The Cryosphere Discuss., 6, C236–C260, 2012
www.the-cryosphere-discuss.net/6/C236/2012/
© Author(s) 2012. This work is distributed under
the Creative Commons Attribute 3.0 License.

The Cryosphere Discussions

Interactive comment on "Spatial patterns of North Atlantic Oscillation influence on mass balance variability of European Glaciers" *by* B. Marzeion and A. Nesje

B. Marzeion and A. Nesje

ben.marzeion@uibk.ac.at

Received and published: 5 April 2012

Reply to the reviews of Anonymous Referees 1 to 4 on our manuscript "Spatial Patterns of North Atlantic Oscillation Influence on Mass Balance Variability of European Glaciers"

We would like to thank Jon Ove Hagen for obtaining the reviews, and we would like to thank the reviewers for their detailed, constructive, and helpful comments on our manuscript. We were able to address all their points (as detailed below), and their suggestions have lead to significant improvements of our manuscript.

Response to Anonymous Referee 1

Comments on the details of the MS

1. (a) **Comment:** Was the termini elevation regarded as fixed over the time domain? Some glaciers have almost 50 years of time series, and have retreated, so termini elevation has changed. How was this treated in the modeling? Was the termini position a static point?
Response: Terminus elevation was treated as constant, i. e., the dynamic response of the glacier to mass changes was neglected. We included a footnote stating this explicitly, and a reference to Marzeion et al. (2012), where the implications of this assumption are discussed in detail. In short: the error introduced to the mass balance model by neglecting terminus elevation change is contained in the total error of the mass balance model determined by the (cross) validation, since measured mass balances of course do respond to terminus elevation changes. It is therefore not an additional, unquantified error of the model. Particularly on interannual (i.e., NAO-relevant) time scales, we would argue that the error has to be small necessarily, since terminus elevation does not change much from year to year.

   (b) **Comment:** The vertical data in the CRU grids were used to calculate lapse rates, was this used to get the surface air temperature (SAT) at the glacier termini? What is the vertical resolution in the CRU data, and is that vertical resolution enough to get the SAT needed?
Response: There is no direct information on vertical lapse rates contained in the CRU data. therefore, we "estimate vertical lapse rates of temperature for all glaciers in the WGI-XF data set by regressing temperature in 3×3

CRU CL 2.0 grid points around the location of the glacier onto elevation." (page 7 line 22 of the discussion paper). The question is therefore rather, whether the *horizontal* resolution of the CRU data is high enough. Since the correlation between CRU elevation and temperature is very high, and above the 95 % confidence level for all glaciers (page 8, line 3 of the discussion paper), we are confident that the resolution is good enough to obtain reasonable estimates of SAT at terminus elevation.

(c) **Comment:** Further, is the SAT given exact where ice meets bare ground, or where exact is the point denoting termini elevation. The ice fronts are sometimes like a cliff, and the glacier close to front, and the ground ca differ 50 m or so...
Response: Terminus elevation (which we obtain from WGI-XF) should be the elevation where ice meets bare ground, but we can safely assume that the error introduced by potentially near-vertical termini would be smaller than the error we introduce by neglecting terminus movements, and also already is contained in the model error determined by the (cross) validation (see response to comment 1. (a) above.)

(d) **Comment:** How was the precipitation calculated? Normally mass balance models use elevation bands to integrate stepwise change in T and P (e.g Radic, V. and R.Hock. 2011. Regionally differentiated contribution of mountain glaciers and ice caps to future sea-level rise. Nat.Geosc., 4(Feb), 91-94). But, as I understand this MS one point on the glacier was used to derive Pi (Solid precipitation). Is this correct? In that case, which elevation was used to calculate Pi for the glacier? Perhaps elevations bands was used, but it is not well described in the text. If you used a single point, like the termini for Pi you must argument why you used it, since Pi probably differs quite a bit over the elevation span.
Response: The answer to this question differs between the three models used – but the general issue is that the glaciers are much smaller than the

CRU grid cells, implying that for each glacier, only one precipitation value is available in the data, but that at the same time, as the reviewer points out, precipitation varies with elevation.
In the individually trained and mean model, the parameter a (eq. 1 in the discussion paper) is included, which "can be understood as a parameter representing effects of a precipitation lapse rate, aeolian transport of 15 snow, and avalanching" (page 5, line 13 of the discussion paper). So for these two models, the effect of a vertical precipitation lapse rate is estimated during the model training (parameter determination). So while P_i is determined from a single point, the model does include effects of precipitation changes with altitude.
For the climatologically derive model, we prescribe a precipitation lapse rate of 2 % for every 100 m elevation gain (page 8, line 4 of the discussion paper), and use this lapse rate to calculate precipitation at the mean elevation of the glacier (this information was missing in the discussion paper, and was included in the new version). Therefore, also for this model effects of precipitation changes with altitude are included.

(e) **Comment:** If Pi is solid precipitation only, did you not regard water retention from rain fall, and refreezing melt water in the firn pack in the income of the mass balance. Probably not, but in case, argument why you did not use it. In the simplest form water retention can be used in a simple parameterization scheme, eg the Radic and Hock reference above.
Response: Refreezing etc. are not addressed directly, but we would argue that the reason for obtaining the best model results when using a rather high temperature threshold (2°C, see page 8, line 16 of the discussion paper) has to do with the effects that liquid precipitation at low temperature may have on the mass balance. We included a sentence to point this out.

(f) **Comment:** The best way to get precipitation is probably to do a dynamically downscale gridded data. My experience is that gridded data often underes-

timate precipitation close to large topographical object, i.e. underestimate the orographic component of precipitation. This modeling is though out of the scoop of this present work, but may be a remark worth to bring in mind.
Response: We agree – both on that it would be a better solution, that it is out of scope here, and that it should be mentioned. We therefore added a paragraph to the discussion section along the lines suggested by the reviewer.

(g) **Comment:** on page 10, li 10 and thereabouts it is a discussion of how to increase the skill score, and correlations between model and data. One thing we must remember is that the measured data still is an estimate of reality. To measure the exact mass balance is a difficult task on most glaciers, and almost an impossible task in a few settings. By this I think we must always understand that there is a deviation of measured mass balance and real mass balance, giving that modeled mass balance could even be a better proxy for real mass balance than measured mass balance in a few (lucky) cases.
Response: The problem is that while it is quite straightforward to quantify the model's error, it is hardly possible to strictly quantify the error associated with the measurements. But we fully agree with the reviewer and thank him for pointing this out.

(h) **Comment:** A final comment I have is that the models likely would give even better performance if daily data is used instead of the very crude average over a whole month; but data handling and availability in this specific data base probably puts the frame of this issue. Perhaps a few words can be used to discuss possible improvement in future work using daily data from example ERA40, or ERA :Interim.
Response: We agree - probably higher temporal resolution would increase model performance, but it is not alone the data handling that makes this difficult: Because of the shorter time span over which these data are available,

also the demand on model performance is increased, because shorter time series raise the bar for achieving significant correlations. We added this discussion to the manuscript.

2. **Comment:** The results and comparisons of model data with NAO clearly show how NAO has a regional influence on the mass balance of the European glaciers, with some regions better related, and some less to NAO. My comment here is why NAO is used over the summer months? NAO is not well defined over JJA, due to the flatter pressure field in the summer months (See Hurrell, 1995). This is further shown by the authors by finding DJF has better scores than the full year in their tests. In an older study it was found that regarding Scandinavian glaciers both winter and summer balances were better correlated to a pressure index they called Norwegian Sea Index, being regionally derived between the pressure centers over about West Siberia/Barents Sea and British Isles (Pohjola, V.A. and J.C.Rogers. 1997. Atmospheric circulation and variations in Scandinavian mass balance. Quaternary Research, 47(1), 29-36). Could a development of the present study be expanded into refining the study into more regional pressure centers, instead of the more "global" NAO?
Response: It is indeed true that the NAO has stronger impacts during winter months - this becomes obvious e.g. in Fig. 7 of the discussion paper, where panel a shows that the NAO index does not exhibit a strong seasonal variability in its amplitude, but NAO-related temperature (panel d) and precipitation (panel g) anomalies show a strong amplitude seasonality. The reason why we discuss the NAO "off"-season anyway is the difference between Fig. 8a and Fig. 10a: What is happening during the core NAO season is able to explain the annual mass balance anomalies in the western Alps and southern Scandinavia, but not in the eastern Alps and northern Scandinavia. Here, the weaker NAO off-season influence is able to cancel the core winter effects (eastern Alps), or override them (northern Scandinavia). Of course, the method used here could well be applied

to other circulation anomalies than the NAO – with more regionally relevant patterns implying the loss of relevance in other regions. In this study, the connection between the Alps and Scandinavia was the target. We added a reference to Pohjola and Rogers, 1997.

Detailed comments:

1. **Comment:** Abstract li2. Was it really 7735 individual time series, or 7735 modeled mass balance years? Perhaps best to take out this number, it is not necessary, and not described (as far as I could see) in the further text.
   Response: It really are 7735 individual time series - we added a statement to the Results section to clarify this.

2. **Comment:** The mean model is not described very well. Perhaps add a line or two about it.
   Response: We reformulated the description in order to clarify.

3. **Comment:** I did not understand li10-15, how did you get a r = 03?
   Response: This was a typo - correct is 0.**7**3 – corrected.

4. **Comment:** Table 1 and 2. You need to explain better in head what the parameters in the tables represent. I guess the tables is the comparison model to measurements? Is MB in years? In Table 1 is rows 1-2 individually trained, and rows 3-7 the mean model?
   Response: We extended the description of the table to clarify.

5. **Comment:** In general: Use r2 instead of r, it is a better measure of skill.
   Response: We would prefer to keep using r, for two reasons: (i) For comparability with the results in Marzeion et al. (2012), where r is given and to which

the reader is referred multiple times because of the more detailed discussion of the validation procedure, and (ii) because as a measure of skill, we think the skill score is an even better measure, and this is already included.

Response to Anonymous Referee 2

Specific queries about the methods

1. **Comment:** I'm not sure how the threshold length of 12 years for glacier mass balance measurements (p.9, last sentence) was chosen.
   Response: Ultimately, it is a subjective choice where to draw the limit. Including more glaciers is better, but including less robust parameter estimates is worse. However, we tested how sensitive our results are against varying the threshold length, and found them to be very robust. We added a sentence to clarify this.

2. **Comment:** p.12, line 6 from bottom "the climatologically derived model provides the most reliable results": why should this be given that the other models are trained against mass-balance measurements? - this seems counter-intuitive.
   Response: We agree that this is counter-intuitive at first. But it actually shows that typically, the number of available mass balance measurements is not enough to reliably determine model parameters – and this problem is detected by the cross validation. We added a paragraph to the discussion to point this out.

3. **Comment:** I am concerned that the correlation maps merge mass balance records based on different time periods, which are not clearly specified in the paper and could bias the results if the NAO-mass balance relation is time-dependent

with different spatial correlations and strengths according to the exact timeframe being considered (which it most likely is to some extent).
Response: All the correlation maps shown in the paper (except in Fig. 4) are based on the same time period, i.e. the time period covered by CRU data. We added a sentence to the beginning of the results section in order to clarify this. In Fig. 4, correlations are calculated over the length of the available measured mass balance record - but since this figure only relates to the validation of the model, it is not subject to the (justified) concern of the reviewer that different time periods could be problematic.

4. **Comment:** p.10, ,line 14 (& elsewhere): change "exemplary" to "example".
Response: Text changed accordingly.

5. **Comment:** p.13, l.15. p.11, ll.12-13: "...(mean correlation of 0.74), but the performance suffers only a little when applied over all of Europe (mean correlation 0.3)." - this seems to me to be a big correlation difference, so can the authors reword and/or clarify?
Response: This was a typo – corrected to "...(mean correlation of 0.74), but the performance suffers only a little when applied over all of Europe (mean correlation 0.**7**3)." (the numbers are given in table 1).

Response to Anonymous Referee 3

See above for replies.

Response to Anonymous Referee 4

General comment

Comment: This paper describes the application of a 'minimal' mass balance model to all glaciers in the European Alps and Scandinavia, in order to derive correlations of the computed mass balance series with the NAO. The text is well-structured and written in clear English. However, I have serious concerns about the methods used for the mass balance reconstruction, as explained in the major comments below. In addition, this paper does not present sufficient new results or applications. I therefore do not think this paper meets the quality standards of The Cryosphere.

Response: We appreciate the reviewer's concerns and thank him for providing a such detailed review. His main points of criticism can be summarized as (i) the models are inadequate, and (ii) there are not enough new results.

Regarding (i), we would like to point out that the adequacy of a model always depends on the objective of the study. We completely agree with the reviewer that all three of the models used in manuscript are inadequate if the objective was to model the mass balance of a glacier for which there are substantial amounts of data available, e.g. (distributed) mass balance measurement, locally measured atmospheric data, etc. But this is not the case here: to meet the objective of the study, i.e., modeling mass balance variability for each and every glacier in Europe, we had to find a model that can be applied to each and every glacier in Europe, and that produces a high correlation between observed and modeled mass balances. The need to assess correlation as the measure of model skill follows from the setup of the study – note that this implies that other measures of model skill – e.g., bias and/or variance – are not of primary interest (see also detailed arguments below).

Regarding (ii), we don't agree with the reviewer: There are more than 7000 glaciers in Europe, and far fewer than 1% of these have measured mass balance time series (or other observations) long enough to allow for studies on how interannual climate variability influences these glaciers. All previous studies on NAO-MB relations rely on these glaciers exclusively. The fact (correctly pointed out by the reviewer) that the spatial distribution of these glaciers is similar to the spatial distribution of glaciers covered in our study does not lessen this severe under-sampling problem. And to our knowlegde, no other mass balance model exists that has been shown by independent validation to have similarly high skill in the reconstruction of mass balances of the remaining $>$99% of the glaciers, as the models used by us do. We are convinced that extrapolating the results from previous studies of $\sim$30 glaciers to $>$7000 glaciers is not adequate, unless a strict evaluation shows it to be reasonable. Our study provides such evaluation.

Below, we give detailed replies to all of the reviewers concerns, and we have generally revised the manuscript to clarify the objective of the study.

Major comments

- **Comment:** The influence of the NAO on climate in Europe is well-known and several studies relating European glacier mass balance variations to the NAO have been conducted in the past. Several are mentioned by the authors in their introduction, but also Pohjola and Rogers (1997) and Rasmussen and Conway (2005) provide correlation analyses for Scandinavia. Although these studies only included glaciers with mass balance observations, the spatial coverage is similar to the current study. The authors do not state the additional value of their work. Furthermore, I wonder what applications the results could have in further research; the authors do not provide any.

  Response: We agree that the connection between NAO and glacier mass balance is well established – we also added the references given by the reviewer to the introduction section (and note that in the conclusions, we specifically indicate which of the results have been found previously). But the main point is exactly as the reviewer says: so far, only glaciers with measured mass balances have been used, and while the spatial coverage of previous studies is similar, there is a serious problem of under-sampling (which is valid in mass balance studies of glaciers in general): in Europe, there are 71 glaciers on which mass balances have been measured (i.e., less than 1 % of all glaciers - and note that this is the highest sampling density in the world). Most of these time series are too short to find significant relations to interannual climate variability such as the NAO. The present study gives very strong evidence that the previous results are not spoiled by this under-sampling; additionally, we find effects of the NAO that have previously been overlooked (e.g. in the Eastern Alps). Regarding further applications: the method used here could easily be extended to other glaciated regions in the world.

  All these findings are summarized in the conclusions section.

- **Comment:** The authors use a so-called minimal mass balance model to extend the mass balance records back to the beginning of the 20th century. I doubt whether such a simple model is suitable for the presented application. In this model accumulation is determined from estimated solid precipitation (multiplied with a factor in part of the model variants), while ablation is a linear function of positive air temperatures. The multiplication factors are derived from a calibration with measured annual mass balances or by assuming that the total area-averaged mass balance is zero. Subsequently, the modelled mass balance is correlated with the NAO. Since glaciers in both regions have distinct accumulation and ablation seasons, this principally comes down to correlating winter precipitation and summer temperatures with the NAO. Modelling the mass balance seems

like a redundant intermediate step. The actual glacier mass balance is affected by an interplay of processes, depending on more meteorological variables than only temperature and precipitation. To demonstrate whether their model provides a good measure for the correlation between mass balance and NAO, the authors should first compare the correlations of the measured and the modelled mass balances with the NAO over the same period.
Response: We are not aware of any other model type than a very simple one, that would be suitable for the presented application. We agree with the reviewer that it is always desirable to resolve the energy balance at the glacier surface to determine the mass balance. But this is incompatible with our objective to apply the model to each and every glacier in Europe. And modeling the mass balance with such a simple model is not a redundant intermediate step to correlating winter precipitation and summer temperatures with the NAO: Within the same CRU grid cell (i.e., identical correlation of winter precipitation and summer temperatures with the NAO), multiple glaciers may have quite a different response to the NAO: depending on altitude range, lengths of accumulation and ablation season vary, the amount of precipitation varies, and the fractionation between solid and liquid precipitation varies. All these things are accounted for in our simple model.

Of course, we do compare the correlations of the measured and the modeled mass balances with the NAO, as is suggested by the reviewer: We find that generally speaking, correlations between measured MBs and NAO are similar to correlations between modeled MBs and NAO (there are many references in the manuscript, including the Conclusions section, where we address this). But this alone would not be indicative of good model performance: What is critical is that the model reproduces the observed mass balances – and we demonstrate that this is case by validating the models rigorously, and then selecting the model that performs best in the measure of skill that we have to rely on: correlation.

- **Comment:** It appears that the authors only use measured annual mass balances

to calibrate the two multiplication factors (a and μ) in their model. This is an underdetermined problem, since one variable is used to calibrate two parameters. The parameter set with the smallest error may not reflect the real accumulation and ablation on the glacier, it is more a minimization of the errors in the model formulation and the climate input data. For a proper parameter calibration, the authors should calibrate a with the measured winter mass balance for the set of glaciers, which can easily be obtained from the World Glacier Monitoring Service.
Response: We do not understand why this should be an underdetermined problem. It is well possible to calibrate multiple parameters from one variable, as long as the time series of that variable is long enough. The cross validation not only allows for the estimation of a and μ, but also provides an estimate on the quality of the calibration, i.e. it tests whether the time series is indeed long enough to calibrate the parameters. As shown in Fig. 3 of the discussion paper, we can be confident that the parameter estimate is robust if more than 12 mass balance measurements exist.

- **Comment:** An even more serious problem occurs for the 'climatologically derived' model, where the multiplication factor for precipitation is not included. The authors implicitly assume that the CRU precipitation is representative of the precipitation falling on the glacier. The calibrated values for a_{opt} in Table 1 demonstrate that this is very likely not the case and from my own experience I also know that precipitation in glacier catchments is seriously underestimated in the CRU data.
Response: We agree that it is questionable whether CRU precipitation data are representative of the precipitation falling on the glacier. To be very clear: They don't have to be. What is critical for our purpose is that the **variability** of CRU precipitation is characteristic – since we base our study on correlation, bias and variance do not matter. We show very clearly that the model is able to capture mass balance variability in spite of the problematic CRU data (Fig. 4 of the dis-

cussion paper), and this is all we need.

- **Comment:** The calibration procedure of the three model variants gives unexpected results, for which the authors should provide explanations to convince the reader that these are correct. Most surprising is the result that the mass balance modelled with the least realistic model, the climatologically derived model (no precipitation correction, glaciers in equilibrium, no calibration with measurements) show the highest correlation with the measured records. I would expect that the individually trained model performs best. Does this imply that one should not bother about calibrating a mass balance model with measurements, but just use a parameter set that gives a zero annual mass balance? Secondly, the bias for the individually trained model is close to zero, but when the mean parameter values are used to compute mass balances, the bias becomes considerably negative for all regions. How can that happen? I would expect that mean parameter values would result in overestimated mass balances on one glacier and underestimation at another, so the mean bias would still be similar. Could it be that extreme outliers affect the cross validation? Last, the standard deviation of the biases is very large, in my opinion.
  Response: We agree that it is counter-intuitive at first that the climatologically derived model performs best for our purpose. But it actually shows that typically, the number of available mass balance measurements is not enough to reliably determine model parameters – and this problem is detected by the cross validation. We added a paragraph to the discussion to point this out.

  Of course, this does not imply that one should not bother about calibrating a mass balance model with measurements, but just use a parameter set that gives a zero annual mass balance (and we don't make any suggestions like this anywhere in the paper). As pointed out above, it depends on the objective of the study what one should do – and clearly, other objectives will need other models.

  Because of differing accumulation and ablation season lengths, and differing el-

evation characteristics, it cannot be expected that the bias of the mean model is linearly related to the mean bias of the individually trained model. While the scatter of biases in very large indeed, the reason for the bias of the mean model is therefore not extreme outliers of the individually trained model.

Detailed comments

- **Comment:** Section 2: This type of mass balance model is not able to incorporate effects of changes in glacier area and hypsometry on the mass balance. The authors should note that the mass balance they calculate is with respect to a reference geometry. Such a reference mass balance record is the correct choice when examining the relation between climate changes and mass balance variations (e.g., see Leclercq et al., 2011). However, most glaciers in the sample probably did not have a constant geometry over the period of observations. How might this affect the calibration and validation procedures?
  Response: Both terminus elevation and glacier geometry were treated as constant, i. e., the dynamic response of the glacier to mass changes was neglected. We included a footnote stating this explicitly, and a reference to Marzeion et al. (2012), where the implications of this assumption are discussed in detail. In short: the error introduced to the mass balance model by neglecting terminus elevation and area change is contained in the total error of the mass balance model determined by the (cross) validation, since measured mass balances of course do respond to terminus elevation changes. It is therefore not an additional, unquantified error of the model. Particularly on interannual (i.e., NAO-relevant) time scales, we argue that the error has to be small necessarily, since terminus elevation and area do not change much from year to year.

- **Comment:** p8, 2-3: It is not surprising that the correlation between temperature and elevation is very high: the CRU temperature at every grid point has been in-

terpolated from station records, taking elevation into account. Calculating vertical lapse rates from the CRU data practically comes down to determining the lapse rates used to interpolate the temperature records over the grid.
Response: We agree, and it was also our objective to do so: While CRU data contain no explicit information on lapse rates, it is inherently included, and we wanted to extract this information.

- **Comment:** p8, 3-7: The authors do not motivate their choice for a precipitation increase with elevation of 2% per 100 m. They only note that it is lower than vertical gradients derived from the CRU data. Why did they still choose this (relatively small) value? And is it necessary to apply a vertical gradient, is this effect not already captured in the parameter a?
Response: We chose this relatively low value as a conservative estimate - as we pointed out above, absolute precipitation numbers are not of primary importance. We also tested the sensitivity of model performance to changes in the prescribed precipitation lapse rate (i. e., lower or higher than 2% for every 100 m elevation increase), and found that while model performance decreases slightly for both lower and higher lapse rates, the results of our study are insensitive to the choice of precipitation lapse rate. We extended the discussion to clarify this. Note that the climatologically derived model, which is the only one for which the prescribed lapse rate is applied, does not include the parameter a (eq. 1 and 2 of the discussion paper).

- **Comment:** p8, 15-19: It is not clear to me how the solid precipitation fraction is determined. What do the authors mean with 'Within this range...'? Elevation range, temperature range, the range of the fraction [0,1]? If only part of the glacier receives solid precipitation, the glacier hypsometry is needed to compute the correct fraction, but I assume the authors assumed a constant distribution of area with elevation?
Response: The reviewer is correct – we implicitly assume (for the purpose of

precipitation fractionation) that the distribution of area is constant with elevation, and added a sentence to the manuscript to clarify.

- **Comment:** p8, 21: As already mentioned in the major comments, errors in the absolute value of $P_{i,\text{clim}}$ directly affect the value found for μ_{clim}, how can the authors be sure that $P_{i,\text{clim}}$ is representative of the precipitation on the glacier? Optimized values for the parameter are close to 2 instead of 1; a doubling of $P_{i,\text{clim}}$ would also double μ_{clim}.
Response: As detailed above, the absolute values of precipitation are of secondary importance for our study (this is the reason why the models perform well, even though the precipitation amount may be underestimated). Therefore, as a first approximation, it does not matter that a doubling of $P_{i,\text{clim}}$ would also double μ_{clim}. But more importantly, the validation of the model shows that the underestimation of precipitation plays only a minor role (and becomes most apparent in the reduced variance of the modeled mass balances, see below).

- **Comment:** p9, 16: What value is used for t_{lag}?
Response: As we state in the discussion paper, "t_{lag} is the length of autocorrelation in the measured mass balance time series." (page 9, line 17). It varies from glacier to glacier, and typically is one year. We added a brief statement.

- **Comment:** p10, 6-8: It is not clear to me what mse_{ref} represents and how it is calculated, can the authors clarify this?
Response: As we state, "mse_{ref} is the mean square error of a reference model (in this case, we use the mean of the measured mass balances as the reference model)" (page 9, lines 6-8). This means that as a best guess in the absence of a model, one could use the mean of observed mass balances to predict the mass balance of the year in question.

- **Comment:** p10, 15: I would like to see a comparison of μ_{clim} and μ_{opt}, the provided tables and figures do not allow for this. Are values for individual glaciers

very different?
Response: We included an additional column in Table 2 to show the values of μ_{clim}, which allows for a direct comparison with μ_{opt} in Table 1.

- **Comment:** p11, 13-16: I do not understand the explanation why the mean model performs better, what is the 'vastly increased data basis'?
Response: What we mean is that many more measured mass balances values enter the determination of the model parameters of the mean model – we changed the text to clarify this.

- **Comment:** p12, 10-11: For a reliable correlation analysis, the modelled mass balance should not only correlate highly with the observations, but the mass balance variations should have a comparable magnitude as well. This is not the case for any of the three models as shown in Figure 5, especially not for the climatologically derived model. It suggests that the model does not capture all processes that determine the interannual variability in the mass balance. The worst performance of the climatologically derived model seems to be a direct consequence of not including a: accumulation and hence ablation are underestimated.
Response: As we detail above (response to major comments), the reduced variance likely is the result of underestimation of precipitation in the CRU data. Ideally, this problem would be overcome by downscaling precipitation but doing so is out of the scope for this study — especially since underestimated mass balance variance does not affect the results presented here, which are based solely on correlation analysis. We added a paragraph on this to the discussion section.

- **Comment:** p12, 15-16: Figure 6 does not allow for a detailed comparison of the three models' performance. I would suggest the authors to include an additional figure showing the measured and three modelled mass balances in one panel, only covering the period with measurements.
Response: The purpose of Fig. 6 is not the illustration or comparison of model

performance. For this, we have Fig. 4 and 5, and Tables 1 and 2. The prupose of Fig. 6 is to illustrate typical time series of mass balance, precipitation, temperature, and NAO. We included the equivalent of Fig. 6 for all 39 glaciers with MB time series with more than 12 values into the supplementary material.

- **Comment:** p12, 19: As argumented above, I would think the climatologically derived model performs worst and therefore do not understand why the authors select it as the most reliable.
Response: As we explained, it depends on the purpose how to rate the models' performances. For our purpose, correlation is the best indicator – and since the climatologically derived model achieves the highest correlations, we choose this one. Other purposes will need other measures of model skill.

- **Comment:** p13, eq 8 and 9: With this notation, it is not clear which terms are anomalies and which are not. Please use a symbol like Δ, δ or ' to indicate the (two in this case) anomaly terms, also at other places in the manuscript.
Response: Since we have two anomaly terms (anomalies associated with the NAO, and anomalies unrelated to the NAO), we need two different symbols to mark anomalies. As the reviewer suggest, we use a prime (′) for one of them, and the subscript $_{\mathrm{NAO}}$ for the other. We state clearly for each term whether it is an anomaly or a mean value, and we think that the subscript $_{\mathrm{NAO}}$ allows for a more intuitive association with the NAO than another symbol, such as Δ or δ.

- **Comment:** p13, 15-16: Please motivate why these years were selected, the measured mass balance is not captured well for the first year.
Response: The purpose of Fig. 7 is not showing how well (or not) the model captures the observed mass balance - as stated above, this can be seen in Fig. 4 and 5, and the tables. It would also not be helpful to pick a period where the model performs best – those two years seem quite average in terms of model performance. The reason we picked these two years is that they are two con-

secutive years, one associated with a positive and one a with negative NAO, and in both years the amplitude is close to one standard deviation. I.e., all in all two typical years.

- **Comment:** p14, 1-15: Please discuss the results in Figure 7, clearly mainly variations in winter precipitation affect the mass balance on this glacier. Are similar results obtained for the other years in the record and for the other glaciers? Perhaps show a similar figure for a glacier with a small or anti-correlation with the NAO.
  Response: We discuss that winter precipitation is most relevant for glaciers in this area in the results section and the conclusions (page 19, lines 1-5 of the discussion paper). We included the equivalent of Fig. 7 for all 39 glaciers with MB time series with more than 12 values into the supplementary material.

- **Comment:** p15, 3-4: The correlations in Figure 8 are very low, are they significant? One would expect the highest correlations for the most maritime glaciers in southern Norway, but this seems not the case. Could this be because also summer months are included in the analysis?
  Response: As is stated in the legend, values below the 95% confidence intervall have been omitted. And we do indeed see the the highest correlations for glaciers in southern Norway (Fig. 8). Fig. 10 shows that the reason for the most western glaciers not having the strongest relation with the NAO is not caused by seasonal differences – both during the core NAO season (DJF) and the rest of the year, the westernmost glaciers have lower correlations with the NAO than the southernmost glaciers.

- **Comment:** p15, 10-11: Apparently, the authors expected to find the correlation to depend on the terminus elevation. Please elaborate on this, what would the terminus elevation represent, continentality or glacier size perhaps?
  Response: It is apparent that there is no simple relation between terminus ele-

vation and relation to the NAO within each of the three regions. The lower panels were included because in Marzeion et al. (2012), we found that within the Alps, there is a significant dependence of temperature and precipitation sensitivity to terminus elevation (which can be explained by wet glaciers typically extending lower than dry glaciers). One therefore could have expected a similar pattern to emerge in the relation to the NAO, but apparently, things are more complicated (even with this simple model).

- **Comment:** p15, 18-24: Why is exactly this division of the year chosen? The period March- November covers three seasons and different phases in the mass balance seasonal cycle. This period is not suitable to examine the effect of summer temperatures on the mass balance, instead only the summer months or the main melt season should be used. A more logical choice would for example be to use November-March as the winter season and May-September for the summer season. Then the mean NAO over both periods can be correlated with the cumulative mass balance over the same two periods, instead of averaging the values per month. This would practically come down to correlating the winter mass balance with the winter NAO and summer mass balance with the summer NAO.
  Response: The choice for this division is not related to the mass balance year, but to the NAO year. While the NAO itself does not show a strong amplitude modulation throughout the year, its impacts on temperature and precipitation do (this becomes clear in Fig. 7). Dec.-Feb. can therefore be considered the core NAO season. What the reviewer suggests is addressed much more directly in Fig. 11, where influences of precipitation and temperature are disentangled. Doing so has the advantage of not having to rely on a definition of accumulation and ablation period (which may vary over time) at all. Instead, temperature anomalies per definition only show through during the ablation season, and precipitation anomalies only show through when accumulation is occurring (this division is caused by the model formulation).

- **Comment:** p16, 5-10: Do the authors suggest that in parts of Norway summer temperatures correlate positively with the NAO while in other parts there is an anti-correlation? Are summer temperatures that different between those regions or are other factors at play?
  Response: This was not formulated well in the paper: What we meant is core winter (i.e., NAO) season vs. remainder of the year (see comment above). We reformulated to clarify.

- **Comment:** p17, 2-5: Not only the increase with altitude, but also underestimation in the CRU data is not taken into account.
  Response: Text changed accordingly.

- **Comment:** p17, 9-12: The best measure for the altitudinal gradient in precipitation on the glaciers are measured winter balance profiles, which are available for many glaciers in Scandinavia and the European Alps.
  Response: They are available for many glaciers, but essentially for those on which measured mass balances exist. It is not at all obvious that such profiles can be extrapolated to unsampled glaciers, since there are many local factors that influence winter accumulation other than direct precipitation. But as we argue above, the potential underestimation of precipitation in CRU is of secondary importance for our analysis, so that this shortcoming is minor.

- **Comment:** Table 1: What is the unit for μ_{opt}? Preferably, μ_{opt} (Table 1) and μ_{clim} (Figure 2) would be presented with the same unit, to facilitate a comparison.
  Response: We added a column for μ_{clim} in Table 2, and give the values with the same units.

- **Comment:** Table 2: Is the bias for southern Scandinavia negative because the mean measured mass balance was positive, i.e. as a result of the zero mass balance assumption? Please also include the calibrated values for μ_{clim}, to compare with Table 1.

  Response: We added a column for μ_{clim}, and yes: as we state on page 18 line 5 of the discussion paper, "the bias is therefore a measure of how far away the glaciers are from equilibrium with the climatological forcing."

- **Comment:** Figure 1: This figure is impossible to interpret, since different months are used for each grid cell. I would suggest to show panels for both winter and summer conditions.
  Response: Whether this figure is useful depends on the purpose: since temperature and precipitation are accumulated in the mass balance model, it does not matter in which exact month the values plotted in Fig. 1 occur. Additionally, we use this figure mainly for illustrating the position of the glaciers within the NAO-associated temperature and precipitation anomaly patterns in Europe. The question how glaciers' mass balances are influenced by these temperature and precipitation anomalies is addressed in Fig. 11.

- **Comment:** Figure 2: Please use smaller markers (like the green dots in Fig. 1) in this and similar figures, there is too much overlap to see the majority of the glaciers.
  Response: We reduced the marker size in Figures 2, and 8–11.

- **Comment:** Figure 3: Please indicate the value associated with the vertical line (12 years) and preferably extend the x-axis until 0.
  Response: We added the number to the caption. Extending the x-axis only adds white space, so we think the figure is more accessible as it is.

- **Comment:** Figure 8: Please use the same colour scale and range in all correlation figures to facilitate a comparison between figures. Why does Aalfotbreen have such a low correlation, it is one of the most maritime glaciers in the sample?
  Response: The different color scales were chosen deliberately to maximize the information that is conveyed by the figures: in some of the figures (4 and 9), the correlation is always positive by construction. In others, it may be both positive

or negative, but it does not make sense to compare the correlation between two mass balance time series (e.g., Fig. 11) with a correlation between a mass balance time series and the NAO index (Fig. 8). We therefore chose to maximize figure clarity, not comparability.

Aalfotbreen has quite a high correlation with the NAO from Dec to Mar - but a negative correlation during the rest of the year. So on an annual basis, the correlation is small.

Interactive comment on The Cryosphere Discuss., 6, 1, 2012.