Response to Review of "Global streamflow and flood response to stratospheric aerosol geoengineering" by Wei et al.

We first thank the referee for his/her insightful comments, which helped us clarify and greatly improve the paper. In the reply, the referee's comments are in *italics*, our response is in normal and changes to the text are shown in blue.

Anonymous Referee #2

The authors present a suite of simulations from six GCMs that participated in GeoMIP, under two scenarios: RCP4.5 and an SRM scenario from GeoMIP (G4). The authors note that this is the first study to assess how SRM might affect global-scale streamflow. The authors compare the two scenarios to demonstrate what effect SRM might have compared with a non-SRM scenario in a mid-emissions (RCP4.5) future, in terms of high and low flows, mean flow, and return period flows.

General Comments:

1. *Lines 76-80: "Under the Geoengineering Model Intercomparison Project (GeoMIP; Robock et al., 2011; Kravitz et al., 2011, 2012, 2013a), G4 experiment a constant 5Tg per year of SO2 is introduced into the lower tropical stratosphere of climate models during the period of 2020-2069, while greenhouse gas forcing is defined by the RCP4.5 scenario". It would be helpful to readers if a little more information could be said about G4, to provide some context to the results. For example, how does the injection affect the time-series of global-mean temperature (a graph showing global-mean temperature would be quite informative here, or at least a statement of the magnitude of global cooling achieved by G4 relative to RCP4.5).*

   Reply: Yes, we agree that more details are needed on the basic response of G4 compared with RCP4.5. As this is just the introduction, we don't want to introduce figures yet. In fact, we do add a more detailed description of the results from the G4 experiment in a new section 3.1 (Precipitation, evaporation and runoff changes) later. In the Introduction we already state (line 83) "The direct radiative effects mainly result in the sharp reduction of TOA net radiative flux with a significant drop in global surface temperature, and concomitant decrease in global precipitation (Yu et al., 2015).", giving the mechanism why immediately afterwards. We also state a few side-effects of the sulphate injection. We prefer to leave the more detailed analysis to the Results section 3.1 later.

2. *Lines 61-62: "River flood models such as CaMa-Flood (Yamazaki et al., 2011) are important tools for simulating flood hazard." Yes they are, but CaMa-Flood is better described as a "river routing model" as opposed to a "river flood model". The latter implies that hydrological processes are included explicitly but they are not, since routing models take the outputs of hydrological models or climate models and route them through a network.*

Reply: Agreed. We rephrase this sentence as the following:

River routing models, such as CaMa-Flood, are important tools for the future flood hazard projection.

3. *Lines 64-66: "The high-resolution models have contributed to better simulation of river discharge (Yamazaki et al., 2009; Yamazaki et al., 2013 and Mateo et al., 2017)". This is a fair point and it is worth noting that 'offline' (i.e. separate from the hydrological model) routing models, such as CaMa-Flood specifically, have resulted in better agreement between simulated and observed discharge, compared with when the native hydrological model routing methods are used (see Zhao et al., 2017).*

   Reply: Agreed. We rephrase this sentence as the following:

   The high-resolution offline river-routing models, such as CaMa-Flood, have contributed to better simulation of river discharge (Yamazaki et al., 2009; Yamazaki et al., 2013; Mateo et al., 2017). Zhao et al. (2017) use daily runoff from Global hydrological models (GHMs) driving CaMa-Flood to produce monthly and daily river discharge. Zhao et al. (2017) find that this approach results in better agreement between simulated and observed discharge, compared to using native hydrological model routing. The CaMa-Flood model accounts for floodplain storage and backwater effects that are not represented in most GHM native routing methods (Yamazaki et al., 2014; Zhao et al. 2017; Mateo et al., 2017).

4. *Line 79: "while greenhouse gas forcing is defined by the RCP4.5". I appreciate that this scenario can be used with G4 but it is worth noting in the Discussion section, that the general conclusions drawn from this research are based upon these specific scenarios, i.e. G4 and RCP4.5. There are three other emissions scenarios under the RCPs (2.6, 6 and 8.5), which means the simulated offsetting (or otherwise) effects of SRM, particularly in terms of magnitude, could be different if the underlying emissions scenario was different (i.e. RCP2.6, 6 or 8.5).*

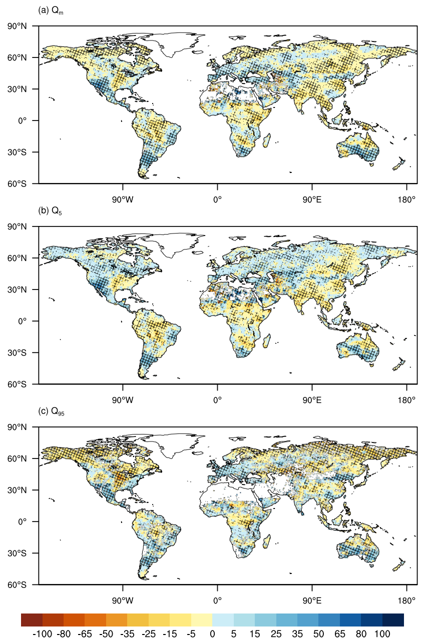
   Reply: This relates to the linearity of response of both greenhouse gas and solar radiation management. This has been explored in several studies over the past 5 years, in particular using control theory methods (e.g. MacMartin et al., 2014). At least in the climate models the responses to most variables such as global temperature, precipitation are surprisingly linear, while the response of others, particularly associated with ice/water phase change, e.g. sea ice extent, are not. This general linear response spans the full set of RCP scenarios and their radiative forcing negation by sulphate injection. Several papers discuss the control theory and linearization in terms of GeoMIP experiments, and also with the CESM1 models using a large ensemble (Tilmes et al., 2018). We add the following paragraph in Discussion section to clarify this point:

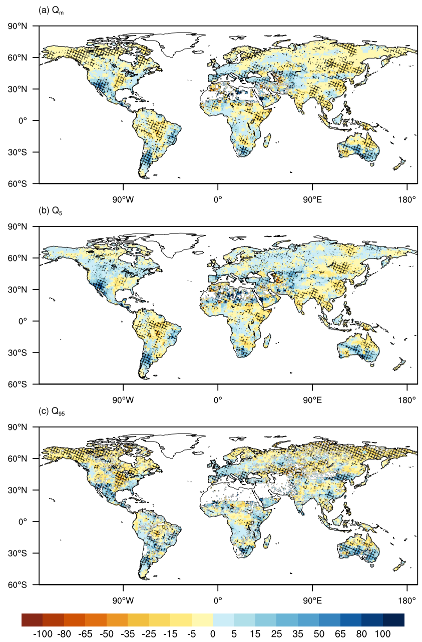
   Our results on streamflow and flood response are based on GeoMIP G4 experiment and its reference rcp45 experiment. The generalizations of the work to other types and extents of solar geoengineering depends on the linearity of the streamflow response to both greenhouse gas and geoengineering. The linearity of response of radiative forcing and global temperatures in particular have been explored in CESM1 stratospheric aerosol Geoengineering Large Ensemble (GLENS, Tilmes et al., 2018). Many climate fields, such as temperature, are surprisingly linear under a very wide range of forcing, potentially allowing standard engineering control theory methods (e.g. MacMartin et al., 2014) to tailor a global response given the freedom to use different latitudinal input locations for the aerosol injection (MacMartin et al., 2018; Kravitz et al., 2018), or combinations of, for example aerosol injection and marine cloud brightening (Cao et al., 2017). Non-linearities are expected for systems that depend on ice/water phase changes, and these could affect global streamflow and flood responses in some regions, especially in the Arctic. Moreover, the type of solar geoengineering might be relevant as well. Ferraro et al. (2014) found that the tropical overturning circulation weakens in response to geoengineering with stratospheric sulfate aerosol injection due to radiative heating from the aerosol layer, but geoengineering simulated as a simple reduction in total solar irradiance does not capture this effect. A larger tropical precipitation perturbation occurs under equatorial injection scenarios (such as G4) than under simple solar dimming geoengineering, or the latitudinal varying injections schemes explored by GLENS, or a mix of different geoengineering strategies (such as aerosol injection and marine cloud brightening, Cao et al., 2017). So the response of streamflow and flood would be expected to differ, to some extent, under different types of solar geoengineering.

5. *Table 1: Several of the GCMs include multiple runs for a single scenario, e.g. 3 runs for the CanESM-2 historical scenario. Can the authors please explain what this means? Is it a perturbed parameter ensemble of three members, or something else? Table 1. In Section 2.1 the authors also need to explain how the multiple runs in this table were dealt with. Was an ensemble mean used where there were three runs for one GCM, or were the calculations performed for each of the three runs in turn? From lines 159-160 it appears as though the runs were averaged, but it would be helpful to clarify this in Section 2.1.*

   Reply: The multiple runs in Table 1 are the number of realizations of the experiment that each model made. BNU-ESM, CanESM2, MIROC-ESM and NorESM1-M all have the same number of realizations for its historical, rcp45 and G4 experiments. Their rcp45 and G4 runs are branched from the end of corresponding historical runs, while their historical runs are branched from each model's pre-industrial control runs that were started with different initial states. MIROC-ESM-CHEM has only one historical run and its three rcp45 and G4 runs are branched from this same historical run with different initialization perturbations. GEOSCCM has no historical run, its rcp45 and G4 runs are forced with sea surface temperature and sea ice concentrations simulated by the CESM rcp45 runs. Equal weight is given to each model in the analysis and streamflow and flood response are calculated for each model before multi-model ensemble averaging is done. For models with multiple realizations, streamflow and flood response are calculated for individual realization and then averaged for each model.

To ensure our analysis is consistent on streamflow and flooding return period changes under rcp45 and G4 scenarios, we now remove the GEOSCCM model due to it lacking corresponding historical runs and also the 2nd and 3rd rcp45 and G4 realizations of MIROC-ESM-CHEM which also have no corresponding historical run. We find our conclusions hold with the reduced ensemble members. The following two figures show the streamflow changes with all models and their realizations included as in the submitted manuscript (left panel) and the streamflow changes with GEOSCCM model and the 2nd and 3rd rcp45 G4 realizations of MIROC-ESM-CHEM excluded (right panel).





We revised Table 1 to reflect these changes:

Table 1: GeoMIP models and experiments used in this study.

Model	*Resolution (degrees lat × lon, level)*	*Number of ensembles*		
		Historical	RCP4.5	G4
BNU-ESM (Ji et al., 2014)	2.8 × 2.8, L26	1	1	1
CanESM2 (Arora et al., 2011; Chylek et al., 2011)	2.8 × 2.8, L35	3	3	3
MIROC-ESM (Watanabe et al., 2011)	2.8 × 2.8, L80	1	1	1
MIROC-ESM-CHEM (Watanabe et al., 2011)	2.8 × 2.8, L80	1	1	1
NorESM1-M (Bentsen et al. 2013, Tjiputra et al. 2013)	1.9 × 2.5, L26	1	1	1

6. *Line 148: "an adaptive time step approach was applied in simulation". Can the authors please explain in detail what this means in practical terms? I had presumed that CaMa-Flood was run at daily temporal resolution for all GCMs, but this text suggests that this is not the case.*

Reply: We rephrase this sentence as the following to clarify:

The spatial resolution of CaMa-Flood is set to 0.25° (~25km at mid-latitudes). An adaptive time step scheme was applied in the model numerical integration leading to a time step of about 10 minutes, while the model outputs at daily temporal resolution.

7. *Line 150: "to the fine resolution hydrological model". I made this point earlier – CaMa-Flood is not a hydrological model, it's a routing model.*

Reply: We rephrase this sentence as the following:

In order to conserve the input runoff mass, an area-weighted averaging method is used in CaMa-Flood to distribute the coarse input to the fine resolution routing model.

8. *Section 2.4. and Figure 4: the authors calculate 30, 50 and 100-year return period levels of flows and then calculate the average across all GCMs. This approach is reasonable but in applying this method the authors overlook two important uncertainties that could influence the results significantly: 1) climate model uncertainty (from using several GCMs); and 2) statistical uncertainty introduced by calculating extreme flows for return periods that are longer than the period used to calculate them (40 years). It is known that climate model uncertainty can result in return period flows that vary more between GCMs than they do between warming scenarios, and that the range in return period flows across GCMs can be significant (Gosling et al., 2017). The authors may therefore like to consider presenting the range across all GCMs (as opposed to just the ensemble median).*

Reply: Thanks for your constructive suggestions, we add the following paragraph in Results section:

Gosling et al. (2017) compared the river runoff output from multiple global and catchment-scale hydrological model under three warming scenarios simulated by global climate models (GCMs) finding that the across-model uncertainty overwhelmed the ensemble median differences between the scenarios. In this study we use the offline hydrological model driven by runoff outputs from GCMs to calculate the streamflow, the uncertainty between GCMs is reflected in the range of return period based on streamflow change. Figure 9 shows the multi-model ensemble range of the 30-year return period level. Regions that have the shorter return period (i.e. higher flood frequency) from historical to future, show a relatively small range among models (e.g. India and Southeastern Asia). Regions that have the longer return period show a large range (e.g. Europe and North America). This reflects larger inter-model uncertainty over dry zones than for wetter ones. The return period change over dry zones is more meaningful when interpreted as the change of drought tendency. 50- and 100-year return period levels flow show larger uncertainty than 30-year return period level, which is expected when estimating the low probability extreme tails of the flow probability density function from relatively short (40 year) sets of results.

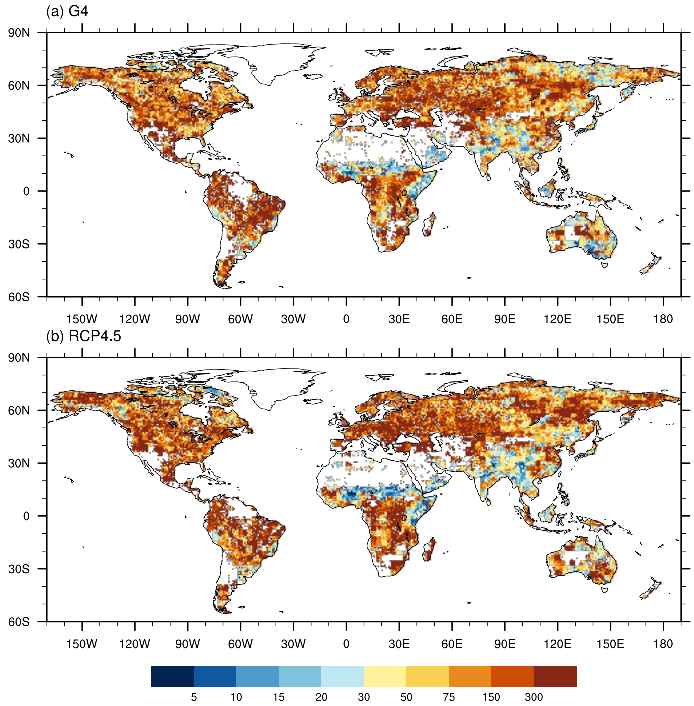


Figure 9: Multi-model ensemble range of return period for discharge that correspond with the 30-year return period in the historical simulation (1960-1999) under (a) G4 and (b) RCP4.5 scenarios, as the difference between maximum and minimum return periods. Grid cells in extremely dry regions, i.e. Q_m < 0.01 mm/day and extreme high value of return period regions were masked out.

9. *Lines 399-402: "Under the G4 experiment, some recent studies (Jones et al., 2018; Sonntag et al., 2018) have pointed out that the increased P–E (difference between precipitation and evaporation) in northern Asia caused by global warming could be partly counteracted by solar geoengineering." It is perhaps also worth noting that the way in which evapotranspiration is estimated is quite important, as this can vary significantly between different models (Wartenburger et al., 2018).*

   Reply: Agreed. We add following sentences to reflect the uncertainties on estimating evapotranspiration:

   The method for calculating potential evapotranspiration (ET) plays a significant role in determining simulated surface runoff changes (Haddeland et al., 2011; Thompson et al., 2013), which would influence the condition of streamflow. A recent study (Wartenburger et al., 2018) compared the ET spatial and temporal patterns simulated by GHMs in second phase of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP2a) also confirmed that the ET scheme used affects model ensemble variance. The ET in this study is calculated by the ESMs (Table 1), not GHMs, and any biases in ET would feed into streamflow. For example, Mueller and Seneviratne (2014) found that climate models which participated in CMIP5 display an overall systematic overestimation of annual average ET over most regions, particularly in Europe, Africa, China, Australia, Western North America, and part of the Amazon region.

10. *Lines 425-427: "The CaMa-Flood river routing model also does not consider anthropogenic effects on rivers (e.g. dams), so the results presented here are for a hypothetical natural condition." This is true but can the authors explain how this may have affected their results? Would the differences be smaller or lager if human impacts were included? This could lead on to an interesting discussion on the relative value of using runoff direct from GCMs compared with inputting precipitation and other variables from GCMs into hydrological models that include human impacts. Recent work with hydrological models shows that including dams etc. within them improves their representation of river flows compared with excluding dams (Veldkamp et al., 2018; Zaherpour et al., 2018), but also that the way human impacts such as dams are presented is quite important (Masaki et al., 2017) – so, does this mean that we should be using hydrological models that include human impacts to assess changes in the hydrological cycle with SRM, or is it reasonable to use naturalised runoff direct from GCMs instead? Clearly there is no straightforward answer but the Discussion chapter could be enhanced by considering this important issue.*

Reply: Thanks for your constructive suggestions. We agree this is an interesting and important topic and so we add the following paragraphs to discuss this issue:

In this study we use runoff direct from ESMs to drive the river routing model CaMa-Flood to study streamflow and flood response. CaMa-Flood does not consider anthropogenic infrastructures, such as dams or reservoirs, which some hydrological models do include. However, estimating future changes in human intervention on the natural system is highly uncertain. Technological advances over the century that may affect anthropogenic changes are by their nature entirely unknown at present. Hence integrating the human dimension into a model of the physical system is fraught with difficulty and uncertainty.

Several studies can be used as a guide to the possible effects of anthropogenic impacts compared with natural changes that are captured CaMa-Flood. Dai et al. (2009) argued that the direct human influence on the major global river streamflow is relatively small compared with climate forcing during the historical period. Mateo et al. (2014) suggested that dams regulate streamflow consistently in a basin study using CaMa-Flood combined with integrated water resources and reservoir operation models. Wang et al. (2017) shows that the reservoir would effectively suppress the flood magnitude and frequency. Recently, analysis of the role of human impact parameterizations (HIP) in five hydrological models and found the inclusion of HIP improves the performance of GHMs, both in managed and near-natural catchments, and simulates fewer hydrological extremes by decreasing the simulated high-flows (Veldkamp et al., 2018; Zaherpour et al., 2018). These studies suggest that the high-flows and flood response under G4 relative to rcp45 might be smaller when human intervention is considered.

As anthropogenic GHG emission increasing, human society would continually adapt to climate change and mitigate the related risk, including building new dams and reservoirs to withstand enhanced strength of global hydrological cycle. How the society would response to future streamflow and flood risk becomes to an important topic in both science research and policy making. This is especially true for the developing world, where many cities are experiencing subsidence due to unsustainable rates of ground water extraction. Subsidence accounted for up to 1/3 of 20th century relative sea level rise in China (Chen, 1991). Sea level has risen fastest in deltas and coastal plains around the coastline of the China Seas largely due to the local subsidence (Chen, 1991; Ren, 1993). Subsidence and sea level rise both increase the flooding risks. In particular, in densely populated regions with long experience of irrigation management, such as Southeast Asia and India, the reduced flood frequency under G4 stratospheric aerosol geoengineering might be further ameliorated.

The accurate assessment of human impacts on flood frequency and magnitude depends not only how human activities are represented in geoengineering scenarios, but also on how anthropogenic effects are parameterized in hydrological models (Masaki et al., 2017). Using the outputs from climate models to drive river routing models or hydrological models is a reasonable way to study how the streamflow and flood response under different climate changing scenarios. River routing models driven by runoff directly from GCMs and hydrological models considering human impacts both contribute to better our understanding of how the hydrological cycle would change under solar geoengineering.

References:

Cao, L., Duan, L., Bala, G. and Caldeira, K.: Simultaneous stabilization of global temperature and precipitation through cocktail geoengineering, Geophys. Res. Lett., 44(14), 7429–7437, doi:10.1002/2017GL074281, 2017.

Kravitz, B., MacMartin, D. G., Mills, M. J., Richter, J. H., Tilmes, S., Lamarque, J.-F., Tribbia, J. J. and Vitt, F.: First Simulations of Designing Stratospheric Sulfate Aerosol Geoengineering to Meet Multiple Simultaneous Climate Objectives, J. Geophys. Res. Atmos., 122(23), 12,616-12,634, doi:10.1002/2017JD026874, 2017.

MacMartin, D. G., Kravitz, B., Keith, D. W. and Jarvis, A.: Dynamics of the coupled human-climate system resulting from closed-loop control of solar geoengineering, Clim. Dyn., 43(1–2), 243–258, doi:10.1007/s00382-013-1822-9, 2014.

MacMartin, D. G., Kravitz, B., Tilmes, S., Richter, J. H., Mills, M. J., Lamarque, J.-F., Tribbia, J. J. and Vitt, F.: The climate response to stratospheric aerosol geoengineering can be tailored using multiple injection locations, J. Geophys. Res. Atmos., 1(November), doi:10.1002/2017JD026868, 2017.

Tilmes, S., Richter, J. H., Kravitz, B., MacMartin, D. G., Mills, M. J., Simpson, I. R., Glanville, A. S., Fasullo, J. T., Phillips, A. S., Lamarque, J.-F., Tribbia, J., Edwards, J., Mickelson, S. and Gosh, S.: CESM1(WACCM) Stratospheric Aerosol Geoengineering Large Ensemble (GLENS) Project, Bull. Am. Meteorol. Soc., BAMS-D-17-0267.1, doi:10.1175/BAMS-D-17-0267.1, 2018.

Yamazaki, D., Sato, T., Kanae, S., Hirabayashi, Y. and Bates, P. D.: Regional flood dynamics in a bifurcating mega delta simulated in a global river model, Geophys. Res. Lett., 41(9), 3127–3135, doi:10.1002/2014GL059744, 2014.