Interactive comment on “The Latest Improvements in SURFEX v8.0 of the Safran-Isba-Modcou Hydrometeorological Model over France” *by* Patrick Le Moigne et al.

Anonymous Referee #1

Received and published: 10 March 2020

1 - The colour black is used for the reviewer questions/remarks.
2 - The colour green is used to answer questions raised by the reviewer.
3 - The colour blue is used for the text added to the manuscript.

This paper presents an analysis of a number of improvements to a land surface model. I do appreciate the amount of work the authors have performed, but at this point I also have a number of major comments to the paper.

1 - A first problem is the quality of the writing. There are a large number of grammatical errors that many times make the paper very hard to read. This renders the paper unacceptable for publication.

2 - The manuscript was completely revised and corrected by a native English speaker.

1 - Top of page 2: There is a vast amount of literature on the evaluation of land surface models at various spatial scales using Bowen ratio or eddy covariance data (latent and sensible heat fluxes), ground heat flux data, soil temperature and moisture profiles, infiltration data, ground water levels, etc. Similar comparisons have been performed using remote sensing data. Stating that models have been evaluated using discharge data and ground water levels is blatantly ignoring an enormous amount of literature.

2 - It was not intended to state that the models were evaluated using only flow and groundwater level data in general, but only in this study. However, the authors would like to thank the reviewer for this remark, which shows a lack of references to studies that have been conducted in the past and still are conducted today to evaluate LSMs with different data sets.

The introduction of the manuscript has been modified and the following paragraph has been added (top of page 2 to top of page 3, i.e. lines 35-71):

3 – Land surface models (LSMs), whether or not coupled to hydrology, have been the subject of numerous studies that have improved them over time and have led to a better description and understanding of the key processes governing exchanges at the interface between the surface and the atmosphere and the surface and the subsurface. These studies have made it possible to evaluate surface models, and even certain parameterizations, by comparing simulation results

with different types of observations such as in situ measurements, reanalyses or satellite products. Over time, a number of international measurement campaigns have been organized to evaluate the performance of the models by comparing them with in situ measurements. Simulations were carried out offline, i.e. decoupled from the atmosphere, to avoid introducing atmospheric biases into the surface schemes. The first international model intercomparison projects were the Project of Intercomparison of Land surface Parametrization Schemes (PILPS), described in Henderson-Sellers et al. (1996), which began with forcing from atmospheric simulations (Pitman et al., 1993) and, in a second stage, forcing from local observations (Chen et al., 1997). The successive phases also focused on different issues, such as snow and frost parameterization (Schlosser et al., 2000), river flow assessment (Wood et al., 1998; Bowling et al., 2003). In the spirit of PILPS, GSWP (Global Soil Wetness Project, Dirmeyer, 2011) was initiated with global scale simulations. The results of this project are the first global offline multi-model simulations of LSMs. Other more specific intercomparison projects have been carried out such as SnowMIP (Etchevers et al., 2004) to study snow-related processes, ALMIP (Boone et al., 2009), focusing on critical surface processes in West Africa at regional scale, or Rhône-AGGrégation (Boone et al., 2004) to study coupling with hydrology. More recently, the PLUMBER project (Best et al., 2015) has attempted to identify how LSMs behave in relation to certain benchmarks and to define performance criteria that LSMs should be able to achieve according to the information available in atmospheric forcing, thus avoiding direct comparison with observations.

In many of these intercomparison studies, the surface models were validated at the local scale and used average parameters that were known fairly accurately. However, these models sometimes have strongly non-linear components, such as the link between root zone moisture and transpiration when the soil dries out (Sellers et al., 1997), so it is necessary to develop sub-grid parameterizations to compensate for the lack of representativeness of the mean parameters. Overgaard et al. (2006) conducted a review of surface models based on energy balance and used for hydrological purposes. They stressed the need to validate the models at the local scale, but also showed the interest of using remote sensing data to evaluate the models. Indeed, the validation of LSMs using river flows alone does not prove that surface fluxes, for example, are well simulated by the model and that there is no error compensation. Furthermore, estimating surface fluxes by remote sensing is not straightforward and requires certain assumptions that are not always valid, and inversion models are used to translate the remote sensing measurement into a model variable equivalent. However, using surface fluxes to validate surface models is also subject to questioning since the energy balance measured at the surface is generally not closed (Foken, 2008) whereas it is an imposed constraint in surface models. The use of international measurement networks such as FLUXNET (El Mayaar et al., 2008; Napoly et al., 2017) is also widely used to evaluate surface models at the point scale. Remote sensing provides a means of observing hydrological state variables over large areas (Schmugge et al., 2002) and can be useful in the case of LSMs coupled to hydrological models, in particular in order to assess evaporation (Kalma et al., 2008; Long et al., 2014; Wang et al., 2015) or soil moisture (Goward et al., 2000; Albergel et al., 2012; Fang et al., 2016). It should be noted that these remote sensing data can be assimilated to correct the model state variables at the initial time as well as during the hindcast (Albergel, et al., 2017).

The list of references has been modified and the following references added:

Henderson-Sellers, A., McGuffie, K. & Pitman, A. The Project for Intercomparison of Land-surface Parametrization Schemes (PILPS): 1992 to 1995. *Climate Dynamics* **12,** 849–859 (1996). https://doi.org/10.1007/s003820050147

Pitman, A., Henderson-Sellers, A., Abramopoulos, F., Avissar, R., Bonan, G., Boone, A., Cogley, J., Dickin- son, R., Ek, M., Entekhabi, D., et al. (1993). Project for intercomparison of land-surface parameteriza- tion schemes (pilps): results from off-line control simulations (phase 1a). *Inter GEWEX Project Office Publ*, 7.

Chen, T. H., Henderson-Sellers, A., Milly, P., Pitman, A., Beljaars, A., Polcher, J., Abramopoulos, F., Boone, A., Chang, S., Chen, F., et al. (1997). Cabauw experimental results from the project for intercomparison of land-surface parameterization schemes. *Journal of Climate*, 10(6):1194–1215.

Schlosser, C. A., Slater, A. G., Robock, A., Pitman, A. J., Vinnikov, K. Y., Henderson-Sellers, A., Speranskaya, N. A., and Mitchell, K. (2000). Simulations of a boreal grassland hydrology at valdai, russia: Pilps phase 2 (d). *Monthly Weather Review*, 128(2):301–321.

Wood, E. F., Lettenmaier, D. P., Liang, X., Lohmann, D., Boone, A., Chang, S., Chen, F., Dai, Y., Dickin- son, R. E., Duan, Q., et al. (1998). The project for intercomparison of land-surface parameterization schemes (pilps) phase 2 (c) red–arkansas river basin experiment: 1. experiment description and sum- mary intercomparisons. *Global and Planetary Change*, 19(1):115–135.

Bowling, L. C., Kane, D. L., Gieck, R. E., Hinzman, L. D., and Lettenmaier, D. P. (2003). The role of surface storage in a low-gradient arctic watershed. *Water Resources Research*, 39(4).

Dirmeyer, P. A. (2011). A history and review of the global soil wetness project (gswp). *Journal of Hydrom- eteorology*, 12(5):729–749.

Etchevers, P., Martin, E., Brown, R., Fierz, C., Lejeune, Y., Bazile, E., Boone, A., Dai, Y.-J., Essery, R., Fer- nandez, A., et al. (2004). Validation of the energy budget of an alpine snowpack simulated by several snow models (snowmip project). *Annals of Glaciology*, 38(1):150–158.

Boone, A., De Rosnay, P., Balsamo, G., Beljaars, A., Chopin, F., Decharme, B., Delire, C., Ducharne, A., Gascoin, S., Grippa, M., et al. (2009). The amma land surface model intercomparison project (almip). *Bulletin of the American Meteorological Society*, 90(12):1865.

Boone, A., Habets, F., Noilhan, J., Clark, D., Dirmeyer, P., Fox, S., Gusev, Y., Haddeland, I., Koster, R., Lohmann, D., et al. (2004). The rhone-aggregation land surface scheme intercomparison project: An overview. *Journal of Climate*, 17(1):187–208.

Best, M., Abramowitz, G., Johnson, H., Pitman, A., Balsamo, G., Boone, A., Cuntz, M., Decharme, B., Dirmeyer, P., Dong, J., et al. (2015). The plumbing of land surface models: benchmarking model per- formance. *J. Hydrometeor.*, 16(3):1425–1442.

Sellers, P., Dickinson, R., Randall, D., Betts, A., Hall, F., Berry, J., Collatz, G., Denning, A., Mooney, H., Nobre, C., et al. (1997). Modeling the exchanges of energy, water, and carbon between continents and the atmosphere. *Science*, 275(5299):502–509.

Overgaard, J., Rosbjerg, D., and Butts, M. B.: Land-surface modelling in hydrological perspective – a review. Biogeosciences, European Geosciences Union, 2006, 3 (2), pp.229-241. hal-00297556

Foken T. (2008). THE ENERGY BALANCE CLOSURE PROBLEM: AN OVERVIEW. Ecological Society of America, Volume 18, Issue 6, pages 1351-1367. https://doi.org/10.1890/06-0922.1

El Maayar, M., Chen, J. M., Price, D. T.: On the use of field measurements of energy fluxes to evaluate land surface models. ecological modelling, Volume 214, pages 293–304, 2008.

Napoly, A., Boone, A., Samuelsson, P., Gollvik, S., Martin, E., Seferian, R., Carrer, D., Decharme, B., and Jarlan, L.: The interactions between soil–biosphere–atmosphere (ISBA) land surface model multi-energy balance (MEB) option in SURFEXv8 – Part 2: Introduction of a litter formulation and model evaluation for local-scale forest sites, Geosci. Model Dev., 10, 1621–1644, https://doi.org/10.5194/gmd-10-1621-2017, 2017.

Schmugge, T. J., Kustas, W. P., Ritchie J. C., Jackson, T. J., and Rango, A.: Remote sensing in hydrology. Advances in water resources. Volue 25, Issue 8-12, pages 1367-1385, 2002.

Kalma, J. D., McVicar, T. R.,and McCabe, M. F.: Estimating Land Surface Evaporation: A Review of Methods Using Remotely Sensed Surface Temperature Data. Surv Geophys, 29:421–469, https://doi.org/10.1007/s10712-008-9037-z, 2008.

Long., D., Longuevergne, L., and Scanlon, B. R.: Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites. Water Resources Research. Volume 50, Issue 2, pages 1131-1151, https://doi.org/10.1002/2013WR014581, 2014.

Wang, S., Pan, M., Mu, Q., Shi, X., Mao, J., Brümmer, C., Jassal, R. S., Krishnan, P., Li, J., and Black, T. A.: Comparing Evapotranspiration from Eddy Covariance Measurements, Water Budgets, Remote Sensing, and Land Surface Models over Canada. Journal of Hydrometeorology, Volume 16, 1540-1560, 2015.

Goward, S. N., Xue, Y., and Czajkowski, K. P.: Evaluating land surface moisture conditions from the remotely sensed temperature/vegetation index measurements. An exploration with the simplified simple biosphere model. Remote Sensing of Environment, Volume 79, Pages 225–242, 2000.

Albergel, C., de Rosnay, P., Gruhier, C., Muñoz-Sabater, J., Hasenauer, S., Isaksen, L., Kerr, Y., and Wagner, W.: Evaluation of remotely sensed and modelled soil moisture products using global ground-based in situ observations. Remote Sensing of Environment, Volume 118, Pages 215–226, 2012.

Fang, L., Hain, C. R., Zhan, X., and Anderson, M. C.: An inter-comparison of soil moisture data products from satellite remote sensing and a land surface model. International Journal of Applied Earth Observation and Geoinformation, Volume 48, Pages 37-50, 2016.

Albergel, C., Munier, S., Leroux, D. J., Dewaele, H., Fairbairn, D., Barbu, A. L., Gelati, E., Dorigo, W., Faroux, S., Meurey, C., Le Moigne, P., Decharme, B., Mahfouf, J.-F., and Calvet, J.-C.: Sequential assimilation of satellite-derived vegetation and soil moisture products using SURFEX_v8.0: LDAS-Monde assessment over the Euro-Mediterranean area, Geosci. Model Dev., 10, 3889–3912, https://doi.org/10.5194/gmd-10-3889-2017, 2017.

1 - We need more information on how the point data were upscaled to allow comparisons with the model grid results. Comparing point data to spatially averaged model results does not make much sense. Furthermore, some grids may contain more stations than others. This will impact the results of the comparison, and needs to be explained better.

2 - It is not clear what data the reviewer is referring to. In any case, the answer we propose is valid for all types of observations used in the study. First, the model results are not spatially averaged when compared to point data. Rather, the reverse is true, since it is a downscaling from a resolution of a few tens of kilometres to a finer resolution of 8 km for the surface model and up to 1 km for the hydrogeological model. The method is as follows: the SAFRAN analysis is performed on homogeneous zones in terms of horizontal gradients. The analysed fields are then spatially interpolated to a regular 8 km grid taking altitude into account. Thus, the comparison of IR radiation is made between the SAFRAN analysis interpolated at 8km and point data (GLACIOCLIM or LSAF). The horizontal variability of IR radiation at 8 km is small enough to allow a direct comparison with in situ observations. Second, the ISBA model outputs of ground temperature and snow depth profiles are relatively sparse and only a direct comparison between the model outputs and the observations is possible. Finally, with respect to river flows, the MODCOU model grid varies in the range of 8 km to 1 km near the riverbed, and the comparison between the model output and the observed flow is made by considering the flow at the river outlet and the corresponding model grid point in the 1 km hydrological network grid.

The section 3.2 of manuscript was modified and the following paragraph was added at the end, after the description of the different datasets (lines 312-321):

3 – The SAFRAN analysis is performed on homogeneous zones in terms of horizontal gradients, and the analysed fields are spatially interpolated to a regular 8 km grid taking altitude into account. Thus, the comparison of infrared radiation (IR) is made between the SAFRAN analysis interpolated at 8 km and the local observation. The horizontal variability of IR radiation at 8 km

is small enough to allow a direct comparison with in situ observations. Moreover, the ISBA model outputs of ground temperature and snow depth profiles are relatively sparse and only a direct comparison between the model outputs and the observations is possible. Finally, with respect to river flows, the MODCOU model grid varies in the range of 8 km to 1 km near the riverbed, and the comparison between the model output and the observed flow is made by considering the flow at the river outlet and the corresponding model grid point in the 1 km hydrological network grid. This way of locally validating models by comparing the observation to the corresponding model grid point is not new and has been used in many studies in France and elsewhere (Habets et al., 2008; Decharme et al., 2013; Lafaysse et al., 2011; Vergnes et al., 2014).

1 - Please make it crystal clear how the 470 discharge stations were selected. If there are more stations, why were they not used?

2 - Only discharge stations with observations for at least half of the days over the total period were kept.

The following sentence was added in section 3.2 (lines 297-298):

3 - Only gauging stations with observations for at least half of the days over the total period were kept.

1 - Line 351: some basins are anthropized. I assume this means "urbanized", because agricultural crops are now in the new vegetation classes. If basins are urbanized or semi-urbanized, and the model does not simulate this, then these basins cannot be used to evaluate the model.

2 - The term anthropized doesn’t seem appropriate. It is partly urbanized as highlighted by the reviewer, which means that the presence of urbanization will affect for example the surface runoff by reinforcing the imperviousness. This effect is accounted for in the model by replacing the urban areas by rocks that will facilitate surface runoff. On the other hand, it’s true that agricultural crops are part of the vegetation classes and that the model does not represent explicitly the agricultural practices such as irrigation. Moreover, the model does not account for dams whereas some rivers are highly affected by their presence. But we do think that these basins have to be part of the evaluation to identify the weaknesses of the model and put the efforts on developing methods to account for irrigation or the presence of dams.

“whereas some basins are anthropized.” was changed in the manuscript section 4.2 as follows (lines 366-371):

3 – while some basins are influenced by human activity. In some basins, the human footprint on the landscape is characterized by an increase in urban and agricultural areas and the presence of dams. In the model, urban areas have been replaced by rocks, a type of natural surface, to represent the presence of urban areas that enhance surface runoff. However, the model does

not explicitly represent irrigation or the impact of the presence of dams on river flow. The basins impacted by human activity are of great interest for the evaluation as they allow quantifying errors in the system and proposing improvements.

1 - Line 417: "can probably be attributed to a deficit in the incoming shortwave ration ... or geothermal heating". This seems speculative, it could simply be that the soil or land cover parameters are wrong.

2 - It is true that there is no evidence that geothermal heating is a source of error in the ground temperature. This was removed from the manuscript and the original sentence in section 4.6:

“As in the previous study of Decharme et al. (2013), a global cold bias (here of about -0.8 K) is observed at each depth, which can probably be attributed to a deficit in the incoming shortwave radiation at the surface and/or to the none representation of the deep earth geothermal heating.“

Was changed into:

3 - “As in the previous study by Decharme et al. (2013), a cold bias (here about -0.8 K) is observed at each depth, which can probably be attributed to a deficit of incoming short-wave radiation at the surface and/or to an incorrect specification of soil physical properties or surface parameters.”

1 - Section 4.7: remote sensing data could have been used to substantiate the results here. At this point, unless I am misunderstanding, no data are used to validate the conclusions.

2 - It is true that remote sensing data could have been used to consolidate the results. But the choice was first to focus on the evaluation against river discharges, snow depth and soil temperatures, and second to propose a climatological comparison of the Bowen ratio and the evaporation to precipitation ratio of the system before and after the changes. These two objectives are very important for the applications that are used downstream of the system, especially in the departments at MF in charge of hydrology and agriculture.

On the other hand, remote sensing data does not cover the entire period and a fair comparison to the model climatology is not possible.

The beginning of section 4.7 was modified as follows:

3 - This section compares the climatology of the SIM system before and after the changes made. The aim is to qualitatively identify the impact of the new model on the distribution of energy flows, which is important for certain hydrological or agriculture-related applications.

Some minor issues:
1 - Add units to the list of numbers at the bottom of page 3.

2 - The list was changed as follows:

3 - 0.01 m, 0.04 m, 0.1 m, 0.2 m, 0.4 m, 0.6 m, 0.8 m, 1 m, 1.5 m, 2 m, 3 m, 5 m, 8 m, 12 m

1 - Top of page 4: be specific, do not state "few tens of centimeters" and "a few meters".

2 - In section 2.2, the sentence “Heat transfer is resolved over the total depth, while moisture transfer is resolved only over the depth of the roots, which depends on the type of vegetation, a few tens of centimetres for crops and a few metres for forests.” was changed into:

3 - Heat transfer is resolved over the total depth, while moisture transfer is resolved only over the depth of the roots, which depends on the type of vegetation and its geographical location: a maximum of 1.5 m for type C3 crops and 2.5 m for forests in France.

1 - Line 215-220: this explanation is very unclear.

2 - In section 2.5, the paragraph “In SAFRAN the analysis is performed on homogeneous areas of several hundreds of square kilometres and an explicit vertical discretization is applied so that the analysis is done every 300 meters. For each grid box *i*, the analysed variables $X^{:}(i)$ are then interpolated on a horizontal 8 km grid, accounting for the averaged elevation of each grid box, and used as input to the ISBA model.” was changed into:

3 - In SAFRAN, the analysis is performed on homogeneous zones of several hundred square kilometres and the vertical component is explicitly considered with to a 300-metre slicing along the vertical. For each grid cell *i*, the analysed variables $X^a(i)$ are then interpolated on an 8 km horizontal grid, considering the average altitude of each grid cell. The analysed variables are then used as input to the ISBA surface model.

1 - Section 4.7: Bowen ration -> Bowen ratio.

2 - corrected

1 - There are 17 figures in the paper. This seems like a bit much to me. Can this perhaps be reduced?

2 - Figure 2 and Figure 7 have been deleted and Figures 16 and 17 have been grouped together.