Note: reviewer comments are in italics and the authors' responses and manuscript revisions are in normal face.

Comment: *This paper presents how Global Sensitivity Analysis (GSA) can be used to discuss snowpack model sensitivity and to identify factors affecting model results in terms of snow depth and surface energy balance (SEB). The authors apply the GSA method to punctual snowpack simulations carried out a high Arctic glacier (Kongsvegen Glacier). The topic of this paper is important for scientists involved in snowpack modelling since it proposes a technique to quantify model sensitivity including interactions between parameters. This method can be extended to other geophysical models. The results are really interesting and should be published in GMD. Prior to publication, major revisions should be made to better illustrate the reliability of the results obtained with the GSA. The description of the reference simulation should also be improved. They are listed below (General comments) followed by more specific and technical comments.*

Thank you very much for reviewing this paper.

Comment: *1. The conclusions of the authors in terms of model's sensitivity rely on the parameters uncertainties and their distribution used to generate the ensemble and described in Table 2. The authors should evaluate if their ensemble represents correctly the model uncertainty. This could be achieved comparing the ensemble dispersion (the standard deviation of the members relatively to their average) to the model RMSE (computed using observations and the ensemble average). The error can be computed for snow depth but also for other measured parameters such as albedo and surface temperature. The ensemble dispersion is expected to be of the same magnitude as the error. If the ensemble dispersion is too large compare to the error, this would mean that one or several error distributions are too large and not appropriate.*

The aim of this study is to estimate the model uncertainty based on the accuracy given by the manufactures. In real life applications it is not possible to reduce the uncertainty more than the specified accuracy. Hence, the analysis can be seen as a conservative estimation. The idea is to include all measurement errors and check whether a forcing factor is sensitive or not. We agree that modellers need to check for overdisperison when doing ensemble predictions. This is not the intention of this study since we are rather interested in the complete uncertainty range. We have carefully reworked the text and point out in the introduction (p4L15-p4L17) and the discussion sections (p29L11-p29L17) that the study focuses on systematic biases (given by the sensor's accuracy) in the forcing data.

This comparison requires having a reliable estimation of the model error. One year of simulation at KNG8 may be not sufficient. Therefore, the authors should consider extending their analysis to other years. If the data are not available at KNG8, a good alternative would be to use meteorological data available at KNG6 (Karner et al, 2013). The differences between their conclusion and those of Karner et al (2013) are mentioned several times in the paper (P 2826 l 2-3, P2828 l 3-5). Applying the GSA method at KNG6 would also allow the authors to discuss more in details these differences. Extending the GSA method to other years will probably require reducing the number of members of the ensemble. The authors could follow a two-step approach. They could first present the results of the analysis of the original version of the paper (one year at KNG8 with 16000 members) and then extend the analysis to other years at KNG8 or KNG6 using an ensemble with a restricted number of members.

A longer period would certainly be of benefit. Unfortunately, reliable data is not available for a longer period at KNG8. However, we have followed the suggestion of the reviewer and analysed the sensitivity pattern at station KNG1 located in the ablation zone of the Kongsvegen glacier. In contrast to KNG6, this station has a more distinct characteristic from KNG8. For both station the sensitivity indices have been estimated from 20000 ensemble members. The number of ensemble members has been increased in order to better estimate the accuracy of the indices by bootstrap sampling. We discuss in detail the differences of the sensitivity pattern in the Result and Discussion section.

Comment: *2. The authors include the aerodynamic roughness length, z0, in the sensitivity analysis (Tab. 2). As mentioned in Tab 1, this corresponds to the roughness length for momentum. In land surface models, the roughness length for heat exchanges z0H is generally deduced from z0 following z0H=z0/K. K is a constant equal to 10 by default in SURFEX (Mascart et al, 1995). This is the case in the reference run (Tab. 1). The authors should explicitly mention if z0H is modified or not when generating the ensemble members using the uniform distribution for z0. If not, it probably reduces the mean sensitivity measures for z0 in the GSA. In this case, it would be really interesting to account for the dependency of z0H on z0 in the GSA. Note that this general comment may be not relevant if this dependency is already taken into account in the numerical experiments described in the paper. In any case, z0H should be mentioned in Tab. 2 with the type of distribution and the range of values.*

The z0H is modified when generating the ensemble members using the uniform distribution. However, the factor K is equal for each member. We have added the following sentence to the penultimate paragraph in Section 2.5 (p14L21-p14L22): "The roughness length for heat zh0 is derived from the roughness length for momentum using the relation zh0 = z0/10.

Comment: *3. Section 3.1 contains the description of the results of the reference simulation. For this simulation, the authors do not consider using a spin up to generate the initial profile of snowpack properties (especially temperature) contrary to the method followed by Karner et al (2013). Are meteorological data from previous years available at KNG8 and could they be used to improve the initial snowpack following the same method as Karner et al (2013)? What would be the impact on the simulations, at least for the reference run?*

The vertical temperature profile in the snowpack is directly measured by a chain of temperature sensor at KNG8. However, the exact depth of the sensors could not be reconstructed for the beginning of the simulation period. The simulations start at the end of the ablation period and the snowpack temperature is close to the melting point. We have performed several simulations with various initial boundary conditions, but the effect of the initial temperature profile on the simulation results was very small.

Comment: *4. The description of the results (P2819 l 25 to P 2820 l 15) is rather short and could be more clearly identified using separated paragraphs for example. Then, the authors should discussed more in details the physical processes behind the model results. For example, the difference of snow temperature close to the surface may also be due to higher snowpack thermal conductivity because of the higher snowpack density close to the surface. On contrary, the sentence P 2820 l10-11 suggesting a "surprisingly" direct link between surface albedo and snow surface density should be explained.*

We have carefully restructured the paper and better described and discuss the results. The first paragraph of the Section "Reference run" and "Uncertainty estimation" have been moved to the methods section as suggested by the reviewers. A new chapter "Reference run setup" has been added describing the initial and boundary conditions of the reference run. Furthermore, the results section has been restructured and consists now of four subsection: (I) Reference run, (II) Integrated model uncertainty and (III) Mean total-order sensitivity indices and (IV) Temporal evolution of the total-order sensitivity indices. In the Reference Run section we examine the accuracy of the reference runs in more detail (Note, we have included a second station, KNG1, for comparison). In this revised version we focus on the total-order indices rather than on the first-order indices. The mean values and the evolution of the indices are described in the last two sections of the results chapter. The discussion has been rewritten and should not infiltrate the discussion anymore..

Specific Comments:

Title The title should better reflect the content of the article. In particular, it should be mentioned that the GSA method was applied at one high Arctic site, which somewhat reduces the generality of the title. Abstract The Abstract is too vague and does not allow the reader to extract the method followed by the authors and the main findings of this study. For example, it should clearly mention where the study has been carried out (the current version only mentions "in a high Arctic environment") and which factors affect the results. It would *also be useful to state the name of the snowpack model in the abstract.*

We followed the suggestion of the reviewer and changed the title to: "Assessing the uncertainty of glacier mass balance simulations in the European Arctic based on variance decomposition". We also modified the abstract and mentioned where study has been carried out and which factors contribute to the model variance. Furthermore, we have mentioned the snowpack scheme Crocus.

P2813, l 5, the sentence "measured input except of outgoing infrared" is not clear. Indeed, when using a detailed snowpack model such as Crocus, incoming shortwave and longwave radiations are provided as atmospheric forcing. They can be measured values such as in this study. However outgoing shortwave radiation is not a measured value suggested by this sentence

We have deleted this part of the sentence.

P2813, l 12, the terms P and M of Eq. 3 should be defined. In Armstrong and Brun (2008), the term E of Eq 3 has a different definition than the one used in this study. In Armstrong and Brun (2008), E is defined as the sublimation and evaporation rate at the surface.

Thanks for this hint. Indeed, E denotes the sum of sublimation and evaporation rate and the annotation of Eqn. 3 was correspondingly corrected.

P 2814, l 12-13, the term "roughness length for fresh snow" is not precise enough. Please specify which roughness length (see general comment 2).

We have replaced this expression by "roughness length for momentum ...". In the Section GSA (penultima paragraph) we have specified how the roughness length for heat is derived (p14L21-p14L22).

P 2815, l 20, adaptations of Crocus's parameterisation for fresh snow density have been already proposed in polar environment (Dang et al, 1997, Libois et al, 2014) and could be compared to the modification proposed by the authors.

We have included the work of Libous et al, 2014 in this section (p9L27).

P 2820, l 16-29, precise the time period considered as the "ablation season" and the "accumulation season".

Ablation and accumulation periods are of different length at KNG1 and KNG8. To enable comparison of the GSA analysis we now consider a central summer (JJA) and a winter period (DJF), respectively.

P 2821, l 1-2, Karner et al. (2013) give the 10-year average surplus as KNG6. It would be interesting to give the inter-annual range of surplus so that the reader can better realize the differences between KNG6 and KNG8.

The present study is based on simulations of a one year period. Therefore, a straightforward comparison of inter-annual ranges of the energy balance calculated at KNG6 and KNG8 is not possible. Please note, however, that due to now considering an additional site (KNG1), Table 1 and related discussion allowed for extended consideration of the issue. Comparison is possible for the ablation period which is commonly defined as JJA. The main characteristic are that net radiation decreases with elevation along the glacier (KNG1/KNG6/KNG8: 83/38/13), sensible heat flux decreases (KNG1/KNG6/KNG8: 21/8/3) as well as latent heat flux (KNG1/KNG6/KNG8: 21/-2/3). But even that comparison is hampered by comparing a 1 year period (KNG1 and KNG8) with decadal averages (KNG6), though conforming to the expectations regarding the elevation gradients along glaciers (e.g. Oerlemans J., Björnsson H., Kuhn M., Obleitner F., Palsson F., Smeets P., Vugts H.F. and de Wolde J., 1999: A glaciometeorological experiment on Vatnajökull, Iceland, Boundary-Layer Meteorol., 92, 1, 3-26.)

P2821 l 25, P 2822, l 19-20, it would be useful if the authors could precise how the ensemble members are generated. For a given member, are the perturbations of the parameters fixed for the whole duration for the simulation (for example, RH is always increased by 2.

The Section “Global Sensitivity Analysis” has been extended and two new sections “Measurement error characteristics” and “Sobol sampling” have been added. At the end of the Section 2.4 (p14L14-p15L9), we have added a new paragraph describing the calculation of the sensitivity indices, sampling strategy and how the accuracy of the indices is determined.

P2827, l 25, the impact of PVOL on Crocus simulations has been previously discussed in Gascon et al. (2014) and could be mentioned in this discussion part.
Conclusion The conclusion is written as a single paragraph and would be more clear if the authors could divide it into several paragraphs. Table 1: The definitions of ALB0.3, ALB0.8 and ALB1.5 are erroneous. These 3 parameters refers to the spectral albedo for surface ice for 3 spectral bands and not to the absorption coefficient for 3 spectral bands. These parameters are not used to compute snowpack albedo (see Vionnet et al (2012) for more details concerning the computation of snowpack albedo in Crocus). They are only used to compute albedo when the snow density is above the ice density threshold (830 kg m-3 in this study).

Thank you very much for this reference. We have referenced the this article in Section “Crocus model setup” (p8L3) and in the discussion part (p29L4).

We have completely rewritten and shortened the conclusion section so that no paragraphs are necessary.

The definitions of the spectral albedo for surface ice has been corrected.

P2810, l 19, (and after), the authors use sometimes “snowpack model” or “snow model”. Please be consistent. I personally recommend “snowpack model” to avoid confusions with the modelling of snow in the atmosphere as a microphysical specie.

We follow the recommendation of the reviewer and use the term snowpack model throughout the text.

P2810, l 19, (and after), use Crocus instead of CROCUS.

We have changed the expression.

P2812, l 4, remove “of” between “the” and “snow grains”

We have removed the “of”.

P2816, l 10, use “snow depth variations smaller than...” rather than “ snow depth smaller than ...”

Changed accordingly.

P2817, l 7, define SD

Changed accordingly.

P2818, l 6, the sum goes from “i to k” rather than from “i to r”

Changed accordingly.

P 2819, l 20, “shortwave radiation” is not used to evaluate the reference simulation

Thats true, we have removed “shortwave radiation”.

P2819, l 23, the date of the snow pit profile is 6 April “2011” instead of “2010”.

Changed accordingly.

P2820, l 4-5, Fig. 3 shows the vertical profile of temperature and not the “vertical temperature gradient”.

Changed accordingly.

P 2820, l 18 (and after), energy fluxes are in W m-2

The unit has been changed throughout the text.

P2181, l1, close the bracket after Fig. 1

Changed accordingly.

Table 2: remove “%” after the range of values for PVOL (0.03-0.05).

The “%” has been removed.

Note: reviewer comments are in italics and the authors' responses and manuscript revisions are in normal face.

This paper tackles an interesting topic that is seldom properly covered and is based on a robust methodology. This paper has the potential to greatly benefit the community once its shortcomings would be addressed. Such shortcoming include the following points:

Thank you very much for reviewing this paper and the helpful comments.

Comments:

The paper should be reworked to ensure better clarity (including some parameter definitions that are missing). Quite a few things should be rephrased for clarity and/or grammar and some sections are badly structured (see in the detailed comments).

***Comment**: 1. The reference scenario should be properly shown: very few information are given. A graph giving an overview of the forcing together with the temporal evolution of the snow height could really help the reader to make up his mind about this scenario. A graph showing the evolution of the energy balance components could also prove very useful when linked with the impact of the uncertainty on various parameters.*

The reference runs of KNG1 and KNG8 are shown in Figure 3 together with the confidence intervals. We added a new Table 3 with the mean and standard deviation of the meteorological variables and energy balance components for the two stations (KNG8 and KNG1) considered. We believe the table is informative than a graph showing the temporal evolution of the components.

***Comment**: 2. The graphs showing the uncertainty provide some kind of a worst case scenario (multiple parameters combining their worst case values). This is very interesting but also very surprising at first because the amplitudes of the effects of such uncertainties are beyond common expectations and experience. I even set up a similar simulation with the SNOWPACK model in order to check the numbers because this seemed so surprising compared to regular simulations (and finally SNOWPACK shows very similar results to the CROCUS results shown here). To my understanding, even simulations with very poor datasets tend to fare better than the worst case combinations as presented here because some errors compensate each other (for example the Undercatch would be compensated by Incoming Long Wave parametrizations leaning toward clear skies). I think the surprisingly large amplitude of the uncertainty of the results should be better explained/demonstrated and potentially compared to real life data sets. For example, a graph showing the min/max/avg snow height development when only one parameter is changed; or showing how a few low quality datasets would compare to the findings presented here (although this would involve quite some work and would be based on other locations where both low quality and high quality data are available). An alternative approach would be to synthetically generate degraded parameters out of your data set mimicking the data quality issues of real, low quality sites. If these suggestions are impractical, in any case the authors should consider how they could bridge the gap between the common perception of model users (even when dealing with low quality data sets) and their findings*

The uncertainty estimation is based on the uncertainty given by the manufacture's specified accuracy. In real life applications it is not possible to reduce the uncertainty more than the specified accuracy. Hence, the analysis can be seen as a conservative estimation. The compensating effect is still present since the data is only systematically perturbed by small biases. In fact, a synthetically generated dataset with the same characteristics as the measurements could be used for the analysis instead. We have carefully reworked the text and point out in the introduction and the discussion sections that the study focuses on systematic biases in the forcing data, e.g " … identifies how systematic measurement errors (biases) and uncertainties of some critical factors influence our confidence in glacier mass balance simulations." (p4L15-p4L17). We analyse in detail the interaction of variables in the Section 4 and relate the sensitivity pattern to physical processes (p22L19-p22L20).

The authors did not mention if (or how much) the CROCUS team was involved. Since the new snow density was tweaked to better fit the results, one is left to wonder if there was any discussion with the CROCUS authors on this topic (although this matches a similar value for Arctic conditions in the SNOWPACK model).

We are closely cooperating with the CROCUS developer team and discussed the modifications. In particular, the modification of the water flow/refreezing module and the snow density calculations are the results of the close collaboration. We have mentioned the collaboration in the acknowledgements.

The authors emphasize the effects of the interactions although these only represent 7% of the ensemble spread... Doesn't this mean that first-order, linear effects are by far dominants (pages 2824 and 2828)?

This was a serious mistake. The number refers to the maximum contribution of linear effects on the model variance (SHC) and not on the time averaged value. We have now given the averaged first-order indices for each target metrics (see Figure 4). The average values first-order indices are significantly smaller than the 93% given in the text and vary between 0.69 and 0.82.
In the discussion section we (p22L7-p22L12) we write: "The overall results of this work show that on average about 80% of the total variance of SHC and SEB can be explained by first-order effects (Fig. 4). This means that the remaining 20% of the variance is due to non-linear interaction effects. There is no significant difference between the two sites at the glacier. This is in partial contrast to the findings of Raleigh et al. (2015), who performed similar investigations for different snow regimes and found that first- and total-order indices are of comparable magnitude."

The figure 8 is very interesting and therefore should be better explained and emphasized in the text, the last paragraph of section 3.3 should be expanded.

In contrast to the old version of this paper, the Figures 5 and 6 show now the total-order effects and not the first-order effects anymore. Since the GSA permits to recover the complete variance and not only first-order effect, we believe these indices are of more interest. At the end of Section 2.4 a new paragraph has been added describing in more detail how the indices and confidence intervals are estimated (p14L15-p15L9). Additionally, we have added a new Section 3.4 "Temporal evolution of the total-order sensitivity indices" describing the temporal variability of the total-order indices shown in Figure 5 and 6.

I would suggest writing equations (1), (2) and (3) in a more consistent way, making sure all parameters are described properly and maybe considering basing them on a positive energy change instead of negative.

We have reworked the entire chapter and believe it is more consistent now.

One single year of validation data is a little short. Would it be possible to expand the reference period?

A longer period would certainly be of benefit. Unfortunately, reliable data is not available for a longer period. As an alternative we used the data from a second station, KNG1, located in the ablation zone and discuss the differences (see Reviewer 1). However, a one year period is sufficient to estimate the sensitivity pattern of the snowpack model.

Detailed comments: *most of these comments relate to sentences that are not very well written and should be rephrased in a more natural manner.*

the title should be improved to mention the Arctic conditions

We have changed the title to "Assessing the uncertainty of glacier mass balance simulations in the

European Arctic based on variance decomposition"

page 2809, rephrase lines 12 ("through to detailed"), 13 ("Besides all advantages"), 19 , 26-27, 29 (replace "apportioned" by "distributed" or something similar)

We have changed “through to detailed” by “to detailed”, “Besides all advantages” to “Due to the increasing complexity of detailed models ...”, and replaced “apportioned” by “assigned”.

page 2810, rephrase line 5

The phrase now reads as “To achieve a full understanding of the sensitivity pattern of highly interconnected and nonlinear models, ...”

page 2813, rephrase lines 5 ("measured input except of outgoing infrared") and 13-15

We have deleted this passage, since the input data is specified in the subsection “Model input/output”.

page 2814, the glacier flows north-westwards!

The passage has been changed to “… the glacier flows towards the north-west coast of the archipelago.”.

page 2815, rephrase line 4-5

The text has been changed to “When the surrounding stations had missing values, the values were estimated by a stochastic nearest-neighbour resampling conditioned on the remaining variables (Beersma and Buishand, 2003).” (p9L13-ppL15).

page 2816, line 16, remove the word "changes"

The word “changes” has been removed.

page 2819, the same things are said twice in the same paragraph

The paper has been carefully restructured as proposed by Raleigh. The first paragraph of the Section “Reference run” and “Uncertainty estimation” have been moved to the methods section. A new chapter “Reference run setup” has been added describing the initial and boundary conditions of the reference run. We believe the paper is better structured now and makes a clear distinction between methods and results.

page 2819, replace "starts at ..." by "starts on ..." and similarly for "ends at …"

The passages now read as: “At both sites, the simulations start at the end of the ab- lation season, with the lowest recorded snow depth (defined by the minimum recorded surface height), and they are forced by hourly measurements.” (p11L20-p11L22).

page 2819, line 20, consider specifying that the measurements are hourly?

The text reads now as “The Crocus model is forced by air temperature (T), relative humidity (RH), wind speed (U), incoming shortwave radiation (SW), incoming longwave radiation (LW), precipitation rate (P) and atmospheric pressure (see Sect. 2.2). These time-dependent parameters were measured at both sites and are provided to the model by Netcdf-file for hourly time steps.” (p8L17-p8L24).

page 2819, line 21 "these data", which ones?

We have deleted this sentence.

page 2819, rephrase lines 24-25

The sentence has been changed to “In terms of water equivalent, the accumulated mass during the winter amounts to +0.76 m, compared to +0.82 m having been observed.” (p16L18-p16L19).

page 2821, line 1, a ")" is missing

The “)” has been added.

page 2821, rephrase line 4

The passage was rephrased to “The overall impact of individual error sources on the sensitivity pattern varies for different zones on the glacier.” (p29L17-p29L19).

page 2821, rephrase line 25

We have rewritten this chapter and the sentence does not exist anymore.

page 2822, rephrase lines 15-16, 19-20 and fix spelling on line 25

We have rewritten this chapter and the sentence does not exist anymore.

page 2824, line 15 replace "proof" by "prove"

Has been changed accordingly.

page 2825, rephrase line 26

We have rewritten this chapter and the sentence does not exist anymore.

page 2826, rephrase lines 7-8

The lines has been changed to “Hence, the sensitivity of net 5 shortwave radiation (∂G) to measurement errors (∂ESW) is given by $\partial G/\partial ESW = 1 - \alpha$, with α denoting albedo.” (p28L4-p28L6).

page 2828, line 20 replace "firstly" by "first" and line 21 "secondly" by "second

Have been changed accordingly.

page 2829, line 1 replace "proofed" by "proved" and line 2 replace "provides" by "provided". Remove "by" on line 3

Have been changed.

page 2829, rephrase lines 4, 7, 10, 22 and 25-26

We have rewritten this chapter and the sentence does not exist anymore.*"*

page 2829, line 13 "considerably" and "This lower proportion"

We have rewritten this chapter and the sentence does not exist anymore.

page 2829, please define Q on line 20

Q has been changed to RH throughout the text and is defined in Section 2.1 (p8L18).

page 2830, rephrase line 3-4

Changed.

page 2838, fig. 1, rephrase "Map demonstrating"

Changed to: "A map showing the location of the Kongsvegen glacier and the position of the automatic weather stations KNG8, KNG6 and KNG1 (Norwegian Polar Institute, 2014).*" (p40).*

page 2841, fig 4, "snow albedo at the KNG8 location"

Changed.

page 2842, fig 5, idem

The caption now reads as: "Spread of the ensemble simulation at KNG8 (upper panel) and KNG1 (lower panel) due to propagating uncertainties in the model inputs. The black lines represent the reference run. The intervals show the 99, 95 and 75% quantiles estimated from the quasi-random Monte-Carlo runs (20000 ensemble members). Note the different horizontal and vertical scales.*"*

page 2844, fig 7, please define the parameters. I also don’t find this graph very clear, improvements would be welcomed

The figure has been replaced by a new figure. We hope the figure is more clear now.

Note: reviewer comments are in italics and the authors' responses and manuscript revisions are in normal face.

Comment: *The results demonstrate that large uncertainties in modeled snow depth can emerge from relatively conservative input uncertainties. Longwave radiation uncertainty had a strong control on all four model outputs while precipitation uncertainty had only a substantial control on snow depth uncertainty. The other factors were typically less important (with sensitivity indices usually less than 0.25), although during episodic wind storms, uncertainties in wind speed and aerodynamic roughness increased in importance. The authors do not generalize their results for other locations but suggest their approach can be applied in other locations and for other models. I think the study demonstrates the value of variance decomposition sensitivity analysis for understanding how input uncertainty matters when modeling cryospheric processes at a high latitude glacier. A growing number of studies in hydrology and earth sciences are applying variance based sensitivity methods to better understand model behavior, and thus this topic is timely and relevant. Despite this potential, I think there are a number of areas where the paper needs to be improved before being considered for publication, and so I recommend that the authors consider these suggestions for strengthening their contribution.*

> Thank you very much for your detailed comments and constructive criticism of the original manuscript. We have carefully revised the paper and considered your suggestions to improve our contribution.

Comment: *1. First, I would like to gently make the authors aware that my colleagues and I have recently presented a very similar framework for assessing the impact of forcing uncertainty on modeled snow variables with variance based global sensitivity analysis. I refer the authors to this paper (Raleigh et al., 2014) and include the citation at the end of my review. Our contributions are different in that mine focuses more on how specific error characteristics (i.e., error types, probability distributions, and magnitudes) in the model forcings matter to the outputs, and I examine different sites and a different snow model. However, I note there are some similarities between our studies in terms of the experimental setup (e.g., I also consider a scenario of specified measurement uncertainty) and some results (e.g., the importance of longwave uncertainty). In any case, I suggest that it might be appropriate to consider the connections between the results of our independent experiments in your discussion section.*

> We have to admit that we were not aware of this paper at the time when we came up with the draft. We are really sorry that we have overlooked this paper. Indeed this is a very interesting work and take up for the very first time this relevant topic. We think your study and our work complement each other very well. It is exciting and encouraging to see that some results are in line with our findings. Whenever possible we refer to your paper and contrast findings in the results and discussion section.

Comment: *2. There are numerous grammatical problems and awkwardly phrased sentences throughout the manuscript. These were distracting, though I was usually able to determine what the authors meant. I recorded many of these issues in the "Technical Corrections" section (see below), but I certainly did not catch all problems. The paper would thus strongly benefit from a more thorough grammar and language review to ensure the English usage is correct and clear throughout the manuscript.*

> The manuscript has been proof-read by a professional editor. We hope all the grammatical problems awkwardly phrased sentences have been corrected, now.

Comment: *3. The organization of the paper needs more attention. Specifically, the results section is actually a mix of methods, results, and discussion and would therefore benefit from careful restructuring. As an example of these elements, you can see aspects of methods (e.g., page 2821, lines 6-27) and discussion (e.g., page 2820, line 2; page 2820, lines 7-8) infiltrating into the results section. There needs to be a more clear division between sections to present a more logical exposition of the analysis.*

> We have carefully restructured the paper. The first paragraph of the Sections "Reference run setups" and "Uncertainty estimation" have been moved to the methods section as suggested by the

reviewer. The "Reference run setups" chapter describes the initial and boundary conditions of the reference run. Furthermore, the results section has been restructured and consists now of four subsection: (I) Reference run, (II) Integrated model uncertainty and (III) Mean total-order sensitivity indices and (IV) Temporal evolution of the total-order sensitivity indices. In the Reference Run section we examine the accuracy of the reference runs in more detail (Note, we have included a second station, KNG1, for comparison). In this revised version we focus on the total-order indices rather than on the first-order indices. The mean values and the evolution of the indices are described in the last two sections of the results chapter. The discussion has been rewritten and should not infiltrate the discussion anymore.

Comment: *4. The global sensitivity analysis (section 2.3) needs to be described in more detail. While the conceptual equations are provided (equations 6-9), it is not clear how the variances are actually calculated for the first-order and total-order sensitivity indices, and whether any* ***bootstrapping*** *was conducted to assess the confidence in the indices. Because there are several methods available for the variance calculation (see Saltelli et al., 2010), this needs to be clarified. Also, it would be helpful to include more detail in section 2.3 about how the* ***sampling*** *was done, how the* ***errors were selected and assigned****, and i****nformation about the convergence rates*** *of the sensitivity indices.*

The Section "Global Sensitivity Analysis" has been extended accordingly. At the end of the Section we have added a new paragraph (p14L1 – p15L9) describing the calculation of the sensitivity indices, sampling strategy and how the accuracy of the indices is determined. We have performed new simulations for both stations (KNG1 and KNG8) with N=2000 base samples. This results in 20000 model runs for each station. The accuracy of the sensitivity indices was assessed from 1000 empirical bootstrap samples. Similar to Raleigh et al. (2015) we estimated the 95% confidence regions for each total-order index (see Figure 4).

Comment: *5. It is not clear why the authors considered a mix of meteorological forcings and just two model parameters in their uncertainty/sensitivity analysis. Why were only the aerodynamic roughness and maximum liquid water hold capacity parameters considered, and not the other parameters in Table 1?*

The user provides the meteorological input data as a NetCDF file. Relevant model parameters are given by a separate option file. This file contains parameter such as roughness length, pore volume, initial temperature profile, land surface categories, initial grain types, and a few more internal snow parameters. The roughness length and pore volume are the only two parameters in this file which are relevant for the calculation of the surface fluxes. We had to change the code of Crocus so that these parameters can be provided by the NetCDF file in order to perform the Monte-Carlo runs. Other relevant parameters are defined within the model code and it turned out to be difficult to include them in the input file. Furthermore, including more parameters would have gone beyond the scope of the computational possibilities. We agree that more parameters need to be considered for a more complete variance analysis.

Specific Comments:

- *Page 2812, Lines 4-5: What are these "indicators of snow grain history" exactly?*

We have specified the indicator of snow grain history in the paragraph on microstructure (p6L1-p6L9): "Snow layers are described through bulk physical properties (thickness, density, temperature, liquid water content) and microstructure parameters. The latter characterize the state of the snow crystals in terms of dendricity, sphericity, grain size and snow grain history. The parameter of snow grain history indicates whether there once was liquid water or faceted crystals in the layer (Brun et al., 1992; Vionnet et al., 2012). The changes in the morphological shape of snow crystals depends on snow metamorphism in response to atmospheric forcing and internal processes. To adequately treat the internal processes, the model employs a number of parametrizations derived from specific field and laboratory experiments."

- *Page 2812, line 12: The authors usually use the term “longwave radiation” (e.g.,Page 2814, line 25) throughout the manuscript, but here they use the term “infrared radiation”. Please pick one convention and be consistent everywhere.*

The term “infrared radiation” has been changed to “longwave radiation” throughout the text.

- *Page 2813, Line 5: Earlier (page 2812, Lines 11-12) you noted that incoming solar radiation was a model input, but this line suggests that net solar radiation is used. Please clarify. How does the model calculate albedo?*

The phrase “Net radiation itself is composed of the sum of incoming and outgoing solar- and langwave radiation” describes the term NR in Equation 1 and does not refer to the model input. The model is forced by incoming solar radiation.

We have added the following paragraph to clarify the albedo calculation (p7L4-p7L8):
“Crocus treats solar radiation in three spectral bands ([0.3-0.8],[0.8-1.5] and[1.5-2.8] μm) and empirical coefficients (0.71, 0.21, and 0.08) describing spectral albedo as a function of the near surface snow properties (Vionnet et al., 2012). For each band the spectral albedo is computed as a function of the snow properties (microstructure). The remaining energy absorbed by the snowpack is assumed to decay exponentially with snow depth.”

- *Page 2813, Lines 9-10: Instead of “Eq. 3, right hand terms”, I think you might mean “Eq.2, first terms on right hand” for the phase change. Referencing Eq. 3 does not make sense in this context because it is the mass balance.*

Yes, the reference has been changed accordingly.

- *Page 2815, Line 18 (equation 4): Please provide a citation for the new snow density equation. What study does CROCUS cite for this equation?*

Vionnet et al. (2012) does not cite any reference. The snow density equation has been proposed by the original paper of Brun et al. (1992). We included both references for this equation.

- *Page 2816, Lines 24-25: This sentence is misleading, as not all sensitivity analysis methods rely on variance decomposition. Please rephrase.*

We have deleted the part of the sentence which says “... by decomposing the variance of the model output.”.

- *Page 2819, Lines 3-22: Much of the text here is more appropriate for section 2 (data and methods) and not the results section.*

We have added a new chapter “Reference run setup” in section 2 (p11L5-p12L4). This chapter describes the boundary conditions and initial conditions for the reference run. Parts of the text from section 3 has been moved the new section.

- *Page 2819, Lines 25-27: You should state somewhere in section 2 what you assumed the snow emissivity was in order to calculate snow surface temperature from upwelling longwave radiation.*

We paragraph (p16L19-p16L21) has been changed as follows: “Fig. 2 shows the comparison of simulated snow surface temperature with observational data. The simulated snow surface temperature is derived from upwelling longwave radiation assuming a snow emissivity of 0.99.” The snow emissivity has also been added to Table 1.

- *Page 2820, Line 4: Can you clarify what you mean by “diverse model uncertainties”?*

We have changed “diverse model uncertainties” by “structural model uncertainties”. Structural uncertainty refers to model inadequacy which comes from the general lack of knowledge. This uncertainty depends how accurately the snow model describes the natural/physical processes.

- *Page 2820, Line 7: It appears that the metrics used to describe model performance for the temperature variables (snow surface temperature and temperature profile) are inconsistent. For the former, they report the deviations in terms of a range (i.e., 95% within 1.1 K for snow surface temperature), but in the latter they use RMSE for the temperature profile. Please consider using more consistent evaluation metrics.*

We changed the evaluation metrics of the snow surface temperature to RMSE. The metrics should be consistent for all variables, now.

- *Page 2820, Lines 14-15: These stated albedo ranges are for the measured values, correct? Please clarify.*

The corresponding phrases have been revised to read as follows (p17L10-p17L14): ”The The RMSE between the measured and modelled albedo over the entire simulation period is 0.06 [-]. Note that the measured albedo ranges between 0.65 in the ablation period and 0.92 in the accumulation period, which is characteristic for a site in the accumulation region (Armstrong and Brun, 2008; Greuell et al., 2007).”

- *Page 2821, Line 25: Please clarify whether this means 16 000 model simulations were evaluated, or if this means that you selected 16 000 points in the uncertainty space. These two are not equivalent, because the latter will result in at least N(k+2) model evaluations where N is the number of points in the input uncertainty space.*

We have updated the complete section on “Global Sensitivity Analysis” and added a new paragraph (p14L15-p15L9) describing in more detail the estimation of the sensitivity indices, the sampling and the number of base samples. We have repeated the simulations with 20000 model runs (k=8, N=2000) to provide better accuracy of the indices. The accuracy of the sensitivity indices was assessed from 1000 empirical bootstrap samples. Similar to Raleigh et al. (2015) we estimated the 95% confidence regions for each total-order index (see Figure 4).

- *Page 2824, Lines 15-19: This argument depends strongly on whether one considers 7% interaction variance as a significant contribution. Also your argument would be strengthened if you examined your own results in Figure 7 and gave a specific example of how wrong conclusions might be drawn if only the first-order effects were considered (as is the case in SA methods that are designed for linear models). Currently this all seems like conjecture and the argument is not compelling.*

This was a serious mistake. The number refers to the maximum contribution of linear effects on the model variance (SHC) and not on the time averaged value. We have now given the averaged first-order indices for each target metrics (see Figure 4). The average values first-order indices are significantly smaller than the 93% given in the text and vary between 0.69 and 0.82.
In the discussion section we (p22L7-p22L12) we write: “The overall results of this work show that on average about 80% of the total variance of SHC and SEB can be explained by first-order effects (Fig. 4). This means that the remaining 20% of the variance is due to non-linear interaction effects. There is no significant difference between the two sites at the glacier. This is in partial contrast to the

findings of Raleigh et al. (2015), who performed similar investigations for different snow regimes and found that first- and total-order indices are of comparable magnitude."

- *Page 2828, Lines 7-8: The logic of this argument is unclear to me. Because you have calculated the first order indices, I would argue that you actually can compare your results to the first order effects found in Karner et al. (2013) and Obleitner and Lehning (2004). So this begs for additional discussion and comparison between the results here (which found low first-order sensitivity values for T and SW in the SEB) and those previous results (which had higher first-order effects from T and SW). Why do you suppose the first order effects are different?*

Karner et al. (2013) uses a different snow model, where measured albedo values are provided as input and do not depend on the snow microstructure. The station is also located in a different glaciological regime (equilibrium line altitude vs accumulation area, see also Table 3). In order to support our statement we have extended our analysis and simulated and analysed a second station, KNG1, on the Kongsvegen glacier located in the ablation zone. Consideration of sites other than KNG6 was motivated by the availability of correspondingly suitable data. We have rewritten and extended the discussion section and compare in detail our findings with the results of Karner et al. (2013) – see p23L27-p23L29; p25L26-p25L29; p28L17-p28L18.

In the conclusion we emphasized the differences between the two stations, KNG1 and KNG8 (p29L17-p29L25): "The overall impact of individual error sources on the sensitivity pattern varies for different zones on the glacier. In the accumulation zone, precipitation and longwave radiation are key factors for the evolution of the snowpack and contribute most to the model uncertainty. The significance of precipitation decreases with altitude, while other factors, such as wind velocity and surface roughness, gain importance. Uncertainties in the measurement of incoming shortwave radiation and air temperature have little influence on the model outcome, the former being biased by the specific, i.e. Arctic, conditions. The calculated seasonal sensitivity patterns are similar overall at both study sites."

- *Page 2829, Lines 5-8: I would disagree that your analysis supports this conclusion. As I understand the selected input uncertainty ranges, these are informed by manufacturer's specified accuracy. Hence, it may not be possible to reduce the LW uncertainty to anything better than +/-10%, but the results show that even in this "best case" (of having LW measurements instead of parameterizing the flux), the model outputs are still strongly controlled by the measurement uncertainty of LW.*

We agree that it may not be possible to reduce the LW uncertainty anything better than +/-10%. The phrase "..., however, can be gained more easily by using direct measurements of LW, ..." was misleading and has been removed.

Technical Corrections:

- *Page 2808, Line 15: Replace "there" with "their".*
Changed accordingly.

- *Page 2809, Lines 1-3: This first sentence reads awkwardly, perhaps because you use "which" three times in the sentence. Please rephrase.*
The sentence has been removed.

- *Page 2809, Lines 21-23: This sentence reads awkwardly. I recommend rephrasing to " ...scientists to quantify the uncertainty in the model outcome, and to provide information on its robustness."*
Changed.

- *Page 2810, Line 7: Replace "have been" with "has been".*

Changed.

- *Page 2810, Lines 7-9: This sentence reads awkwardly. Please rephrase.*
The sentence has been rephrased: “In recent years there have been increasing efforts to quantify the parametric and predictive uncertainty of mass and energy balance models … “.

- *Page 2812, Line 4: Replace “of the of” with “of”.*
Changed.

- *Page 2813, Line 8: What do you mean by “according changes”? This does not make sense.*
The expression according changes has been changed to “The change of internal energy ...”

- *Page 2813, Line 18: This is not grammatically correct. It should read “We refer to superimposed ice as ...”.*
Changed.

- *Page 2813, Line 20: The comma after “showed” is not necessary.*
Changed.

- *Page 2814, Lines 6-7: This sentence is not grammatically correct. Please rephrase.*
The sentence now reads as: “This prevents percolation of water into the glacier ice and increases the refreezing potential at the snow-ice interface.”

- *Page 2814, Line 17-18: Based on the inset map of Figure 1, this statement (“located in north-eastern Svalbard”) does not appear to be correct. Please correct.*
*“north-eastern” was c*hanged to “north-western”.

- *Page 2814, Line 20: Based on Figure 1 (which shows the glacier flowing northwest), this description (“flows north-eastwards”) does not appear to be correct. Please correct.*
Changed.

- *Page 2815, Line 5: Reverse the order here (“also showed”) to make this sound less awkward.*
Changed.

- *Page 2815, Line 9: “Based to the distance” does not make grammatical sense*
Changed to: “Then the missing value has been replaced by randomly drawing on out the 20 most analogous days.”

- *Page 2815, Line 11: It should be “physically” instead of “physical”.*
Changed.

- *Page 2816, Line 9: It should be “arbitrarily” instead of “arbitrary”.*
Changed.

- *Page 2816, Line 17: Spelling error: it should be “sensor” instead of “senor”.*
Changed.

- *Page 2817, Line 8: Please define “SD” for clarity.*
SD changed to “standard deviation”.

- *Page 2817, Line 20: Remove the comma after “that”.*

  Changed.

- *Page 2817, Line 23: Add “the” before “following” and add a colon after “expression”.*

  Changed.

- *Page 2819, Line 24: Add “to” after “amounts”.*

  Changed.

- *Page 2820, Line 3: It should read “associated with” not “associated to”.*

  Changed.

- *Page 2820, Line 7: It should be “measurement shortcomings” instead of “measurements shortcomings”.*

  Changed.

- *Page 2820, Line 9: Why is Figure 5 reference before Figure 4? Consider renaming the figures to reflect the order in which they are introduced, or rewrite the text here to introduce the albedo figure before the density figure.*

  The order of the figures has been changed.

- *Page 2820, Line 16: “Following we indicate” does not make grammatical sense. Please revise.*

  The sentence was removed.

- *Page 2820, Lines 16-29 (and elsewhere): In all cases where the authors report energy fluxes, they omit the negative sign in the meters squared term. It should read “W m-2” but they consistently report it as “W m2”.*

  Changed.

- *Page 2820, Line 29: Reporting the mean annual energy surplus to the hundredths place is probably not warranted or useful.*

  Changed.

- *Page 2821, Line 1: Missing a closing parentheses “)” after “Fig. 1”.*

  Changed.

- *Page 2821, Line 3: Add “for” before “considering”.*

  Changed.

- *Page 2821, Line 10: Add “the” before “case”*

  Changed.

- *Page 2822, Line 6: Add a comma after “simulation period”.*

  Changed.

- *Page 2822, Line 7: It should be “reaches” instead of “reach”.*

  Changed.

- *Page 2822, Line 15: Replace “are decisive” with “control”.*

  Changed.

- *Page 2822, Line 19: Remove the comma after “mind”.*
Changed.

- *Page 2824, Line 1: This should read something like “to remind the reader” or “to note”. Also, the comma after “remind” is unnecessary.*
Changed.

- *Page 2824, Line 15: Replace “proof” with “prove”.*
Changed.

- *Page 2825, Line 4: Delete the comma after “showed”.*
Changed.

- *Page 2826, Line 23: Either say “a negative feedback” or just “negative feedbacks” (no “a”).*
Changed to “negative feedbacks”.

- *Page 2828, Line 18: It should read “scientists” (plural) instead of “scientist”.*
Changed.

- *Page 2829, Line 1: Replace “proofed” with “proved”.*
Changed.

- *Page 2829, Line 20: Why is humidity labeled as RH elsewhere but here it is Q*
Label has been changed to RH.

Table and Figure Comments

- *Table 2: For the units of PVOL, are you sure there should be a “%”? Because it is a fraction and because this value is usually around 5%, I think the uncertainty range should just read 0.03-0.05. Please confirm.*
That's right. The % has been deleted.

- *Figure 2: Due to the large number of points in the scatterplot, it is difficult to understand the distribution of points. Consider using a scatterplot with a density color scheme.*
The scatterplot uses a density color scheme. The original figure has a high resolution and should be easy to read.

- *Figure 3: This figure would be improved by also plotting markers on each line that show the measurement and model nodes. This can be inferred from changes in the slope of the temperature profile, but it is not clear in all cases where the nodes are in the case of more subtle variations in slope.*
Markers have been added for both lines.

- *Figure 3: Snow temperature is described in negative Kelvin, which is physically impossible. I think the label should either be in degrees centigrade or the numbers on the x-axis should change to be in Kelvin.*
The axis has been changed to degrees centigrade.

- *Figure 6: It is difficult to read the text in this figure and it is not clear how well this will display in the final GMD paper format. Is it possible to change the aspect ratio and/or resolution and/or font sizes of the figure?*
We have improved the figure.

- *Figure 7: Is Q supposed to represent RH in this plot? Is Q now a different humidity metric (such as absolute* *or specific humidity)? If this is still relative humidity, then it is best to remain consistent with the acronym usage and just keep RH. Please clarify.*

The acronym Q has been changed to RH.

- *Figure 7: Because they are all the same, you could safely remove the legends from three of the four panels and just leave one.*

The legends have been removed from three of the four panels.

- *Figure 8: Please specify at what temporal frequency (e.g., daily?) these sensitivity indices are calculated for snow depth.*

Caption changed to: "Evolution of the 6-hourly first-order ...".

- *Figure 8 caption: Table 1 does not include the indicated uncertainty factors. Do you mean Table 2?*

Changed to Table 2.

Manuscript prepared for Geosci. Model Dev. Discuss.
with version 2014/05/30 6.91 Copernicus papers of the LaTeX class copernicus.cls.
Date: 3 August 2015

Assessment of Assessing the uncertainty of snowpack glacier mass balance simulations in the European Arctic based on variance decomposition

T. Sauter[1,2] and F. Obleitner[1]

[1]Institute of Meteorology and Geophysics, University of Innsbruck, Innsbruck, Austria
[2]Institute of Geography, University of Erlangen-Nuremberg, Erlangen, Germany

Correspondence to: T. Sauter (tobias.sauter@uibk.ac.at)

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

Abstract

State of the art numerical snow

State-of-the-art numerical snowpack models essentially rely on observational data for initialization, forcing, parametrization and validation. Such data are available in increasing amountamounts but the inherent propagation of related uncertainties on the in simulation results has received rather limited attention so far. Depending on their complexity, even small errors can have a profound effect on simulations, which dilutes our confidence in the results. This paper quantifies the aims at quantification of the overall and fractional contributions of some archetypical measurement uncertainties on key simulation results in a high Arctic environment. The contribution of individual factors on the model snowpack simulations in Arctic environments. The sensitivity pattern is studied at two sites representing the accumulation and ablation area of the Kongsvegen glacier (Svalbard) using the snowpack scheme Crocus. The contribution of measurement errors on model output variance, either alone or by interaction, is decomposed using *Global Sensitivity Analysis*. The work focuses on global sensitivity analysis. This allows for investigation of the temporal evolution of the fractional contribution of different sources on the model uncertaintyfactors on key model output metrics, which provides a more detailed understanding of the model'model

s sensitivity pattern. The decompositions demonstrate, that the impact of measurement errors on calculated snow depth and the analysis demonstrates that the specified uncertainties in precipitation and longwave radiation forcings had a strong influence on the calculated surface height changes and surface energy balance componentsvaries significantly . The model output sensitivity patterns also revealed some characteristic seasonal imprints. For example, uncertainties in the longwave radiation trace on the calculated surface energy balance were continuous throughout the year . Some factors show episodically strong impacts, although there overall mean contribution is low while others constantly affect the results. However, these results are not yet to be generalized imposing the need to further investigate the issue for and occured at both study sites, while precipitation exerted the most influence during the winter and at the upper site. Such findings are valuable for identifying critical parameters and improving their measurement, and correspondingly, updated simulations may shed new light on the confidence of results from snow or glacier mass and energy balance models. This is relevant for many applications in the fields of, e.g.other glaciological and meteorological settings.

, avalanche and hydrological forecasting.

1 Introduction

Snow is a key component of the earth system, which has aand it has a vital importance for the structure and dynamics of the atmospheric boundary layer by modifying, e.g., the exchange processes between the atmosphere and the underlying ground. Bridging the gap between the inherent microphysical snow processes and the exchange processes at the snow surface, e.g. their effect on bulk properties or the exchange of energy and matter, still constitutes major challenges to scientists.

With this in mind, snow scientists have done much research during the last years improving our knowledge and understanding of the associated processes. Our theoretical understanding is largely derived from observations, which provide the basis for numerical models which frequently have to employ parametrizations of processes which yet cannot explicitly be treated at the model grid or in terms of the prognostic variables. Sophisticated snow models summarize our

Sophisticated snowpack models summarize the present knowledge and prove themselves to be a useful tool in simulating the spatial and temporal evolution of snowpacks. As reported in many studiesThus, snow models have been successfully applied and implemented for climate impact studies (e.g. Durand et al., 2009), adopted for avalanche forecasting (e.g. Bellaire et al., 2013; Durand et al., 1999; Lehning et al., 1999), glacier modelling (e.g. Obleitner and Lehning, 2004; Gallée et al., 2001)and , hydrological research (e.g. Magnusson et al., 2014; Lehning et al., 2006; Liston and Elder, 2006; Bernhardt et al., 2010). Snow models currently used , and climate impact studies (e.g. Durand et al., 2009). The currently used snow models can be roughly classified by their degree of complexity, ranging from simplified bulk or single-layer models through to detailed physical snowpack models (Etchevers et al., 2004; Feng et al., 2008; Rutter et al., 2009). Besides all advantages of detailed models, the increasing complexity leads inevitably to higher demands on the kind and quality of data required to force these models. However, the ’true’ In general, the development and use of higher order models also induces a need for more and better data to constrain the initialization, forcing, parameterizations and results of the simulations. However, the quality of relevant data (model and observations) is still difficult to assess. In that sense, the "true" value of a measured quantity is rather a theoretical concept, remains a rather theoretical concept and can often not be determined. In view of this uncertainty, we usually estimate One therefore usually estimates a range of values within which the true value is likely to fall. Ideally, these uncertainties should be considered in the modelling process and when interpreting results - not at least to ensure good scientific practise. However, in practise it is not always easy to derive a reliable probability density function describing the inherent Probability density functions are widely recognized as appropriate measures for describing the uncertainty of input data and model parameters(see Section , and they are used in this study (see Sect. 3.2). Taking into account systematic measurement errorsallow scientists quantifying the uncertainty in the In practise, however, these can often not be specified in a straightforward manner due to the complex nature of, e.g., measurement errors. It is nevertheless a major methodical issue to account for best estimated measurement errors, which allows scientists to objectively quantify their impact on the model outcome, and providing information on its robustness . A kind of minimum approach is through Monte Carlo analysis by randomly drawing to provide information on the robustness of the results. A corresponding approach is based on Monte-Carlo methods considering randomly drawn samples for each input factor from previously derived distribution functions. From the model we can compute first First and higher moment statistics can be computed to quantify the integrated model uncertainty. In this context, integrated is understood as the total effect of all measurement or parameter uncertainties on the model's variability. At this point, there is still no information on how uncertainty in the model output can be apportioned assigned to different sources of uncertainty in the input data set or parameter setting. Since with increasing degree of model complexity sub-routines or modules become highly coupled, unambiguous allocation of For example, interaction effects make it difficult to unambiguously allocate the uncertainty of model parameters and forcing data on the model's varianceis hampered by interaction effects. To achieve a full understanding of the model's sensitivity pattern , in particular sensitivity pattern of highly interconnected and nonlinear models, such as sophisticated snow models, it is necessary to decompose the complete variance of the model results.

In recent years Following this line, there have been an increasing awareness of the issue yielding increasing efforts to quantify the uncertainty associated with the various sources of error in the parameter setting to assess parametric and predictive uncertainty (e.g. Franz et al., 2010; He et al., 2011; Schmucki et al., 2014; Gurgiser et al., 2013). Efforts to assess the climate sensitivity of snow and glaciers based on of mass and energy balance models go into the same direction (e.g. Gerbaux et al., 2005; Fujita, 2008; Radić and Hock, 2006; Greuell and Oerlemans, 1986; Oerlemans, 1992; Braithwaite and Zhang, 2000). However, thorough investigation of (e.g. Franz et al., 2010; He et al., 2011; Schmucki et al., 2014; Gurgiser et al., 2013; Gerbaux et al., 2005; Fujita, 2008; Radić and Hock, 2006; Greuell and Oerlemans, 1986; Oerlemans, 1992; Braithwaite and Zhang, 2000). Some of these also consider the investigation of effects on the specific influence of uncertainties related to model input and its effect on energy and mass balance calculations received rather scant attention of glaciers or ice sheets (e.g. Karner et al., 2013; Van de Wal and Oerlemans, 1994; Greuell and Konzelmann, 1994).

Raleigh et al. (2015) were the first to explore how different error types and distributions influence the physically based simulations of snow variables in snow-affected catchments. Their approach to testing the model sensitivity to co-existing errors in the forcing was based on Sobol's global sensitivity analysis. The present study intends to contribute to our understanding on how was developed independently and follows a similar concept to identify how the systematic measurement errors (biases) and uncertainties of some critical factors influence our confidence in snowpack glacier mass balance simulations. We study this effects using the snowpack model CROCUS, which is applied at a study site on the Kongsvegen Glacier in Svalbard(see Section the seasonal evolution of the energy and mass balance of snow and ice at two sites on the Arctic glacier Kongsvegen (Svalbard) (see Sect. 2.2). CROCUS has been These sites are chosen to represent conditions in the accumulation and ablation area of the glacier, thus addressing different mass and energy balance regimes. Crocus was originally developed and is still used for operational snow avalanche warning (Brun et al., 1992; Durand et al., 2009) , warnings (Brun et al., 1992; Durand et al., 2009) and has been applied to various research problemsstudies, e.g. Brun et al. (2013); Fréville et al. (2014); Carmagnola et al. (2013); Wang et al. (2013); Phan et al. (2014); Gallet et al. (2014); Castebrunet et al. (2014). Vionnet et al. (2012) give aprovide a comprehensive review of CROCUS Crocus and its implementation in SURFEX, i.e. a model platform for simulation of which is a comprehensive platform for simulating earth surface processes.

The results of our study may not yet

This study is the first to address the uncertainty of simulations using Crocus, and it may be generalized due to the rather local nature of our simulations, but local application and possible specific influences due to, e.g., the Arctic environmental conditions. However, it may be useful for other studies using sophisticated snow models in similar environmental settings. A better understanding of the model's sensitivity can be very helpful to establish priorities in research, identify critical regions in the input space and even for policy assessmenthelpful to demonstrate the benefits of the applied method, to identify critical issues concerning model input and parametrization and to establish future priorities in corresponding research. An attractive approach to estimate for estimating sensitivity measures independently of the degree of linearity (model-free) is based on the Global Sensitivity Analysis global sensitivity analysis (GSA), which is introduced in Section 2.4. Before finally dealing with the decomposition of the model uncertainty in Section ??, we first perform a common Sect. 2.4. We then developed reference runs which are validated by key observations at the two glacier sites. Based on these reference runs and the specification of the uncertainties of key variables and parameters, we performed Monte-Carlo uncertainty estimation on a validated reference run (see Section 3.1 and 3.2). In the last section we discuss the information gained from the analysis, limitations of linear sensitivity measures, general problems of sensitivity analysis and what can be learned from this analysissimulations. The results are presented in (Sect. 3.1 and are mainly discussed regarding the impact of key drivers in terms of first and total order indices and inherent limitations as well as regarding differences concerning the two sites at the glacier.

2 Data and Methodsmethods

2.1 CROCUS Crocus model setup

CROCUS is aphysical,

Crocus is a physically based finite-element and one-dimensional multilayer snow scheme implemented in the land-surface model ISBA of the surface modelling platform SURFEX. Snow is considered as aa porous material whose properties are determined by the microstructure characteristics - basic microstructure characteristics – grain size, dendricity, and sphericity. These properties mainly describe The time and depth dependent evolution of these parameters describe bulk properties like porosity, diffusivity, heat conductivity, viscosity , or and extinction of radiation. The evolution of the microstructure characteristics is closely linked to Changes to snow microstructure characteristics are strongly driven by the prevailing environmental conditions and the related exchange processes. Snow metamorphism laws for the evolution of types and size of the snow grains have been derived from empirical observations and are implemented by parametrizations (Brun et al., 1989).

The model is extensively described elsewhere (Vionnet et al., 2012; Brun et al., 1992) (Vionnet et al., 2012; Brun et al., 1992), and we therefore give just a simply provide a basic description and note modifications important for this study. CROCUS is a Crocus is a one-dimensional snow snowpack model which simulates the evolution of the physical and morphological snow properties and related processes depending on the atmospheric and basal boundary conditions. It thereby considers the conservation changes of energy and mass within layered control volumes and the associated processes (molecular conduction layers due to molecular conductivity, radiative transfer, turbulent exchange of sensible and latent heat, phase changes and gravitational water transport). Snow layers are described through bulk physical properties (thickness, density, temperature , and liquid water content) and microstructure parameters(. The latter parameter characterizes the state of the snow crystals in terms of dendricity, sphericity, grain size and indicators of the of snow grain history. The parameter of snow grain history indicates whether there once was liquid water or faceted crystals in the layer (Brun et al., 1992; Vionnet et al., 2012). The latter enables CROCUS to describe the changes in the morphological shape of snow crystals depending depend on snow metamorphism in response to atmospheric forcing and internal processes. To adequately treat these the internal processes, the model employs a number of parametrizations derived from specific field and laboratory experiments. The governing equations are numerically solved in a vertical domain with space and time varying grid distances (which are necessary in order to cope with e.g. accumulation or settling processes). The model is forced by the basic meteorological parameters (air temperature, humidity, wind speed and precipitation rate as well as incoming solar and infrared longwave radiation) and is initialized by the vertical profiles of the key physical properties of snow and its underlying substrate. Model output comprises outputs comprise the vertical profiles of the bulk physical (snow temperature, density, liquid water content) and structure parameters as well as and prognostic time series of surface temperature, snow depth and energy- and mass balance components, the latter two being coupled. Following, e.g., Armstrong and Brun (2008), the change model treats layerwise changes of internal energy according to

$$-\frac{dE}{dt} = NR + SHF + LHF + R + G$$

$$= L_{li}(R_f - R_M) - \int_{z=0}^{HS} \left[\frac{d}{dt}(\rho_z c_p T_z)\right] dz,$$

$$\mathrm{SEB} = \mathrm{NR}+\mathrm{SHF}+\mathrm{LHF}+\mathrm{R}+\mathrm{G} \tag{1}$$

$$= \mathrm{L_{li}}(\mathrm{R_f} - \mathrm{R_M}) - \int_{z=0}^{\mathrm{HS}} \left[\frac{\mathrm{d}}{\mathrm{dt}}(\rho_\mathrm{z} \mathrm{c_p} \mathrm{T_z})\right] \mathrm{dz}. \tag{2}$$

of the snowpack depends on the surface energy budget (SEB)Therein, SEB denotes the surface energy budge, i.e. the sum of net radiation NR(NR), the turbulent fluxes of sensible (SHFSHF) and latent heat (LHFLHF), the heat transfered transferred by precipitation and blowing snow (R), and by conduction from the underlying material G (glacier ice in our case). Thus available The associated changes in energy can be used for changes in the cold content of the snow pack throughout its total depth HS (right-hand term in Eqn.(HS) (second term in Eq. 2) or phase changes (melt or freeze; first-hand term in Eqn.first term in Eq. 2). R_f and R_M R_f and R_M are the freezing and melting rate, L_{li} L_li is the latent heat of the fusion of ice ($3.34 \times 10^5\,\mathrm{J\,kg^{-1}}$), c_p is the specific heat capacity of ice (2.1×10^3), , $\mathrm{J\,kg^{-1}\,K^{-1}}$) and ρ_z and T_z denote the density and snow temperature at depth z, repsectively. Net radiation itself is composed of the sum of incoming and outgoing solar- and infrared radiation(measured input except of outgoing infrared) and the turbulent longwave radiation. Crocus treats solar radiation in three spectral bands ([0.3-0.8],[0.8-1.5] and [1.5-2.8] µm) and empirical coefficients (0.71, 0.21 and 0.08) describing spectral albedo as a function of the near-surface snow properties (microstructure) (Vionnet et al., 2012)). For each band, the spectral albedo is computed as a function of the snow properties (microstructure). The remaining energy absorbed by the snowpack is assumed to decay exponentially with snow depth. Turbulent fluxes are parametrized following the standard micrometeorological framework based on Monin-Obukhov similarity theory.

the Monin–Obukhov similarity theory, which employs a bulk-transfer approach and some specific modifications considering, e.g., treatment or surface roughness or turbulent exchange at stable stratification of the atmospheric surface layer.

The according changes in available Layerwise changes in internal energy induce either varying cold content (warming/cooling; last term of Eqn.second term of Eq. 2) or phase changes of individual snow layers (Eqn.3, right hand termsEq. 2, first term). Melt water refreezing and/or sublimation/evaporation rates (E) as well as runoff ($R_{runoff}R_{\mathrm{runoff}}$) couple the energy- and the mass budget of a snow pack according to

$$\frac{dM}{dt} = P \pm E - R_{runoff}.$$

$$\frac{\mathrm{dM}}{\mathrm{dt}} = \mathsf{P} \pm \mathsf{E} - \mathsf{R}_{\mathrm{runoff}}. \quad (3)$$

Key The key parameters of this coupled system will also be addressed in this study, too. CROCUS . Crocus has not yet been applied to Kongsvegenbefore. The following paragraphs summarize the main modifications and setup used in this study in order to develop reference runs properly reproducing the seasonal evolution of snow and ice at the two glacier sites.

Water flow and refreezing. . Superimposed ice is a common feature The formation of superimposed ice is an important factor for the mass balance of Arctic glaciers, and a better understanding of the relevant processes is currently an active field of researchan appropriate treatment is also important in this study. We refer as superimposed ice to superimposed ice as all water which percolates through the snowpack and refreezes on the glacier surface (Wright et al., 2007; Brandt et al., 2008; König et al., 2002). Obleitner and Lehning (2004) and Karner et al. (2013) showed , that on Kongsvegen glacier that on the Kongsvegen glacier, the superimposed ice layer can reach a thickness of several decimetres in some years. The water percolation and refreezing routine in the current CROCUS Crocus version basically simulates the gravitational water flow through the snowpack (Gascon et al., 2014). The energy available for refreezing is calculated at the beginning of each iteration step. If the snow layer temperature is below the melting point, water refreezes and the residual liquid water is retained up to a maximum holding capacity. The maximum liquid water holding capacity $PVOL$ is usually assumed to be , which is difficult to determine. Default Crocus assumes a value of 5% of the total pore volume . The refreezing process increases pore volume and reproduces an increase of the average density and mass of concerned layers of layers affected by the refreezing of water (Vionnet et al., 2012). This implementation does not account for superimposed ice, since the water percolates through the glacier ice is appropriate for the simulation of snow evolution in, e.g., Alpine terrain but it fails to reproduce the formation of superimposed ice because water can percolate through glacier ice as well. To overcome the issue, all water exceeding the maximum liquid water holding capacity at an impermeable snow-ice snow–ice interface is assumed to contribute to the runoff, and the water flow to the next layer is simply set to zero. This modification avoids, that the scheme removes too much water to deeper layers, which limits the refreezing potential prevents percolation of water into the glacier ice and increases the refreezing potential at the snow–ice interface. This approach has been successfully applied in a similar setting using a different snow model (Obleitner and Lehning, 2004).

Model input/output. *The CROCUS* . The Crocus model is forced by air temperature (T), relative humidity (RH), specific humidity , wind speed , incoming radiation, precipitation rate (U), incoming shortwave radiation (SW), incoming longwave radiation (LW), precipitation rate (P) and atmospheric pressure (see Section Sect. 2.2). These time-dependent factors are provided parameters were measured at both sites and are provided to the model by Netcdf-file . Besides the time-dependent forcing file, a constant model parameters are provided by a static option file. In order to perform Monte-Carlo simulations we have included for hourly time steps. The input file was modified to include the roughness length for fresh snow momentum as well as the fraction of total pore volume in the Netcdf-file, used to calculate the maximum holding capacityin the forcing file . Constant model parameters are provided by a static option file describing the initial and boundary conditions and the basic model parameters.

2.2 Input Datadata

To run the snowpack scheme CROCUS

To run Crocus, we use meteorological and glaciological observations on the Kongsvegen Glacier (78.75N, 13.33E, 668 asl), located in northeastern from two sites at the Kongsvegen glacier, located in north-western Svalbard. The Kongsvegen glacier currently covers a total surface area of ~100 $\sim 100\,\mathrm{km}^2$ and extends over a total length of $26\,\mathrm{km}$. From the highest point (750asl $\mathrm{m\,a.s.l.}$) in the east, the glacier flows north-eastwards towards the north west coast towards the north-west coast of the archipelago. Several automatic weather stations are were operated along the flow line of the glacier, of which this study only . The study makes use of the station two of them: KNG8 operated (78.75° N, 13.33° E, 668 $\mathrm{m\,a.s.l.}$), located in the accumulation zone(see Figure , and KNG1 (78.84° N, 12.66° E, 162 $\mathrm{m\,a.s.l.}$), located in the ablation zone (see Fig. 1). Due to computational limitations, we had to restrict our simulations and error analysis to a one-year period.

The station is equipped with state of the art The stations are equipped with state-of-the-art sensors for air temperature, relative humidity, wind speed, and direction as well as the shortwave and longwave radiation components 2. Surface height changes were measured by an ultrasonic ranger. Karner et al. (2013) performed a comprehensive data quality assessment and correction of unreliable observations. The processed data are available as hourly averages and enhanced quality checking of the data suggested to apply some further correctionsComprehensive quality control of the recorded data was performed following the method of Karner et al. (2013). The data have been further corrected:

Filling remaining data gaps. For shorter gaps, the missing values have been were estimated by linear regression from surrounding stations, where it was possible. In cases this was not possible, e.g. because When the surrounding stations showed also gaps, the missing values have been had missing values, the values were estimated by a stochastic nearest-neighbour resampling conditioned on the remaining variables (Beersma and Buishand, 2003). This was achieved by first calculating the euclidean Euclidean distance between the present day and all other days without gaps. Based to the distance one out of the Then the missing value was replaced by randomly drawing of one out the 20 closest days have been stochastically selected and the missing value has been replaced by the corresponding valuemost analogous days. This approach is convenient for small gaps and guarantees physical physically consistent fields.

Discussion Paper
Discussion Paper
Discussion Paper

Conversion of snow depth changes to water equivalent. Snow precipitation rates were derived from surface height changes measured by the ultrasonic ranger, and which needed to be converted to snow water equivalent (SWE) for input to the model. The density of freshly fallen snow ρ_{new} ρ_{new} was calculated according to the equation used by CROCUSdefault parametrization used by Crocus, which is a function of wind speed U, and air temperature $T_{air}T_{\mathrm{air}}$, given as

$$\rho_{\mathrm{new}} = \mathrm{a}_\rho + \mathrm{b}_\rho \cdot (\mathrm{T}_{\mathrm{air}} - 273.16) + \mathrm{c}_\rho \cdot \sqrt{\mathrm{U}}, \tag{4}$$

$$\rho_{new} = a_\rho + b_\rho \cdot (T_{air} - 273.16) + c_\rho \cdot \sqrt{U},$$

where $a_\rho = 300\,\mathrm{kg\,m^{-3}}$, $b_\rho = 6\,\mathrm{kg\,m^{-3}\,K^{-1}}$, and $c_\rho = 26\,\mathrm{kg\,m^{-7/2}\,s^{-1/2}}$ (Brun et al., 1992; Vionnet et al., 2012). Note, that in the original model version the default value for a_ρ is set to 109. We modified this value since the CROCUS model underestimated the initial $\mathrm{kg\,m^{-3}}$ (Libois et al., 2014). The modification of this parameter accounts for the systematic underestimation of simulated settling and compaction of the upper snow layers , and has revealed best results concerning the simulated density profile. According to point snow-cover data from near-surface snow layers compared to repeated snow-pit studies, observations. The latter reveal that the mean density of the snowpack in the upper few centimetres usually lies near surface snow layers is usually in the range of $100-200$ 100–200 $\mathrm{kg\,m^{-3}}$. It was further necessary to reduce the amount of noise in the original snow records in order to avoid erratic precipitation events, which lead to unrealistic unrealistically high accumulation. The main factors that affect the sensor signal are blowing snow, intense snowfall, uneven snow surfaces, extreme temperatures and snow crystal type(low density). Blowing and drifting snow are frequent processes in the European Arctic and often result in the formation of sastrugi(Sauter et al., 2013). The , which introduce additional surface variability not associated with precipitation events (Sauter et al., 2013). In principle, the associated small scale variability is can be usually reduced by moving average filter, but the very different event durations make it sometimes sometimes make it difficult to determine an appropriate fixed subset size. We therefore decided to take the mean saltation trajectory height as ameasure criterion of the uncertainty, which is assumed to be proportional to the surface shear stress u_*^2 u_*^2 $[\mathrm{m^2\,s^{-2}}]$ (Pomeroy and Gray, 1990),

:

$$h_{salt} = \frac{1.6 \cdot u_*^2}{2 \cdot g},$$

$$\mathrm{h_{salt}} = \frac{1.6 \cdot \mathrm{u}_*^2}{2 \cdot \mathrm{g}}, \tag{5}$$

where g $[\mathrm{m\,s^{-2}}]$ is the gravitational acceleration. The surface shear stress has been estimated from the by assuming a logarithmic wind profile and an arbitrary arbitrarily chosen constant roughness length of $z_0 = 0.02\,\mathrm{m}$. Finally, snow depth smaller than $0.8 \cdot h_{salt}$ variations smaller than $0.8 \cdot h_{\mathrm{salt}}$ were considered as noise. The factors z_0 and 0.8 are were used for calibration and to determine how much signal were was removed from the original time series. Filtering out the small scale variability reduced the total precipitation amount at At KNG8by 29%, and yields a, for example, this procedure yields a simulated end-winter snow accumulation which is well validated by independent stake observations.

Large amplitude spikes. Large amplitude data spikes in recorded snow depth changes can occur during intense snowfall events when snow particles obstructs obstruct the propagation of the senor sensor signal (ultra-sonic pulses). Sudden snow depth changes greater than in excess of $50\,\mathrm{mm\,h^{-1}}$ are assumed to belong to this class of events, and were simply ignored. Transition from rain to snow was assumed to take place in the range from $0°\mathrm{C}$ to $1°\mathrm{C}$ with half of 0 to 1 °C, with half the precipitation falling as snow, and the other half as rain. There was no direct information available to determine this thresholdbetter, which leaves a relative large uncertainty better constrain this threshold. Input of calculated changes in precipitation water equivalent are considered as part of the calibration procedure of the reference runs and yield overall satisfactory reproduction of the independently observed end winter snow height, i.e. accumulation at both sites.

2.3 Reference run setups

The reference runs serve as basis for the subsequent assessment of the uncertainty of the simulation results (see Sect. 3.2), and the corresponding decomposition of the model variance (see Sect. 3.4). The modified Crocus model (see Sect. 2.1) is forced with the pre-processed and adjusted input data introduced in Sect. 2.2. The most relevant model parameters are given in Table 1. The initial snowpack is assumed to be isothermal with 273.16 K, and a constant base temperature of 271 K has been applied to the bottom of the model domain. The maximum number of snow layers is set to 50 in order to obtain a detailed snowpack stratigraphy. The initial grid spacing is increased from 0.01 m at the surface to 10 m at the bottom. The number of grid cells and their spacing is updated during the simulation according to the accumulation, temperature, density and melt. KNG8 is located in the accumulation zone of the glacier, where the near surface layers consist of perennial snow rather than bare ice (Björnsson et al., 1996; Brandt et al., 2008). Following Björnsson et al. (1996) and Brandt et al. (2008), the model is initialized with an isothermal firn layer with a mean density of $600\,\mathrm{kg\,m^{-3}}$ and a total thickness of 20 m. KNG1 is located in the ablation area of the glacier, where surface conditions are characterized by less snow accumulation during winter, stronger melt during summer and a corresponding prevalence of bare ice at the surface. At both sites, the simulations start at the end of the ablation season, with the lowest recorded snow depth (defined by the minimum recorded surface height), and they are forced by hourly measurements. Simulation results are stored every 6 h for analysis. At the lower station, KNG1, the glacier ice reappears at the surface in the course of the ablation season. To represent the site-specific condition, the initial density is set to $830\,\mathrm{kg\,m^{-3}}$, which is corroborated by observations from ice cores. Measurements of surface temperature and albedo are used for validation only and are considered as key indicators to judge the model's ability to calculate the energy balance. Validation of mass balance calculations is performed by comparing simulated and observed snow temperature and density profiles. Note, however, that the reference simulations were not optimized to fully reproduce the available observations. This would have required further tunings, which are not necessary for the purpose of this methodical study. There is no doubt, however, that the two reference runs truly reflect the basic characteristics of the seasonal evolution of snow and ice at the two considered sites.

2.4 Global Sensitivity Analysis sensitivity analysis (GSA)

In general, sensitivity analysis (SA) permits inferences on the different sources of uncertainty in model inputs by decomposing the variance of the model output (Sauter and Venema, 2011)(Sauter and Venema, 2011). This section gives an overview of how model-free sensitivity measures can be derived from variance-based methods. For the purpose of illustrationlets assume a, let's assume a generic model f

$$\mathbf{Y} = f(\mathbf{X_1}, \mathbf{X_2}, \cdots, \mathbf{X_k}),$$

$$\mathbf{Y} = \mathsf{f}(\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_\mathsf{k}), \tag{6}$$

with the model output $\mathbf{Y}$, the input quantity $\mathbf{x_k}$, $\mathbf{X}_k$ and the corresponding total or unconditional variance $V(\mathbf{Y})$. Most common SA measures are based on local derivatives $\partial\mathbf{Y}/\partial\mathbf{X_k}$ $\partial\mathbf{Y}/\partial\mathbf{X}_k$ to estimate the relative importance of individual quantitiescomponents. It is convenient to normalize the derivatives by the standard deviations, deviation so that the measures are weighted and sum up to one. In this context it It is also interesting to note , that in case of linear modelsin this context that in linear models, the normalized derivatives coincide with the well known well-known standardized (linear) regression coefficients (Saltelli et al., 2006). Obviously, both measures rely on the assumption of linearity, which makes them unsuitable for complex models. This is in particular particularly true when interaction effects become important, a which is a characteristic property of nonlinear and non-additive models. Such effects are captured However, such effects may be addressed by so-called model-free measures, which can be effectively estimated by the Global Sensitivity Analysis (GSA) GSA method described here.

If one forcing input $\mathbf{x_i}$ $\mathbf{X}_i$ is fixed at aparticular value x_i^* particular value x_i^*, the resulting conditional variance of $\mathbf{Y}$ is accordingly $V_{X\sim i}(\mathbf{Y}|\mathbf{X_i} = x_i^*)$given by $V_{X\sim i}(\mathbf{Y}|\mathbf{X}_i = x_i^*)$. This measure characterizes the relative importance of the factor $\mathbf{x_i}\mathbf{X}_i$, since the conditional variance will be less than the unconditional variance. The fact that , this sensitivity measure depends on the value of x_i^* x_i^* makes it rather impractical. Taking instead the average of this measure over the uncertainty distribution of $x_i^*x_i^*$, the undesired dependence will disappear (Saltelli et al., 1999, 2006). We can obtain following The expression

$$V(\mathbf{Y}) = E_{X_i}(V_{X\sim i}(\mathbf{Y}|\mathbf{X_i} = x_i^*)) + \\ V_{X_i}(E_{X\sim i}(\mathbf{Y}|\mathbf{X_i} = x_i^*)),$$

$$\mathsf{V}(\mathbf{Y}) = \mathsf{E}_{\mathsf{X_i}}(\mathsf{V}_{\mathsf{X\sim i}}(\mathbf{Y}|\mathbf{X}_\mathsf{i} = \mathsf{x}_\mathsf{i}^*)) + \mathsf{V}_{\mathsf{X_i}}(\mathsf{E}_{\mathsf{X\sim i}}(\mathbf{Y}|\mathbf{X}_\mathsf{i} = \mathsf{x}_\mathsf{i}^*)), \tag{7}$$

where the second conditional variance on the right hand side is called the decomposes the total variance $V(\mathbf{Y})$ into the first-order effect of $\mathbf{X_i}$ on $\mathbf{Y}$(second right-hand-side term) and higher-order (first right-hand-side term) contributions. The corresponding first-order sensitivity index of $\mathbf{X_i}$ $\mathbf{X}_i$ is given by

$$S_i = \frac{V_{X_i}(E_{X\sim i}(\mathbf{Y}|\mathbf{X_i} = x_i^*))}{V(\mathbf{Y})}.$$

$$\mathsf{S_i} = \frac{\mathsf{V}_{\mathsf{X_i}}(\mathsf{E}_{\mathsf{X\sim i}}(\mathbf{Y}|\mathbf{X}_\mathsf{i} = \mathsf{x}_\mathsf{i}^*))}{\mathsf{V}(\mathbf{Y})}. \tag{8}$$

This sensitivity index indicates the importance of individual factors without considering interactions effects. In case When the model belongs to the class of additive models, the first-order terms add up to one, e.g. $\sum_{i=1}^{r} S_i = 1$$\sum_{i=1}^{k} S_i = 1$. If this is not the case, the remaining variance must be explained by the higher-order effects (interaction) between induced by the interaction of input factor uncertainties. Interactions represent an important feature , especially, of nonlinear non-additive models. The total sensitivity S_{Ti} of afactor $\mathbf{X_i}$ S_{T_i} of a factor $\mathbf{X}_i$ is made up of the first- and all higher order terms where agiven factor $\mathbf{X_i}$ given factor $\mathbf{X}_i$ is participating, consequently giving information on the non-additive character of the model. The S_{T_i} S_{T_i} can be computed using

,

$$S_{T_i} = \frac{E(V(\mathbf{Y}|\mathbf{X}_{\sim \mathbf{i}}))}{V(\mathbf{Y})},$$

$$\mathsf{S}_{\mathrm{T_i}} = \frac{\mathsf{E}(\mathsf{V}(\mathbf{Y}|\mathbf{X}_{\sim \mathsf{i}}))}{\mathsf{V}(\mathbf{Y})}, \tag{9}$$

where $\mathbf{x}_{\sim \mathbf{i}}$ $\mathbf{X}_{\sim i}$ indicates that all factors have been fixed and only $\mathbf{x}_{\mathbf{i}}$ $\mathbf{X}_i$ varies over its uncertainty range. This approach permits, even for non-additive models, to recover the the recovery of the complete variance of $\mathbf{Y}$. The sum of S_{T_i} S_{T_i} is equal to one for perfectly additive models, otherwise it is always greater than one. The difference between S_i and S_{T_i} is aS_{T_i} is a useful measure of how much each factor is involved in interactions with any other factor (Saltelli et al., 2010).

The

First and total-order indices can be efficiently computed by Monte-Carlo based numerical procedures (Saltelli et al., 2010; Sobol et al., 2007). (Saltelli et al., 2010; Sobol et al., 2007). Estimating the conditional variances, such as $V_{X_i}(E_{X\sim i}(\mathbf{Y}|\mathbf{X}_i = x_i^*))$, is computationally expensive, but Saltelli et al. (2010) provide an efficient algorithm for the simultaneous computation of S_i and S_{T_i}. The calculation requires two independent sampling matrices $\mathbf{A}$ and $\mathbf{B}$, with the elements a_{ji} and b_{ji}. The subscript i runs from one to the number of factors k, while j runs from one to the number of samples N. A third matrix ,$\mathbf{A}_{\mathbf{B}}^{(\mathbf{i})}$, is introduced, where all columns are taken from $\mathbf{A}$, except for the i-th column, which is from $\mathbf{B}$. The first-order effect can then be computed as

3 Results

$$V_{X_i}(E_{X_{\sim i}}(\mathbf{Y}|\mathbf{X}_i = x_i^*)) = \frac{1}{N}\sum_{j=1}^{N} f(\mathbf{B})_j (f(\mathbf{A}_{\mathbf{B}}^{(\mathbf{i})})_j - f(\mathbf{A})_j), \tag{10}$$

2.1 Reference run

The reference run serves as basis for the uncertainty estimation of the simulation results (see Section 3.2), and the corresponding decomposition of the model variance (see Section **??**). The modified CROCUS model (see Section 2.1)is forced with the pre-processed and corrected input data introduced in Section 2.2. Most relevant model parameters are where $f(\cdot)_j$ denotes the model output of the j-th row. Similarly, total effect can be estimated by

$$E(V(\mathbf{Y}|\mathbf{X}_{\sim i})) = \frac{1}{2N}\sum_{j=1}^{N} (f(\mathbf{A})_j - f(\mathbf{A}_{\mathbf{B}}^{(\mathbf{i})})_j)^2. \tag{11}$$

The indices are estimated at a total cost of $N \cdot (k+2)$ model runs with N, a sufficiently large number of base samples. In this study, we performed 20000 model runs with $k = 8$ factors and $N = 2000$ base samples, which proved to be a reasonable compromise between computational feasibility and robustness of the results. The base samples were generated from quasi-random Sobol sequences (see Sec. 2.2). The Sobol sequence generates quasi-random numbers in a range between [0,1]. The random numbers are then mapped to match the uncertainty distributions given in Table 1. The initial snowpack is assumed to be isotherm with 273.16 , and a constant base temperature of 271 .

The maximum number of snow layers is set to 50 in order to get a detailed snowpack stratigraphy. The initial grid spacing increases from 0.01 at the surface to 10 at the bottom. The number of grid cells and their spacing is updated during 2 (see also Sec. 2.1). The roughness length for heat z_{h_0} is derived from the roughness length for momentum using the relation $z_{h_0} = z_0/10$ as its a default setting for Crocus. The snowpack model is forced with each of the simulation according to the accumulation , temperature, density and melt. The KNG8 is located in the accumulation zone of the glacier where the near surface layers consist of perennial snow rather than bare ice (Björnsson et al., 1996; Brandt et al., 2008). Following Björnsson et al. (1996) and Brandt et al. (2008), the model is initialized with an isothermal firn layer with amean density of 600 and a total thickness of 20.51 . The starting date is chosen to be the end of the ablation season, with the lowest recorded snow depth.Based on this initialisation set up, the one-year simulation period starts at 11th August 2010 and ends at 10th August 2011. The model is forced by hourly data, whereas results are saved every 6 hour for analysis. Measurements of surface temperature, shortwave radiation, albedo, 20000 parameter combinations.

Sensitivity indices are computed from the 6-hourly model output of these Monte-Carlo runs and are analysed with regard to snow depth, surface energy balance, turbulent heat flux and latent heat flux. The calculations are based on the reference runs performed at the two glacier sites. Therefore, this strategy allows for the study of the detailed temporal evolution and distinction of patterns during summer and winter and in different mass balance regimes of the glacier (accumulation and a snow pit profile in spring are available for validation. Note, that these data have not been used as model inputablation area), respectively. The accuracy of the sensitivity indices was assessed from 1000 empirical bootstrap samples being randomly drawn with replacement from the original dataset. The indices S_{T_i} are calculated for each of the bootstrap datasets, and the 95 % confidence regions are estimated.

2.1 Measurement error characteristics

The model uncertainty is estimated from a set of quasi-Monte Carlo sequences (see Sec. 2.2), based on the calibrated reference runs and specified uncertainty measures of key input factors and model parameters (Table 2). The probability density distributions of the measurement errors are either derived from simultaneous measurements with two sensors (as for air temperature) or by the accuracy of the sensor (given by the manufacturer specifications). When dealing with measurement errors, there is usually insufficient information on how the given uncertainties were determined and how the underlying distribution functions look. Regarding field applications, additional factors come into play that are usually not considered in calibration procedures. For example, temperature measurements may be affected by aging or insufficient shielding from solar radiation, both being crucial in glacier environments, too. To characterize the uncertainty of the measured meteorological parameters used to force the model, we follow a common approach and assign normally distributed errors considering the standard deviation derived from the manufacturer specifications. Roughness length and pore volume fraction are assumed to vary uniformly in a pre-defined range, which appears justified by observational evidence indicating high local and temporal variability of snow surface conditions due to, e.g., the formation of sastrugi or melt water conduits (Sauter et al., 2013). It therefore also seems reasonable to use a uniform range of pore volume fractions rather than assuming a truncated normal distribution.

2.2 Sobol sampling

Randomly drawn samples from a hypercube space tend to have clusters and gaps. Such sequences are said to have a high discrepancy. Low-discrepancy sequences, also known as quasi-random sequences, are designed to have well-distributed numbers in a multidimensional space, even for small quantities. Quasi-random algorithms bias the selection of points to maintain an even spread across the hypercube. These sequences are commonly used in sensitivity analysis and provide better estimates of the model-free sensitivity measures (see Sect. 2.4). Sobol sequences, which are used in this study, belong to this class of sequences (Sobol, 1998; Sobol et al., 2007).

3 Results

3.1 Reference run

Here we mainly examine the accuracy of the reference run at KNG8, which is representative of the accumulation area of the glacier and prevailing snow conditions. Validation of the reference run for KNG1 (representative of the ablation area of the glacier) reveals similar skills, and so we more or less forego a detailed description of those results. Comparison of the simulation at KNG8 with the snow pit profile from 6th April2010 shows a6 April 2011 shows a difference in snow depth at the end of the winter period of less than 0.1 m . The simulated mass gain amounts +0.76 2.

In terms of water equivalent, which corresponds approximately with the observed mass gain of +the accumulated mass during the winter amounts to +0.76 m, compared to +0.82. Figure ?? m having been observed. Fig. 2 also shows the comparison of simulated snow surface temperature with observational datacomputed . The simulated snow surface temperature is derived from upwelling longwave radiation assuming a snow emissivity of 0.99. Surface temperature is a key variable for flux parametrizations . The and also links the calculated mass and energy balance. Its temporal variability is well captured ($R^2 = 0.93$), and 95% of the absolute deviations are within ± 1.1 a root mean squared error of 2.3 K and conforms to the general skill of most sophisticated snow snowpack models (Obleitner and De Wolde, 1999; Rutter et al., 2009; Etchevers et al., 2004). The spread increases in the wintertime, whichmight in part be associated to , which, for example could be associated with undetected riming of the sensor or diverse structural model uncertainties. The vertical temperature gradient in the snowpack is an important driver of snow metamorphism and is depicted in Figure ??.Fig. 2. In the upper 0.6 m, the observed temperature is slightly higher than modelled and the RMSE=11.2 $^\circ\mathrm{C}$ is in part attributed to measurements shortcomings as well also attributed to measurement shortcomings (Obleitner and De Wolde, 1999). The corresponding density profile confirm confirms that the model is able to simulate the gross snowpack layering (see Figure ??Fig. 2). The relatively large difference within the upper 0.1 m is due to the fact , that the constant a_ρ in Eqn.4 Eq. (4) is set to 300 $\mathrm{kg\,m^{-3}}$. Although this leads to rather high fresh snow densities, the choice is justified when comparing the daily mean snow albedo (see Figure ??Fig. 2). Albedo here denotes the broad-band reflectivity of the snow surface, which is a key parameter determining net radiation. The RMSE of the between the measured and modelled albedo over the entire simulation period is 0.06. Albedo . Note that the measured albedo ranges between 0.65 in the ablation period and 0.92 in the accumulation period.

, which is characteristic for a site in the accumulation region (Armstrong and Brun, 2008; Greuell et al., 2007). Albedo is significantly depleted at the lower site during summer, as is typical for a site in the ablation area due to exposure of darker glacier ice, which has also been confirmed by Greuell et al. (2007).

Following we indicate some gross features of the seasonal evolution of the Table 3 gives a summary of the observed meteorological variables and the calculated energy balance components . The annual longwave radiation budget is negative on average (-18.7), with enhanced losses during early summer . The yearly average of net radiation is slightly negative (-1.7) . An enhanced energy deficit (-13.2) is observed during the accumulation period when the incoming shortwave radiation is zero due to polar-night conditions. The energy deficit by radiation is compensated by an effective average energy input of +4.3 from the turbulent sensible and at KNG8 and KNG1. We thereby distinguish values for consistent summer and winter periods covering the months JJA and DJF, respectively. These must not be interpreted as ablation or accumulation periods, which are, in fact, of different durations at both sites. Air temperatures decrease with elevation and remain negative all over the glacier during the considered winter period, while they are positive during the central summer months. This is basically reflected in the observed surface temperatures, which indicate that the glacier (snow) surface melts during JJA while remaining frozen during the DJF period. Bulk vertical temperature gradients between the 2 m level (nominal) and the surface indicate inversion conditions prevailing throughout the year. Humidity increases with elevation along the glacier as expected. The otherwise observed decrease of vapour pressure with elevation during the summer may be related to low-lying clouds (as suggested by longwave incoming radiation data). The local vertical gradients in vapour pressure are calculated by assuming saturation at the surface, and they reveal higher values in the air. This leads to positive latent heat fluxes . During the accumulation period more energy is lost by the strong negative radiation budget than gained by turbulent fluxes, which leads to an overall negative surfaceenergy balance (SEB, -3.7). In July, the SEB is stongly positive with +37.4 due to the radiation input (+34.3) and turbulent sensible heat flux (+4.5). In contrast, during the ablation season the turbulent latent heat flux is slightly negative (-1.44). In total there is a mean annual surplus of energy of about +2.67 . Karner et al. (2013) demonstrated for another site some providing mass and energy to the surface. Wind speeds are generally higher at KNG1, which is more pronounced during the winter months, when the air is more stably stratified. Katabatic winds play a role in this context, as is obvious from analysis of wind directions (shown in Karner et al. (2013)). With regard to the radiation components, there is virtually no input from solar radiation during the winter months. During summer, global radiation, i.e. the sum of direct and diffuse solar radiation, increases by about 5 $\mathrm{W\,m^{-2}}$ per 100 m below KNG8 (see Figure 1, that the 10-year average surplus is elevation. About 80% of incoming solar radiation is reflected at the higher site (KNG1) during the summer and reflects the persistence of snow. An albedo of about +9.5 . The pronounced local differences in 48% is calculated for the lower site, where snow disappears early in the SEB componentson Kongsvegen emphasizes that the results of this analysis cannot be generalized , which imposes the need considering characteristic zones on the glacier separately.

3.2 Uncertainty estimation

The integrated model uncertainty for snow height is estimated from a set of Monte-Carlo runs, based on the reference run and specified uncertainty measures of key input factors and model parameters (Table 2) . The probability density distributions of spring and exposes darker glacier ice at the surface. Longwave incoming radiation is an important source of energy during both seasons. Its increase with elevation during the winter and the measurement errors are either derived from simultaneous measurements with two sensors, as in case of air temperature measurements, or by the accuracy of the sensor given by the manufacturer specifications. Dealing with measurement errors, there is usually no information on how these uncertainties are distributed and it is not always obvious which uncertainties are taken into account by the manufacturers. In addition, other sources of uncertainty such as aging or radiation effects on temperature sensors are usually not known, but can play a crucial role. Except for the roughness length and the pore volume fraction which are assumed to vary uniformly in the pre-defined range, we follow the common approach and assign normally distributed errors with the standard deviation given by the sensor's accuracy. The uniform distribution of the roughness length is justified by the fact, that throughout the uppermost parts of the Kongsvegen the spatial distribution of snow is strongly influenced by snowdrift that results in frequent sastrugi formation (wind induced dunes) and high local-scale and temporal variability of surface roughness (Sauter et al., 2013). It seems also reasonable to use a uniform range of pore volume fractions rather than assuming a truncated normal distribution. From the distributions a low-discrepancy Sobol sequence has been generated with a total number 16000 ensemble members (Saltelli et al., 2006). These sequences are commonly used in sensitivity analysis and provide better estimates of the model-free sensitivity measures (see Section **??**)decrease during the summer reflects corresponding changes in cloud characteristics (low level fog in winter). These characteristics of the radiation components induce a decrease of net radiation with elevation, with overall negative values during the winter and positive ones during the summer. Sensible heat fluxes are generally directed towards the surface, which is more pronounced during the winter and at the lower site KNG1. This is also true with regard to latent heat fluxes, which by magnitude equally contribute to the calculated surface energy budget. The latter is characterized by negative values during the winter, when small gradients along the glacier occur. During the considered summer months, the energy budget is strongly positive and fosters melt at both sites. Naturally, this is more effective at the lower site, which mainly can be traced back to stronger input from solar radiation (lower albedo) and turbulent fluxes. There were corresponding developments of the mass balance at both sites. Note that further energy and mass balance components were calculated by the simulations, which on average were small by magnitude. Therefore, they are not considered in the overall context of this study, which does not aim at a detailed investigation of the individual fluxes and associated processes themselves.

Figure ??

3.2 Integrated model uncertainty

Fig. 3 shows the time series of snow depth for the reference run as well as of the quantiles estimated from the ensemble simulations . The for KNG8 and KNG1. At KNG8, the 95% quantile range can be clearly divided into two regimes: (i) the build up of the snow pack when the 95% interquantile interquartile range increases towards ± 1.2 ± 1.2 m until the end of June, and (ii) the melt period when the interquantile interquartile range experiences an additional increase. At the end of the one-year simulation period, the uncertainty (95% quantile range) in snow depth caused by the systematic measurements errors reach measurement errors reaches more than 3 m. Note , that the interquantile that the interquartile range shows a clear asymmetry which is more pronounced after June 2011. At this time the snowpack contains higher fraction of liquid water which decreases the albedo and increases the compaction by This marks the onset of effective melt, which induces a higher liquid water content of the near-surface snow layers. The associated wet snow metamorphism . Obviously, the system becomes more sensitive once the old firn i.e. drives a decrease in albedo. The development is enhanced upon exposure of snow from the previous year, with higher densities and lower albedo , re-appears at the surfacewhich is characterized by even lower albedo and higher density. Sporadic snowfall events (depending on the temperature threshold) in Augustin August 2011 also lead led to an increase of the upper 99% quantile bound. The simulation is also very sensitive in simulations are also quite sensitive to disturbances during the first two months, when the amounts of snowfall are small. Then, uncertainties in the input quantities are decisive whether the new snow remains on the ground or disappears. This indicates a high sensitivity to the model's treatment of new snow processes (deposition and wind induced erosion). The overall evolution and the final characteristics of the ensemble variability at KNG1 are similar to that at KNG8. Note, however, that the accumulation season is significantly shorter (beginning in November compared to August at KNG8) and is characterized by a smaller ensemble spread compared to KNG8. The latter reflects that precipitation at this elevation mostly comes in the form of rain. Throughout the accumulation season, the ensemble spread is low and is related to small snowfall amounts and widens significantly in the ablation season, when the glacier reappears at the surface. The point at which the glacier ice reappears depends on the maximum snow depth and can occur between May and the beginning of July. The final uncertainty (95% quantile range) in snow depth due to measurement errors is almost 5.5 m at the end of the ablation season.

While the Monte-Carlo runs offer a good and practical way to quantify the model uncertainty regarding snow height simulations, it provides no qualitative information on the contribution of each input factor. We should also keep in mind, that all factors are independently varied at the same time and interactions are likely to be important. This issue is addressed by taking advantage of the ensemble runs and further decompose the ensemble variability by GSA.

3.3 Mean total-order sensitivity indices

3.4 Decomposition of the model uncertainty

To understand the contribution of individual factors to the ensemble variability, the complete sensitivity pattern need to be considered. In the following section, different sources of uncertainty are estimated using the variance-based GSA method introduced in Section 2.4. For all factors the first- and total-order indices are calculated. Figure **??** Fig. 4 shows the mean contribution of the S_{T_i} of individual factors on the variability of calculated snow depth changes variance of calculated surface height changes (SHC), surface energy balance (SEB), and the turbulent heat fluxes for three month periods which roughly correspond to seasonsthe turbulent sensible (SHF) and latent heat (LHF) flux at KNG8 and KNG1. Recall that first-order indicesS_i measure individual factor contributions total-order indices, S_{T_i}, measure the contribution of each factor to the ensemble variance, while the total-order indices S_{Ti} also include including all interaction effects.

The results show that first-order impacts on calculated snow height are dominated by uncertainties of

At KNG8, SHC is mainly affected by uncertainties in precipitation P (0.58) and incoming longwave radiation LW (high S_i values0.29). The remaining factors are very likely to have little impact. In the period from May to October, the SEB, SHF and LHF are most sensitive to LW explains 50-60% of the variance, while , with S_{T_i} values ranging from 0.53 to 0.77. Of note is the sensitivity of SEB to P explains around 35-45%. During the accumulation period precipitation becomes the dominant factor and shows first-order indices between 55-70%. Over the entire simulation period, individual variables account on average for 93% (sum of first-order indices) of the total ensemble variance , and thus the remaining 7% is due to interaction effects. In (0.25) and z_0 (0.4). Hence, z_0 is the second-most important parameter for SEB and SHF, and it even explains most of the LHF variance (0.27). A smaller share of SHF and LHF uncertainty is explained by U and RH. In particular, RH is important for LHF. In order to make an important contribution to the ensemble spread, the total-order indices should exceed the 0.05 limit (Saltelli et al., 2006). Following this criteria, some factors (T, Q, SW and $PVOL$) can be designated as insensitive and with little influence on the simulated snow depth changes. Moreover, there is a clear evidence that uncertainties in SHC, SEB, SHF, and LHF. The averaged first-order indices vary between 0.66 and 0.82, depending on the considered model output (see Fig. 4). The sensitivity pattern at KNG1 differs from that at KNG8. SHC is sensitive to LW, P and z_0. In contrast to KNG8, P has less influence on SHC variability than LW by far comprise most to the uncertainty in calculating the SEBcomponents (see Figure **??**) at KNG1. In total, the model is less affected by the uncertainty of z_0. However, RH ($S_{\mathrm{T}_i} = 0.1$) explains slightly more of the LHF variability at KNG1 than at KNG8.

Surprisingly shortwave radiation

3.4 Temporal evolution of the total-order sensitivity indices

Fig. 5 and 6 show the temporal evolution of the S_{T_i} values with respect to SHC, SEB, SHF, and LHF. The S_{T_i} values are calculated for each time step using the 20000 Monte-Carlo runs.

The variability of SHC at KNG8 is mainly caused by the uncertainty of P and LW. From November to May, almost all uncertainty is attributed to P, with S_{T_i} ranging between 0.7 and 0.9. During the ablation season, LW becomes a dominant factor. Other factors, such as U, z_0 and SWonly exceeds the 0.05 limit in spring, while in summer values are very low. In this period U and , make less of a contribution (< 0.2) to overall SHC variability, even though they can have an intermittently strong impact (> 0.3) on the variance of SEB, SHF and LHF. Errors in LW and z_0 are the only factors besides the have a strong impact on SEB all year, while P is only relevant in the winter season. During the spring, SW has an increased influence on SEB and drops to zero during the arctic winter. RH and U contribute most to SEB variability in the period from August to March. Along with LW with noticeable impact on the model uncertainty, U and z_0 have a significant effect on both SHF and LHF variance. The uncertainty in T and $PVOL$ do not have an influence on either SHC or SEB.

The mean seasonal indices can be somehow misleading and impacts might be underestimated in some cases. For example, according to Figure **??** one might conclude that the At KNG1, the contribution of P and LW is lower and not as consistent as at KNG8. In August and September, z_0 has hardly any impact on snow depth changes and even little effect on the SEB. Having a closer look at the temporal evolution of the indices derived for the SEB (see Figure **??**)temporarily contributes (up to 0.8) to the SHC variability. In contrast to the sensitivity pattern at KNG8, other factors (z_0, however, provides some interesting insights. In the summer season sporadic episodes of strong wind events lead to sudden jumps of the first-order indices of U, RH, SW) contribute substantially to SHC. SEB is by far most sensitive to errors in LW. SW gains importance for a short period in May, with S_{T_i} up to 0.4, although most of the time the contribution is very low, which is also true for RH, T and $PVOL$. In general, the sensitivity pattern of SHF and LHF is similar to the pattern observed at KNG8. Here again, z_0 and U, in which these factors explain together up to 50% of the total model uncertainty. U temporarily explain a large share of the variability in turbulent fluxes (> 0.75) in the summer. Errors in RH mainly impact LHF variability.

4 Discussion

For the following discussion we like to remind, that measurement uncertainties are independently sampled and do not possess any correlation structures. Consequently, the approach can not be used to investigate the response of snow or ice depending on e. g.changes in the environmental (climate)conditions. There, some factors show strong coherences, such as LW and T. In order to study climate sensitivity, the input factor set needs a more sophisticated sampling strategy to obtain the same correlation structure as those observed in nature. We investigated the seasonal pattern of the sensitivity of snow model output to uncertainties in input data and some key model parameters. Eight metrics characterizing forcing uncertainties (LW, P, $PVOL$, RH, SW, T, U and z_0) and four metrics characterizing the model response (SHC, SEB, SHF and LHF) have been considered. The introduced uncertainties represent the typical measurement errors of data used to force the model. The presented results are based on Monte-Carlo simulations and subsequent application of Sobol's sensitivity analysis to decompose first- and higher-order effects on the resultant variance. Simulations and analysis were applied to two sites at an Arctic glacier to address characteristics in different mass balance regimes.

However, The results from the reference simulations at the two sites allow for an in-depth discussion of the typical meteorological conditions at the two study sites and the related surface exchange processes which are reflected in the constellation of the energy and mass balance components. Such a study has not been performed at the Kongsvegen glacier thus far, but related aspects will nevertheless not be discussed in more detail here due to the rather methodical outline of this study. Note, however, that Obleitner and Lehning (2004) and Karner et al. (2013) have already studied this issue at another site close to the average equilibrium line of the glacier (ca. 537 m asl.). This location is only about 137 m below KNG8, but the results concerning energy and mass balance are not directly comparable because the sites are located in different glaciological regimes (equilibrium line altitude vs. accumulation area). Some common features may be inferred from Table 3 though, which in part is addressed in Sec. 3.1. Consideration of sites other than KNG6 was mainly motivated by the availability of correspondingly suitable data. It is to be noted in this context that for the purpose of this study, the reference runs were not fully calibrated towards the observations, which no doubt would have been possible for investigations in other directions. The overall results of this work show that on average about 80% of the decomposition of the model uncertainty by GSA turned out to be an efficient way to provide an enhanced understanding of the model's sensitivity patternin response to input and model parameter uncertainties. The results are very helpful to establish priorities in research to constrain influencing factors which need to be measured more accurately in order to reduce the total model uncertainty total variance of SHC and SEB can be explained by first-order effects (Fig. 4). This means that the remaining 20% of the variance is due to non-linear interaction effects. There is no significant difference between the two sites at the glacier. This is in partial contrast to the findings of Raleigh et al. (2015), who performed similar investigations for different snow regimes and found that first- and total-order indices are of comparable magnitude. However, the results cannot be directly compared because they analysed different model output variables and used a simpler (i.e. bulk model), which possibly enhances the interaction effects. The performed sensitivity analysis further demonstrates that the considered model output metrics respond most sensitiveley to uncertainties in the forcings of longwave incoming radiation, precipitation and surface roughness (Figs. 4, 5, and 6). According to the analysis, about 93% of the ensemble spread can be explained by linear effects (first-order) , while the remaining part is due to factor interactions. The results clearly proof, that linear methods such as sigma-normalized derivatives are insufficient to recover the entire variance as they neither account for interactions nor for non-additivity.

In some cases this could lead to an underestimation of the factor's importance, and wrong conclusions may be drawn. As shown by this study, first-order indices may be very close to zero, but they still can make an important contribution to the model's variability by interactions. Based on the GSA outcomes, the following conclusions can be drawn for this specific high Arctic site: Considered in more detail, however, each of these three factors exerts specific footprints depending on season and site, which will be discussed, along with the occasionally emerging impact of the remaining factors. As far as is possible, we try to relate the statistical findings to physical processes in the near-surface snow layers. Longwave incoming radiation depends on column integrated air temperature humidity and cloudiness and is the dominant source of energy for the glacier, independent of site and season. This is typical for glacier environments (Greuell and Smeets, 2001) and is enhanced in the Arctic, where input from shortwave insolation is missing during the polar night conditions (e.g. Obleitner and Lehning, 2004; Van den Broeke et al., 2011; Karner et al., 2013). Variability in LW therefore directly impacts NR and hence SEB. This also holds true for corresponding measurement uncertainties, which are comparatively large. The traces in the sensitivity analysis showing that about 50% of SEB variance can be explained by total-order effects due to LW (Fig. 4). The effect is slightly reduced at the higher site (KNG8), which may be related to the general decrease of longwave incoming radiation with elevation (Tab. 3). Neither study site shows a pronounced seasonal variability in the corresponding SEB sensitivity pattern, which may be related to the rather continuous nature of longwave incoming radiation and its dominance for NR (Figs. 5 and 6). LW uncertainty also strongly impacts on the variance of the calculated turbulent fluxes. Yearly averaged total-order indices are somewhat lower than for SEB, ranging at about 0.3 (KNG8) and ca. 0.5 (KNG1), respectively. The sensitivity analysis further reveals a stronger impact on SHF and an outstanding seasonal dependency of the sensitivity of the turbulent fluxes (Figs. 5 and 6). Feedback related to surface temperature provides a key for understanding these features, which couples the (longwave) radiation budget and the turbulent fluxes. The stronger input by longwave radiation, the more positive NR is, which in part is absorbed at the surface and increases surface temperature. Hence, surface temperature fluctuations are larger than those of air temperature and respond very sensitively to changes (uncertainties) in LW. This in turn effectively changes the stability of the near-surface air, which drives turbulent exchange therein. This feedback is most effective during dry snow, i.e. winter conditions, with large total-order sensitivity indices from autumn until spring (Fig. 5). In the ablation season, when the surface temperature is more or less at the melting point, SHF and LHF are no longer sensitive to uncertainties in LW. LHF is affected too (though to a lesser extent) because of the associated changes in vapour pressure at the surface. LW also strongly impacts SHC variability, which is more pronounced at KNG1 (Fig. 4) and during the summer (Figs. 5 and 6) when LW uncertainty explains more than 80% of SHC variability. This might be related to links between precipitation, cloud cover and snow metamorphism. For example, LW flux increases during snowfall events and supplies more energy to the snow surface. Crocus sets the temperature of freshly fallen snow to the surface temperature. The inherent metamorphism laws describing the evolution rate of the type and size of the snow grain depend on the vertical temperature gradient (Vionnet et al., 2012), and a steeper gradient might accelerate the evolution of the grains and the compaction of the snow pack. A further issue is that the specified input uncertainties (±10%) are larger when the LW flux is greater and thus also during snowfall events. To put these findings in a broader context, Karner et al. (2013) applied another snow model to data from KNG6 (Fig. 1) and also identified LW uncertainty as the most influencial factor on calculated mass balance and SEB. However, their study is based on consideration of single-order effects only. Another hint regarding the outstanding influence of uncertainties in LW is provided by Raleigh et al. (2015), who systematically explored the propagation of forcing uncertainties to snow model output based on Sobol's sensitivity analysis. Their results confirm the importance of LW uncertainty, but a straightforward comparison to our results is hampered due to the different metrics used for input uncertainties and model output.

Precipitation. Precipitation measurements are usually fraught with large uncertainties either by Uncertainty in the quantification of precipitation for input to the simulations is another influential factor on the variance of snow model output. This mainly concerns the simulated surface height changes (which is considered as a metric of calculated mass balance) and surface energy balance. Total-order sensitivity indices are particularly high during the winter and at KNG8 (Fig. 4).

In these higher regions of the Kongsvegen glacier, recurrent snowfall events may occur year round, which results in a deeper snow pack (2.2 m) and a longer accumulation period (October through April). This is evident from Fig. 3 and the corresponding SHC sensitivity patterns. Snowfall occurs comparatively infrequently and is overall inefficient at the glacier tongue and during the summer months. This is mainly an effect of temperature lapse-rate determining the rain-snow transition and the tendency of cloud formation at the crest of mountains. Similar to Raleigh et al. (2015), we find that P uncertainty is a critical factor for the snow disappearance in the ablation zone (see Fig. 3). Depending on the winter conditions, the reappearance of glacier ice typically occurs between May and July. However, we find little evidence that ablation rates are significantly controlled by P. Note that in our study, precipitation was derived from ultrasonic sensors and corresponding uncertainty was specified from the manufacturer specifications. Frequently, however, snow precipitation is derived from standard gauges. As previously mentioned, even small errors due to wind-induced under-catch , or by the or conversion of snow depth changes to precipitation rates in terms of SWE (see also Section 2.2) Sec. 2.2) might thus have a significant impact on the simulations. According to Eqn.4 Eq. (4), the conversion is sensitive to air temperature ($\partial\rho/\partial T_{air} = b_\rho$$\partial\rho/\partial T_{\text{air}} = b_\rho$) and wind velocity ($\partial\rho/\partial U = c_\rho/(2\cdot\sqrt{U})$). Obviously, the fresh snow density calculations are in particular $\partial\rho/\partial U = c_\rho/(2\cdot\sqrt{U})$). This demonstrates that the conversions are particularly sensitive to measurement errors at low wind speed. As shown in Section **??** the input uncertainty related to precipitation has a strong impact on the calculated snow depths all year. Increasing the accuracy of the measurements would drastically (by 50-70%) reduce the uncertainty in the accumulation season, and even by 30-50% in the ablation season. Schmucki et al. (2014) showed , However, precipitation measurements at higher wind velocities usually show a systematic under-catch. Schmucki et al. (2014) showed that for standard precipitation measurements, a correction of under-catch may reduce the mean absolute percentage error by 14% for snow depth at high alpine stations. Førland and Hanssen-Bauer (2000) demonstrated the importance of this issue for Svalbard environments. The SEB results are also strongly affected by uncertainties in the specification of precipitation input, which explains about 25% of the total variance. Basically, this may be related to changes in the physical properties of snow during precipitation events. For example, fresh snow is characterized by dendritic structures, which are parameterized in Crocus and have a direct impact on, e.g., albedo and hence on net radiation. The projected impact on SEB variance is strongest during the winter season due to the more frequent snow fall events. At the lower part of the glacier, fresh snow events are infrequent and inefficient. During the summer in particular, it usually melts within a short period without leaving a significant impact on SEB, which can explain the correspondingly low impact of the corresponding uncertainties. At both stations, turbulent fluxes are only little sensitive to uncertainty in P forcing. This indicates that the contribution of P on SEB is mainly due to interactions with LW via cloud cover.

Our analysis reveals that uncertainty in the specification of surface roughness has a strong impact on the turbulent fluxes and hence on SEB variances. Overall, this is due to the associated processes and their parameterizations (Vionnet et al., 2012). The sensitivity is particularly pronounced regarding SHF and at the upper study site (KNG8), where total-order sensitivity indices reach 0.3 on average throughout the year (Fig. 4) and are highest during the period from April until June. As this period corresponds to the late accumulation period, the findings may reflect an influence of wind drift, which is certainly more pronounced at the upper site. Interestingly enough, KNG1 experiences the most pronounced impact of z_0 uncertainty on SEB during the period from July until September, which constitutes the main ablation period at this site (Fig. 3). This feature is attributed to the concurrent appearance of bare ice and the accordingly parameterized increase of surface roughness, which represent the formation of, e.g., melt water channels. Uncertainty in z_0 also impacts the simulations of SHC, which is most pronounced at the lower site and during the summer. This again is related to the overall increased roughness (factor 10) and accordingly enhanced turbulent fluxes contributing the surface melt. These findings basically conform to the first-order sensitivity studies by Karner et al. (2013), changing z_0 by an order of magnitude. However, as a straightforward comparison is difficult due to the choice of error range, which can have a strong influence on the results (Raleigh et al., 2015). The impact of the specified uncertainties in the basic meteorological forcing data (U, too. Snowfall eventsare less frequent in summer time due to the temperature dependence (interaction with temperature), and thus lead to a drop of S_i values. However, episodic snowfall events in summer temporarily do have an impact on the SEB T, but the overall contribution is low.

Longwave radiation. Weather stations rarely directly measure the longwave radiation, RH and the flux often needs to be parametrized by measured quantities such as temperature, humidity, shortwave radiation or cloudiness. The uncertainty in long-wave incoming radiation determines 80-87SW) on the considered model output metrics is small overall. On average, such uncertainties can explain more than 10% of the ensemble variance of the SEB and only a minor contribution comes from the remaining factors. The is mainly due to total variance (Fig. 4), and there is no significant difference between the two study sites. In general, SHC is less affected compared to SEB, which is reasonable considering the role of those input data in the parameterizations of the associated processes. Regarding the seasonal sensitivity patterns, however, each factor can have an episodically strong impact. Hence, the specified U uncertainty impacts significantly the calculated turbulent fluxes during the period from April-June (KNG8) and July-September (KNG1). It is notable that the latter periods correspond to those when z_0 uncertainties exert the most influence, indicating combined effects. We therefore attribute their impact on SEB and SHC mainly to their direct involvement in the strong link between LW and the snow surface temperature, which in turn directly affects the calculation of the turbulent fluxes . Between 60-85% of and the parameterization of changes in grain shape due to drifting snow. Recall that Crocus takes the latter into account using a driftability index. The largest sensitivity of U is associated with lower wind speeds (Tab. 3). This conforms to the findings of Dadic et al. (2013), who found higher sensitivity of the turbulent fluxes with respect to wind speed in the range of 3–5 $\mathrm{m\,s^{-1}}$. The effect of local wind velocity variations on turbulent fluxes and ablation rates has been also addressed by other studies (Mott et al., 2013; Marks et al., 1998). Air temperature may be expected to strongly influence the calculation of the turbulent fluxes and, therefore, SEB. However, this is not seen in the results of the sensitivity analysis, which at both stations does not show significant impacts on any of the considered model output metrics (SHF, LHF, SEB and SHC). Further, this result must be considered in light of the variances rather than absolute values. The driving temperature gradients between the surface and the air are in the order of 2-5 K (Tab. 3), which reduces the sensitivity of the calculated fluxes due to the uncertainty in sensible heat flux and 40-65% of the latent heat flux can be attributed to errors in comparatively small measurement errors that have been assumed ($\pm$0.3 K). Further, Raleigh et al. (2015) found that T-forcing biases had a stronger impact on ablation rates (which may be considered as measure of summer SEB) compared to random errors, while peak snow water equivalent (comparable to winter SHC) was hardly affected. Similarly, Karner et al. (2013) found a strong impact of T-biases on the calculated SEB. The seasonal T-sensitivity patterns on SEB and its components are characterized by relatively strong impacts in the spring and autumn. During this period, temperature is crucial whether precipitation is considered as snow or rain. Feedback related to albedo or LW (see Figure **??**). Better estimates can be expected using measured snow surface temperature as direct model input, as suggested by Lehning et al. (1999). Depending on the application, such replacement of prognostic variables by observations may be considered as a methodical step backwards. While the SEB is very sensitive to LW throughout the whole year, its impact on snow depth changes shows a pronounced seasonal cycle. This cycle is related to the variations in the LW mean intensity, varying from $255\ Wm^{-2}$ in summer to $226\ Wm^{-2}$ in winter.This also emphasizes the importance of LW for melting processes which hitherto has been underestimated generally may play an additional role there. The sensitivity study was performed based on standard, i.e. laboratory specifications given by the manufacturers. However, the actual uncertainties of air temperature measurements can be much larger depending on, e.g., the efficiency of the used radiation shields or ventilation devices. Relevant to this study, Karner et al. (2013) did not find significant biases between measurements employing ventilated or unventilated sensors. However, this result may not be considered as generally valid and corresponding corrections are proposed.

Shortwave Radiation. During the arctic winter shortwave radiation is zero, and so are the first-order influences. The only noticeable contribution is observed in the period from May to July with S_i values up to 4%. Indeed, this makes SW the second most important The impact of humidity forcing errors on the simulation metrics was analysed concerning the directly measured variable (relative humidity). By definition, however, the latter combines humidity and temperature information and is therefore not an ideal metric, which may be considered in forthcoming studies. Irrespective of that, our results reveal that on average RH uncertainty has an overall small but somewhat stronger impact on calculated SEB compared to, e.g., U (Fig. 4). The impact is less pronounced at the lower site and during the summer (Figs. 5 and 6). The overall variability of the seasonal SEB sensitivity pattern is small, however, and is difficult to interpret due to the inherent temperature effects. There are indications of stronger impacts in the spring when conditions are favourable for sublimation due to high saturation deficits occurring simultaneously with strong winds and moderate temperatures (Sauter et al., 2013; Obleitner and Lehning, 2004; Karner et al., 2013). Calculated RH-sensitivity is generally stronger regarding LHF compared to SHF, which is reasonable.

Shortwave incoming radiation is a strongly influential factor on the SEB in summer, but its impact on SEB is too little to have a significant influence on the calculated snow depth changes. This is not in line with former studies (Karner et al., 2013) and contrasts intuition. The reason for that can be deduced from a simple analysis, whereby the energies of snow and ice, and corresponding uncertainties are expected to have a strong impact on respective simulations. Contrasting this general anticipation, our sensitivity analysis reveals that on an annual basis, only a small amount of the total SEB variance can be explained by the assumed uncertainty of SW input data (Fig. 4). This concerns both sites and basically reflects that in the Arctic, the anticipated effect is generally reduced due to the lack of solar insolation during winter. Recall that in the Kongsvegen environment, the polar night conditions last from late October to early February. This is also reflected in the seasonal sensitivity pattern, which do not show any signal during the winter. There is, however, a significant influence on SEB in the spring and autumn (Fig. 5 and 6). This might be related to the previously mentioned influence of intermittent fresh snow on older surfaces with lower albedo, whose effectiveness also depends on SW and its variability (and temperature as addressed above). Another reasoning is based on the consideration of energy supplied by uncertainties in LW and SW measurements are put in relation. The sensitivity of the compared to those in LW. Hence, the sensitivity of net shortwave radiation (∂Gdue) to measurement errors ∂E_{SW} (∂E_{SW}) is given by $\partial G/\partial E_{SW} = 1-\alpha$$\partial G/\partial E_{\mathsf{SW}} = 1-\alpha$, with α denoting albedo. Obviously, the effect on the net shortwave radiation flux by small errors in the measurement is solely a function of the albedo. The ratio Therefore, the ratio ($R$of the sensitivties of the) of the sensitivities of incoming longwave radiation and the available available net shortwave radiation at the ground is therefore $R = 1/(1-\alpha)$. By multiplying R with the error ratiowe obtain the , we obtain a properly scaled ratio $\hat{R} = (E_{LW}/E_{SW})\cdot(1/(1-\alpha))$$\hat{R} = (E_{\mathsf{LW}}/E_{\mathsf{SW}})\cdot(1/(1-\alpha))$. Assuming a 10% error of typical daytime values in summer ($E_{SW} = 40$, and $E_{LW} = 26$ the summer ($E_{\mathsf{SW}} = 40\,\mathrm{W\,m^{-2}}$, and $E_{\mathsf{LW}} = 26\,\mathrm{W\,m^{-2}}$) and a $\alpha = 0.75$, we obtain $\hat{R} = 2.6$. This means the energy supplied by that the changes in energy due to the measurement uncertainty of LW is are about 2.6 times greater than the energy supplied by measurement uncertainty of that for SW. In the spring and autumn, the ratio becomes larger due to an increasing albedo and decreasing incoming shortwave radiation. This also leads to the conclusion , that increasing the accuracy of SW measurements by a few percent would not increase our confidence in simulations of snow depth or the SEB components.

Temperature. Although the turbulent heat flux is parametrized by measured air temperature differences between the observation and the snow surface temperature, small measurements errors ($\pm$0.3 K) have almost no impact on the calculated turbulent fluxes, and hence on calculated snow depth changes. In part this may also be related to a negative feedbacks. Thus, higher air temperature induce enhanced energy transport towards the surface, leading to higher surface temperature. The latter is effectively counterbalanced by enhanced emission of longwave radiation. The only amplifying interaction is most likely with precipitation when temperatures are close to the phase transition threshold. Notable however, measurement uncertainties can be much larger using e.g. less effective (i.e. unventilated) radiation shields for the measurement of air temperature which is still common practise (Karner et al., 2013; Smeets, 2006).

Humidity. The turbulent latent heat flux is parametrized by the difference of the atmospheric humidity in the surface layer and the saturation specific humidity above the snow surface, which is a function of the snow surface temperature. The weak seasonal variability of total-order indices (see Figure **??** upper panel) can be The sensitivity of SHC on uncertain specification of shortwave radiation SW is negligible overall, except in summer the latter being more pronounced at the lower site. This again reflects a coupling to albedo, which is lower at KNG1. The results conform to Karner et al. (2013) showing that the overall influence of SW is strikingly smaller compared to that of LW. As was pointed out by Raleigh et al. (2015), overall, this is attributed to the interplay between saturation deficit, temperature and wind speed. Particularly in spring, conditions are favourable when high saturation deficits occur simultaneously with strong winds and moderate temperatures (Sauter et al., 2013; Obleitner and Lehning, 2004; Karner et al., 2013). Nevertheless, values are very low and a better accuracy would not reduce much the ensemble spread of the snow depth simulations.

high albedo of snow (reducing absorbed energy and the associated impact of uncertainties) and the non-linear (amplifying) interactions of LW, which through surface temperature is coupled to the calculation of the turbulent fluxes.

Windspeed and roughness length. As discussed in Section **??**, the mean sensitivity measures are not very meaningful for U and z_0. Both, the mean S_i and S_{Ti}, are rather low, but temporarily the factors turn out to be most dominant as shown in Figure **??**. The accuracy of both factors are decisive for the estimation of the turbulent fluxes. Together, the quantities explain about 20% of the uncertainty in the sensible heat flux in summer, and more than 35% in latent heat flux in winter. More accurate measurements of both quantities could reduce the ensemble spread by almost 8-10% in the period from August to January. The largest sensitivity is associated with low wind velocities. This lines up with the finding from Dadic et al. (2013), who found highest sensitivity of the turbulent fluxes with respect to wind speed in the range of 3-5 . Furthermore, the effect of local wind velocity variations on turbulent fluxesand the net melt calculations have been demonstrated by several other studies (Dadic et al., 2013; Mott et al., 2013; Marks et al., 1998).

While the turbulent fluxes are sensitive to uncertainties of both z_0 and U all year, the impact on the SEB almost vanishes in the summer time due to different signs of the turbulent fluxes (see Section 3.1).

Maximum liquid water holding capacity. Liquid The liquid water holding capacity of snow is difficult to measure and $PVOL$ strongly depends on snow microstructure and related surface/subsurface developments throughout the winter season. Fortunately however, our results indicate that the , and it is difficult to measure (Armstrong and Brun, 2008). However, investigation of the propagation of corresponding uncertainties in the snow model results was hardly addressed and therefore was considered in this study. According to our results, the uncertainty in specifying liquid water holding capacity of snow makes only a small the least contribution to the total model variance of virtually all considered output metrics, mainly by interactions. In fact, total-order indices are slightly higher in the melting season but the overall The seasonal $PVOL$ sensitivity pattern reveals some enhanced impact on SEB and snow depth changes is negligible.

Karner et al. (2013) and Obleitner and Lehning (2004) likewise estimated the effect of measurement uncertainties on the energy and mass balance at KNG6 on the Kongsvegen glacier (see Fig. 1). In contrary to our findings, they identified SW and T to be very influential factors for the SEB. U and z_0, on the other hand, had no variability in the spring and autumn, which is more pronounced at the upper site (KNG8). Tentatively, this feature could is attributed to the percolation of rain or melt water and subsequent refreezing. However, it remains to be investigated whether the associated release of energy can explain the observed variance pattern. Gascon et al. (2014) remarked that the Crocus percolation scheme tends to favour near-surface freezing and insufficient refreezing at depth, which could be another factor in this context. Overall, the assumption of default values (as in this study) does not have a significant impact on the model's uncertainty. However, their estimates were based on consideration of plain first order effects and are therefore not directly comparable to the results given here. It is nevertheless important to note that different sensitivity patterns are likely to exist at different elevation zones of individual glaciers. Further investigation of this issue was beyond the scope of this work

calculated mass-balance (SHC).

5 Conclusions

As this study shows, conservatively estimated measurement errors We investigated the seasonal pattern of the sensitivity of snow model output to uncertainties in input data and some key model parameters. A set of eight metrics characterizing forcing uncertainties and four metrics characterizing the model response have been considered. The introduced uncertainties characterize typical measurement errors of data used to force a state-of-the-art snow model, and the presented results are based on Monte-Carlo simulations and subsequent application of Sobol's GSA. Simulations and analysis are applied to two sites at an Arctic glacier to address characteristics in different mass balance regimes. The results clearly demonstrate that even conservatively estimated input uncertainties can lead to a significant loss of confidence in snowpack simulations. In our example, the 95% interquantile range of key simulation results concerning the surface energy and mass budget. The overall impact of individual error sources on the sensitivity pattern varies for different zones on the glacier. In the accumulation zone, precipitation and longwave radiation are key factors for the evolution of the snowpack and contribute most to the model uncertainty. The significance of precipitation variability decreases with altitude, while other factors, such as wind velocity or surface roughness, gain importance. Uncertainties in the measurement of incoming shortwave radiation and air temperature have little influence on the model outcome, the former being biased by the specific, i.e. Arctic, conditions. The calculated seasonal sensitivity patterns are similar overall at both study sites. The most temporally continuous influence on model output is exerted by variance of longwave radiation and surface roughness. Precipitation tends to have the strongest impact during the winter, and the ensemble members showed a spread of approximately 3 at the end of the simulation period, solely caused by key input and parametrization errors. For example, accurate observations of snow depth changes or associated water equivalents are in the rarest cases available. In remote areas scientist usually rely on snow depth measurements by ultrasonic sensors. Unfortunately, this kind of observation has some unavoidable disadvantages: Firstly, these measurements are affected by blowing snow, intense snowfall, or extreme temperatures; and secondly, snow depth changes need to be converted to snow water equivalent. Besides the inherent errors by the sensor itself, the environmental boundary conditions introduce a considerable amount of noise, which needs to be reduced. Small-scale fluctuations are usually reduced by filtering techniques, or sometimes even by more sophisticated approaches.Nevertheless, the accuracy of automatic observations will always contain a significant amount of uncertainty, and it is remains difficult to make any statement about its reliability. Nevertheless, the GSA proofed to be a useful tool to decompose the variance of the snow model, and provides clear evidence on the impact of uncertainties from individual factors as well as by their interaction. The present analysis clearly demonstrates that up to 70% of the model uncertainty could be reduced, in case a better accuracy in precipitation observations is achieved. More confidence in the simulationsremaining ones mainly impact in the summer or transitional seasons. The results thus allow fir the identification of the most critical parameters and environmental conditions, which together with the consideration of relevant model parameterizations, provide directions for future improvements. The analysis is based on rather conservative though commonly used uncertainty estimations. These are mostly based on manufacturer specifications and hence on laboratory settings. In field applications, however, can be gained more easily by using direct measurements of LW, rather than parametrizing this flux with other measured quantities (which is often necessary but are affected by larger uncertainties). Even if direct measurements are available, up to 60% of the snow depth uncertainty is caused by LW measurement errors. The impact on calculated snow depth is related to uncertainties in the SEB, which is determined by approximately 82% by the LW flux. Although the accuracy of the incoming SW measurement is in the same order of magnitude as the LW measurements, its contribution to the uncertainty of the simulation results is considerable less. The lower proportion is related to the year-round high albedo values at this site, and the associated lower net shortwave radiation flux. As follows from the GSA, errors related to wind measurements and roughness length show episodic effects on the SEB (up to 10%) due to their impact on the turbulent fluxes. Especially in wintertime failure of wind measurements are frequent and data gaps need to be filled in order to perform year round simulations. Together with missing information about the roughness length, the associated error propagation can significantly diminish the confidence in the modelled SEB. Other quantities, such as T and Q are often measured directly with higher accuracy and hence do not affect significantly the model results. It is finally noted again that the relative impact of individual error sources is very likely to vary for different zones on the glacier, and may show a different sensitivity pattern for other climatic regions. Investigation of this issue is one of the obvious applications of GSA in the futureeffective uncertainty is likely enhanced but is difficult to quantify. Moreover, we did not systematically consider effects of different uncertainty types (bias vs. random), different probability distributions or their combined propagation effects. Correspondingly set-up ensemble simulations fed by sampling from quasi-random sequences are therefore recommended for future investigations. Overall, the performed decomposition of the snow model output sensitivity by GSA proved valuable for enhancing our understanding of key snow model output sensitivity patterns in response to uncertainties in forcing data. The key findings either confirm or complement those derived from a few other studies employing GSA. GSA itself proved a promising tool to entangle the sensitivity of snow models and inherent critical parameters. The revealed importance of longwave radiation input may be considered as a trend-setting example. No doubt, however, more common efforts are necessary to further test and improve the method. This concerns, e.g., enhanced consideration of the effects of different combinations of error types and probability distributions, thereby also putting an enhanced focus on the propagation of uncertainties related to uncertainties in the specification of key parameterization values, which are mostly even less constrained than measurement errors. Detailed consideration of the parameterization of albedo in Crocus is suggested for the future, which was not addressed in this study. The presented approach is universal and can be applied to earth systems system models in general . Limitations from the and may be applied to snow and glacier mass- and energy balance modelling in all climate regions. From a practical and methodical point of viewconcern the , the main limitations of this study are the high computational effort and proper specification of the probability density functions of the parameter uncertainties. Finally, we would like to note that measurement uncertainties are independently sampled and do not possess any correlation structures. Consequently, the approach cannot be used to investigate the response of snow or ice depending on systematic changes in the environmental (climate) conditions. This requires appropriate sampling strategy to obtain the same correlation structure as those observed in nature.

Acknowledgements. This work was basically supported by the Austrian Science Fund (FWF, grant I 369-B17). Field work at Kongsvegen was performed in cooperation with the Norwegian Polar Institute (TromsoeTromsøe, J. Kohler) and University of Oslo (Ch. Nuuth). The achievement measurement of snow data was supported by Österreichische Polarforschungsgesellschaft, Julius-Payer Stipendium 2010. F. Karner, F. Bilgeri and W. Steinkogler are thanked for performing the field work and pre-evaluation of the data used within this study. We thank M.S. Raleigh and both anonymous reviewers for their detailed comments and constructive criticism of the original manuscript. We would also like to thank Samuel Morin and Matthieu Lafaysse for their support in applying the Crocus snowpack model.

References

Armstrong, R. L. and Brun, E.: Snow and climate: physical processes, surface energy exchange and modeling, Cambridge University Press, 2008.

Beersma, J. J. and Buishand, T. A.: Multi-site simulation of daily precipitation and temperature conditional on the atmospheric circulation, Clim. Res., 25, 121–133, 2003.

Bellaire, S., Jamieson, J. B., and Fierz, C.: Corrigendum to “Forcing the snow-cover model SNOWPACK with forecasted weather data” published in The Cryosphere, 5, 1115–1125, 2011, The Cryosphere, 7, 511–513, doi:10.5194/tc-7-511-2013, 2013.

Bernhardt, M., Liston, G. E., Strasser, U., Zängl, G., and Schulz, K.: High resolution modelling of snow transport in complex terrain using downscaled MM5 wind fields, The Cryosphere, 4, 99–113, doi:10.5194/tc-4-99-2010, 2010.

Björnsson, H., Gjessing, Y., Hamran, S.-E., Hagen, J. O., Liestøl, O., Pálsson, F., and Erlingsson, B.: The thermal regime of sub-polar glaciers mapped by multi-frequency radio-echo sounding, J. Glaciol., 42, 23–32, 1996.

Map demonstrating the location of Kongsvegen glacier within Svalbard and the position of the automatic weather stations KNG8 (red dot) and KNG6 (black circle). The orange outline shows the approximate Kongsvegen extent. (optical LandSat 7 image from July 1999, UTM 34N, WGS84).

Braithwaite, R. J. and Zhang, Y.: Sensitivity of mass balance of five Swiss glaciers to temperature changes assessed by tuning a degree-day model, J. Glaciol., 46, 7–14, 2000.

Comparison of the mean 6-hourly modelled and measured snow surface temperatures at the location KNG8.

Brandt, O., Kohler, J., and Lüthje, M.: Spatial mapping of multi-year superimposed ice on the glacier Kongsvegen, Svalbard, J. Glaciol., 54, 73–80, 2008.

Observed (orange) and modelled (blue) snow temperature profile on 06 April 2011 17 UTC at the location KNG8.

Brun, E., David, P., Sudul, M., and Brunot, G.: A numerical model to simulate snow-cover stratigraphy for operational avalanche forecasting, J. Glaciol., 38, 13–22, 1992.

Daily mean observed (orange) and modelled (blue) snow albedo at the location KNG8.

Brun, E., Martin, E., Simon, V., Gendre, C., and Coleou, C.: An energy and mass model of snow cover suitable for operational avalanche forecasting, J. Glaciol., 35, 333–342, 1992.

Observed (orange) and modelled (blue) snow density profile on 06 April 2011 at the location KNG8.

Brun, E., Vionnet, V., Boone, A., Decharme, B., Peings, Y., Valette, R., Karbou, F., and Morin, S.: Simulation of northern eurasian local snow depth, mass, and density using a detailed snowpack model and meteorological reanalyses, J. Hydrometeorol., 14, 203–219, 2013.

Carmagnola, C. M., Morin, S., Lafaysse, M., Domine, F., Lesaffre, B., Lejeune, Y., Picard, G., and Arnaud, L.: Implementation and evaluation of prognostic representations of the optical diameter of snow in the SURFEX/ISBA-Crocus detailed snowpack model, The Cryosphere, 8, 417–437, doi:10.5194/tc-8-417-2014, 2014.

Uncertainty of the model simulation due to propagating uncertainties in the model inputs. The black lines represents the reference run. The intervals show the 99%, 95% and 75% quantiles estimated from the Monte-Carlo runs (16000 ensemble members).

Castebrunet, H., Eckert, N., Giraud, G., Durand, Y., and Morin, S.: Projected changes of snow conditions and avalanche activity in a warming climate: the French Alps over the 2020–2050 and 2070–2100 periods, The Cryosphere, 8, 1673–1697, doi:10.5194/tc-8-1673-2014, 2014.

Mean impact of measurement uncertainties for different seasons on snow depth changes, surface energy balance (SEB), sensible heat and latent heat flux.

Dadic, R., Mott, R., Lehning, M., Carenzo, M., Anderson, B., and Mackintosh, A.: Sensitivity of turbulent fluxes to wind speed over snow surfaces in different climatic settings, Adv. Water Resour., 55, 178–189, 2013.

Temporal evolution of the first-order sensitivity indices affecting modelled snow depth changes, surface energy balance (SEB), sensible and latent heat flux at KNG8. Refer to Table 1 for explanation of the indicated uncertainty factors.

Durand, Y., Giraud, G., Brun, E., Mérindol, L., and Martin, E.: A computer-based system simulating snowpack structures as a tool for regional avalanche forecasting, J. Glaciol., 45, 469–484, 1999.

Durand, Y., Giraud, G., Laternser, M., Etchevers, P., Mérindol, L., and Lesaffre, B.: Reanalysis of 47 years of climate in the French Alps (1958–2005): climatology and trends for snow cover, J. Appl. Meteorol. Clim., 48, 2487–2512, 2009.

Etchevers, P., Martin, E., Brown, R., Fierz, C., Lejeune, Y., Bazile, E., Boone, A., Dai, Y.-J., Essery, R., Fernandez, A., Gusev, Y., Jordan, R., Koren, V., Kowalczyk, E., Nasonova, N.O., Pyles, R.D., Schlosser, A., Shmakin, A.B., Smirnova, T.G., Strasser, U., Verseghy, D., Ymazaki, T., Yang, Z-L.: Validation of the energy budget of an alpine snowpack simulated by several snow models (SnowMIP project), Ann. Glaciol., 38, 150–158, 2004.

Feng, X., Sahoo, A., Arsenault, K., Houser, P., Luo, Y., and Troy, T. J.: The impact of snow model complexity at three CLPX sites, J. Hydrometeorol., 9, 1464–1481, 2008.

Førland, E. J. and Hanssen-Bauer, I.: Increased precipitation in the Norwegian Arctic: true or false?, Climatic Change, 46, 485–509, 2000.

Franz, K. J., Butcher, P., and Ajami, N. K.: Addressing snow model uncertainty for hydrologic prediction, Adv. Water Resour., 33, 820–832, 2010.

Fréville, H., Brun, E., Picard, G., Tatarinova, N., Arnaud, L., Lanconelli, C., Reijmer, C., and van den Broeke, M.: Using MODIS land surface temperatures and the Crocus snow model to understand the warm bias of ERA-Interim reanalyses at the surface in Antarctica, The Cryosphere, 8, 1361–1373, doi:10.5194/tc-8-1361-2014, 2014.

Fujita, K.: Effect of precipitation seasonality on climatic sensitivity of glacier mass balance, Earth Planet. Sc. Lett., 276, 14–19, 2008.

Gallée, H., Guyomarc'h, G., and Brun, E.: Impact of snow drift on the Antarctic ice sheet surface mass balance: possible sensitivity to snow-surface properties, Bound.-Lay. Meteorol., 99, 1–19, 2001.

Gallet, J.-C., Domine, F., Savarino, J., Dumont, M., and Brun, E.: The growth of sublimation crystals and surface hoar on the Antarctic plateau, The Cryosphere, 8, 1205–1215, doi:10.5194/tc-8-1205-2014, 2014.

Gascon, B., Sharp, M., Burgess, D., Bezeau, P., Bush, B.G., Morin, S., and Lafaysse, M.: How well is firn densification represented by a physically based multilayer model? Model evaluation for Devon Ice Cap, Nunavit, Canada, Journal of Glaciology, 66, 222, 694–704, 2014.

Gerbaux, M., Genthon, C., Etchevers, P., Vincent, C., and Dedieu, J.: Surface mass balance of glaciers in the French Alps: distributed modeling and sensitivity to climate change, J. Glaciol., 51, 561–572, 2005.

Greuell, W., Kohler, J., Obleitner, F., Glowake, P., Melvold, E.B., and Oerlemans, J.: Assessment of interannual variations in teh surface mass balance of 18 Svalbard glaciers from the Moderate Resolution Imaging Spectroradiometer/Terra albedo product, J. Geophys. Res., 112, 2007.

Greuell, W. and Konzelmann, T.: Numerical modelling of the energy balance and the englacial temperature of the Greenland ice sheet. Calculations for the ETH-Camp location (West Greenland, 1155 m asl), Global Planet. Change, 9, 91–114, 1994.

Greuell, W. and Oerlemans, J.: Sensitivity studies with a mass balance model including temperature profile calculations inside the glacier, Z. Gletscherkd. Glazialgeol., 22, 101–124, 1986.

Greuell, W. and Smeets, P.: Variations with elevation in the surface energy balance on Pasterze (Austria), Journal of Geophysical Research, 106, 717–727, 2001.

Gurgiser, W., Mölg, T., Nicholson, L., and Kaser, G.: Mass-balance model parameter transferability on a tropical glacier, J. Glaciol., 59, 845–858, 2013.

He, M., Hogue, T. S., Franz, K. J., Margulis, S. A., and Vrugt, J. A.: Characterizing parameter sensitivity and uncertainty for a snow model across hydroclimatic regimes, Adv. Water Resour., 34, 114–127, 2011.

Karner, F., Obleitner, F., Krismer, T., Kohler, J., and Greuell, W.: A decade of energy and mass balance investigations on the glacier Kongsvegen, Svalbard, J. Geophys. Res.-Atmos., 118, 3986–4000, 2013.

König, M., Wadham, J., Winther, J.-G., Kohler, J., and Nuttall, A.-M.: Detection of superimposed ice on the glaciers Kongsvegen and midre Love'nbreen, Svalbard, using SAR satellite imagery, Ann. Glaciol., 34, 335–342, 2002.

Lehning, M., Bartelt, P., Brown, B., Russi, T., Stöckli, U., and Zimmerli, M.: SNOWPACK model calculations for avalanche warning based upon a new network of weather and snow stations, Cold Reg. Sci. Technol., 30, 145–157, 1999.

Lehning, M., Völksch, I., Gustafsson, D., Nguyen, T. A., Stähli, M., and Zappa, M.: ALPINE3D: a detailed model of mountain surface processes and its application to snow hydrology, Hydrol. Process., 20, 2111–2128, 2006.

Libois, Q., Picard, Ghislain, Arnaud, L., Morin, S., and Brun, E.: Modeling the impact of snow drift on the decameter-scale variability of snow properties on the Antarctic Plateau, J.G.R, 11, 662–681, 2014.

Liston, G. E. and Elder, K.: A distributed snow-evolution modeling system (SnowModel), J. Hydrometeorol., 7, 1259–1276, 2006.

Magnusson, J., Gustafsson, D., Hüsler, F., and Jonas, T.: Assimilation of point SWE data into a distributed snow cover model comparing two contrasting methods, Water Resour. Res., doi:10.1002/2014WR015302, 2014.

Marks, D., Kimball, J., Tingey, D., and Link, T.: The sensitivity of snowmelt processes to climate conditions and forest cover during rain on snow: a case study of the 1996 Pacific Northwest floow, Hydrol. Process., 12, 1569–1587, 1998.

Mott, R., Gromke, C., Grünewald, T., and Lehning, M.: Relative importance of advective heat transport and boundary layer decoupling in the melt dynamics of a patchy snow cover, Adv. Water Resour., 55, 88–97, 2013.

Norwegian Polar Institute. Kartdata Svalbard 1:100000 (S100 Kartdata). Tromso, Norway: Norwegian Polar Institute. https://data.npolar.no/dataset/645336c7-adfe-4d5a-978d-9426fe788ee3.

Obleitner, F. and De Wolde, J.: On intercomparison of instruments used within the Vatnajökull glacio-meteorological experiment, Bound.-Lay. Meteorol., 92, 25–35, 1999.

Obleitner, F. and Lehning, M.: Measurement and simulation of snow and superimposed ice at the Kongsvegen glacier, Svalbard (Spitzbergen), J. Geophys. Res.-Atmos., 109, doi:10.1029/2003JD003945, 2004.

Oerlemans, J.: Climate sensitivity of glaciers in southern Norway: application of an energy-balance model to Nigardsbreen, Hellstugubreen and Alfotbreen, J. Glaciol., 38, 223–232, 1992.

Phan, X. V., Ferro-Famil, L., Gay, M., Durand, Y., Dumont, M., Morin, S., Allain, S., D'Urso, G., and Girard, A.: 1D-Var multilayer assimilation of X-band SAR data into a detailed snowpack model, The Cryosphere, 8, 1975–1987, doi:10.5194/tc-8-1975-2014, 2014.
Pomeroy, J. and Gray, D.: Saltation of snow, Water Resour. Res., 26, 1583–1594, 1990.
Radić, V. and Hock, R.: Modeling future glacier mass balance and volume changes using ERA-40 reanalysis and climate models: a sensitivity study at Storglaciären, Sweden, J. Geophys. Res.-Earth, 111, doi:10.1029/2005JF000440, 2006.
Raleigh, M. S., Lundquist, J. D., and Clark, M. P.: Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework, Hydrol. Earth Syst. Sci., 19, 3153–3179, doi:10.5194/hess-19-3153-2015.
Rutter, N., Essery, R., Pomeroy, J., Altimir, N., Andreadis, K., Baker, I., Barr, A., Bartlett, P., Boone, A., Deng, H., Douville, H., Dutra, E., Elder, K., Ellis, C., Feng, Xia, Gelfan, A., Goodbody, A., Gusev, Y., Gustafsson, D., Hellstöm, R., Hirabayashi, Y., Hirota, T., Jonas, T., Koren, V., Kuragina, A., Lettenmaier, D., Li, W-P., Lice, C., Martin, E., Nasanova, O., Pumpanen, J., Pyles, R.D., Samuelsson, P., Sandells, M., Schädler, G., Shmakin, A., Smirnova, T.G., St/"ali, M., St'öckli, R., Strasser, U., Su, H., Suzuki, K., Takata, K., Tanaka, K., Thompson, E., Vesala, T., Viterbo, P., Wiltshire, A., Xia, K., Xue, Y., Yamazaki, T.: Evaluation of forest snow processes models (SnowMIP2), J. Geophys. Res.-Atmos, 114, doi:10.1029/2008JD011063, 2009.
Saltelli, A., Tarantola, S., and Chan, K. P.: A quantitative model-independent method for global sensitivity analysis of model output, Technometrics, 41, 39–56, 1999.
Saltelli, A., Ratto, M., Tarantola, S., and Campolongo, F.: Sensitivity analysis practices: strategies for model-based inference, Reliab. Eng. Syst. Safe., 91, 1109–1125, doi:10.1016/j.ress.2005.11.014, 2006.
Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., and Tarantola, S.: Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index, Comput. Phys. Commun., 181, 259–270, 2010.
Sauter, T. and Venema, V.: Natural three-dimensional predictor domains for statistical precipitation downscaling, J. Climate, 24, 6132–6145, 2011.
Sauter, T., Möller, M., Finkelnburg, R., Grabiec, M., Scherer, D., and Schneider, C.: Snowdrift modelling for the Vestfonna ice cap, north-eastern Svalbard, The Cryosphere, 7, 1287–1301, doi:10.5194/tc-7-1287-2013, 2013.

Schmucki, E., Marty, C., Fierz, C., and Lehning, M.: Evaluation of modelled snow depth and snow water equivalent at three contrasting sites in Switzerland using SNOWPACK simulations driven by different meteorological data input, Cold Reg. Sci. Technol., 99, 27–37, 2014.

Smeets, C.: Assessing unaspirated temperature measurements using a thermocouple and a physically based model, in: The Mass Budget of Arctic Glaciers, Workshop and GLACIODYN planning meeting, 29 January-3 February 2006, IASC Working group on Arctic Glaciology, Institute for Marine and Amtospheric Research, Utrecht, p. 99, 2006.

Sobol, I. M.: On quasi-monte carlo integrations, Mathematics and Computers in Simulation, 47, 2, 103–112, 1998.

Sobol, I. M., Tarantola, S., Gatelli, D., Kucherenko, S., and Mauntz, W.: Estimating the approximation error when fixing unessential factors in global sensitivity analysis, Reliab. Eng. Syst. Safe., 92, 957–960, 2007.

Van de Wal, R. and Oerlemans, J.: An energy balance model for the Greenland ice sheet, Global Planet. Change, 9, 115–131, 1994.

Van de Broeke, M.R., Smeets, C.J.P.P., and van de Wal, R.S.W..: The seasonal cycle and interannual variability of surface energy balance and melt in the ablation zone of the west Greenland ice sheet, The Cryosphere, 5, 377–390, 2011.

van Pelt, W.J.J., Oerlemans, J., Reijmer, C.H., Pohjola, V.A., Pettersson, R., and Angelen, J.H.: Simulating melt, runoff and refreezing on Nodenskiöldbreen, Svalbard, using a coupled snow and energy balance model, The Cryosphere, 6, 641–659, 2012.

Vionnet, V., Brun, E., Morin, S., Boone, A., Faroux, S., Le Moigne, P., Martin, E., and Willemet, J.-M.: The detailed snowpack scheme Crocus and its implementation in SURFEX v7.2, Geosci. Model Dev., 5, 773–791, doi:10.5194/gmd-5-773-2012, 2012.

Wang, T., Ottlé, C., Boone, A., Ciais, P., Brun, E., Morin, S., Krinner, G., Piao, S., and Peng, S.: Evaluation of an improved intermediate complexity snow scheme in the ORCHIDEE land surface model, J. Geophys. Res.-Atmos., 118, 6064–6079, 2013.

Wright, A., Wadham, J., Siegert, M., Luckman, A., Kohler, J., and Nuttall, A.: Modeling the refreezing of meltwater as superimposed ice on a high Arctic glacier: a comparison of approaches, J. Geophys. Res.-Earth, 112, doi:10.1029/2007JF000818, 2007.

Table 1. Model parameters used for the reference run.

Parameter	Value	Description	
z_0	0.002	m	Roughness length for momentum
zh_0 zh_0	0.0002	m	Roughness length for heat
HC_{LW} $\mathrm{HC}_{\mathrm{LW}}$	0.05	–	Max. liquid water holding capacity
$ALB_{0.3}$ $\mathrm{ALB}_{0.3}$	0.38	–	Absorption coefficient Ice albedo for spectral band 0.3-0.8 –0.8 mm
$ALB_{0.8}$ $\mathrm{ALB}_{0.8}$	0.23	–	Absorption coefficient Ice albedo for spectral band 0.8-1.5 –1.5 mm
$ALB_{1.5}$ $\mathrm{ALB}_{1.5}$	0.08	–	Absorption coefficient Ice albedo for spectral band 1.5-2.8 –2.8 mm
ρ_{thres} ρ_{thres}	830	$\mathrm{kg\,m^{-3}}$	Ice density threshold
$\mathrm{Rain}_{\mathrm{thres}}$	1	°C	Rain threshold temperature
$\mathrm{Snow}_{\mathrm{thres}}$	0	°C	Snow threshold temperature
ϵ	0.99	–	Snow emissivity

Table 2. Specification of basic model input uncertainties and assigned probability density functions. The Sobol sequence has been generated from the distributions given in the last column($\mathcal{N}$ - , where $\mathcal{N}(\mu,\sigma)$ is a Normal distribution ;$\mathcal{U}$ - with mean μ and standard deviation σ and $\mathcal{U}(lb,ub)$ is a Uniform distribution in the interval $[lb,ub]$).

Parameter	Description	Uncertainty	Distribution
T_{air} T_{air}	Air temperature	±0.3 ±0.3 K	$\mathcal{N}(0.00,0.30)$
RH RH	Relative humidity	$\pm3.0$$\pm3.0$ %	$\mathcal{N}(0.00,3.00)$
SW SW	Shortwave incoming radiation	$\pm10.0$$\pm10.0$ %	$\mathcal{N}(0.00,0.10)$
LW LW	Longwave incoming radiation	$\pm10.0$$\pm10.0$ %	$\mathcal{N}(0.00,0.10)$
U	Wind speed	±0.3 ±0.3 $\mathrm{m\,s^{-1}}$	$\mathcal{N}(0.00,0.30)$
P	Precipitation	$\pm25.0$$\pm25.0$ %	$\mathcal{N}(0.00,0.25)$
z_0	Aerodynamic roughness length	$0.001-0.10$ 0.001–0.10 m	$\mathcal{U}(0.001,0.10)$
$PVOL$ PVOL	Pore volume fraction for maximum liquid water holding capacity	$0.03-0.05$0.03–0.05	$\mathcal{U}(0.03,0.05)$

Table 3. Mean and standard deviation (brackets) of the meteorological variables and energy balance components for the summer months (JJA) and winter months (DJF) at KNG8 and KNG1.

	KNG8		KNG1	
Variable	DJF	JJA	DJF	JJA
Air temperature (2 m) [$°C$]	253.5 (8.0)	271.5 (2.6)	261.1 (6.9)	277.1 (1.6)
Surface temperature [$°C$]	252.3 (9.1)	270.1 (3.9)	255.9 (9.4)	273.1 (0.3)
Rel. humidity (2 m) [%]	97 (6)	91 (7)	83 (12)	83 (9)
Water vapour pressure (2 m) [hPa]	3.3 (0.8)	5.3 (0.5)	2.1 (1.3)	6.9 (0.7)
Wind speed (2 m) [$m\ s^{-1}$]	1.2 (1.9)	0.8 (1.9)	4.6 (3.6)	1.6 (2.6)
SW-incoming radiation [$W\ m^{-2}$]	0.1 (0.9)	239.5 (167.9)	1.0 (7.0)	209.3 (157.5)
SW-outgoing radiation [$W\ m^{-2}$]	0.1 (0.8)	193.6 (137.4)	0.8 (5.6)	99.6 (86.4)
LW-incoming radiation [$W\ m^{-2}$]	223.4 (43.9)	268.5 (39.6)	200.7 (55.2)	288.2 (35.7)
LW-outgoing radiation [$W\ m^{-2}$]	231.5 (32.4)	301.8 (16.9)	245.1 (35.3)	315.2 (1.9)
Sensible heat flux [$W\ m^{-2}$]	3.1 (36.0)	3.5 (17.1)	21.0 (16.3)	12.1 (15.2)
Latent heat flux [$W\ m^{-2}$]	2.9 (4.5)	-1.1 (6.1)	20.8 (17.2)	20.5 (33.4)
Surface energy balance [$W\ m^{-2}$]	-1.9 (14.5)	14.8 (34.3)	-2.2 (20.7)	101.1 (86.2)

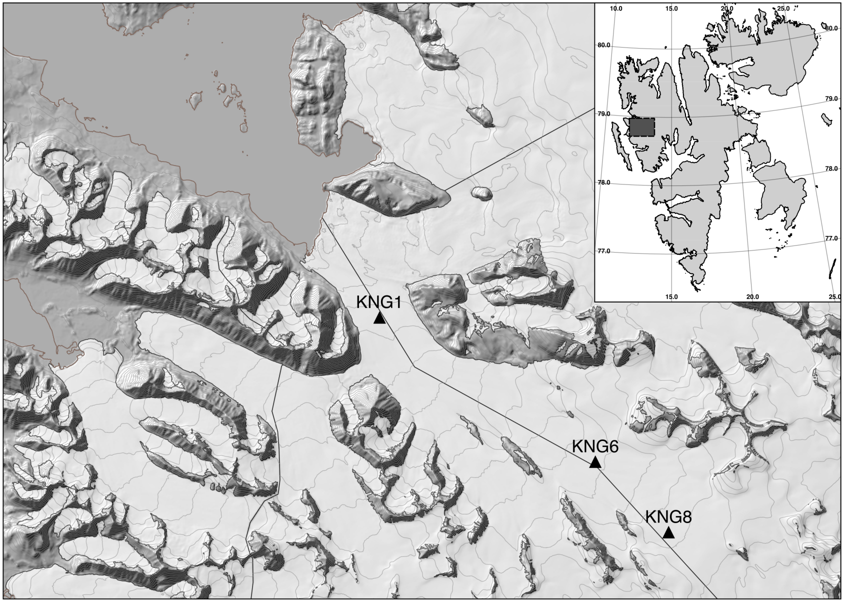

Figure 1. A map showing the location of the Kongsvegen glacier and the position of the automatic weather stations KNG8, KNG6 and KNG1 (Norwegian Polar Institute, 2014).

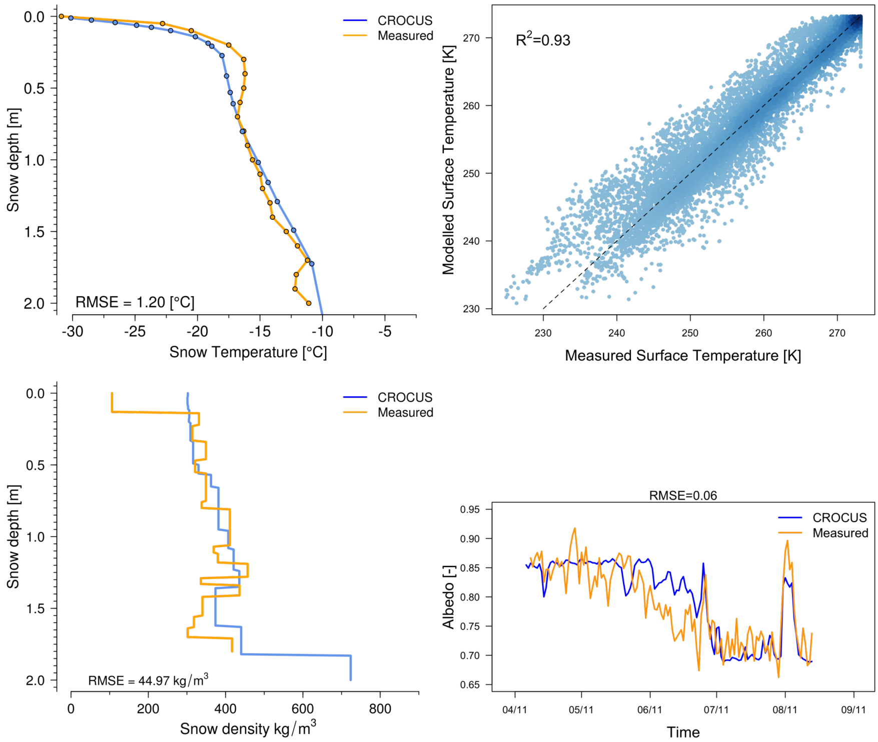

Figure 2. Comparison of the modelled and measured snow temperatures (upper left), snow density (lower left), snow surface temperature (upper right) and snow albedo (lower right) at the location KNG8.

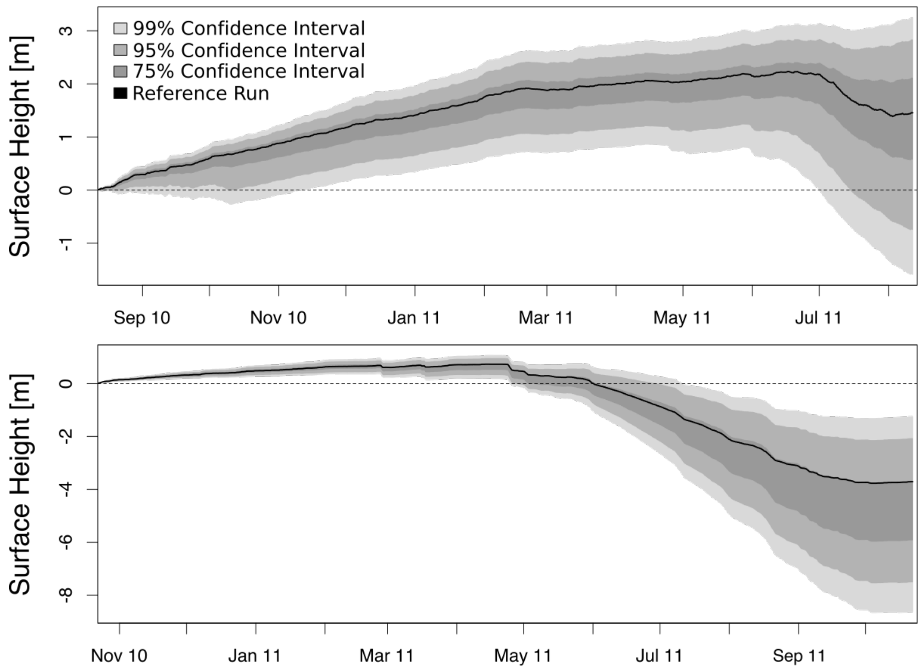

Figure 3. Spread of the ensemble simulation at KNG8 (upper panel) and KNG1 (lower panel) due to propagating uncertainties in the model inputs. The black lines represent the reference run. The intervals show the 99, 95 and 75% quantiles estimated from the quasi-random Monte-Carlo runs (20000 ensemble members). Note the different horizontal and vertical scales.

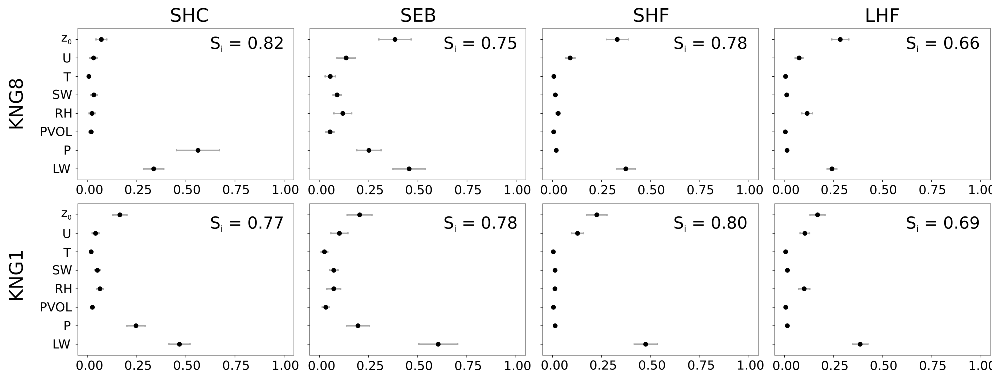

Figure 4. Yearly averaged total-order effects of factors (see Table 2) on surface height change (SHC), surface energy balance (SEB) and sensible heat and latent heat flux (HF) at KNG8 and KNG1. The whiskers show the 95% confidence interval derived from 1000 empirical bootstrap samples. The mean (taken over the whole period) of the 6 hourly first-order sums (linear effects) are given in the upper right corner.

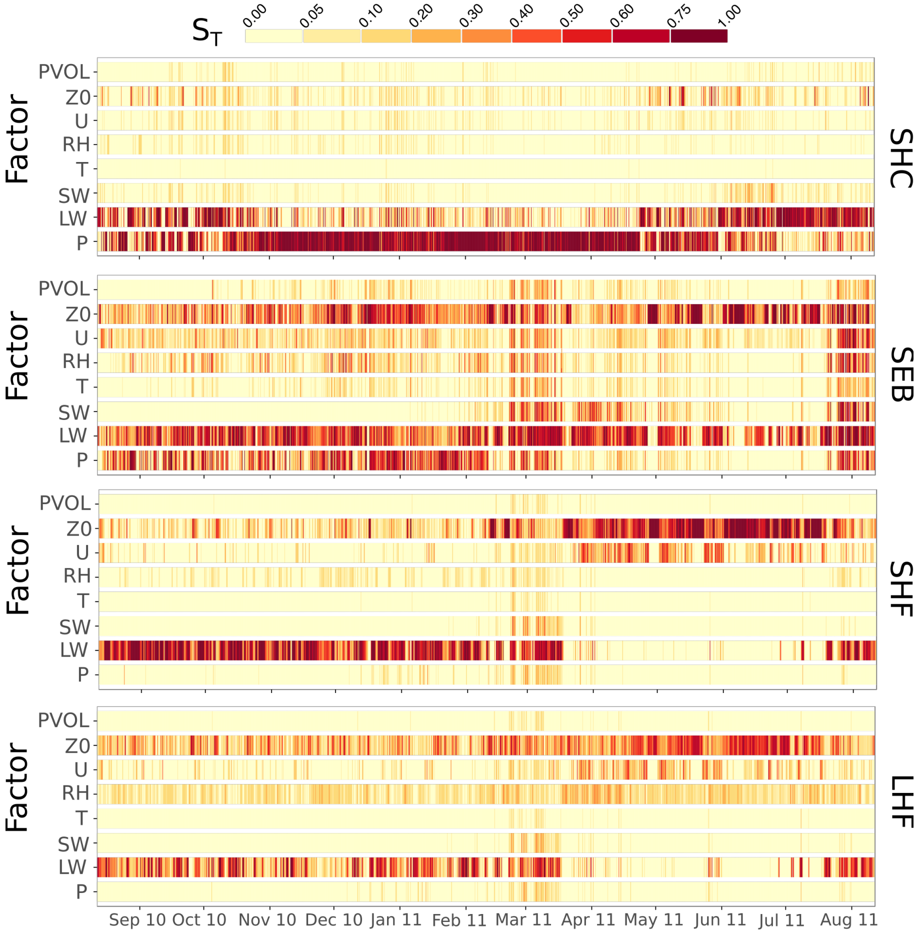

Figure 5. Evolution of the 6 hourly total-effect indices affecting modelled surface height change (SHC), surface energy balance (SEB), sensible heat flux (SHF) and latent heat flux (LHF) at KNG8. Refer to Table 2 for the explanation of the indicated uncertainty factors.

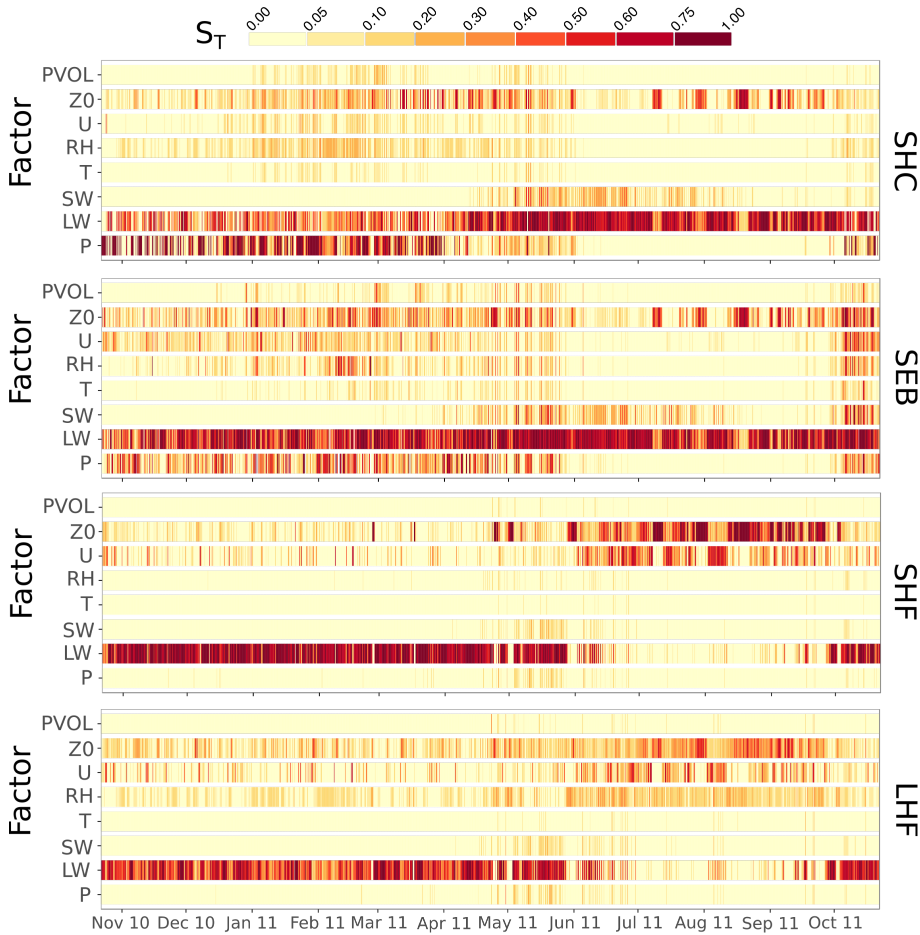

Figure 6. Evolution of the 6 hourly total effect indices affecting modelled surface height change (SHC), surface energy balance (SEB), sensible heat flux (SHF) and latent heat flux (LHF) at KNG1. Refer to Table 2 for the explanation of the indicated uncertainty factors.