Comment on gmd-2020-399
Anonymous Referee #3

The authors present a model ‘STREAM’ that is used to derive river discharge and runoff. The STREAM model is conceptually and computationally simple, and uses inputs of precipitation, total water storage, soil moisture as well as air temperature to provide estimates of global runoff. The results are tested for the Mississippi River basin in the United States and indicate good agreement. Enhancing the ability to model distributed runoff has important applications for hydrology. However, further justification of the methods used and applicability to other climate regimes and regions is needed.

For example, the authors should comment on the sensitivity of the model to the hydrological inputs of precipitation, soil moisture and total water storage anomaly. The authors should comment on the contribution of using these inputs, and whether results are improved or not by using all three. Otherwise it is not immediately clear to the reader the contribution of each to the estimation of runoff.

The authors also only test their model in the Mississippi River basin, however it would be interesting and informative to address the performance of this model in different regions including more arid basins, snow-dominated, lots of topography, heavily managed, etc. The study indicates it is a ‘global’ model so more discussion of its applicability worldwide is needed.

Specific comments:

Line 208 - Can the authors provide the depth of the soil moisture used in this study.
Given that the authors only validate in the Mississippi basin, can they comment on how different climate regimes could impact the accuracy of the modeled runoff in particular more snow-dominated basins. Can the authors comment on the validation of the snow module.

Can the authors provide additional information on how they resolve the differences in spatial and temporal scale between the various input data sets provided. In particular, the coarser scale of the GRACE data.

Can the model be run with same input precipitation as GRUN for the validation purposes? Or can the authors comment on precipitation differences between either product?

Does using GRACE data for water storage (which captures both human and natural processes) address this? GRACE can indicate human activity and water extraction practices, which I think could help improve purely ‘natural’ estimates of runoff.