We would like to thank the reviewer very much for their feedback. In response to your comments we made the following changes to the manuscript: Added a discussion about the limitation of having a fixed terminus elevation and cited Marzeion et al 2014. A discussion on the negative Nash-Sutcliffe numbers and possible explanation for the model bias in the seasonal bass balance validation. We explored the impact of bias correcting the present-day annual mass balance, on future volume loss projections. We also re-ran the future simulations when the model is calibrated by minimising the bias and RMSE and maximising the correlation coefficient. Please see our detailed relies to your comments below.

Reviewer comment: The model presented here is a very timely and relevant contribution to the growing group of global glacier models since it is, to my knowledge, the only global model that has been developed in the framework of a land surface model. It thus follows a concept that is very different from most other glacier models, adding important diversity and opening a potential path to coupled modelling of glaciers (i.e., allowing interactions between glaciers and ocean/atmosphere, e.g. through the changed freshwater balance), as well as a more integrated perspective on hydrologic impacts of glacier mass
loss.
In principle, I therefore think the authors present a very valuable contribution. However, there are numerous results (particularly concerning calibration and validation) that require much more in-depth analysis and discussion than presented in the manuscript.
This concerns particularly (i) the substantial (and very consistently negative) bias; (ii) modelled negative winter and positive summer mass balances that are currently uncommented; (iii) the lack of a representation of the terminus altitude-mass balance feedback which might also contribute to stronger mass loss projections, and (iv) the Nash-Sutcliffe coefficients that are included in the tables, are mostly negative in the validation, but not at all explained or discussed in the text. If these issues – outlined below in more detail – are appropriately addressed, I would be able to recommend the manuscript for publication.

Specific comments:
Abstract and Introduction: I understand the desire to express the considered CMIP5 projections in terms of temperature above pre-industrial, which has become a quite common measure with the formulation of the goals in the Paris Agreement. However, most readers will be familiar with the RCP scenarios. I was wondering until page 10 whether a "mixed" scenario was used, e.g. selecting CMIP runs based on warming. A quick statement in the abstract and the introduction that RCP8.5 is used will make things a lot clearer.

Changes to manuscript: In the abstract we included: The CMIP5 models use the RCP8.5 climate change scenario and were selected on the criteria of passing 2°C global average warming during this century.

Changes to manuscript: In the introduction we included: "The CMIP5 models use the Representative Concentration Pathways (RCP) RCP8.5 climate change scenario for high greenhouse gas emissions. "

Reviewer comment: P2 L5: In some regions this is true, but since glacier water release is not causally related to demand for water, this sentence should be reformulated.

Reply: We removed the part "when demand for water is high"

Reviewer comment: P2 L7: Most of the authors are native speakers, which makes me doubt myself – but I learned that when used as a compound adjective, sea level should be hyphenated (as in "sea-level rise")?

Changes to manuscript: We have hyphenated instances of the adjective "sea-level rise"

Reviewer comment: P3 L14ff: It would be nice if the authors could comment on how strongly this limitation affects the usability of JULES. I.e., is it realistic that the new glacier scheme would be used in a default setup of JULES, or would the limitations inflicted on the other surface classes be too strong?

Reply: It is feasible to use the glacier scheme in the default configuration of JULES with a caveat. The model is set up so that glacier surfaces cannot share a gridbox with other surface types. This means that running the model with glaciers switched on, should not affect the other surface types. The caveat is this: in order to get sufficient accumulation in the mass balance profile, we had to lapse rate correct the precipitation. This is a standard procedure in mass balance modelling, but the consequence is that gridbox mean precipitation over glacier gridboxes is not conserved. We tested scaling the precipitation, while conserving the gridbox mean i.e. reducing the precipitation near the surface and increasing it at height, but this did not yield enough precipitation to get a good agreement the mass balance profiles. If you wanted to run JULES with the glacier model switched on, then it is worth bearing in mind that water will no longer be conserved. If the model is being used to simulate river discharge in glaciated catchments, then the precipitation lapse rate could be used as a parameter to calibrate the discharge.

Changes to manuscript: We have added this point to the model description section 2.2.3 in case the reader also has the same question.

P4 L5f: If I understand correctly, this implies that the negative feedback between terminus elevation and mass balance is missing, and that the only way for a melting glacier to reach equilibrium with climate is by melting completely (similar to Slangen & van de Wal, 2011, https://doi.org/10.5194/tc-5-673-2011). This is a major limitation that should be discussed in greater depth. E.g., Marzeion et al. (2014, https://doi.org/10.5194/tc-8-59-2014) find that depending on scenario and accumulated mass loss, this may contribute a few tens of mm SLE (their Figs. 9 to 11). The differences in Tab. S3 could be
plausibly explained by this alone.

I also think that the discussion of the lack of a parameterization of ice dynamics (end of Sec. 4.4) is flawed with respect to this feedback.

Reply: Yes, you understand correctly here. The negative feedback between terminus elevation and mass balance is missing and the only way for a melting glacier to reach equilibrium with climate is by melting completely.

Changes to manuscript: We have added the following text to the results section.

"Another explanation why our model predicts more volume loss than Radic et al (2014) and Huss and Hock (2015) is because there is no retreat of the glacier terminus represented in the model. The only way for glaciers to reach equilibrium with climate is by melting completely. A study by Marzeion et al (2014) showed that models predict more mass loss when the terminus elevation is fixed, than when it

is allowed to vary. This is because when the terminus retreats, the area available for melting is reduced, leading to less mass loss. Marzeion et al (2014) found that neglecting terminus elevation changes resulted in an extra few tens of mm SLE depending on RCP scenario."

Changes to manuscript: We removed the flawed section regarding the lack of ice dynamics (end of Sec. 4.4)

Reviewer comment: P4 L19ff: Units of the constants are partly missing or wrong.
Changes to manuscript: Units corrected and added.

Reviewer comment: P4 L28: Goff-Gratch (typo), and Landolt-Bordstein 1987 is incomplete in the references, and probably should be Landolt-Börnstein.
Changes to manuscript: Corrected

Reviewer comment: P5 L25: It would be better to say that the energy balance calculated in JULES includes the sensible heat flux, since the snow melt is not a direct (or separate) consequence of the sensible heat flux alone.
Changes to manuscript: Corrected –Instead we say "A component of the energy for melting the snowpack in JULES comes from the sensible heat flux"

Reviewer comment: Eq. 13: I think this equation is a bit problematic if justified by katabatic winds, since the katabatic winds should not be expected to be proportionate to the large-scale wind field.

Reply: We agree. Reviewer #2 also identified this as an unrealistic way to represent katabatic winds (Please see our response to this). We scaled the wind speed to increase the sensible heat flux compared to observations. Although our approach is not a physically realistic, we thought it was better to include a simple way to increase wind speed over glacier gridboxes, than excluding this.

Reviewer comment: P6 L25: "at the beginning" (typo)
P9 L 11: In Marzeion et al. (2012), it is 3 %/100 m.

Changes to manuscript: Corrected

Reviewer comment: Tab. 2: Since the Nash-Sutcliffe efficiency coefficient is included, it should be discussed in the text. It would be particularly good to address the reasons for the numerous negative values – are they caused by bias, too great/small variance, etc.?

Reply: We calculate the Nash-Sutcliffe efficiency coefficient equation using the following equation

$$NS = 1 - \frac{\sum_{i=1}^{n}\left(Y_i^{obs} - Y_i^{model}\right)}{\sum_{i=1}^{n}\left(Y_i^{obs} - Y_i^{mean\ obs}\right)}$$

where the numerator is the mean square error (or the bias) and the dominator is the variance. The negative numbers arise because the bias is greater than the variance of the observations. The negative bias indicates that melting is overestimated in the summer and accumulation is underestimated during the winter (Table 4).

Changes to manuscript: We added the following text in the model validation section 3.2 to explain the possible reasons for the bias.

"For all regions, except Scandinavia in the summer, negative Nash-Sutcliff numbers are calculated for winter and summer elevation-dependent mass balance (Table 4). The negative numbers arise because the bias in the model is larger than the variance of the observations. There are negative biases for nearly all regions implying that melting is overestimated in the summer and accumulation is underestimated in the winter.
Some, but not all, of the bias is due to the partitioning of rain and snow based on an air temperature threshold of 0°C. The 0°C threshold is likely too low, resulting in an underestimate of snowfall. When precipitation falls as rain or snow it adds liquid water or ice to the snowpack. The specific heat capacity of the snowpack is a function of the liquid water (W_k) and ice content (I_k) in each layer (k)

$$C_k = I_k C_{ice} + W_k C_{water} \qquad (17)$$

where C_{ice} = 2100 $JK^{-1}kg^{-1}$ and C_{water} = 4100 $JK^{-1}kg^{-1}$.
The liquid water content is limited by the available pore space in the snowpack, therefore changes in the ice content control the overall heat capacity. The underestimate in the ice content reduces the heat capacity which causes more melting than observed.
Other modelling studies have used higher air temperature thresholds; 1.5°C (Huss and Hock 2015, Giesen and Oerlemans 2012), 2°C (Hirabayashi et al 2010) and 3°C (Marzeion et al 2012). An improved approach would use the wet-bulb temperature to partition rain and snow which would include the effects of humidity on temperature. Alternatively, a spatially varying threshold based on precipitation observations could be used. Jennings et al (2018) showed by analysing precipitation observations, that the temperature threshold varies spatially. Jennings et al (2018) showed by analysing precipitation observations, that the temperature threshold varies spatially and generally higher for continental climates than maritime climates.

Increasing the temperature threshold only reduces the bias slightly, therefore another explanation is that the precipitation in the WFDEI data is too low. Although we have included the variation in precipitation with height, if the gridbox mean precipitation is too low then snowfall on the elevated tiles will be underestimated. We did not bias correct the precipitation before applying the lapse rate correction unlike other studies do (Marzeion et al. 2012, Huss and Hock 2015). The quality of the WFDEI precipitation maybe poor because the data is constrained by rain gauge observations which are sparse in high mountains regions and often biased towards low elevation levels. Even when observations are available snowfall at higher altitudes is often difficult to accurately measure and susceptible to undercatch by 20–50% (Rasmussen et al. 2012). The biases listed in Table 4 are larger in the summer than in the winter. It is likely that the simple albedo scheme, which relates albedo to the density of the snowpack surface, does not perform particularly well in the ablation zone. "

Reviewer comment: Tab. 3: Please add global mean values.
Changes to manuscript: Global values are added. Also, we noticed the number of observations listed in column 5 was incorrect, so we updated this.

Reviewer comment: P10 L1f: The tropical glaciers are really small; there are probably numerous more likely explanations for a warm bias than glaciers lacking in the model.

Figs. 3 and 4, and discussion around them (also P10 L14f): I'm not convinced the Pyrenean glacier is to blame for the low correlation. How much does the correlation change if you exclude it? I'm also wondering why the point cloud for Central Europe in Fig. 3 looks different from the one in Fig. 4?

Reply: There was an error in Figure 4 which we have corrected, and the point cloud now matches the data for Central Europe in Figure 3.

Changes to manuscript: We added the following "In Central Europe some of the poor correlation with observations is caused by the Maladeta glacier in the Pyrenees (Fig. 4) which is a small glacier with an area of 0.52 km^2 WGMS (2017). When this glacier is excluded from the analysis the correlation coefficient increases from 0.26 to 0.35 and the RMSE decreases from 1.99 to 1.73 meters of water equivalent per year."

Reviewer comment: Tab. 4: Please add global mean values.
Changes to manuscript: Values are added

Reviewer comment: Tab. 4: Again, it is necessary to discuss the negative NS-values. There is only one positive value in the table, which indicates that only for summer mass balance in Scandinavia, the model has better predictive skill than taking the mean of the observations. Also, given that all the biases are negative (with the exception of one that is close to zero), the implications for the projections need to be addressed. E.g., if the bias was compensated for in the projections: how would that change the results? Could the differences to previously published projections be explained by the global mean bias?

Reply: Please see above for a discussion on the negative NS-values.

We explored the impact of correcting the bias in the annual mean mass balance on the volume projections. The differences to previously published projections cannot be explained by the bias alone, but it does account for why we have larger volume losses in the Southern Andes, where the bias was particularly large.

Changes to manuscript: The text below has been added to section 4.5 (Comparison with other studies).

"We estimate the end of century global sea-level contribution, excluding Antarctic glaciers, to be 215 ± 20mm which is higher than 188mm (Radic et al. 2014) and 136±23mm (Huss and Hock 2015) caused mainly by greater contributions from Alaska, Southern Andes and the Russian Arctic. These three regions are discussed in turn.

For the Southern Andes our estimates are approximately double (14.4mm) that of the other studies (5.8mm (Huss and Hock 2015), 8.5mm (Radic et al. 2014)). This region has the largest negative bias in the calibrated present-day mass balance (-2.87 $m.w.eq.yr^{-1}$ see Table 3). To explore the effects of correcting the calibration bias on the ice volume projections, we subtract the bias values listed in Table 3 from the future annual mass balance rates. Each gridbox is assumed to have the same regional mass balance bias. The bias corrected volume losses are listed in Table S3 in the supplementary material. For the Southern Andes, the volume losses are much closer to the other studies (7.6mm) when the bias is corrected. The impact is less for the other regions where the biases are smaller. For the Russian Arctic our volume losses are higher than the other studies but that should be interpreted with caution because there were no observations available in this region to get a tuned parameter set (global mean *parameters where used instead). In Alaska the bias in annual mass balance is small (0.06 m.w.eq.yr^{-1}) so correcting the bias has little effect on the volume loss projection for this region. Applying the bias correction increases the global volume loss from 215 ± 20mm to 222.5±20.1mm, therefore the difference between our model and the other studies cannot be explained by the bias in the calibration."*

Reviewer comment: Fig. 5: the RMSE mentioned in the caption is missing in the panels.

Changes to manuscript: The caption has been modified so it no longer says the RMSE error values are shown on the plots. The RMSE values are listed in Table 4.

Reviewer comment: Fig. 6: It is surprising that the model produces many substantially negative winter mass balances (and not quite as many positive summer mass balances). This behaviour should be looked into and discussed in the text.

Reply: This suggests there is a warm bias in the winter but no equivalent cold bias in the summer. We think the presence of a winter warm bias in the model should not necessarily mean there is an equivalent cold bias in the summer.

Reviewer comment: Caption of Fig. 6: "number of glaciers" – isn't that the number of grid boxes?

Reply: This is the number of glaciers for which there are observations.

Reviewer comment:
P10 L23: "sensitivities" (typo)
P11 L12: "Arctic" (typo)

Changes to manuscript: Corrected

Reviewer comment:
Fig S7: It is hard to see anything here; perhaps just leave out the East African and Indonesian glaciers (which is sad for sentimental reasons, but I think they are mostly irrelevant for sea level and water availability).
Changes to manuscript: The figure is modified to only show South and Central America.

Reviewer comment: P11 L16:
Changes to manuscript: Figs. S1 to S7.

Reviewer comment: Sec. 4.3: It is good to see that the model results appear to be robust; on the other hand, this may indicate the negative bias may be hard to overcome. In the calibration, minimizing the RMSE was used for identifying the best parameter set(s). Another way of looking into "parametric" uncertainty (in a wider sense) would be to minimize the bias, or to maximize the correlation or the NS coefficient. These experiments might give valuable insights into the causes of the sometimes problematic model performance, which needs to be better explained.

Reply: We investigated the effect of using multiple performance metrics on the calibrated parameters and glacier volume projections

"Another way to explore the uncertainty in the volume projections caused the calibration procedure, is to use different performance metrics to identify best parameters sets. In addition to using RMSE, we calculate best parameter sets by (1) minimising the absolute value of the bias and (2) maximizing the correlation coefficient. The best regional parameter sets are different depending on the choice of performance metric used (See Tables S2 and S3 in the Supplementary material). For twelve regions, minimising the bias results in higher precipitation lapse rates, than when RMSE values are used to select parameters. This suggests the bias in many regions is caused by underestimating the precipitation lapse rates. As discussed above, this could be due to the fact the gridbox mean WFDEI precipitation was not bias corrected. Glacier volume projections are generated by repeating the simulations using these two additional performance metrics to identify best parameter sets. The uncertainty in the global volume loss when the extra performance metrics are used, is approximately double the uncertainty arising from the different climate forcings (Fig. 16, Table 7). When extra performance metrics are used, the upper bound volume loss increases to 281.1 mm sea-level equivalent by the end of the century."

Changes to manuscript: We changed Figure 16 to show the large spread in the global volume loss when extra performance metrics are used in the calibration.

Extra columns are added to Table 7 to list the regional volume losses when we minimise the bias and RMSE and maximise the correlation coefficient.

Tables are added to the Supplementary Information listing the best parameter selected by minimising the bias (Table S2) and maximining the correlation coefficient (Table S3).

Reviewer comment:

P13 L2: "equivalent" (typo)
P14 L27: Reduction in glacier mass, not necessarily in mass balance.
P15 L18: "periphery" (typo)

Changes to manuscript: Typos corrected

Giesen, R. H. & J. Oerlemans (2012) Calibration of a surface mass balance model for global-scale applications. *Cryosphere,* 6**,** 1463-1481.

Hirabayashi, Y., P. Doll & S. Kanae (2010) Global-scale modeling of glacier mass balances for water resources assessments: Glacier mass changes between 1948 and 2006. *Journal of Hydrology,* 390**,** 245-256.

Huss, M. & R. Hock (2015) A new model for global glacier change and sea-level rise. *Frontiers in Earth Science,* 3.

Jennings, K. S., T. S. Winchell, B. Livneh & N. P. Molotch (2018) Spatial variation of the rain–snow temperature threshold across the Northern Hemisphere. *Nature Communications,* 9.

Marzeion, B., A. H. Jarosch & M. Hofer (2012) Past and future sea-level change from the surface mass balance of glaciers. *The Cryosphere Discuss.,* 6**,** 3177-3241.