Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2016-508-SC1, 2016

Hydrology and Earth System Sciences Discussions

Open Access

HESSD

Interactive comment

Interactive comment on "A Hydrological Prediction System Based on the SVS Land–Surface Scheme: Implementation and Evaluation of the GEM-Hydro platform on the watershed of Lake Ontario" *by* Étienne Gaborit et al.

E. Gaborit

etienne.gaborit.5@gmail.com

Received and published: 21 December 2016

Please see the attached file (supplement file) for a version of the manuscript tracking the changes listed in the answers to reviewer (AR):

General comments

The manuscript presents interesting topic, but in its current form it is very difficult to read and reads more like a technical report rather than a scientific paper. I have several

Printer-friendly version

Discussion paper

comments, which might be considered for a revision:

1) The formulation of research hypothesis and novel scientific contribution is not clear. The first objective (as formulated in the manuscript) is to propose a methodology for calibrating distributed hydrologic model, but the results refer to an a priori selected number of model runs (in model calibration), without testing whether this approach is better (and in which aspects) than some methodologies used for model calibration.The introduction refers to numerous previous studies but does not clearly indicate what unanswered question is investigated here. Just implementation reads more as a technical than a scientific question. In its current form it is formulated as a case study. Why it should be interesting for international audience of the journal? What can be learned/generalised from the results which will be interesting/relevant also for other regions in the world?

AR: the introduction (and abstract) was modified to make the novelty of this contribution clearer. Therefore, the reader should now better understand that the aim of the study is not to find the best way to calibrate distributed models (which can always be improved and has almost an infinity of possibilities, making it quite impossible to find anyway), but the proposition of an efficient and reliable way to implement a certain type of distributed models over a large area with ungauged parts. This could be applied anywhere on Earth. This is a novel approach because global calibration is rarely employed in the Litterature, and because we propose a new way of calibrating Land Surface Schemes, i.e., by getting rid of the time-consuming routing part of distributed hydrologic models. Another research objective is to contribute to the GRIP framework, by comparing different models with the same forcings and implementation methodology in the Great Lakes region. In regard of the long introduction, we actually believe that it gives a good overview of hydrological research in the Great Lakes region, which may be of interest to a relatively large audience. However, an effort of synthesis was still put into the introduction and it was slightly shortened. The new introduction is pasted below:

" Given the continuous increase in precipitation forecast skill of Numerical Weather Prediction (NWP) systems, as documented for example over the United States (US) by Sukovich et al. (2014), it is becoming possible to obtain skillful runoff forecasts directly from NWP model outputs, and streamflow forecasts by routing these gridded runoff fields. Indeed, modern NWP models all simulate to some extent the snow, vegetation, and soil processes that contribute to the generation of runoff and streamflow. However, many limitations are still associated with the representation of such processes in NWP systems, and are documented in Clark et al. (2015) and Davison et al. (2016). Hydrological processes in land-surface models used for NWP are improving quickly (Balsamo et al., 2009; Masson et al., 2013; Alavi et al., 2016; Wagner et al., 2016), as soil water content and snow are recognized as important sources of their predictability that remain to be fully tapped into (Koster et al., 2004; Entekhabi et al., 2010). Environment and Climate Change Canada (ECCC), the Canadian department that provides operational weather and environmental forecasts, is in the process of implementing a major upgrade to the LSS used by its NWP model, the Global Environmental Multi-scale model (GEM). This new scheme, named SVS for Soil, Vegetation and Snow, has been devised in order to assimilate space-based soil moisture retrievals as well as surface data, and has proven efficient at simulating soil moisture and brightness temperature (Alavi et al., 2016; Husain et al., 2016). SVS will replace the Canadian version of the ISBA scheme (Interaction Sol-Biosphère-Atmosphère) that has been used operationally since 2001 (Bélair et al., 2003). One of this paper's objectives is to present the first evaluation of the capabilities of the new SVS scheme for hydrological prediction in Canada. GEM's LSSs can be run either two-way coupled to the atmospheric model or offline, using GEM or other observed atmospheric forcing. The platform for running GEM offline is known as GEM-Surf (Bernier et al., 2011). Runoff obtained from the LSS can then be routed to the outlet of the watershed using the WATROUTE routing scheme (Kouwen, 2010). This configuration is known as GEM-Hydro. Our current evaluation of GEM-Hydro focuses on the Lake Ontario watershed for many reasons including (1) the socio-economic impacts that improvements to streamflow and lake level prediction skill can have on a region of Canada that is quite populated and industrialized; (2) the large amount of data available for model set up, calibration, and validation, compared to other regions of Canada; and (3) the fact that this is a Canada-USA transboundary watershed which is co-managed by ECCC and US Army Corps of Engineers (USACE) staff, in accordance with water level management rules set by the International Joint Commission (IJC) for each control structure, including the Moses-Saunders power dam at Cornwall, the outlet of Lake Ontario (Fig. 1). Different cascades of interconnected models have been developed over the years to simulate the Great Lakes water levels and thermodynamics, such as Wiley et al. (2010), Deacu et al. (2012), and Gronewold et al. (2011), the latter describing the Advanced Hydrologic Prediction System (AHPS), a seasonal water supply and water level forecasting system developed by the National Oceanic and Atmospheric Administration (NOAA) Great Lakes Environmental Research Laboratory (GLERL) in the mid-1990s that has since been employed operationally (with few changes in methodology) by the USACE and regional hydropower authorities. Recently, ECCC has implemented a short-term (84-h) operational water cycle prediction system for the Great Lakes and St. Lawrence River (WCPS-GLS) that uses coupled atmospheric, hydrologic, and hydrodynamic models (Durnford et al., in preparation). This system makes use of the same platform used in this study, GEM-Hydro, but relies on the simpler ISBA LSS. To our knowledge, the AHPS and WCPS systems are the only two systems that provide inflow forecasts for each of the Great Lakes on both sides of the Canada-US border, and neither relies on very sophisticated hydrological models. The need for improving simulations and forecasts of runoff to the Great Lakes is recognized by both agencies (Gronewold and Fortin, 2012). Multiple additional hydrologic models are indeed available (Coon et al., 2011), however their spatial domains are typically constrained to either the US or Canada. Before embarking on an upgrade of operational systems, GLERL and ECCC agreed to perform a number of intercomparison studies under the umbrella of the Great Lakes Runoff Intercomparison Project (GRIP), in order to better understand

the status of existing systems, and to set a benchmark for model performance against which future models could be compared. The first study was conducted on the Lake Michigan (GRIP-M) watershed by Fry et al. (2014) who compared historical runoff simulations from dissimilar hydrologic models using different calibration frameworks and input data. Amongst the models compared were GLERL's Large Basin Runoff Model (LBRM; Croley and He, 2002) that is part of the AHPS, the NOAA National Weather Service model (NWS; Burnash, 1995), and ECCC's MESH distributed model (Modélisation Environnementale – Surface and Hydrology; Pietroniro et al., 2007; Haghnegahdar et al., 2014). A second configuration of MESH was also included, based on Deacu et al. (2012), from which evolved the configuration of GEM-Hydro used by Durnford et al. (in preparation) for the operational WCPS-GLS system. The NWS model performed best in terms of Nash-Sutcliffe skill, but was positively biased, perhaps because of its typical use as a flood forecasting tool. Overall, it was difficult to attribute any difference in model results to the model structure, given that different forcing data and calibration procedures had been used by each contributor to the project. The GRIP project was extended next to Lake Ontario (GRIP-O) by Gaborit et al. (in Press), who compared two lumped models, namely LBRM and GR4J (modèle du Génie Rural à 4 paramètres Journalier; Perrin et al., 2003), with the exact same forcing data and calibration framework. Two precipitation datasets were used as input: the Canadian Precipitation Analysis (CaPA; Lespinas et al., 2015), and a Thiessen polygon interpolation of the Global Historical Climatology Network - Daily (GHCND; Menne et al., 2012). CaPA is a near real-time quantitative precipitation estimate product from ECCC that is available on a 10-km grid for all of North America: (http://collaboration.cmc.ec.gc.ca/cmc/cmoi/product_guide/submenus/capa_e.html).
The main findings of the first GRIP-O study are that the performance of the models was very satisfactory (average Nash $\sqrt{}$ in validation of 0.86 (over all subbasins and configurations), whatever the precipitation database used, for all tributaries of Lake Ontario, despite the fact that most tributaries have a regulated flow regime. This satisfactory performance justifies the use of CaPA as a precipitation forcing dataset

in later studies, especially for distributed models which require gridded precipitation as input. The performance of lumped models also provides a reference level of performance when evaluating distributed hydrological models (see for example Figure 3, where we can see that GEM-Hydro and GR4J performances are very similar). The present work is an extension of the first GRIP-O study but focused on distributed hydrological models. Compared to the first GRIP-O study which mainly aimed at identifying the performances that it is possible to reach for the area and the local relevance of the CaPA analysis for hydrological modeling, this study mainly aims at finding a methodology to implement the distributed GEM-Hydro model over the whole Lake Ontario watershed, including its ungauged parts, in an efficient manner. Distributed models are more complicated to implement and more computationally-intensive than lumped ones, but have a broader range of applications. Moreover, GEM-Hydro can estimate the Lake Ontario Net Basin Supplies (or NBS, the sum of lake tributary runoff, overlake precipitation, and overlake evaporation: Brinkmann 1983). A second objective is to compare GEM-Hydro with two other distributed models (which is this study's contribution to GRIP-O research purpose) in order to identify avenues to further improve GEM-Hydro. In our view, this study's main contribution to hydrological research is the efficient and reliable methodology proposed to implement a sophisticated distributed model made of a LSS and a routing scheme over a large area, including ungauged parts of the domain. "

2) The number of model runs does not seem to be very large (i.e. adequate), so some more deep analysis and justification of such setup is needed. E.g. I wonder whether 400 runs/combinations for 16 model parameters are enough representative.

AR: this sentence was added close to page 8, line 20, to further justify the methodology used here: " The similarity of the performances obtained with GR4J and GEM-Hydro-UH (Fig. 3) supports the choice of the methodology used here, as GR4J was implemented with a maximum of 2000 model runs, three distinct calibration trials, and had an even lower number of free parameters (6, see Gaborit et al., in Press). "

3) The 2. objective of the manuscript is to compare different models, but it is not clear why not ISBA and SVS are compared, as in the Introduction it is referred to the replacement of ISBA by SVS. This would allow to better demonstrate the potential of the new model/implementation.

AR: As stated in the manuscript close to page 5, line 4, the comparison between ISBA and SVS has been performed and results are described in the technical note mentioned in the manuscript. Results are not shown here for the sake of brevity and because the comparison of ISBA and SVS is, to us, of a lesser importance to the community than the comparison between different models (MESH, WATFLOOD and GEM-Hydro-UH).

4) The manuscript refers to many different other studies within the study region, but not all are relevant to the main objectives, so it distracts the reader from the main story line. Moreover, it is than not quite clear, in which respects is this study novel, so a more clear discussion of the novelty would be helpful. There are numerous references to studies in press or preparation, which does not allow to justify to what extent the study overlaps with previous/recent studies. I would suggest to consider streamlining the text flow, and do not refer much to studies which are not directly linked/relevant with the research questions studied here. For example references to lumped modelling results in the Study area section are not placed/relevant (well) here. Please consider also not to use so many abbreviations, because some parts are then very difficult to understand (e.g. p3., l14).

AR: most of this comment has been treated in the answer to the first reviewer's comment. The reference to the first GRIP-O study will be accessible as the corresponding paper will be published very soon. References to lumped modeling are important to support the methodology used here to calibrate GEM-Hydro-UH with a small number of model runs, given that it leads to similar performances as with the lumped GR4J model which was implemented with 2000 model runs and even fewer free parameters. Moreover, it has to be emphasized that this study is also the second part of the GRIP-O study. GR4J is also used in the objective function used for global calibration, as missing GEM-Hydro performances obtained with local calibration were replaced with GR4J ones (see page 10, line 16). Some abbreviations were removed from the manuscript.

Specific comments

1) Abstract: Please consider to be more specific (i.e. provide numbers, efficiencies, etc), particularly when referring to results found. The context part for the research does not have to be such long. AR: the abstract was changed into the following (see word document with track changes): This work describes the implementation of the distributed GEM-Hydro runoff modeling platform, developed at Environment and Climate Change Canada (ECCC) over the last decade. The latest version of GEM-Hydro combines the SVS (Soil, Vegetation and Snow) land-surface scheme and the WATROUTE routing scheme. in order to provide streamflow predictions on a gridded river network. SVS is designed to be two-way coupled to the GEM (Global Environmental Multi-scale) atmospheric model exploited by ECCC for operational weather and environmental forecasting. Although SVS has been shown to accurately track soil moisture during the warm season, it SVS has never been evaluated from an hydrological point of view, which is done here before for hydrological prediction. This paper presents a first evaluation of its ability to simulate streamflow for all major rivers flowing into Lake Ontario. The skill level of GEM-Hydro is assessed by comparing the quality of simulated flows to that of tTwo established hydrological models are confronted to GEM-Hydro, namely, MESH and WATFLOOD, which share the same routing scheme (WATROUTE) but rely on different land-surface schemes. All models are calibrated using the same meteorological forcings, objective function, calibration algorithm, and watershed delineation. Results show that GEM-Hydro performs well and is reveals competitive with MESH and WATFLOOD: Nash $\surd$ of 0.83 for MESH and GEM-Hydro in validation on the Moira river basin, and of 0.68 for WATFLOOD). A computationally efficient strategy is proposed to calibrate the land-surface model of GEM-HydroSVS: a simple unit hydrograph is used for routing instead of its standard distributed routing component-

WATROUTE. The distributed routing part of the model can then be run in a second step to estimate streamflow everywhere inside the domain. Global and local calibration strategies are compared in order to estimate runoff for ungauged portions of the Lake Ontario watershed. Overall, streamflow predictions obtained using a global calibration strategy, in which a single parameter set is identified for the whole watershed of Lake Ontario, show skills comparable to the predictions based on local calibration: the average Nash $\sqrt{}$ in validation and over 7 subbasins is of 0.73 and 0.61, respectively for local and global calibrations. Hence, global calibration provides spatially consistent parameter values, robust performance at gauged locations, and reduces the complexity and computational burden of the calibration procedure. This work contributes to the Great Lakes Runoff Inter-comparison Project for Lake Ontario (GRIP-O) which aims at improving Lake Ontario basin runoff simulations by comparing different models using the same input forcings.

2) Introduction (p.4., l.5-10): It will be important to clearly formulate in which respect is this study new in comparison to the first GRIP-O study. Please consider also discuss/ show how specific was the model performance and how similar/different it is with respect to this study. AR: this part of the introduction was changed into the following (see file with track changes): The main findings of the first GRIP-O study are that the performance of the models was very satisfactory (average Nash $\sqrt{}$ in validation of 0.86 (over all subbasins and configurations), whatever the precipitation database used, for all tributaries of Lake Ontario, despite the fact that most tributaries have a regulated flow regime. This satisfactory performance justifies the use of CaPA as a precipitation forcing dataset in later studies, especially for distributed models which require gridded precipitation as input. The performance of lumped models also provides a reference level of performance when evaluating distributed hydrological models (see for example Figure 3, where we can see that GEM-Hydro and GR4J performances are very similar). The present work is an extension of the first GRIP-O study but focused on distributed hydrological models. Compared to the first GRIP-O study which mainly aimed at identifying the performances that it is possible to reach for the area and the local rele-

vance of the CaPA analysis for hydrological modeling, this study mainly aims at finding a methodology to implement the distributed GEM-Hydro model over the whole Lake Ontario watershed, including its ungauged parts, in an efficient manner. Distributed models are more complicated to implement and more computationally-intensive than lumped ones, but have a broader range of applications. Moreover, GEM-Hydro can estimate the Lake Ontario Net Basin Supplies (or NBS, the sum of lake tributary runoff, overlake precipitation, and overlake evaporation: Brinkmann 1983). A second objective is to compare GEM-Hydro with two other distributed models (inter-comparison study) in order to identify avenues to further improve GEM-Hydro.

3) P.5, l2.: Please consider be more specific about the calibration strategy of Haghnegahdar et al. (2014). AR: the paragraph was changed into the following, as the calibration strategy is explained in more details in the section "calibration strategy". The version of MESH used in this study relies on version 3.6 of the Canadian LAnd Surface Scheme (CLASS). Each grid cell is subdivided in a number of tiles, and each tile is classified as belonging to one of the five grouped response units (GRUs), based on its land-use/soil type combination. In this paper, we follow the local calibration strategy advocated by Haghnegahdar et al. (2014) for MESH (see section on calibration strategy).

4) p.5, l.14-18: This part is messy and not clear. Please consider to revise. AR: this paragraph was modified into this: The same was shown for WATROUTE which produces outputs of similar quality be it implemented at a low (10 arcmin for MESH and WATFLOOD) or high (0.5 arcmin with GEM-Hydro) resolution, as long as results are evaluated for large enough catchments (i.e., catchments which spread over at least a few grid cells). However, the high-resolution WATROUTE version is preferred in GEM-Hydro for consistency with the WCPS-GLS (Durnford et al., in preparation) recently developed at ECCC. Hence, the higher resolution GEM-Hydro's routing scheme is not expected to give GEM-Hydro any advantage in comparison to SA-MESH and WATFLOOD.

5) P.5, l19-21: Why are the used time-steps different? Has it some implications for interpreting results? AR: Generally, the internal time-step is adjusted in order to be maximized while preserving numerical stability, which hence depends on the equations and numerical schemes (implicit or explicit) of a given model. Of course, the internal time-step however has to remain lower than the output frequency. This is why WATFLOOD uses a time-step of 60 min. (results are being output on an hourly basis). This does not rise any issue in regard of results' interpretation, because we here assess the performances on a daily basis, which is a temporal resolution well resolved by all considered models. In order to make it more clear to the reader, the following sentence was added: " The internal time-step of a model is generally maximized up to the desired output interval, provided that it satisfies numerical stability. In the GEM-Hydro version used in this study, a 10-min. time-step was required to achieve numerical stability, but a newer version now allows to increase it."

6) P.6, l.27: Why is lumped model mentioned here? Are the findings (good performance) for the right reasons? The reference of Gaborit is not accessible so it is difficult to see. AR: Yes, results indicate that the good performances of the models are obtained for the right reasons, or in other words that simulations closely follow the observed streamflow dynamics with realistic parameter values. Moreover, as explained in the first GRIP-O study, the parameters obtained by GR4J on regulated catchments (which contain reservoirs) reveal that the model increased the routing storage capacity, which makes sense. These results obtained with lumped models during the first GRIP-O study are mentioned because they allow to comfort the fact that regulation in the area of the Lake Ontario watershed generally involves simple management operations, and are close to the behavior of natural lakes. The reference will soon be published in the Journal of Great Lakes Research (final corrected proof has been sent) and the final version of this paper will include a reference which is possible to access. To make it more easy to read and more logical, the paragraph was reformulated into the following: "Most of the rivers are regulated in some ways, mainly for hydropower and flood mitigation, but regulation generally consists of reservoirs with a simple weir

at their outlet (i.e., static control). Therefore, this did not prevent lumped models from reaching good performances in the former GRIP-O study of Gaborit et al. (in Press). As a consequence, no effort was made to represent in a detailed manner the artificial structures of the region in WATROUTE." 7) P.7, l.6, l.9: Which hydraulic parameters? Is the maximum soil depth calibrated for each grid cell or entire domain? AR: In Table 3, GEM-Hydro hydraulic parameters for the soil consist of GRKM, KVMO, PSMO, BMOD, WMOD, and to a lesser significant degree, RTD (see definitions in Table 3). The Maximum soil depth is currently fixed over the whole domain in GEM-Hydro. It is different for each of the 5 GRUs in SA-MESH. To make it clearer, the following sentence was modified into this: "The maximum soil depth is calibrated in GEM-Hydro and SA-MESH (Table 3 to Table 5). However, GEM-Hydro relies on a constant soil depth for the entire model, while SA-MESH uses a different soil depth value for each of its five GRUs."

8) P.7, l.19: what is RDPS? AR: It is the Regional Deterministic Prediction System, as defined close to page 7, line 14. 9) P8, l.3: how many runs has typically local calibration? AR: we used 400 with GEM-Hydro-UH, as mentioned close to page 8, line 17. The following sentence was extended as follows: "Calibration cost did not allow models to be calibrated locally for all GRIP-O subbasins (Fig. 1). One local calibration takes between 2 and 5 days of computation (400 model runs, see below)." 10) P.8, l.14: Please be more specific how were the values constrained? AR: the corresponding sentence was modified this way: " More precisely, soil water content thresholds and albedo (Table 3) cannot be higher than 1. Therefore, these values were constrained to realistic ranges after they were adjusted by the calibration algorithm by imposing them a minimum of 0 and a maximum of 1." 11) P.9, l.1-6: This part is not clear. On how many points is then the model verified/ compared /calibrated? AR: The model is verified against observed streamflows for tributaries entering Lake Ontario, at the location of the "most-downstream" gauge stations. In the case of subbasins having only one manor trbutary entering the Lake, the model was verified against the true observations of the most-downstream gauge of this tributary. But in the case of subbasins having more than one main tributary entering the Lake, all most-downstream observed

flows were summed up to end up with a unique time-series of observed flows, as if the catchment had only one major tributary. This was done in order to simplify the implementation of the lumped models of the first GRIP-O study, which cannot simulate streamflow at several points at a time. Therefore, lumped models were implemented on subbasins with more than one major tributary as if it had only one. Moreover, in such a case was this synthetic observed time-series extrapolated to the whole sub-basin area, including its ungauged areas, using a simple area-ratio method (ARM). The same framework was followed with GEM-Hydro UH for this type of subbasins for consistency with the lumped models (i.e., in order to be able to compare results of the different models in a fair manner), and because GEM-Hydro UH also can only simulate streamflow at one point at a time. This methodology is clearly explained in the first GRIP-O study cited in the paper, and is here briefly explained. However, a clarification effort was made and results in commented paragraph ending up as follows: " Finally, some subbasins in Fig. 1 have more than one major tributary flowing into Lake Ontario. In this case, the most-downstream observed flows on independent tributaries are summed and then extrapolated to the whole subbasin using the Area Ratio Method (ARM; Fry et al., 2014). The resulting "synthetic" flows were considered as observations for GEM-Hydro-UH calibration over the whole subbasin. This methodology was applied to all subbasins with more than one most-downstream gauge (identified with the "N/A" mention for the station attribute in Table 6) for consistency with the calibration experiments performed in the first GRIP-O study (see Gaborit et al., in Press), and because lumped models (and GEM-Hydro-UH) can only estimate streamflow at one location. For these subbasins, the true gauged fraction is specified in Table 6." 12) Strategy for ungauged basins: Typically, the prediction in ungauged basins is verified by leave-one-out approach. How do the results compare with such method? Please consider to discuss. AR:This method would have been too costly in terms of computational time given that it would require a number of global calibrations equal to the number of GRIP-O subbasins (i.e., 14). Given that it takes between 10 and 14 days for each global calibration to be performed, it can easily be noticed that the time required for using the "leave-one-out" approach would be too long in this case. The "leave-one-out" methodology is generally performed when accompanied with local calibration which, despite requiring a distinct calibration for each of the subbasins, which is also very time-consuming, is much less costly than when using global calibration (local calibration takes about 4 days per subbasin to complete). In our case, we therefore use a completely different approach than those generally described in the Litterature for example by Razavi and Coulibaly (2012), and Parajka et al. (2013). The "leave-one-out" method does make sense when there's a high variability of the parameter values (as is the case with local calibration), which can not only be due to the model and the calibration algorithm themselves, but also to a high variability in catchment characteristics such as elevation, geology, land-use, soil texture, etc. Here, we have neither of these drawbacks: because we use global calibration, we end up with a unique, spatially consistent parameter set, and the area under study is relatively homogenous in terms of the aforementioned characteristics. Of course, a truly objective assessment of the parameter transfer proposed here would indeed require the "leave-one-out" approach, but this is not the objective of the study, which is instead of proposing a reliable and efficient manner of estimating streamflows for the ungauged areas of the watershed. The reliability of the method employed here (global calibration) is demonstrated by the fact that with a unique parameter set, the simulation performances are satisfying for all of the subbasins, which therefore logically leads to the assumption that it is satisfying for the ungauged areas as well, based on spatial and hydrologic proximity of the catchments. Moreover, temporal robustness of the method is also very good. Therefore, it is argued that the "leave-one-out" approach is not required in our case study to demonstrate the reliability of global calibration and parameter transfer to nearby ungauged catchments. In order to include these considerations in the paper, the following paragraph was added in section 2.3: " Despite a comprehensive assessment of the reliability of the methodology used here for parameter transfer would require the "leave-one-out" framework (see Razavi and Coulibaly, 2012), the satisfying performances and temporal robustness obtained for all GRIP-O subbasins with global calibration, along with the spatial consistency of the unique final parameter set, the homogeneity of the area under study and the spatial proximity of ungauged catchments together justify the relevance and a priori reliability of the methodology employed in this study. This statement if moreover supported by the evaluation performed further down for the whole watershed."

13) P.11, l.28-29: Please consider to show some results supporting this statement. AR: this was done with the addition of the following paragraph along with the addition of the reference to NOHRSC (2004): " Calibration also improves GEM-Hydro-UH Snow Water Equivalent (SWE) simulations but to a lesser degree than for the streamflow. For example, the NSE values for SWE simulations over the 4 consecutive winters of the GRIP-O period improved from -0.12 to 0.42 for the Genessee subbasin, and from 0.49 to 0.68 for the Black River subbasin, respectively before and after calibration (the SWE variable was not used in the computation of the objective function). SWE observations come from the SNow Data Assimilation System (SNODAS, see NOHRSC 2004)." 14) P.16, l.17: Please consider to update XXXX. AR: this will be done at the final stage of the publication process. 15) P.21, l.5: “the most-downstream flow gauges” is not clear. AR: This legend was updated as follows: " Dots (blue for natural flow regimes and red for regulated regimes) are the most-downstream flow gauges (i.e., the main tributaries’ gauges which are closest to Lake Ontario’s shoreline) selected for model calibrations. " 16) Table 1: Please be more specific what is radiative forcings AR: it was specified as follows: " R:, radiative forcings (short- and long-wave incoming radiations)"

Please also note the supplement to this comment:
http://www.hydrol-earth-syst-sci-discuss.net/hess-2016-508/hess-2016-508-SC1-supplement.zip

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-508, 2016.