Response to Referee #2

Our responses to the comments of Referee #2 are organised as follows: comment 2.x from Referee #2; authors' response to 2.x and proposed changes in the manuscript, where added text is formatted italic green. Page and line numbers refer to the first submission. Added references are reported at the end of this document.

2.1 Comment: In the manuscript entitled "Hydrological assessment of atmospheric forcing uncertainty in the Euro-Mediterranean area using a land surface model", the authors evaluate the sensitivity of simulated top layer soil moisture, leaf area index (LAI), and streamflow by the SURFEX-CTRIP model system over Europe given five meteorological forcing data sets. The SURFEX-CTRIP model system is based on the land-surface model ISBA-A-gs, which is based on a biochemical model to simulate the interaction between the soil, biosphere, and atmosphere. The goal of the study is to assess the uncertainty in model simulation that can be attributed to the forcings. The forcing data sets are ERA-Interim reanalysis (ERA-I), ERA-I with precipitation bias-corrected to monthly values by GPCP (P-ERA), WFDEI, PGF, and a reference data set based on several observational data.

2.1 Response: We thank Referee #2 for her/his review of the manuscript and useful comments, which we address in the following.

We would like to clarify that the model was forced using four atmospheric datasets. Furthermore, the observational reference datasets were used to assess the impacts of atmospheric forcing uncertainties through comparisons with model simulations (from page 3, line 30 to page 4, line 15).

2.2 Comment: The authors report, in general, little impact by the different forcing data sets on bias and correlation of simulated values. Exceptions are PGF for simulated soil moisture and P-ERA for streamflow.

2.2 Response: We regard the relatively small sensitivities to forcing uncertainty as indicators of: model robustness, if model performance is satisfactory; or model limitations (in case of unsatisfactory performance), for which we suggest several research directions in Section 6 ("Discussion"). Concerning river discharge, we find that forcing uncertainty has a larger impact on the mean and standard deviation than on the timing, shape and inter-annual variability (page 18, lines 15-25).

2.3 Comment: The motivation of the study is to assess the modelling tool for planning human activities involving freshwater resources (i.e., integrated water resources management). I think this motivation is odd because normally, hydrologic models are used for this purpose instead of land surface models (LSMs). The former are designed for this purpose, they are more conceptual than the latter, and typically, parameters can be calibrated to achieve a satisfying representation of the terrestrial hydrologic cycle. On the contrary, LSMs are more physically-based, incorporate a wider range of processes (e.g., CO_2-cycle), and have been developed to provide the lower boundary condition for coupled atmosphere-land-ocean models over land.

2.3 Response: We agree that the motivations for using a land surface model (LSM) were not properly discussed.

To better frame our work in the climate change context, which requires physically consistent descriptions of water, energy and carbon cycles, we propose replacing the first sentence of the abstract (page 1, lines 10-11):

"The understanding of land surface hydrology is critical for planning human activities involving freshwater resources."

with:

"*Physically consistent descriptions of land surface hydrology are crucial for planning human activities that involve freshwater resources, especially in light of the expected climate change scenarios.*"

In Section 1 ("Introduction"), we overlooked discussing the design differences between global hydrological models (GHMs) and LSMs , where we outline the motivations and the model choice for this study (page 2, lines 4-24). Therefore, we propose replacing the text on page 2, lines 10-17 ("Large-scale hydrology […] (van Beek et al., 2012; van Vliet et al., 2012; Yearsley, 2012).") with the following paragraphs:

"*Large-scale hydrology can be simulated using several approaches, ranging from lumped water balance models to distributed global hydrological models (GHMs) and land surface models (LSMs). LSMs and GHMs are used to study a wide range of water-related problems: hydrological and agricultural droughts (Dirmeyer et al., 2006; Szczypta et al., 2012; Szczypta et al., 2014); floods (Decharme et al., 2012; Pappenberger et al., 2012; Hirpa et al., 2016); agricultural production and irrigation (Rost et al., 2009; Jägermeyr et al., 2016); surface freshwater temperature and its impact on energy production (van Beek et al., 2012; van Vliet et al., 2012; Yearsley, 2012).*

LSMs, which were originally designed to provide lower boundary conditions and vertical fluxes for atmospheric global circulation models (AGCMs), generally simulate the diurnal course of energy exchanges, vegetation and carbon dynamics, and hydrology. GHMs, which were developed to estimate and compare freshwater availability with anthropic requirements, typically simulate processes that are relevant to water resources assessments such as, e.g., hydrodynamic routing, reservoir operation, and water demands (Bierkens, 2015). Recent model intercomparison exercises have evaluated the suitability of several GHMs and LSMs for simulating the continental water cycle globally (Haddeland et al., 2011; Schellekens et al., 2016; Beck et al., 2017a). While parameters of GHMs are generally calibrated using observed river discharge time series, those of LSMs are usually determined a priori from maps of land surface properties or expert judgement. Therefore, calibrated

GHMs tend to reproduce river discharge better than uncalibrated LSMs (Beck et al., 2017a). However, models calibrated on river discharge are not guaranteed to perform better than uncalibrated models when considering other key hydrologic variables such as e.g. evapotranspiration (Schellekens et al., 2016). Since LSMs aim at reproducing several land surface fluxes and states, multi-criteria approaches may provide more robust parameter calibration and sensitivity analysis frameworks compared to traditional single-objective methods (Bastidas, 1998; Gupta et al., 1998; Gupta et al., 1999; Bastidas et al., 2006). Acknowledging that LSMs should be assessed using multiple criteria adds further complexity to the parameter calibration problem.

Parameter calibration may cause model over-fitting, when the parameter set optimised over the calibration period proves to be sub-optimal in other periods (Andréassian et al., 2012). Over-fitting a physically based model may hinder the detection of process misrepresentations and therefore model improvement. Over-fitting is of particular concern for climate change simulations, as models may be applied under climatic conditions that are very different from those of the calibration period (Knutti, 2008). Although model performance under climate change scenarios cannot be validated, uncalibrated models based on physically consistent process descriptions (such as LSMs) might be more robust than models based on pameterisations calibrated under past climate regimes and land uses (such as calibrated GHMs). For example, most GHMs do not simulate explicitly the surface energy fluxes that are crucial for evaporation, while most LSMs solve water and energy balances together (Haddeland et al., 2011; Pokhrel et al., 2012). Physically consistent descriptions of such fluxes are necessary for modelling land surface feedbacks into the atmosphere, which can be represented by coupling LSMs with AGCMs (Koster et al., 2004; Betts, 2007; Campoy et al., 2013).

The outlined calibration trade-off should be carefully considered when choosing the modelling approach. In this study, we chose a LSM for assessing the hydrological impacts of atmospheric forcing uncertainties in the Euro-Mediterranean area. LSMs can be coupled to AGCMs (Widén-Nilsson et al., 2007) or run off-line, forced by gridded atmospheric datasets that are generally obtained from AGCM reanalysis or simulation."

2.4 Comment: Nevertheless, I think that the evaluation of a LSM modelling system by different forcing data sets is a welcome contribution to the field of hydrology, but there are several criteria that have to be met to provide a meaningful analysis. The most important criteria is that the model satisfactorilly reproduces the terrestrial hydrological cycle. Often, streamflow is used for this assessment and the authors also compare simulated streamflow against observations at 35 gauges (locations are shown in Figure 1). The median KGE over all gauges is at most 0.4 with at least 20\% of the gauges having a negative KGE, indicating a poor representation of the surface hydrology. This holds for all forcing data sets, which indicates that the poor performance is independent of these.

2.4 Response: Simulated streamflow is evaluated using 5 scores: the relative errors in the mean (RE_μ) and standard deviation (RE_σ); the monthly (MC) and the anomaly (AC) correlation coefficients; and KGE, which aggregates the information of RE_μ, RE_σ and MC (section 4.3). While an aggregate score may be useful to simplify model calibration (by reducing it to a single-objective problem), it does not provide useful information for understanding the causes of poor model performance, which should be evaluated using multiple criteria (Gupta et al., 2009). KGE is defined as 1 minus the Euclidean distance of the components [RE_μ, RE_σ, 1 - MC] from the ideal point [0, 0, 0] (Equation 2).

Therefore, KGE cannot be larger than the smallest of its components, so the "worst" score is a limiting factor. Therefore, we believe that river discharge simulations should not be evaluated only by means of KGE. The other scores should be considered as they inform about the causes of better or worse model performance.

We observe that the obtained KGE scores are strongly impacted by the relative errors, in particular by RE_σ (page 14, lines 25-32; page 15, lines 20-22; Supplement page 1, lines 33-36; Tables 1 and 2; Figures 7 and S1). The large RE_σ values indicate that the model tends to overestimate the amplitude of the discharge annual cycle (page 16, lines 8-10; page 18, lines 21-25; page 20, lines 7-8).

The correlation scores (MC and AC), which relate to the timing and shape of the streamflow time series, show a relatively good agreement between the simulations. In contrast, relative error scores (RE_μ and RE_σ) show larger inter-forcing spreads (page 14, lines 18-25). Thus, forcing uncertainties have the largest impacts on the river discharge mean and standard deviation (page 15, lines 22-24).

To better account for these observations, we propose:

- Adding the following paragraph after the definition of KGE (page 11, line 18): "*An aggregate score is often used as unique criterion in model calibration, because it allows applying efficient single-objective optimisation algorithms. However, model calibration is an inherently multi-objective problem (Gupta et al., 1998). Thus, aggregate scores are not as informative as their individual components for evaluating model performance (Gupta et al., 2009).*"

- Modifying the text on page 14, lines 25-27: "The relative errors *(in particular RE_σ)* have a strong impact on the summary curves of the aggregate score KGE, according to which the P-ERA simulation is dominated at almost all frequencies. Instead, a clear KGE ranking cannot be established among the other three simulations. *While KGE does not provide diagnostic information on the causes of better or worse model performance, the individual scores show that forcing uncertainties have larger impacts on the mean and standard deviation of the simulated discharge than on shape, timing and inter-annual variability.*"

2.5 Comment: As can be seen in the bottom row of Figure S1 in the supplements, most of the better performing gauges are nested sub-catchments of the Danube, Rhine, and Elbe river. These represent the same humid conditions and the good performance over the same area is "double counted". Other areas such as the cold region in North-Eastern Europe and catchments in the Mediterranean show significantly poorer performance. These also represent a large fraction of the European area. The North-Eastern part of Europe also experiences very poor agreement with soil moisture observations (Figure 3) and LAI (Figure 5). This poor agreement in this region does not allow any assessment of the forcing uncertainty because the model might lack important processes to reproduce the terrestrial hyrdological cycle there. This might also be due to the fact that the authors use the parameters of previous work that has been only validated in the Rhone catchment (Decharme and Douville, 2006).

2.5 Response: The redundant accounting of scores computed for river basins with multiple gauges is acknowledged on page 15, lines 1-5. For this reason, we also show and discuss the river discharge

scores and annual cycles at the most downstream gauges of each catchment (page 15, lines 8-24; Tables 1 and 2).

The KGE ranges of the downstream gauges of the Danube, Rhine and Elbe basins are: [0.06, 0.70], [0.13, 0.55] and [-0.62, 0.65], respectively (Tables 1 and 2). Similar ranges are spanned by the downstream gauges of other basins: Vistula [-0.07, 0.51], Rhone [0.32, 0.65], Ebro [0.34, 0.69], Po [0.52, 0.72], Tejo [0.40, 0.83], Daugava [0.29, 0.62], Duero [-0.03, 0.81], Kemijoki [0.24, 0.57], Glama [0.41, 0.58], Weser [0.16, 0.71], Tornealven [0.52, 0.77], and Meuse [0.67, 0.83]. Some of the latter are located in North-Eastern Europe (Daugava, Kemijoki, Tornealven), in the Mediterranean area (Rhone, Ebro, Po), or in the Iberian Peninsula (Ebro, Tejo, Duero). Catchments and regions characterised by poorer model performance are discussed in Section 5.3 (page 15, lines 8-19) and in the Supplement (page 1, lines 22-36).

Small or negative temporal correlations between simulated surface soil moisture (SSM) and satellite retrievals in North-Eastern Europe are discussed on page 12, lines 11-19. We agree that the mismatches may be due to model misrepresentations of land surface processes that are particularly relevant in cold and mountainous areas. Therefore, we propose replacing the following text (page 12, lines 17-18):

"Possible causes for these low values may be vegetation density, topography and soil frost occurrences, all of which are known to negatively affect SSM retrievals."

with:

"*Possible causes for these low values may be: vegetation density, topography and soil frost occurrences, all of which are known to negatively affect SSM retrievals; and model misrepresentation of surface hydrological processes that are particularly relevant in cold and mountainous regions.*"

Several hypotheses for the satellite/model SSM mismatches and possible research directions are discussed in greater detail in Section 6.2. Concerning possible model misrepresentations of hydrological processes at high latitudes, we propose adding the following suggestion for future work (page 17, line 2):

"In light of the presented results, we suggest the following research directions: […] (Stoffelen et al., 2017)*; investigating the improvement or inclusion in LSMs of physical processes that are relevant to topsoil hydrology at high latitudes, e.g. snowmelt, flooding and ponding (Gouttevin et al., 2013).*"

LAI monthly temporal correlation coefficients in North-Eastern Europe are between 0.4 and 0.9, with most values above 0.7 (top-left map in Figure 5). In contrast, anomaly correlations are significantly lower: all values are below 0.5 and the majority is below 0.3 (top-right map in Figure 5). Based on these values, the agreement between simulated and remotely sensed LAI in North-Eastern Europe appears consistent with the rest of the domain.

The model parameters used for this study are not those of Decharme and Douville (2006a), whom Referee #2 cites. For example, Decharme and Douville (2006a) used a 3-layer soil description, while we simulate 14 soil layers; moreover, the equations describing the unsaturated soil water and energy fluxes are different. Our model configuration and parameters are described in Section 2 (from page 5, line 13 to page 6, line 8). The CTRIP parameters are derived from Decharme et al.

(2010), who performed a global evaluation of simulated river discharge. Land surface and vegetation parameters are provided by the ECOCLIMAP-II database (Faroux et al., 2013) and were used by Szczypta et al. (2014) and Albergel et al. (2017), who compared simulated SSM and LAI with satellite retrievals. Furthermore SURFEX-CTRIP was included in recent model intercomparison exercises (Schellekens et al., 2016; Beck et al., 2017a). Thus, we disagree with the statement "[…] the authors use the parameters of previous work that has been only validated in the Rhone catchment (Decharme and Douville, 2006)".

2.6 Comment: In conclusion, it is not demonstrated that the little differences seen among the forcing data sets are not due to the chosen model parameters (e.g., soil water parameters like porosity, saturated hydraulic conductivity and d_ice etc.), or lacking hydrologic processes in the model. In other words, it is very likely that the results are an artefact of poor model performance over large parts of Europe. The authors have to investigate the sensitivity of model parameters at the different hydro-climatic regimes in Europe and choose parameters that lead to a better representation of the terrestrial hydrologic cycle before they can assess the influence of different forcing data sets. This calibration exercise should be conducted using the observation-based reference data set.

2.6 Response: Unsatisfactory model performances that appear to be forcing-independent (thus systematic) may be due to model limitations, in addition to other causes such as forcing inter-dependence or satellite retrieval uncertainties. These are acknowledged in several parts of the manuscript, where generally we also identify possible research directions to tackle systematic model errors (page 13, lines 1-6; from page 13, line 27 to page 14, line 2; from page 16, line 22 to page 17, line 2; page 17, lines 12-19; from page 17, line 31 to page 18, line 6; from page 18, line 21 to page 19, line 16). Although this critical model evaluation aims at contributing to model improvement, we believe that the acknowledged specific limitations should not be generalised to the whole model. Moreover, SURFEX-CTRIP outputs were successfully compared to several observation datasets in previous studies (Decharme et al., 2010; Szczypta et al., 2012; Szczypta et al., 2014; Albergel et al., 2017) and included in recent model intercomparison exercises (Schellekens et al., 2016; Beck et al., 2017a), as stated in response 2.5.

Sensitivity analyses can be applied to LSMs for assessing parameter uncertainties and addressing the dimensionality of the parameter estimation problem (Bastidas et al., 2006). Since LSMs are designed to represent the complex and interconnected processes underlying the energy, water and carbon cycles, they should be evaluated by comparing a relatively large number of simulated fluxes and state variables with observations. Thus, multi-criteria (multi-objective) approaches are necessary to carry out meaningful parameter sensitivity analysis and calibration exercises (Bastidas, 1998; Gupta et al., 1998; Gupta et al., 1999; Bastidas et al., 2006). Considering the implied complexity, we believe that parameter sensitivity analysis and calibration are beyond the scope of this study. Decharme et al. (2010) performed sensitivity analyses on the groundwater delay and river flow velocity CTRIP routing parameters. Moreover, we maintain that parameter calibration is beyond the purposes of our study according the following considerations:

1. To model the energy, water and carbon cycles consistently, LSMs generally use parameters obtained from observations and measurements of land surface properties. The underlying

assumption is that uncalibrated LSMs may behave more robustly under climate and land use change scenarios than models based on pameterisations calibrated under past climate regimes and land uses. Furthermore, over-fitting a physically based model may hinder the detection of process misrepresentations and therefore model improvement (see response 2.3). Calibrated models can be good for wrong reasons (e.g. error compensation).

2. Calibrating model parameters based on several (four in our case) forcing datasets would entail: either obtaining an optimal parameter set for each forcing dataset; or defining a unique parameter set by weighting the objective function values computed for each forcing dataset. In both cases, forcing uncertainty would be translated into parameter uncertainty, thus deviating from the objectives of our study.

3. As LSM parameter calibration is an inherently multi-objective problem (Gupta et al., 1998), it would yield multiple optimal parameter sets. While this would inform about parameter sensitivity, it would also introduce parameter uncertainty in the analysis, which is out of the scope of our study.

These considerations motivate the choice of our approach. However, we are aware of the trade-off between calibrated and uncalibrated models, which we discuss in response 2.3.

2.7 Comment: Furthermore, main conclusions are not supported by the results. I would like to give two examples for these.

2.7 Response: We understand Referee #2 refers to the suggestions for model improvement that are summarised in Section 7 ("Conclusions"), on page 20, lines 17-28. These stem from the discussions of model limitations carried out in several parts of Section 6 ("Discussion"). Although our suggestions may be questionable, we do not agree with the statement "[…] main conclusions are not supported by the results", because most conclusions (from page 19, line 18 to page 20, line16) summarise the results presented in Sections 5 ("Results").

We address the specific examples provided by Referee #2 in the following responses (2.7.1 and 2.7.2).

2.7.1 Comment: 1.) p. 20, l. 21 ff: The authors conclude that LSMs may progressively integrate groundwater modelling to improve the simulation of river discharge. This does not relate to the findings of this study. All the runs presented in this study include a linear reservoir to represent groundwater storage (see p. 5, l. 32f). The authors would have to present results that do not use this groundwater component to make this conclusion.

2.7.1 Response: CTRIP uses a simplified linear reservoir to mimic the delay of soil drainage contributions to river discharge. This model component is not meant to represent actual groundwater processes (page 6, lines 1-2). We argue that including more physically-based groundwater schemes in LSMs may be beneficial for the simulation of freshwater availability.

Vergnes and Decharme (2012) showed that coupling a distributed groundwater scheme with the CTRIP river channels improved the simulations of discharge (assessed with GRDC data) and total water storage variations (assessed with GRACE satellite retrievals). To better account for these findings, we propose adding the following text on page 18, lines 32-33:

"Vergnes and Decharme (2012) implemented a simple groundwater scheme coupled to the river channel in each model grid cell allowing bidirectional water exchanges through the riverbed*, showing improvements in the simulations of total water storage variations and river discharge*.".

To clarify our suggestion in Section 7 ("Conclusions"), we propose replacing the following sentence (page 20, line 21-24):

"To improve the simulation of river discharge and to assess the impacts of water abstractions, LSMs may progressively integrate groundwater modelling (see e.g. Vergnes et al., 2014) and lake and reservoir regulation (see e.g. Hanasaki et al., 2006; Pokhrel et al., 2012)."

with:

"*To improve the simulation of freshwater availability, LSMs may progressively integrate more physically based groundwater schemes (see e.g. Vergnes and Decharme, 2012; Vergnes et al., 2014) as well as lake and reservoir regulation (see e.g. Hanasaki et al., 2006; Pokhrel et al., 2012).*"

2.7.2 Comment: 2.) p. 20, l. 24f: The authors conclude that: "Due its relevance for crop yield and water demand prediction, large-scale irrigation schemes should be embedded into LSMs". This is contradicting the findings of this study. Section 6.3 shows that the model best reproduces correlations to LAI during the maximum phenological development phase (p. 17, l. 29f). The authors further state in this section that: "This is an encouraging result for the simulation of crop yield and, in general, of the primary production of land surface ecosystems." (p. 17, l. 30f). The authors also demonstrate that LAI is systematically overestimated during the maximum phenological development. While I do agree that it is important to include irrigation in LSMs, it is not supported by these findings. In the case of the ISBA-A-gs model, including irrigation would lower the water stress of plants during summer (i.e., the maximum phenological development phase) and lead to even more increased GPP and LAI. The authors have to conduct additional analysis and perform runs including an irrigation scheme to make this conclusion.

2.7.2 Response: We agree that the inclusion of irrigation schemes in LSMs deserves a complex discussion that is beyond the scope of this study. Therefore we propose deleting the following sentence from Section 7 ("Conclusions"), on page 20, lines 24-25: "Due its relevance for crop yield and water demand prediction, large-scale irrigation schemes should be embedded into LSMs (Jägermeyr et al., 2016).".

2.8 Comment: Overall, the authors have to either investigate which processes are lacking in ISBA-A-gs to better represent the hydrologic cycle in cold regions and semi-arid ones. Alternatively, they

have to conduct a comprehensive parameter calibration study. Both of these avenues are beyond the scope of this study. Therefore, unfortunately, I have to recommend to reject this manuscript.

2.8 Response: We understand this last comment of Referee #2 summarises previous ones, which we address in detail in the corresponding responses. Here we report the main points of the latter.

Regarding the misrepresented or lacking physical processes in the model, our work contributes to their identification with constructive suggestions (see responses 2.2, 2.5, 2.6 and 2.7). However, we believe that further work aimed at rigorously testing the inclusion of new process representations is beyond the scope of this study, which focuses on the assessment of the hydrological impacts of atmospheric forcing uncertainties.

SURFEX-CTRIP simulations have already been successfully compared to observations/retrievals of river discharge, surface soil moisture and leaf area index either globally or in the Euro-Mediterranean area by previous studies (Decharme et al., 2010; Szczypta et al., 2012; Szczypta et al., 2014; Albergel et al., 2017). Moreover, SURFEX-CTRIP has been included in recent global model intercomparison exercises (Schellekens et al., 2016; Beck et al., 2017a), together with other state-of-the-art land surface and global hydrological models (see responses 2.5 and 2.6).

The reasons for choosing an uncalibrated physically-based modelling approach, as well as those for excluding parameter calibration from the scope of this study, are thoroughly discussed in responses 2.3 and 2.6.

References

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., 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, Geoscientific Model Development 10, 3889–3912, doi:10.5194/gmd-10-3889-2017, 2017.

Andréassian, V., Le Moine, N., Perrin, C., Ramos, M. H., Oudin, L., Mathevet, T., Lerat, J., and Berthet, L.: All that glitters is not gold: the case of calibrating hydrological models, Hydrological Processes, 26, 2206–2210, doi: 10.1002/hyp.9264, 2012.

Bastidas, L. A.: Parameter estimation for hydrometeorological models using multi-criteria methods, PhD dissertation, Department of Hydrology and Water Resources, University of Arizona, Tucson, 1998.

Bastidas, L. A., Hogue, T. S., Sorooshian, S., Gupta, H. V., and Shuttleworth, W. J.: Parameter sensitivity analysis for different complexity land surface models using multicriteria methods, Journal of Geophysical Research, 111, D20101, doi:10.1029/2005JD006377, 2006.

Beck, H. E., van Dijk, A. I. J. M., de Roo, A., Dutra, E., Fink, G., Orth, R., and Schellekens, J.: Global evaluation of runoff from 10 state-of-the-art hydrological models, Hydrol. Earth Syst. Sci., 21, 2881-2903, https://doi.org/10.5194/hess-21-2881-2017, 2017a.

Betts, A. K.: Coupling of water vapor convergence, clouds, precipitation, and land-surface processes, Journal of Geophysical Research, 112, D10108, doi:10.1029/2006JD008191, 2007.

Bierkens, M. F. P., Global hydrology 2015: State, trends, and directions, Water Resources Research, 51, 4923–4947, doi:10.1002/2015WR017173, 2015.

Campoy, A., Ducharne, A., Cheruy, F., Hourdin, F., Polcher, J., and Dupont, J. C.: Response of land surface fluxes and precipitation to different soil bottom hydrological conditions in a general circulation model, Journal of Geophysical Research: Atmospheres, 118, 10, 725–10,739, doi:10.1002/jgrd.50627, 2013.

Decharme, B., and Douville, H.: Introduction of a sub-grid hydrology in the ISBA land surface model, Climate Dynamics, 26, 1, 65–78, doi:10.1007/s00382-005-0059-7, 2006a.

Decharme, B., Alkama, R., Douville, H., Becker, M., and Cazenave, A.: Global evaluation of the ISBA-TRIP continental hydrologic system, Part II: Uncertainties in river routing simulation related to flow velocity and groundwater storage, Journal of Hydrometeorology, 11, 601–617, doi:10.1175/2010JHM1212.1, 2010.

Faroux, S., Kaptué Tchuenté, A. T., Roujean, J.-L., Masson, V., Martin, E., and Le Moigne, P.: ECOCLIMAP-II/Europe: a twofold database of ecosystems and surface parameters at 1 km resolution based on satellite information for use in land surface, meteorological and climate models, Geoscientific Model Development, 6, 563–582, doi:10.5194/gmd-6-563-2013, 2013.

Gouttevin, I., Bartsch, A., Krinner, G., and Naeimi, V.: A comparison between remotely-sensed and modelled surface soil moisture (and frozen status) at high latitudes, Hydrology and Earth System Sciences Discussions, doi:10.5194/hessd-10-11241-2013, 2013.

Gupta, H. V., Bastidas, L. A., Sorooshian, S., Shuttleworth, W. J., and Yang, Z. L.: Parameter estimation of a land surface scheme using multi-criteria methods, Journal of Geophysical Research, 104, 19,491–19,504, 1999.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of Hydrology, 377, 1–2, 80–91, doi:10.1016/j.jhydrol.2009.08.003, 2009.

Gupta, H. V., Sorooshian, S., Yapo, P. O.: Toward improved calibration of hydrologic models: multiple and noncommensurable measures of information, Water Resources Research 34, 4, 751–763, 1998.

Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N.W., Bertrand, N., Best, M., Folwell, S., Gerten, D., Gomes, S., Gosling, S. N., Hagemann, S., Hanasaki, N., Harding, R., Heinke, J., Kabat, P., Koirala, S., Oki, T., Polcher, J., Stacke, T., Viterbo, P., Weedon, G. P., and Yeh, P.: Multimodel Estimate of the Global Terrestrial Water Balance: Setup and First Results, Journal of Hydrometeorology, 12, 5, 869–884, 2011.

Hanasaki, N., Kanae, S., and Oki, T.: A reservoir operation scheme for global river routing models, Journal of Hydrology, 327, 1–2, 22–41, doi:10.1016/j.jhydrol.2005.11.011, 2006.

Knutti, R.: Should we believe model predictions of future climate change?, Philosophical Transactions of the Royal Society A, 366, 4647–4664, doi:10.1098/rsta.2008.0169, 2008.

Koster, R., Dirmeyer, P., Guo, Z., Bonan, G., Chan, E., Cox, P., Gordon, C. T., Kanae, S., Kowalczyk, E., Lawrence, D., Liu, P., Lu, C.-H., Malyshev, S., McAvaney, B., Mitchell, K., Mocko, D., Oki, T., Oleson, K., Pitman, A., Sud, Y. C., Taylor, C. M., Verseghy, D., Vasic, R., Xue, Y., and Yamada, T.: Regions of strong coupling between soil moisture and precipitation, Science, 305, 5687, 1138–1140, 2004.

Pokhrel, Y., Hanasaki, N., Koirala, S., Cho, J., Yeh, P. J.-F., Kim, H. Kanae, S., Oki, T., Incorporating anthropogenic water regulation modules into a land surface model, Journal of Hydrometeorology, 13, 1, 255–269, doi:10.1175/JHM-D-11-013.1, 2012.

Schellekens, J., Dutra, E., Martínez-de la Torre, A., Balsamo, G., van Dijk, A., Sperna Weiland, F., Minvielle, M., Calvet, J.-C, Decharme, B., Eisner, S., Fink, G., Flörke, M., Peßenteiner, S., van Beek, R., Polcher, J., Beck, H., Orth, R., Calton, B., Burke, S., Dorigo, W., and Weedon, G. P.: A global water resources ensemble of hydrological models: the eartH2Observe Tier-1 dataset, Earth System Science Data, 9, 389–413, doi:10.5194/essd-9-389-2017, 2017.

Szczypta, C., Decharme, B., Carrer, D., Calvet, J.-C., Lafont, S., Somot, S., Faroux, S., and Martin, E.: Impact of precipitation and land biophysical variables on the simulated discharge of European and Mediterranean rivers, Hydrology and Earth System Sciences, 16, 3351–3370, doi:10.5194/hess-16-3351-2012, 2012.

Szczypta, C., Calvet, J.-C., Maignan, F., Dorigo, W., Baret, F., and Ciais, P.: Suitability of modelled and remotely sensed essential climate variables for monitoring Euro-Mediterranean droughts, Geoscientific Model Development, 7, 931–946, doi:10.5194/gmd-7-931-2014, 2014.

Vergnes, J.-P. and Decharme, B.: A simple groundwater scheme in the TRIP river routing model: global off-line evaluation against GRACE terrestrial water storage estimates and observed river discharges, Hydrology and Earth System Sciences, 16, 3889–3908, doi:10.5194/hess-16-3889-2012, 2012.

Vergnes, J.-P., Decharme, B. and Habets, F., Introduction of groundwater capillary rises using subgrid spatial variability of topography into the ISBA land surface model, Journal of Geophysical Research: Atmospheres, 119, 19, 11065–11086, doi:10.1002/2014JD021573, 2014.