

## Reply on RC1

Emmanuel Dubois et al.

---

Author comment on "Simulation of long-term spatiotemporal variations in regional-scale groundwater recharge: contributions of a water budget approach in cold and humid climates" by Emmanuel Dubois et al., Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2021-71-AC1>, 2021

---

Dear Referee,

We would like to thank you for your comments and review aimed at improving our paper to make it more interesting to a wider audience. We will modify our paper following your recommendations. We agree that the results need to be presented in a broader, less local-context way and we will include modifications to position our paper within the context of cold and humid climate research on groundwater recharge (GWR). To reflect this, the title will be modified to "Simulation of long-term spatiotemporal variations in regional-scale groundwater recharge: Contributions of a water budget approach in cold and humid climates". The introduction will be modified to position the study within the recent research on GWR in cold and humid climates, such as Grinevskiy et al., 2021; Kløve et al., 2017; Nygren et al., 2020; and Pozdniakov et al. 2020 (complete references at the end of our response). As well, the sub-sections 5.1 and 5.2 will be modified to compare the regional-scale results to GWR estimates obtained in likewise cold and humid climates such as in Scandinavia and northern Russia. It will include a comparison of the GWR trends within the warming climate observed in these regions since the middle of the 20<sup>th</sup> century. Thus, it will be possible to discuss further the contributions and limitations of water budget approach in these climates. Hereafter we describe the main modifications that will be made to the document based on your comments. The general comments are addressed first (labelled C#), followed by the minor remarks (labelled mC#).

*C1: Lines 172-175. The Authors mention three basic hypothesis. I have serious concerns on the third one: the watershed response time is shorter than one month, thus compensating for the absence of water routing. This is a very strong hypothesis, as it allows neglecting transient dynamics of the aquifer. Before proceeding with any computation, Author should demonstrate and convince the reader that this a reliable assumption.*

A1: We agree with the Referee that this hypothesis needed to be justified. However, monthly time steps are frequently used in groundwater flow modelling and are considered a reasonable timescale for groundwater flow processes in cold and humid climates (Dripps and Bradbury, 2007; Guay et al., 2013). To avoid misleading the reader, L172-177 will be rephrase: "The HydroBudget model is calibrated on superficial watersheds, based on the hypotheses that: 1) surface watersheds match hydrogeological watersheds, 2) the rivers drain unconfined aquifers. Under these conditions, for any given watershed, potential GWR

should be similar to river baseflow at the outlet, and the sum of runoff and potential GWR should be equal to the total flow at the outlet. Dripps and Bradbury (2007), and Guay et al. (2013) have shown that results of GWR water budget models (SWB and HELP, respectively), compared better to groundwater flow models (GFLOW and CATHY, respectively) at the monthly time step than at a daily time step. Thus, runoff, AET, and potential GWR were computed with HB with a daily time step, and the results were aggregated on monthly time steps. Calibration and validation of the model and analysis of the results were all performed on the monthly aggregated results."

*C2: The model accounts for 8 parameters to be calibrated. It seems to me that some of them are set of parameters. For example the Runoff factor (as explained by the Authors "Partitioning between runoff computed with the RCN method and infiltration into the soil reservoir") do depend also on the land cover or not? in the first case, you are calibrating each runoff factor for each soil. Am I wrong? Maybe I do not understand correctly.*

A2: We thank the Referee for this comment, and we agree that there is a need for clarification. The runoff computed with the RCN indirectly depends on the land cover since simple land cover classes are used to compute RCN values that are later used to compute runoff for each grid cell. However, a single value of the runoff factor parameter ( $f_{runoff}$ ) is used for the entire watershed (one value for all grid cells). As detailed in Dubois et al. (2021; eq(1) p<sup>o</sup>9) :  $R = f_{runoff} \times R_{RCN}$  with R the runoff for each grid cell and  $R_{RCN}$  the runoff computed for each grid cell. To ensure that this is clear, we will change the description of  $f_{runoff}$  in Table 2 for "Correction factor for runoff computed with the RCN method for the partitioning between runoff and infiltration into the soil reservoir". As well, L150 will be rephrased as follows: "It is based on simplified process representations and is driven by eight parameters that need to be calibrated. These parameters are held constant for all the grid cells and through time (Table 2)."

*C3: The coupling of time steps (daily time step for soil modelling and monthly time step for GWR) is not clear to me. Please, give more details on that.*

A3: We thank the Referee for pointing out the missing information concerning the aggregation on a monthly time-step of the outputs computed with a daily time-step. The model uses daily precipitation and temperature data to compute daily runoff, evapotranspiration, and potential GWR. These daily results are aggregated at a monthly time-step to perform 1) calibration and validation and 2) results analysis. We integrated a clarification of this point in the modified L172-177 presented in A1.

*C4: The targets for calibration are both the total surface flow and the baseflow. However, the first is observed, whereas the second is estimated (through the Lyne and Hollick filter). In my opinion this is a weak point of the whole calibration procedure: generally speaking I do not like to calibrate a model using output of another model. Even if they match each others, what can I say on the reliability of both? The Authors should at least convince the reader on the reliability of the baseflow estimates.*

A4: We thank the Referee for this comment and agree that it is a sensitive point in our research. In eastern Canada and in other similar geo-climatic environments, the topography-driven water table make streams and rivers discharge areas of groundwater systems. In these conditions, baseflow estimated with recursive filters have long been used as proxies of groundwater recharge for model calibration. Although it is an

approximate proxy, the lack of local data/studies that would demonstrate its relevance and limits encourage authors to continue using river baseflows in GWR and other groundwater-related studies (Rivard et al., 2009). In section 5.4 on method limitations, paragraph L472-483 acknowledges the use of baseflows as an imperfect proxy for calibration. This paragraph also underlines the usefulness of using a baseflow estimate based on recursive filters for simulation in an area where mainly long-term river flow measurements are available. Nevertheless, to avoid any misleading, L178-180 will be rephrased as follows: "The daily flow rates at the outlets were those from the 51 gauging stations (Fig. 1). Baseflows were estimated from the flow rate time series calculated following the proposed approach of Ladson et al. (2013) based on the Lyne and Hollick (1979) recursive filter, using a stochastic calibration and 30 filter passes. As aquifers generally discharge into superficial water bodies in southern Quebec, baseflows estimated with recursive filters are considered a satisfactory, although approximate, proxy of GWR dynamic."

*C5: The previous one is a very important point, also because the objective function (eq. 1) is a linear combination of two different metrics, one referring to the total flows, the second one only to the baseflows. In my opinion, dependence of the calibration on the weights adopted to define the objective function should be explored much more in details. The explanation you gave for your choice (lines 203-205) is too simplistic.*

A5: We thank the Referee for pointing this out and we agree that our explanation about the weights on the objective functions was too simplistic and may have been misleading. The caRamel algorithm performs the multi-objective optimization using the objective functions separately, here  $KGE_{qtot}$  and  $KGE_{qbase}$ , and plots the evolution of the Pareto front (model runs function of the objective functions). Therefore, the weights used to compute the  $KGE_{mean}$  are only used to extract the best compromise from the ensemble of model runs computed during the optimization and do not impact much the choice of the best compromise. To illustrate this and provide a convincing argument, supplementary table A1 will be completed to values of the objective function based on other combinations of  $KGE_{qtot}$  and  $KGE_{qbase}$  (see below). L203-205 will be rephrased as follows: "The weights attributed to each objective function in  $KGE_{mean}$  were arbitrarily chosen to select the calibrated parameter set that maximizes the quality of simulated baseflows, considered as a proxy for GWR ( $KGE_{qbase}$ ), without losing the benefits of the multi-objective optimization. Supplementary Table A1 shows that the objective functions and the calibrated parameters were moderately sensitive the weights."

**Calibration**

**Validation**

**KGE<sub>mean</sub>**

**KGE<sub>qtot</sub>**

$KGE_{mean} = 0.4 \times KGE_{tot} + 0.6 \times KGE_{base}$  (retained solution)

**Best compromise**                      0.76<sub>1</sub>                                      0.80<sub>7</sub>

**25 best compromises**                      [0.75<sub>8</sub> ; 0.76<sub>1</sub>]                                      [0.78<sub>1</sub> ; 0.82<sub>0</sub>]

$KGE_{mean} = 0.5 \times KGE_{tot} + 0.5 \times KGE_{base}$

**Best compromise**                      0.77<sub>2</sub>                                      0.81<sub>9</sub>

**25 best compromises**                      [0.77<sub>0</sub> ; 0.77<sub>2</sub>]                                      [0.80<sub>5</sub> ; 0.82<sub>5</sub>]

$KGE_{mean} = 0.6 \times KGE_{tot} + 0.4 \times KGE_{base}$

**Best compromise**                      0.78<sub>2</sub>                                      0.82<sub>5</sub>

**25 best compromises**                      [0.78<sub>0</sub> ; 0.78<sub>2</sub>]                                      [0.81<sub>4</sub> ; 0.83<sub>5</sub>]

**Optimized parameters\***

*KGE<sub>mean</sub> = 0.4 x KGE<sub>qtot</sub> + 0.6 x KGE<sub>qbase</sub> (retained solution)*

**Best compromise**

## **25 best compromises**

$$KGE_{mean} = 0.5 \times KGE_{tot} + 0.5 \times KGE_{base}$$

## **Best compromise**

## **25 best compromises**

$$KGE_{mean} = 0.6 \times KGE_{tot} + 0.4 \times KGE_{base}$$

## **Best compromise**

## **25 best compromises**

*\* $F_T$  is held constant at the nominal value of 16.4 d*

*C6: I found the sensitivity analysis performed on the W6 group of gauging stations very interesting and potentially the core business of the paper (which otherwise it is only a modelling study, important for Quebec, but not interesting for the international reader of HESS). Why the Authors decided to carry out the sensitivity analysis only on one group of gauging stations. In my opinion, it could be much more interesting to perform the same analysis to several groups of gauging stations to explore the possible differences among watershed and the dependences on climate forcing, soil, land cover etc In other words, has the ranking proposed for W6 a general validity? why?*

A6: This is an excellent suggestion. We are currently carrying out the sensitivity analysis for the rest of the gauging station groups. As soon as the runs will be over and the results are interpreted, we will post here a supplementary answer with the update. We thank the Referee for this helpful comment.

*C7: Presentations of the results are sometimes not easily readable. For example, in section 4.3 one can find several information already shown in figure 6 and table 4). I think that the description of the results (sections 4.2, 4.3, 4.4) can be much simplified. On the contrary, results on the temporal evolution (analysis of trends) deserves for much more attention and a dedicated picture to presents results.*

A7: We agree with the Referee that results presentation would benefit from being simplified. To make the manuscript clearer and more concise, we will withdraw fig. 6 since the presented information are redundant with table 4. As suggested, we will develop the trend analysis since the length of the time series would allow to follow the GWR evolution since the 1960s. We will include our proposition of the new sections in our future answer, along with the sensitivity analysis.

*C8: The entire discussion on the temporal patterns of groundwater recharge relates the time variation of meteorological forcing to discharge, neglecting possible changes in the land cover. I do not know if such an assumption is correct, however it should be explicitly stated*

A8: This is a very pertinent comment. We have worked under the assumption that no major land cover change occurred during the simulation period. However, we did not mention this in the text and we thank the Referee for pointing this out. The following sentence will be added at the end of the paragraph L172-177: "As well, the calibration and validation of the model were performed under the hypothesis that no major land use change occurred during the simulation period (*i.e.* land use was held constant)."

*C9: Results of figure 6 are surprising: proportion between runoff, AET, GWR does not depend on the watershed (differences among watersheds are in the order of 1-2%). Figure 5(b) shows different patterns: for example over W2 GWR/P spans between 0.1 and 0.15 for most of the grids; on the contrary, over W4, it seems that GWR/P is >0.3 for half*

*of the cells. Maybe I missed something, but it does not seem to me that Fig 5 and Fig 6 are coherent.*

A9: We thank Referee for this comment, and we have double-checked our results. Considering the spatial variations in precipitation within the study area (952 mm/yr in W1, 1 039 mm/yr in W2, and 1 123 mm/yr in W4; table 1; fig. 2), variations are found between absolute values of potential GWR (109 mm/yr in W1, 119 mm/yr in W2, and 147 mm/yr in W4; table 4). The ratios of average potential GWR/average precipitation for these three watersheds are 0.11<sub>4</sub>, 0.11<sub>5</sub>, and 0.13<sub>1</sub>, respectively. When computed from the spatially-distributed GWR/P ratio presented in fig. 5b, the averaged ratios for the same three watersheds are 0.11<sub>4</sub>, 0.11<sub>3</sub>, and 0.13<sub>1</sub>, respectively. Results are equal and we think that the difference in precipitation explains that the two figures are coherent although it seems that the GWR rates would be higher in W4.

Minor comments:

mC1: *Line 102-109 All the information given here are summarized in Table 1. This lines appear to me not useful.*

L97-108 will be rephrased as follows: "The study area is located in the province of Quebec (humid continental climate), between the St. Lawrence River and the Canada-USA border, and between the Quebec-Ontario border and Quebec City (35 800 km<sup>2</sup>) (Fig. 1). It is comprised of the watersheds of eight main tributaries of the St. Lawrence River (numbered 1 to 8 from west to east) (Table 1). Watersheds W1 (Châteaugay River), W2 (Richelieu River), and W4 (Saint-François River) are partially located in the USA (42%, 83%, and 15% of their total areas respectively). Topography is flat with low elevation areas close to the St. Lawrence River and higher elevations in the Appalachian mountain range, associated with steeper slopes. Land cover includes agriculture, forest, wetlands, urban uses, and surface water (Fig. 2a). Agriculture dominates in the watersheds located in the St. Lawrence Platform, while forest occupies most of the Appalachian watersheds."

mC2: *Line 112. In my opinion it would be better to add also the mean bias as metrics of the goodness of interpolation, to avoid systematic under or overestimation.*

L110-112 will be rephrased as follows : "The high density of measurements during this period generated minimal error on the interpolated data, with a root mean square error (RMSE) of 3 mm/d for precipitation (mean bias of +0.1 mm), 2.5°C for minimal temperature (mean bias of -0.5°C), and 1.5°C for maximal temperature (mean bias of +0.1°C; Bergeron, 2016)."

mC3: *Figure 1. In my opinion the map in the middle is not useful: I suggest to eliminate it, retaining the location map and the watersheds map.*

Fig. 1 will be modified as presented in the supplementary PDF file.

mC4: *Figure 4. As GWR is your main output, I would present several graphs as the panel (b) for several stations (one as an example is not useful)*



Fig. 4 will be as presented in the supplementary PDF file.

The associated paragraph L247-267 will be rephrased as follows: "The calibrated HB model was used to simulate potential GWR for the entire study area on a 500 m x 500 m grid for the 1961-2017 period. Examples of the resulting monthly vertical inflows, baseflows, and potential GWR are illustrated in Fig. 4 for the downgradient stations in W3, W7, and W8. The simulated potential GWR compared favorably with baseflows estimated using the Lyne and Hollick (1979) digital filter. Maximum values were reached simultaneously in April, during the spring month(s) of maximum VI. A second baseflow and GWR peak was observed and simulated in November-December of most years. Lowest values were reached in July to September (high AET rates) and February (minimum VI). Similar matching results (timing and amplitude) were obtained for the river flow (not presented here). Simulated AET was null in the winter until the spring thaw, after which it quickly reached its highest value (> 100 mm/month) in July and decreased at the end of August, reaching null values again in November. Comparable results were obtained for the other gauging stations."

#### References:

Grinevskiy, S. O., Pozdniakov, S. P., and Dedulina, E. A.: Regional-Scale Model Analysis of Climate Changes Impact on the Water Budget of the Critical Zone and Groundwater Recharge in the European Part of Russia, 13, 428, <https://doi.org/10.3390/w13040428>, 2021.

Kløve, B., Kvitsand, H. M. L., Pitkänen, T., Gunnarsdottir, M. J., Gaut, S., Gardarsson, S. M., Rossi, P. M., and Miettinen, I.: Overview of groundwater sources and water-supply systems, and associated microbial pollution, in Finland, Norway and Iceland, *Hydrogeol J*, 25, 1033–1044, <https://doi.org/10.1007/s10040-017-1552-x>, 2017.

Nygren, M., Giese, M., Kløve, B., Haaf, E., Rossi, P. M., and Barthel, R.: Changes in seasonality of groundwater level fluctuations in a temperate-cold climate transition zone, *Journal of Hydrology X*, 8, 100062, <https://doi.org/10.1016/j.hydroa.2020.100062>, 2020.

Pozdniakov, S. P., Vasilevsky, P. Yu., Grinevskiy, S. O., Lekhov, V. A., Sizov, N. E., and Wang, P.: Variability in spatial-temporal recharge under the observed and projected climate: A site-specific simulation in the black soil region of Russia, *Journal of Hydrology*, 590, 125247, <https://doi.org/10.1016/j.jhydrol.2020.125247>

Please also note the supplement to this comment:

<https://hess.copernicus.org/preprints/hess-2021-71/hess-2021-71-AC1-supplement.pdf>