General Comments:

The paper estimates the shortwave (SW) radiative forcing at the surface from the injection of sulfur in to the stratosphere. Their method is applied to the GeoMIP G4 models where 5 Mt SO2 is injected annually in a transient RCP4.5 scenario. A single layer model for SW radiative transfer is used to calculate the forcing of the aerosol layer, rapid adjustments and reactions from changes in water vapor, cloud amount, and surface albedo. The simple model is a simplification but it allows to differentiate between different effects and provide useful information relevant to human activities and interests and the assumptions made are clearly stated. The paper will be a valuable addition to current literature on the subject after some minor revisions.

A number of scatter plots are included in the analysis where means over 3 decades are used. These single values should be split into decadal means such that you have three values to base the analysis on instead of one. This will not only provide more data points for the regression, but also indicate more clearly the steadiness of the differences in climates over this period of the simulations.

> Throughout the manuscript there is content that belongs only in the figure captions and not in the main text. E.g. "(shown by red symbols)" and similar is unnecessary in the main text; more interesting to read about the meaning of the results in the figures than color coding etc.

You should make some concluding remark on what your findings imply for human activities at the surface, as this is your initial motivation for the paper, and put your findings into a wider context.

Specific comments:

- The brief 'review' of solar constant experiments p. 2 lines 6-14 is not very informative nor important for this paper. I suggest removing it.
- P2, l19: 'arctic' should be 'Arctic'.
- P2 l17-26: you mention a number of sulfate aerosol papers, but you do not say how or why they or their findings are relevant or important. Either just list them and say they cannot be compared or find something relevant. Otherwise, it is merely "stuffing".
- P3 l18-32: you discuss factors explaining the spread in the climate response in G4. It would be interesting if you could also note down the spread in the RCP4,5 models (for the same models as in the G4 study referred to) for comparison. Is the spread larger for G4?
- P3 l34: 'behaivour' -> 'behaviour'.
- P3 l34 and P4 l4, l8: you refer to several, recent and many studies; please include some citations.
- P4 l11: radiation at the surface is also important for oceanic processes like biogeochemistry and ocean carbon cycle. This should be mentioned too. Human activities, like fisheries, might also be affected at sea.
- P5 l4: "In addition, the microphysics of the tropospheric sulphate aerosols is not calculated in MIROC-ESM-CHEM-5 AMP to avoid drift in the simulated climate." Why would this cause a drift? Please note a brief explanation in the text.
- P5 l6: The experiment has also been done with NorESM – and maybe IPSL and MPI-ESM(?). CSIRO-Mk3L has done 'G4S', which is not G4, hence no need to mention.
- P6, l12: 'cloud effects': do you mean feedbacks due to clouds?
- P7: on what time-scales are these assumptions valid?

- P7: E_c –cloud amount: this can also be output as a variable from the models. Do any of the models account for injected sulfate interactions with clouds? e.g. Kuebbeler et al. ?
- Figure 2: Why was year 2020 of RCP4.5 used as 'baseline'? Why not center the baseline on a 5 or 10 year period around 2020, to account for variability of the climate considering 2020 might be a particularly warm or cold year. It would be a cleaner comparison, particularly for BNU-ESM who has a lot of year-to-year variability.
- P8 l11-12: "For all models, T in G4 decreases or remains at the 2020 level for a few decades and begins increasing from around 2040 or earlier": This is a bit inaccurate. For models a-c, yes, but not really for the MIROC models.
- Figure 2 discussion; you're applying a fixed magnitude forcing every year, whilst the anthropogenic forcing in RCP4.5 keeps increasing. Hence the limitation to the cooling evolution you describe on page 8. Include some comment on this in the discussion of the results.
- Why is there hardly any cooling in MIROC-ESM-CHEM?
- P8 l 17: " … and then 25 returns to the RCP4.5 level in each model": HadGEM2-ES has not returned to RCP4.5 levels at the end of the run as this model has a larger temperature response to the forcing. You may include the comment that the stronger the temperature response to the forcing is, the longer it takes to return to the otherwise temperature path and the rate of change of the climate system to sudden termination would become more drastic.
- Figure 3: You find a correlation coefficient of 0.88. Please include the regression line in the figure. Also you select data from three decades as the temperature differences between G4 and RCP4.5 are steady over this period. I suggest you therefore use one value from each decade in every run in the figure. Then you have more data points and information.
- Figure 4: Fsrm varies by ~1 Wm-2 throughout G4 in HadGEM2-ES. Why is this so much more variable than the other model that accounts for formation and transportation of the aerosols (MIROC-ESM-CHEM-AMP)?
- Section 3.3: Considering the focus on the last three decades you should break the 3 decade mean into 1 mean for each decade, as mentioned before. The scatter plots have to be updated accordingly.
- "HadGEM2-ES does not calculate the sulphate aerosols in the tropospheric and stratosphere: separately" is mentioned several times. Please explain. Not clear what is meant by this at all.
- P10, l23-24: please say "… considered having performed the G4 simulation" – or similar at end of sentence. I.e. point to G4.
- P11, end of section 3.3: can you explain more clearly why the cloud amount is strongly dependent on the initial conditions? To my belief the clouds are even more so dependent on the cloud parameters. (perturbed initial conditions ensembles versus perturbed physical parameter ensembles.)
- Figure 7: I am not sure how much sense it makes to draw regression lines for Ec. There is clearly little correlation.
  Also; which models years are included in the figure? 2020 – 2069? Please make a note in caption.
- Figure 7 cont.: Have you tried to plot annual means for years 1-10 of the simulations and then decadal means for the remaining 4 decades? This might be more representative for gauging fast vs slow response.
- Section 3.5 Figure 8: good if you could remind readers at this stage if each model has been weighted by number of ensemble member in these figures.

- Figure 8 clearly indicates that something is going on with the South Pacific Convergence Zone. Detailed study of this is beyond the topic of your study; however, if this is discussed elsewhere in the literature, it would be great to point to it in the text.
- P13, l25: “However, the situation at TOA is also of interest.” Please say why.
- Section 4.2: doesn’t the LW heating from the aerosols also increase water vapor in the stratosphere, contributing to ozone losses? This aspect might be mentioned here briefly, as stratospheric O3 amounts do indeed impact human activities at the surface.