in the sequel.

#### **Major comments**

**Comment:** Please include some of the major finding in the abstract. The abstract could be improved by including less information on the introduction and the necessity of the study and more about the approach and results.

**Replay:** We modified the abstract taking into account the suggestion of the reviewer. Here the new abstract

*“During the last decade the opportunity and usefulness of using remote sensing data in hydrology, hydrometeorology and geomorphology has become even more evident and clear. Satellite based products often provide the advantage of observing hydrologic variables in a distributed way that can help to understand and model the hydrological cycle. Moreover, remote sensing data are fundamental in scarce data environments. The use of satellite derived Digital Terrain Model (DTM), which are globally available (e.g. from Shuttle Radar Topographic Mission, SRTM), have become standard practice in hydrologic model implementation, but other types of satellite derived data are still underutilized. As a consequence there is the need of developing and testing techniques that allows exploiting the opportunities given by remote sensing data to parameterize hydrological model and improving their calibration.*

*In this work Meteosat Second Generation Land Surface Temperature (LST) estimates and Surface Soil Moisture (SSM) available from EUMETSAT H-SAF are used together with streamflow observations to calibrate the Continuum hydrological model that computes such state variables in a prognostic mode. The first part of the work aims at proving that the use of satellite observations can reduce uncertainties in parameters calibration by reducing their equifinality. In the second part four parameter estimation strategies are implemented and tested: i) a multi-objective approach that includes ground observations and satellite data as an attempt to use different source of data to add constraints to the parameters ii) two approaches based only on remotely sensed data that reproduce the case of scarce data environment where streamflow observations are not available, iii) a standard calibration based on streamflow observation used as benchmark.*

*Two Italian catchments are used as the test bed to verify the model capability in reproducing long-term (multi-year) simulations.*

*The results of the analysis evidences that model parameters are not always directly and strongly related to the observations, but prove the usefulness of using data from both ground stations and satellite to add constrains to the parameters in the calibration process and reducing the number of equifinal solutions.”*

**Comment: A clear objective is missing from the introduction. Please indicate more clearly the objectives of this study.**

**Replay: We tried to rephrase the paragraph of the introduction in which we show the objective of the study:**

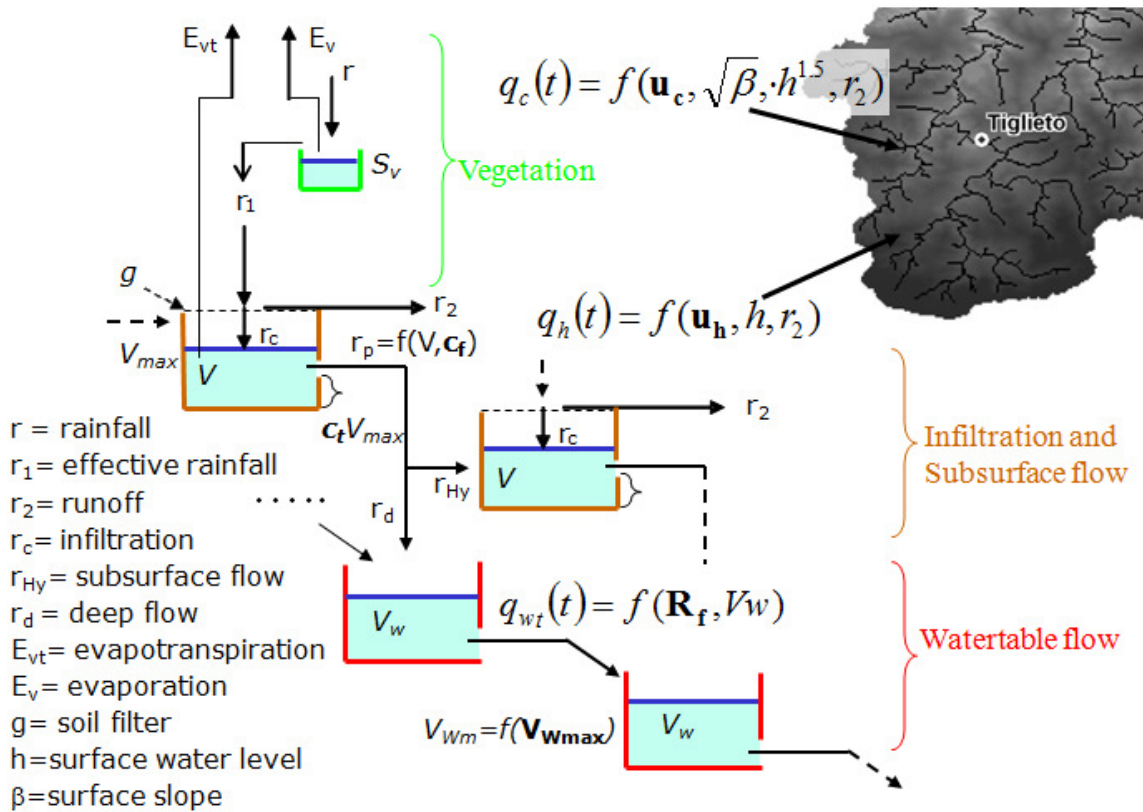
*“The objective of this work is to analyze the improved calibration skill of a distributed continuous hydrologic model by augmenting the model constraints with satellite-retrieved data. As a first analysis, in the context of a classical uncertainty analysis (Beven and Binley, 1992; Shen et al. 2012) it is shown that using satellite data together with ground stations observations can reduce the parameters’ uncertainty and equifinality. Than two simple calibration methods were developed and applied in order to exploit the advantages of utilizing multi-sensor observations. The first method lies...”*

**Comment: Page 6220, how is the downscaling of soil moisture to DEM resolution done? Since ASCAT soil moisture observations, typically, have a much coarser resolution than a DEM.**

**Replay: The soil moisture observations (SM-OBS-1), derived by ASCAT sensor, have a resolution of 25 km and are provided on an irregular grid, which changes with each satellite pass. Each single map has been therefore re-gridded on the regular grid of Continuum (DEM resolution) using the nearest neighbor technique.**

**Comment: Section 2: A Figure containing the set-up of the hydrological model would significantly help the reader to understand the text and the calibration process. Please include.**

Replay: We inserted a new figure with the set up of the model in the new version of the paper as suggested by the reviewer. See below



Caption: Representation of the different processes described in Continuum model and connection between cells. Surface flow is described by non-linear and linear equations respectively on channels and hillslopes.

Comment: A lot of different calibration and validation periods are included and not always overlapping while performance of calibration scenarios are evaluated in different periods as well. Please provide a Table with the different scenario, basins, calibration periods and validation periods. Explain why different calibration periods are selected for different scenarios, is there a good reason or just a random selection. It seems a little bit arbitrary at the moment and is rather confusing for the reader. Please state clearly the different case studies, in the current form it is too confusing for the reader. Are the calibration and validation period overlapping (as suggested by page 6223). Why not separate these periods for a more fair comparison.

Replay: We added a table with the calibration period for each variable. We added a paragraph as introduction of sections 6 to justify our choice:

“.....In both the test cases a calibration period has been chosen for each variable (Streamflow, LST, SWI). Using a same period for all the variable is not always the best option, in fact they describes components of the hydrological cycle that need to be sampled in different periods of time, moreover in certain cases data are available on different periods that not always overlap. Furthermore it is interesting to understand what kind of results can be achieved reducing the length of calibration periods but augmenting the number of observed variables. Alternatively, in the case

*data are available, one could work on longer period of times that ensure to catch the seasonality of the hydrological processes but with the disadvantage of lengthening the calculation time.*

*Table 4 summarizes the calibration periods for the two test cases...”.*

**Comment:** Page 6223 Objective 3, how is the rescaling of ASCAT handled and how do the authors deal with the low penetration depth of ASCAT compared to the model (which presumably has a thicker first soil moisture layer). Also the differences in observation times of the satellite and the model simulation times. Please clarify. Please provide more information on the satellite data, overpass times, availability and resolution. It seems an exponential filter is used to convert surface soil moisture to root zone soil moisture. What delay time is used? This is not stated in the manuscript.

**Replay:** the SSM have been re-sampled to the model resolution using the nearest neighbor method. SM-OBS-1 (H07) data with quality flag greater than 15 were discarded. To obtain the saturation degree in the root zone we applied the exponential filter (Wagner et al., 1999) with characteristic time length equal to 10 days. This filter allows relating the surface soil moisture estimates to the profile soil moisture content. It relies on the assumption that the variation in time of the average value of the soil moisture profile is linearly related to the difference between the surface and the profile values. The SWI is then rescaled to model climatology using a min-max correction technique (Brocca et al., 2013). After the rescaling the mean SWI at basin scale has been computed and used to calibrate the model's parameters.

The section 3 has been modified adding the missing information.

In the following further information are given. SM-OBS-1 (H07) (Large-scale surface soil moisture by radar scatterometer) is based on the radar scatterometer ASCAT embarked on MetOp satellites. It consists of European maps of large scale SSM with a spatial resolution of about 25 km. The values of the SSM range from 0 (dry) to 100 (wet). This product has been developed at TU-Wien since long time, based on the instrument predecessor ESCAT on ERS-1/2. For Western Europe, measurements are generally obtained twice a day: one in the morning (descending orbit) and one in the evening (ascending orbit). ASCAT data are processed soon after each satellite orbit completion and are distributed to end-users in near real time mode, by EUMETCast (the EUMETSAT primary dissemination mechanism for the near real-time delivery of satellite data and products).

**Comment:** A sensitivity analysis is missing. Why are these parameters selected for calibration and additionally it is very difficult for the reader to see which parameter has the highest impact or the highest potential after calibration to improve discharge simulations. Please provide a sensitivity analysis, especially for the selected calibration parameters.

Replay: The sensitivity analysis of the parameters and their impact on output has been showed in Silvestro et al. (2013). This is mentioned in the text and in section 4 some comments about the impact of parameters on model outputs are reported.

Comment: More literature on calibration of hydrological models with remote sensing is needed. More papers exist and should be mentioned in the manuscript, including: Crow, W.T., E.F. Wood and M. Pan, "Multiobjective calibration of land surface model evapotranspiration predictions using streamflow observations and spaceborne surface radiometric temperature retrievals," *Journal of Geophysical Research - Atmospheres*, 108(D23), doi:10.1029/2002JD003292, 2003. Santanello, J., C. D. Peters-Lidard, M. E. Garcia, D. M. Mocko, M. A. Tischler, M. S. Moran, and D. Thoma (2007), Using remotely-sensed estimates of soil moisture to infer soil texture and hydraulic properties across a semi-arid watershed, *Remote Sensing of Environment*, 110 (1), 79 – 97, doi:10.1016/j.rse.2007.02.007. Montzka, C., H. Moradkhani, L. Weihermuller, H.-J. H. Franssen, M. Canty, and H. Vereecken (2011), Hydraulic parameter estimation by remotely-sensed top soil moisture observations with the particle filter, *Journal of Hydrology*, 399 (34), 410 – 421, doi:10.1016/j.jhydrol.2011.01.020. Sutanudjaja, E. H., L. P. H. van Beek, S. M. de Jong, F. C. van Geer, and M. F. P. Bierkens (2013), Calibrating a large-extent high-resolution coupled groundwater-land surface model using soil moisture and discharge data, *Water Resources Research*, pp. n/a–n/a, doi:10.1002/2013WR013807.

Replay: We introduced the reference to most of the suggested papers in the new version of the manuscript in section 5.1.

Comment: Figure and Table caption are insufficient and should be more informative to provide the reader with more information on the content of the figures and results shown.

Replay: We modified the captions in order to make them more informative in the new version of manuscript.

### **Minor comments**

Comment: Page 6216 Line 5: please explain "different view" what kind of different view? Line 7 - 8: please explain DTM and SRTM.

Replay: the abstract has been modified and DTM and SRTM explained: *"The use of satellite derived Digital Terrain Model (DTM), which are globally available (e.g. from Shuttle Radar Topographic Mission, SRTM),*

Comment: Page 6217 Line 2: What is a "complete" model? Line 2-3: please improve the readability of the sentences

Replay: With “complete” we refer to models that solve all the main hydrologic processes (interception, infiltration, interflow, runoff, water table dynamic, evapotranspiration). Scale event models does not satisfy this characteristic. Maybe we could change complete with continuous.

Comment: Line 13: how do you know that the data sharing capacity is exponentially increasing? Reference maybe?

Replay: We corrected with “...data sharing capacity is increasing...”

Comment: Line 23: Should river size also should not be included in the hydraulic parameters?

Replay: We added river size

Comment: Line 28 page 6218- line 4 page 6218: Please rephrase

Comment: Page 6218 Line 2: Spatial resolution of what?

Replay: The sentence has been modified and part of it deleted

*“...Simplified models that make use of some of these parameters introduce uncertainties that limit their applicability (see e.g. Bjerklie et al, 2003), moreover the detection of changes in hydraulic parameters has to deal with the spatiotemporal resolution of the satellite sensors. These models (see Bjerklie et al, 2003 for a comprehensive review) are therefore not suitable for detecting changes in discharge for small-scale basins (Brakenridge et al. 2012)”*

Comment: Line 10-12: Incorrect reference order

Replay: Corrected

Comment: Page 6223 Line 2-3: Contradicts with line 17 Line 9-10 Why a separate paragraph?

Replay: Orba basin and Casentino basin have two different simulation periods due to the availability of meteorological data.

Comment: Line 13: h = hour

Replay: Corrected

Comment: Page 6224 Line 21: fifteen =15 Line 23: The model resolution is equal to the DEM resolution, however, the DEM resolution is not mentioned? Please state here

Replay: We added the information in section 3 “The DEM has a resolution of 0.0011 deg (about 100 m).”

Comment: Line 24:computational reasons, is that stability, if so please say so

Replay: We added stability in the sentence "...for computational stability reasons..."

Comment: Page 6225 Line 3L thirty is 30

Replay: Changed

Comment: Line 13: in -> by

Replay: Corrected in the new version

Comment: Page 6226 Equation 3: is  $\leq$  equal to the mean? If so please use overline

Replay: Corrected in the new version

Comment: Line 6-7 Please give formula's of the scores, also indicate what are good scores, high or low, maximum, minimum score and so on.

Replay: Added in the new version

Comment: Line 14: Dotty plots are just simple scatterplots and probably don't need a reference

Replay: Removed the reference in the new version

Comment: Line 17: Please give letters to the subplots. That is easier for the reader.

Replay: Added in the in the new version

Comment: Please also explain subplots in figure caption Line 19-22: "more complex paths" this can not be stated here. Maybe these parameters simply don't have an impact at all. Please provide sensitivities before making these kind of statements.

Replay: In Silvestro et al. 2013 an analysis of the impact of the parameters on hydrographs has been done, showing that changing  $c_t$  and  $c_f$  there is an important impact on streamflow. We did not report this analysis in this manuscript.

Comment: Line 23-24: Please rephrase Line 26: Why did the authors select 15% and not 10 or something else? Why not 50%? This make Figure 4 rather confusing. Is only 15% of the ensemble members between the lines? If so than a very large proportion seems to be between 15 an 90 percent.

Replay: We now assumed 10 % and 90% as more standard references confidence intervals. We can eventually maintain only 90%. The comment has been changed to avoid confusion:

*"..Most of the observed hydrograph, and specifically the peak flows, lays in the 10% limits. ..."*

Comment: Page 6227 Equation 5: please explain L

Replay: Added in the in the new version

Comment: Line 21: I believe this should be Fig 3

Replay: Corrected in the in the new version

Comment: Page 6228 Line 7: August - September 2009, is this period not far to short for Soil Moisture simulations. Only a very short period of time is considered while dynamics in winter and or spring might be very different. Why are only two months selected? Please specify

Replay: In the summer period the remote sensing data are generally more reliable and continuously available without gaps because of bad weather, moreover we used a short simulation time (but that crosses summer and autumn) in the uncertainty exercise in order to reducing computational time. This has been lengthened for calibration.

Comment: Page 6230 Line 1-2: What please clarify. Why can different periods be considered, more explanation is needed.

Replay: We modify the sentence: *"Different periods for LST, saturation degree and Q could be considered in order to choose the most suitable time window in terms of availability of data and good representativeness of the dynamic of the phenomena mostly influenced by the different variables"*.

We added more details regarding this issue even in the introduction of section 6 (See previous answer).

Comment: Page 6230 Line 26-27 "weakly related to processes that influence LST and SWI observations". How could this statement be made without a sensitivity analysis. For LST I agree that it seems natural that there is a weak relationship. However,SWI represents total soil moisture, which should than also be related to deep soil parameters... Due to the exponential filtering nature of the SWI, I agree that it is unlikely that a strong relationship exist. However a sensitivity analysis is needed to confirm this.



Replay: In the schematization of the adopted model SWI is compared with the humidity of the root zone, which involves the subsurface flow module and its parameters (ct and cf). This is in line with other works (e.g. Brocca et al 2011). The deep flow in the model represents the flow in watertable.

The interaction between the two layers in the adopted schematization is generally confined to the cells near the network (channel cells), as a consequence the impact of deep flow on the basin scale soil humidity of the first soil layer (root zone) is negligible.

Comment: Page 6233 Line 12-14: Why not use one single calibration period for all observations?

Replay: We added a paragraph in the intro of sections 6 to justify our choice.

Comment: Page 6234 Line 1-3: How is it possible to obtain routing parameters with Remote sensing of LST and SM? Both of these observations do not contain information on the discharge levels, nor on the routing of the discharge through the channel network.

Replay: We calibrate the routing parameter using the DEM that can be a remote sensing information (SRTM). LST and SM are used to calibrate other .

Comment: Page 6237 Line 4-6: What kind of information can be derived from the DEM?

Replay: The information needed to estimate the characteristic lag time (which is generally function of Area, Slope...). See section 5.2.1

Comment: Figure 1: Improve this figure. Include coordinates place on the world, scale

Replay: Improved in the new version

Comment: Figure 2: Include more information, which case, which time period, number panels. Figure should be self-explaining

Replay: Improved in the new version

Comment: Figure 3: Please explain more, what do we see, why are the strange stripes present (undersampling?), number panel (a, b, c, d)

Replay: A new version of the figure has been inserted in the new version of the paper. We added more details in the caption.

Comment: Figure 4: Why is so much data outside the 15% and not within the 90%, please give ensemble mean, to get a better sense of the distribution

Replay: Added the ensemble mean and used more standard 10% and 90% confidence intervals

Comment: Figure 11: Lower plot around 2007-12-22, why do we see the flat soil moisture lines, is this a model artifact?

Replay: The figure represent the evolution of discharge in time (a hydrograph), in the period around 2007-12-22 the flow is quite constant.

Comment: Figure 12: Lower plot (please number again, a, b, c, ...) half of the time series can not be seen in the plot, please adjust the scale.

Replay: We adjusted a little bit the scale but in some periods the streamflow drops to 0; it is not possible to visualize it in log scale

## **Reviewer 2**

We thank the reviewer for her/his valuable suggestions that helped us to clarify and improve the manuscript. A detailed answer to each comment is reported in the sequel.

### **General comments**

The authors present a modelling experiment on two Italian basins to show the potential of including remotely sensed data to improve the calibration of a hydrological model.

Their work is based on a sensitivity analysis of model parameters and on the assessment of model performance for four calibration strategies using several skill scores.

The main contributions are the methodology to incorporate satellite observations to model calibration and some conclusions extracted from the discussion of results.

The research topic is interesting and the results are relevant. In my opinion, the work deserves publication in Hydrology and Earth System Sciences. However, the paper needs more elaboration before it can be finally published. The manuscript is poorly organized and many details require additional polish. It lacks a clear strategy for presentation of the work and the authors mix methodology and results too frequently.

Some figures need additional work.

The presentation of the work needs improvement. I suggest a clear division between presentation of methodology and results as applied to the two case studies. The authors should clearly state their objectives at the beginning and devote a methodological section to present their strategy, indicating the alternatives to be compared and the metrics that will be used in the comparison.

From the text I gather that the analysis is focused on a sensitivity experiment to learn about parameter uncertainty and a calibration experiment to learn about parameter estimation. Sensitivity analyses are performed on streamflow, land surface temperature and surface soil moisture. The calibration strategies compared are four: 1) Calibration with streamflow alone, SN; 2) Calibration with land surface temperature, RS (LST); 3) Calibration with surface soil moisture, RS (SWI) and 4) Multiobjective calibration, MO. I would have liked to see an orderly presentation of the approach followed on these analyses, together with a clear definition of the metrics used to evaluate the performances. Then the two case studies can be introduced and the results presented, followed by the corresponding discussion.

Apart from the lack of structure, the manuscript is not properly finished. There are too many typographical errors that indicate that the manuscript is still at an early stage of revision. The authors should have checked their manuscript for such evident errors before submission. You may find below several of such errors but the list is not complete.

Reply: We tried to follow most of the suggestion of the reviewer in the new version of the manuscript. We even added a specific section to describe the calibration experiment settings. More

details have been introduced regarding parameters, scores, and model features. Various Figures have been modified and corrected.

### Specific comments

**Comment:** The abstract needs reworking. It should focus on the new material presented in the paper rather than on general comments (first paragraph). The statement that satellite observations “dramatically” reduce uncertainties in parameters is not supported by the results presented in the paper, since the improvement obtained in performance metrics while using remote sensing is not substantial. The paper compares four calibration strategies, not two. The results obtained and the conclusions drawn should also be mentioned in the abstract, since they are the most relevant information for the reader.

**Replay:** Abstract has been modified in the new version of the manuscript follow the suggestions of the reviewer.

The paper does not aim to improve in absolute values a certain metrics (for example the NS on hydrographs) but to identify parameters sets that can produce good metrics on different observations (Streamflow, LST, SWI). This should help in a better overall calibration of the model.

Here the new abstract:

*During the last decade the opportunity and usefulness of using remote sensing data in hydrology, hydrometeorology and geomorphology has become even more evident and clear. Satellite based products often provide the advantage of observing hydrologic variables in a distributed way while offering a different view that can help to understand and model the hydrological cycle. Moreover, remote sensing data are fundamental in scarce data environments. The use of satellite derived Digital Terrain Model (DTM), which are globally available (e.g. from Shuttle Radar Topographic Mission, SRTM), have become standard practice in hydrologic model implementation, but other types of satellite derived data are still underutilized. As a consequence there is the need of developing and testing techniques that allows exploiting the opportunities given by remote sensing data to parameterize hydrological model and improving their calibration.*

*In this work Meteosat Second Generation Land Surface Temperature (LST) estimates and Surface Soil Moisture (SSM) available from EUMETSAT H-SAF are used together with streamflow observations to calibrate the Continuum hydrological model that computes such state variables in a prognostic mode. The first part of the work aims at proving that the use of satellite observations can reduce uncertainties in parameters calibration by reducing their equifinality. In the second part four parameter estimation strategies are implemented and tested: i) a multi-objective approach that includes ground observations and satellite data ass an attempt to use different source of data to add constraints to the parameters ii) two approaches based only on remotely sensed data that reproduce the case of scarce data environment where streamflow observations are not available, iii) a standard calibration based on streamflow observation used as benchmark.*

*Two Italian catchments are used as the test bed to verify the model capability in reproducing long-term (multi-year) simulations.*

*The results of the analysis evidences that model parameters are not always directly and strongly related to the observations, but prove the usefulness of using data from both ground stations and satellite to add constrains to the parameters in the calibration process and reducing the number of equifinal solutions.*

Comment: The basic calibration parameters of the Continuum models should be presented and discussed in more depth. Table 1 presents the six calibration parameters of Continuum, with an indication of the physical process parameterized. However, I would have liked to see a deeper discussion of the physical meaning of parameters themselves. There are three adimensional parameters ( $c_f$ ,  $c_t$  and  $R_f$ ), but these are just numbers. I can speculate on the physical meaning of  $R_f$ , if it describes the ratio between vertical and horizontal conductivity, but I have no clue on the role of  $c_f$  controlling infiltration capacity at saturation or the role of  $c_t$  controlling field capacity. Likewise for the dimensional parameters: flow motion in hillslope is controlled by a parameter with units of inverse time while friction in channels is controlled by a parameter with more complex units. In order to properly understand the discussions, the reader should be presented with a detailed explanation of calibration parameters and their role on model equations. The authors state the Continuum is a distributed model, but calibration parameters are lumped. How are they related to the physical properties distributed over the basin? Do the same calibration parameters apply to forests, grassland and cultivated areas? How do their values affect, for instance, local soil properties? The author is referred to the original paper by Silvestro et al., 2013, but in that reference there is little discussion of the role of these parameters on calibration, which is the central topic of this manuscript. On section 4, four sources of remotely sensed data are presented, but it is not clear to me how Leaf Area Index data were used. Please explain.

Replay: In the new version of the manuscript we add more discussion about calibration parameters and their role. We also added a new figure (reported below), as suggested by reviewer 1, that represents the main setting of the model.

We report the main paragraph:

*“...The hillslope flow motion parameter  $u_h$  influences the general shape of the hydrograph, while the impact of  $u_c$  on the hydrograph shape depends on the length of the channeled paths. These are two lumped parameters:  $u_c$  represents the friction coefficient in the channel motion equations,  $u_h$  accounts for the general characteristics of the hillslope that influence the motion (friction, slope,...) and is more an empirical parameter (see Figure 1- In the text we inserted a new figure).*

*The parameter  $c_t$  is related to the soil field capacity  $V_{fc}$  and identifies the fraction of water volume in the soil that can be extracted only through evapotranspiration. The relationship is:*

$$V_{fc} = c_t V_{max}$$

*Where  $V_{max}$  is the maximum capacity of the soil to storage water.*

The “infiltration capacity” parameter  $c_f$  controls the velocity of subsurface flow (i.e, it is related to saturated hydraulic conductivity), defining the asymptotic minimum infiltration rate for saturated soils  $f_1$  with the following equation:

$$f_1 = c_f \cdot f_0$$

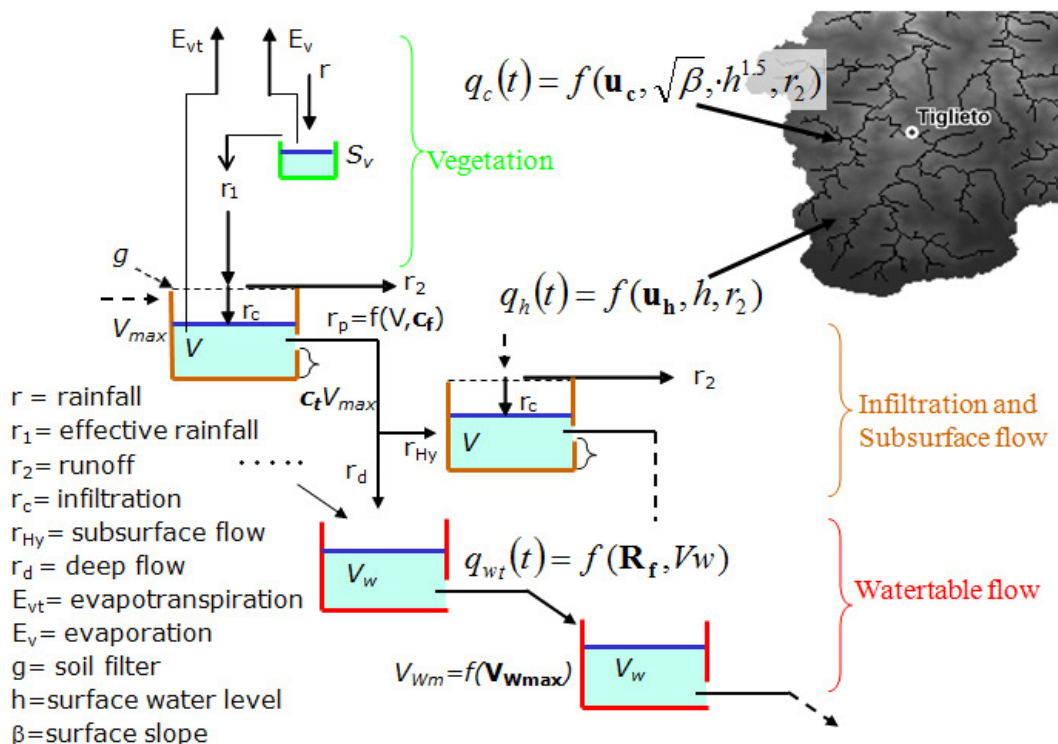
Where  $f_0$  is the maximum infiltration rate for completely dry soils.

Both  $c_t$  and  $c_f$  regulate the dynamics of saturation at cell scale. Since both  $f_0$  and  $V_{max}$  are distributed parameters function of Curve Number (Gabellani et al., 2008) the pattern of  $f_1$  and  $V_{fc}$  is spatially modulated by the pattern of Curve Number maps (Silvestro et al., 2013) which are a synthetic representation of the local soil properties.

The parameters  $V_{Wmax}$  and  $R_f$  govern the deep flow and the water table dynamic (Silvestro et al., 2013).  $V_{Wmax}$  represents the absolute maximum water content of the aquifer on the whole investigated area, the maximum water content on each cell is estimated basing on  $V_{Wmax[SG1]}$  and on the slope (Saulnier et al., 1997).  $R_f$  is a multiplicative factor in the Darcy equation used to estimate the flux per unit area between two contiguous cells and mainly takes care of differentiating the saturated vertical and horizontal conductivity. These two parameters have a reduced influence compared to the other four parameters because of the slow temporal dynamic of the water table. The sensitivity to  $R_f$  increases with the total basin drainage area when the effect of the interaction between the water table and the vadose zone becomes crucial in the formation of the recession curve between the rainfall events (Silvestro et al., 2013).

.... “

“...Leaf Area Index (LAI) is used to parameterize the storage capacity of the vegetation (Kozak et al., 2007).”



Comment: By looking at Figure 7 it appears that LST is as insensitive to parameter  $c_f$  as it is insensitive to parameters  $u_c$  and  $u_h$ . A similar situation occurs in Figure 8 with respect to SWI, although here SWI shows some sensitivity to  $c_f$ . However, LST and SWI are later on used to estimate parameter  $c_f$ . Is this possible? Please discuss.

Replay: Figure 7 shows that there are not clear paths that identify a range of the parameter  $c_f$  where we have optimal values of the statistic independently by the values of other parameters. These optimal values are distributed (more or less along all the range of the parameter). This means we have difficulties to constrain this parameter with LST observations only. In Figure 8 the situation is a little bit better: it is possible to identify a range of parameters where there is a maximum and the Y (metric) range is narrower.

We agree that in the case of LST probably it is not easy to constrain  $c_f$  parameter adequately.

We added more comment in section 6.2 regarding these points when commenting results on Orba basin. See also the answer to the next comment.

Comment: Overall, I find the discussion of results in section 4 quite weak to support the conclusion stated on page 12, lines 7-8. The plots shown in Figures 5 and 6 are very confusing. They indeed convey an image of equifinality. However, the statement that calibration with remotely sensed data can help to reduce equifinality is not clear to me. Optimal values are only apparent for parameter  $c_t$  under LST calibration but the optimal values corresponds to low values of the probability distribution shown in Figure 6. This would mean an additional peak in the joint distribution, increasing equifinality. Optimal values for  $c_t$  under SWI calibration do coincide with peak values in Figure 6, but the maximum for  $c_t$  in Figure is very weak and it is unlikely to reduce equifinality. The discussion of results should be further elaborated if that conclusion is proposed from the analysis.

Replay: We do not have the claim of eliminating the equifinality but we tried to explore ways to reduce it. Although Figure 5 and 6 give an image of the presence of a certain equifinality they show ranges of the parameter combinations with evident maximum probability. We evidenced in section 4 that various maxima of the likelihood function are present. Those figures are quite useful to the discussion and we propose to keep them in the manuscript but if editor and reviewer disagree we can remove them.

We added some comments in section 4 and section 6.2 to discuss the issues of the reviewer.

*“...The maximum of NS  $c_t$  lies in the range 0.45-0.55 of the parameter  $c_t$ ; and a weak, but quite evident independence of  $c_f$  arises with optimal values around 0.015-0.025 (close to the lower limit of the parameter range). In both cases the NS values are in a quite narrow range in correspondence of the aforementioned parameters range....”*



*“..In the case of R.S. (SWI) the calibrated  $c_t$  and  $c_f$  values confirm the finding of section 4; even if the dotted plot in Figure 9 does not show a really strong independence from the other parameters but more a range of the two parameters that always furnish good reproduction of SWI.*

*In the case of R.S. (LST) results are different respect the finding of section 4 (Figure 8). The  $c_t$  optimal parameter has a different value in respect the optimal range found in uncertainty analysis, maybe this could be related to the fact that a longer period for comparison has been used, or to the combination with parameter  $c_f$  that do not show independence from the other parameters in LST analysis (Figure 8). When using the R.S (LST) strategy the benefit of exploiting LST data seems more related to opportunity of doing a calibration in case of lack of streamflow data than in reducing equifinality....”*

*“...M.O. approach appears to be the best way to attempt the equifinality reduction; looking at the dotted plot representations in section built with streamflow, LST and SWI data, it is evident that even when graphs show an independence of a parameter from the others, this independence is not very pronounced. In other cases there is no evident independence. Anyway combining the different objectives showed in section 5.1 we should eliminate those solutions (parameters combinations) that give good values for a certain metric (for example NS on streamflow) but very bad values for another one (for example BIAS on LST). This does not allow to eliminate equifinality but goes in the direction of reducing it and of obtaining an overall better calibration....”*

**Comment: Mathematical rigor is a must in scientific papers. Performance metrics should be mathematically described, even if they are common: On Equation 3, the meaning of  $\langle Q_0 \rangle$  is missing Expressions for the four performance metrics proposed (CM, RMSE, Corr, Rel.Err) should be included. Same for Bias, introduced later on (in absolute values?).**

**Expressions for computing the basin “Saturation Degree” and “Mean land surface temperature” should be provided. Are they temporal or spatial means? What is their time step compared to that of the model? The second term of the objective function presented in Equation 7 (page 12) has dimensions of temperature. How can it be added to the other three adimensional terms? The objective function as shown in Equation 7 is never applied in the paper. Why not define the individual terms  $F_i$  and then integrate them in Equation 8? From the description of the calibration periods (see below) it is not clear if all components of  $F$  are summed over the same time interval from  $t=1$  to  $t_{max}$ .**

**Please clarify.**

**Equation 9 is not clear. What is  $F_{j,min}$ ? How can  $A_i$  depend on  $F_{j,min}$ ?**

**Replay: we added more detail on basin LST and SWI computation and about how they are compared with model output (section 3).**

**We included the corrected definition of the statistics in the new version of the manuscript (section 4).**



We added more details regarding multi-objective function (section 5.1) and the period of data used for each involved variables (section 6.1) as suggested by the reviewer.

There was a mistake in the definition of the corrective factor  $A_i$  in the multi-objective function, the correct one is:

$$A_i = \text{MAX}(F_{j,\min}; j = 1, 2, \dots, n) - F_{i,\min}$$

$A_i$  depends even on  $F_j$  as defined in Madsen (2000), for each term of the objective function a minimum can be found on the  $M$  available parameter sets and this is used to calculate the transformation constant.

**Comment:** How does the normalization of the four components of  $F_{adj}$  work? Do any of them have more weight than the others? A brief discussion is appropriate. How are the observed values of remotely sensed data compared to simulated values in the basin? What is the temporal and spatial resolution in both cases? There are certainly scale problems associated to the procedure. Please explain, including mathematical expressions if possible.

**Replay:** the corrective factor  $A_i$  in the multi-objective function has the role of putting equal weights on the different objectives of the function (Madsen 2000). We explicitly indicated this fact in section 5.1 in the new version of the manuscript.

More details about temporal and spatial resolution of the data and the comparison with model output have been added in section 3

**Comment:** How are time series sorted according to NS to produce Figure 4? How is it related to the estimation of percentiles every time step? Why choosing asymmetric confidence limits (15% and 90%, page 10, line 19) for Figure 4?

**Replay:** We introduced standard confidence intervals 10% and 90% in the new version. For each time step the X% hydrographs with best NS values are considered, and the lower and higher streamflow values are considered.

**Comment:** The periods chosen for the different sensitivity analyses and calibrations in the two case studies differ without a clear justification and it is very difficult to even realize the differences. A systematic listing of such periods would be advisable. They are ambiguously described, both for the Orba and the Casentino basins. There are four calibrations (tables 4 and 8), while the text mentions only three calibration periods.

What is the calibration period for MO? Is it a mixture of the three individual periods or were the four terms of the objective function applied to the same period? Please clarify.

The chosen calibration periods do not even cover a full year, which is surprising considering the strong seasonality of hydrologic processes. If the calibration periods were chosen because of limited data availability, the authors say so. If they were chosen for other reasons, they should be clearly presented. The reference to “different flood and drought regimes” (page 17, lines 5-6) is too vague and does not convey real meaning.

The last paragraph of page 17 is not a conclusion. It should be removed from the conclusions section.

Replay: In the new version of manuscript more details are furnished regarding the choices done in designing the calibration experiment.

A table with the different periods used in the calibration process for Orba and Casentino river has been added. In text we add explicitly mention of the fact that in M.O. periods are merged (in section 6.1)

In new section (6.1) the topic of temporal periods and calibration setting has been faced. We report part of it:

*“.....In both the test cases a calibration period has been chosen for each variable (Streamflow, LST, SWI). Using a same period for all the variable is not always the best option, in fact they describes components of the hydrological cycle that need to be sampled in different periods of time, moreover in certain cases data are available on different periods that not always overlap. Furthermore it is interesting to understand what kind of results can be achieved reducing the length of calibration periods but augmenting the number of observed variables. Alternatively, in the case data are available, one could work on longer period of times that ensure to catch the seasonality of the hydrological processes but with the disadvantage of lengthening the calculation time.*

*Table 4 summarizes the calibration periods for the two test cases.*

*The calibration strategies that have been compared are four:*

- *The M.O. strategy described in section 5.1 (in this case the calibration periods of the different observations are merged)*
- *The R.S. approach described in section 5.2 using as comparison data the satellite LST. Hereafter we will call this strategy R.S. (LST)*
- *The R.S. approach described in section 5.2 using as comparison data SWI estimation derived from satellite SSM. Hereafter we will call this strategy R.S. (SWI)*
- *A standard approach based on the maximization of the Nash Sutcliffe between observed and modeled streamflow, hereafter we will call this method S.N.”*

The last paragraph of page 17 has been removed as suggest.

We even added a sentence in section 3 to clarify the choice of different periods for the two case studies:

*“In both cases the period of data has been chosen based on the data availability and in order to have reliable stage-discharge curves to estimate streamflow.”*

### **Minor details**

Comment: Section numbering is not correct. We find section 1.1 under section 2, sections 1.2 and 1.5 under section 5, sections 1.1.1 and 1.1.2 under section 1.3 and section 1.4 and 1.2 under section 6

Replay: Section numbering has been corrected in the new version of the manuscript

Comment: Figure captions are too schematic. They should be elaborated to identify exactly what is being represented.

Replay: Figure captions have been improved in the new version of the manuscript.

On pages 11-12, lines 27-28 and 1, it is stated that Figure 8 shows the comparison of the model saturation degree and satellite SWI. On the caption, Figure 8 is described as representing NS vs NWI. However, Figure 8 actually shows a plot of NS versus parameter values. Please clarify.

Replay: We changed the sentence:

*“The model saturation degree and the satellite SWI maps have been averaged at basin scale and the resulting time series are used to build a dotty-plot graph”.*

Comment: The first sentence of the second paragraph of page 12 (lines 3-4) is not clear. It mentions the maximum of ct while it may refer to the maximum of NS.

Replay: Corrected.

Comment: On section 5, a subsection is devoted to calibration strategies MO, RS (LTS) and RS (SWI). Another subsection should be devoted to SN.

Replay: We inserted a new subsection for SN

Comment: In some figures it is difficult to interpret the time axis. Rather than showing full dates at random I suggest to choose time tags and the beginning of months or years to facilitate the interpretation of the plots.

Replay: Figures have been modified as requested

Comment: Three calibration strategies are mentioned on page 15, line 19, while Table 4 shows four.

Replay: Corrected.

Comment: On page 15, lines 25-26 a reference is made to hydrographs obtained with the best parameter sets. Where are they shown?

Replay: Corrected in the new version. A inconsistent sentence was erroneously inserted.

Comment: The notation should be consistent. The parameter  $u_h$  is mentioned as  $u_v$  in the text and in several figures. On tables 4 and 8, parameter  $u_c$  is mentioned two times.

Replay: The correct notation is  $u_h$ , we corrected text, figures and tables

Comment: Typographical errors: On page 3, line 8, change "significantly" into "significant"

Replay: Corrected.

Comment: Onpage 7, line 81, change "areas" into "area"

Replay: Corrected.

Comment: On page 11, line 12, change "estimate" into "estimation"

Replay: Corrected.