Comment on gmd-2020-429
Liangjun Zhu (Referee)

Referee comment on "GP-SWAT (v1.0): A two-level graph-based parallel simulation tool for the SWAT model" by Dejian Zhang et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-429-RC2, 2021

General comments:
This manuscript proposed a unified parallelization strategy for both watershed-level and subbasin-level parallelization and developed the GP-SWAT accordingly, which is valuable and useful for both developers and users in the scientific community. Except for the specific comments posted by Anonymous Referee #1 (https://gmd.copernicus.org/preprints/gmd-2020-429#RC1), I have some other specific comments.

Specific comments:
1. The phase "two-layer graph-based parallelization" is ambiguous and inexact. (1) Do you mean "layer" equals "level" since you use "model-level" (or "watershed-level", please unify the terms) and "subbasin-level" in the manuscript? (2) The "graph-based" is the parallelization strategy for "subbasin-level" not for "model-level", since model runs are independent to each other and only their outputs are concerned.
2. In the introduction, the first paragraph stated that both the single model and numerous models require prohibitive execution time. However, the first three of the four methods introduced in the second paragraph are used for model-level applications. In my view, it is more clear to introduce the methods for alleviating computational burden for the model level applications and the single model, separately. And then both focus on parallel computing.
3. In the sentence around No 65, it is more precisely to cite Liu et al. (2016) and/or Zhu et al. (2019) rather than Liu et al. (2014).
4. In the sentence around No 80, "However, these methods have two major limitations: ... complex computational facilities that may not be readily available..." should be reconsidered. In fact, many parallel computing models (e.g., MPI) can also be running on the personal computer and obtain a good speedup ratio.
5. This manuscript focuses on the parallelization of both model-level and subbasin-level but missed existing similar methods such as Zhu et al. (2019).
6. Overall, the introduction failed to raise the scientific issue clearly and precisely, that is, there is no unified parallelization strategy for both watershed-level and subbasin-level parallelization that do not need to reconstruct source code of hydrologic model to handle data communication among subbasins explicitly (e.g., using MPI). If this is correct, the title may also be changed accordingly.
7. What is the phrase "the current computation step" mean in the sentence around No 140? To my understanding, the proposed method is different from the spatial-temporal
discretization proposed by Wang et al. (2013). For example, in Fig 4, Subbasin 1 and Subbasin 2 are executed for the entire simulation period (e.g., 5 years) first, then Subbasin 3 begins to run, and so on. This may lead to a poor load balance, and hence a low speed up ratio, especially for a single model run. Is this right?

8. In Section 4.1, all the results are compared through speed-up ratios, you should also give the actual execution times. I want to know the performance of the subbasin-level parallelization for a single model simulation compared with the original SWAT model.

9. Why use two study areas? From my perspective, the two case studies have no significant difference. We need more information about the study areas.

10. In sentences around No 235, it is weird that the numbers of HRUs per subbasin can be set the same.

**Technical corrections:**

1. Did you mean that the “Spark-SWAT” is the alias of the “GP-SWAT”?
2. In the code, all file paths are specific to the author’s computer. This is not suitable for code distribution. Even so, the modification of these paths should be clarified in the tutorial.

**References:**

