Dear Dr. Markus Hrachowitz,

Thank you very much for your comments and providing the references. I fully agree with you and the reviewer regarding the possibility of groundwater inter-watershed flows. Please see my response in more detail below. The manuscript has been revised by addressing all of the three comments, as shown in the track-change version.

We greatly appreciate your time and help on improving our manuscript.

Sincerely,

Shusen Wang and Co-Authors

Response to reviewer's comments:

1. I cannot agree with the assertion that Eq (1) does not built on the assumption of watershed water balance closure, although it might not be a big issue for this watershed. I did not find solid evidence for the declaration that the subsurface water change Sg only contributes to the baseflow. How do the authors know there are no groundwater systems across two watersheds?

Authors: We agree with the comment, and yes, watershed water imbalance can affect the model performance. If water balance was an issue, for example, due to the inter-watershed groundwater flows, the model would misrepresent the natural hydrological system, particularly for its baseflow. Our previous studies on water balance (Wang et al., 2014a, b) for Canada's watersheds showed that the Albany watershed had a fairly good water closure, suggesting that the data quality issues and groundwater inter-watershed flows were likely small. Discussions on the potential impact of groundwater inter-watershed flows on our modelling are added in Paragraph 5 of the Discussion Section (Page 18). The studies of Bouaziz et al. (2018) and Hulsman et al. (2021) helped the discussions and were cited therein.

2. As declared as physically-based model, the authors define the meaning of the parameters as shown in Table 1. I understand the authors include the variable Tacc to account for the annual/yearly variations of watershed conditions over the freezing period. Clearly, the parameters are assumed to be constant when using the 15-year winter data for calibration. But I do not understand why the derived parameter K0 derived from the winter period can work for the summer period. It did not consider the variation of K0 for summer periods. Therefore, it should not work well in the summer periods. In fact, we can see this in Fig. 8. In comparison to other methods, it yields much higher baseflow coefficients for the summer periods like in years 2005,2006,2010,2011, and so on. In my understanding, this definitely means a systematic bias over a year. Whereas, Fig 9 easily mislead the readers, since the proposed method originates from the 15-year data with a kind of statistical meaning. Therefore, I can not agree with the authors' explanation. The limitation should be discussed clearly.

Authors: The K0 is defined as the K for summer when frozen soil is absent. The baseflow varies in summer due to the water storage changes, rather than the K changes, in our model. Our model is basically a modified linear reservoir model in summer so K is treated as a constant (K0) (see different model comparisons for the Albany basin in Wang, 2019). In winter, frozen conditions limited liquid water transfers and reduced soil/aquifer hydraulic conductivity, and K becomes dynamic with freezing temperatures. The above-mentioned years had dry summers. Our model estimated little surface runoff for these years in summer. This is not a result of higher baseflow coefficients in these summers. Rather, it is the result that our model estimated the watershed water storage was mainly attributed to subsurface water in these summers. The baseflow coefficients estimated from both our model and inversely calculated from observations (Fig. RC2-1) also did not suggest higher values for these years. Discussions on this was added in Paragraph 7 of the Discussion Section (Page 19) in the revised manuscript.

3. The stream flow data in Fig. 8a is different from that in the other three subplots. There needs some explanation. Please scale it for better comparison with other methods.

Authors: Thanks for the comment! The difference is due to that the “months” in GRACE data are slightly different with the calendar months (See Fig 6 for detail). The “GRACE month” varied from year to year so it is not an ideal option to use it to scale other Q data that were based on calendar month (standard). Luckily, the difference between the “GRACE month” and the calendar month is mostly small and it had minor impact on our analyses. Explanation is added in the Figure caption in the revised manuscript.