Response to reviews of manuscript:

Assessment of glacier melt-model transferability: comparison of temperature-index and energy-balance models TC-2010-71

We appreciate the thoughtful and thought-provoking comments of both reviewers and have responded to each comment below. The reviews are copied verbatim and are italicized. We make reference to a revised manuscript, which we would like to submit to *The Cryosphere* subject to an invitation from the Editor.

Response to reviewer N. Arnold

General Comments

This paper presents the results of a set of model intercomparison experiments for 4 glacier melt models, applied to two small glaciers in the St Elias Mountains of Canada, for 2 years, 2008 and 2009. The models tested consist of a simple degree day model, an enhanced degree day model (including the spatial distribution of potential solar radiation) and two energy balance models with different complexities. Rather than a more traditional intercomparison experiment, in which different model experiments are conducted with the same data and with optimal parameter sets for the model and dataset under investigation in order to assess the differences between predictions made by different types of model, this paper focuses on the differences between the results of each model type when run with locally-optimised parameters versus runs with 'nonlocal' parameter sets in order to assess the transferability of models (more specifically in fact, the optimised parameter sets) in both space and time. It therefore assesses the degradation in model performance when parameter sets derived from a different glacier or different year are used to drive each class of model, in comparison with the optimised runs for each glacier for each year.

The key conclusion of the paper is that the most complex energy balance model is the most transferable, and suffers the smallest degradation in model effectiveness when used with non-local parameter values. The other models, however, show no consistent relationship between complexity and transferability. Overall, the experiments presented in the paper provide a useful addition to the literature on modelling glacier mass balance and melt by quantifying the differences produced by a set of models when used with differing, non-local parameter values. Within itself, the paper fulfils the criteria for publication well; it is well written; the methods used are clearly presented and appropriate for the nature of the study, the results are clearly presented (with the possible exception of Figures 3 and 5), and support the conclusions reached.

My main concern with the paper is this very simplicity, however. The model runs conducted with parameter values derived from the 'other' glacier and/or the 'other' year in the study have an intuitive 'logic' as to why those parameter values were selected, but in fact form just 3 experiments from an effectively infinite set of model experiments which could be performed with randomly chosen parameter values.

This is true in the sense that we selected these three experiments from the infinite set of possible experiments, because they are directly motivated by the data collected at the two sites over the two years. The hypothesis we are testing is that melt can be accurately simulated with parameters derived locally but in different years or with parameters derived from another nearby site. We are not explicitly investigating model sensitivity, in which case there would be a stronger argument for experimenting with random parameter values. A more comprehensive study than we were able to conduct would certainly incorporate parameters optimized from other glaciers in the region and for different years. To better acknowledge this we have added the following to section **2.3 Model transferability experiment design** after line 25:

"In transferring parameters between glaciers and/or years we aim to test the hypothesis that melt can be accurately modelled with parameters derived from other sites within the region or derived locally in other years. These transfer tests, while motivated by the data, represent only a tiny slice of parameter space and therefore do not give an indication of overall model sensitivity."

In a similar way, the optimisation strategy (based on minimising the RMS error between model predictions and stake measurements), whilst entirely valid and one which is commonly used within studies of glacier melt, is only one of a possible set of method for optimising model results. Given the stake measurements, the model could have been optimised with the simple R2 value, the Nash-Sutcliffe measure of model efficiency, or the mean absolute error, for example. Alternatively, the stake measurements could have been used to calculate summer mass balance gradients for the given glaciers in the study years and the models optimised to fit this gradient.

The reviewer's point is valid: the choice of optimizing the model by minimizing RMS error instead of any number of other statistics is essentially arbitrary, and we had neglected to acknowledge this in the paper. To rectify this omission we have added the following text to section **2.2.5 Model calibration and tuning** after line 12:

"Optimizing the empirical models by minimizing RMSE as opposed to any other statistic (e.g. R^2, mean absolute error, or Nash-Sutcliffe model efficiency) is ultimately an arbitrary but necessary decision. Minimizing another statistic would change the optimum parameter values by rebalancing the weight assigned to each difference between modelled and measured ablation. Although examining the effect of the chosen validation statistic(s) on model transferability is of interest we leave this to future investigation."

All these would have produced a different set of model results, with different though equally valid sets of parameter values, and could have produced different estimates of model transferability. This limitation could perhaps be acknowledged a little more within the paper; the simplicity and limited nature of the transferability experiments conducted in the study is already partially acknowledged within the paper (e.g. at the end of the introduction), but I think that acknowledging the limitations of single optimisation measures, whilst accepting that they are often necessary in practice, would be also useful.

We hope that the above revisions and the changes to the abstract and conclusions (detailed in response to Dr. Braithwaite's review, below) better acknowledge the limitations of our study.

In spite of this discussion, I am not suggesting that the authors be requested to revise the paper by conducting a 'full' sensitivity analysis or evaluate the range of model parameters produced by different optimisation strategies; however, I do think there is another set of model experiments *which would add to the paper, and which would further the aim of assessing model transferability. These experiments should consist of driving the four models with optimised parameters derived from the complete mass balance dataset (i.e. both glaciers and both years), and then assessing the model skill for each individual glacier and each individual year. In some ways, this 'reverses' the experiments in the paper as it stands, which assess the performance of 'local' parameters (i.e. derived from a single glacier, in a single year) in 'regional' model experiments (i.e. applied for a different glacier, in a different year). The runs I suggest would instead allow the assessment of the performance of 'regional' parameters on 'local' model experiments. Whilst there may be compromises in model performance when compared with the locally-optimised runs (the 'control' experiments in the paper) for the individual glaciers for individual years, it would be instructive to see which class of model can best simulate local mass balance values with regionally-derived optimisations for both glaciers in both years.*

We have conducted this experiment and incorporated the results into the paper. The following content has been added to the manuscript to explain the methods, results and conclusions of this experiment. In addition, Figures 3 and 5, and Table 2 have been updated to include the new results.

To the methods section we have added an additional subsection to explain the regional parameter experiments:

"**2.4 Regional parameter experiment**

In this experiment, we calibrate each model with the full complement of data from both glaciers and both melt seasons. We then evaluate model performance with these "regionally" calibrated parameters. The parameters are derived using the same model calibration methods as in the control runs (see above). Ablation stake records for both glaciers in both years are used in the derivation of parameters for the temperature-index models and the simplified energy-balance model. AWS albedo records from both glaciers in both years are used to derive parameters for both energy balance models. The regional parameter model runs are evaluated using RMSE between the simulated and measured ablation at the stake locations. These are compared to the control runs by comparing the total modelled surface ablation. This experiment has a similar design to some of the earlier work on glacier melt model transferability (e.g. Braithwaite, 1995; Shea et al. 2009)"

To the results section we have added an additional subsection:

"**3.4 Regional parameter results**

An assessment of model performance with the regional parameter values is included in Fig. 3 (R symbol) and summarized in Table 3. With one exception, these simulations produce higher RMSE values than the control runs using local parameters. RMSE values for the regional parameter tests tend to be closer to those for the control runs for South Glacier and for the energy balance models, though large variations are evident. All but two of the RMSE values for the regional parameter tests lie between those of the control runs and those of the spatial transferability tests. This result is sensible considering that the regional parameter values implicitly contain information about both glaciers. The fact that the DEBM for North Glacier 2008 produces a lower RMSE with the regional parameters, as compared to the control parameters, demonstrates that model calibration with a single record of albedo is insufficient to produce the best parameter set for glacier-wide simulations in this case. That RMSE values for the regional parameter tests tend to be closer to those of the control runs for South Glacier simulations suggests a parameter bias toward the glacier with higher ablation.

The absolute deviations between values of A_s estimated in the control runs and those estimated in the regional parameter tests (Fig. 5, R symbol; Table 4) show that the DEBM produces results most similar to the control runs. Deviations in estimated A_s relative to the control runs are lower, on the whole, for 2009 than 2008, with the notable exception of that simulated for North Glacier 2009 with the SEBM. This again suggests a disproportionate influence of locations and periods of higher ablation on regional parameter values."

The updated Figure 3 is show below:

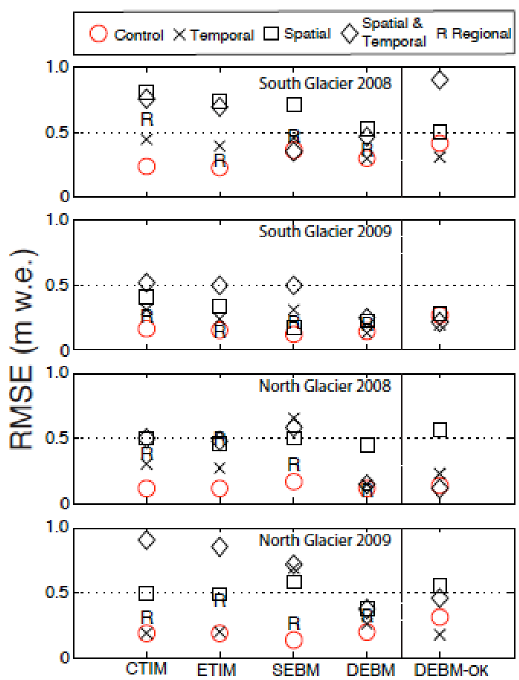

Fig. 3. RMSE for the control runs and the parameter tests for each of the four melt models. Each panel represents one data set as labelled. DEBM-OK is an adapted version of the DEBM which uses the albedo parameterization of Oerlemans and Knap (1998).

The updated Figure 5 is below:

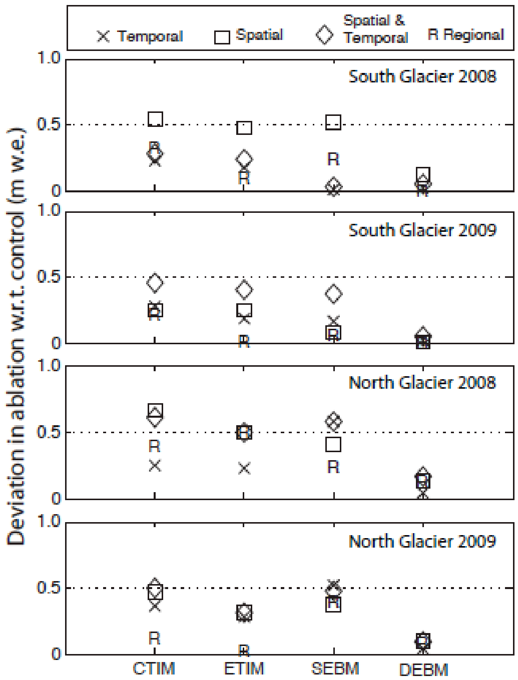

Fig. 5. Deviation between glacier-wide surface ablation estimated in each test relative to that estimated in the control run for each model.

The following has been added to the conclusions after line 13:

"Using “regional” parameter values derived from both glaciers and both years generally increases error in simulated ablation at the stake locations relative to using locally derived parameter values. "

The updated Table 2 including parameter values from the regional experiment.

Table 2. Parameters used in each of the four melt models. S08 is South Glacier 2008, S0 is South Glacier 2009, N08 is North Glacier 2008, N09 is North Glacier 2009 and R is th regional experiment where all data from both glaciers in both years are used to derive mode parameters.

Symbol	Units	Description	S08	S09	N08	N09	R
All Models							
E_f	m	Firn line elevation	2450	2450	2400	2480	2445
CTIM							
DDF_{snow}	w.e. mm d^{-1} K^{-1}	Degree day factor snow	7.0	6.0	2.5	5.0	5.5
DDF_{ice}	w.e. mm d^{-1} K^{-1}	Degree day factor ice	10.5	7.5	4.0	4.5	6.0
ETIM							
MF	w.e. mm d^{-1} K^{-1}	Temperature melt factor	2.0	1.9	1.4	1.7	0.9
r_{snow}	w.e. m h^{-1} K^{-1} W^{-1} μm^2	Radiation melt factor snow	0.66	0.62	0.23	0.46	0.80
r_{ice}	w.e. μm h^{-1} K^{-1} W^{-1} m^2	Radiation melt factor ice	1.5	1.05	0.56	0.72	1.40
SEBM							
C_0	W m^{-2}	Independent radiation constant	-49	-44	-57	-42	-50
C_1	W m^{-2} K^{-1}	Temperature radiation constant	-2.2	2.8	1.8	3.9	2.5
SEBM & DEBM							
α_o	–	Initial albedo of snow	0.85	0.85	0.85	0.85	0.85
$\frac{d\alpha_i}{dZ}$	100 m^{-1}	Change in ice albedo with elevation	0.11	0.11	0	0	0.55
a_1	$\ln(°C)^{-1}$	Albedo rate constant	0.032	0.031	0.042	0.030	0.032
a_2	day$^{-\frac{1}{2}}$	Albedo rate constant	-1.54	-1.68	-1.71	-1.61	-1.76
a_3	–	Albedo rate constant	0.0074	0.0112	0.0104	0.0142	0.0144
a_4	h m^{-1}	Albedo rate constant	44	30	88	60	30
α_i	–	Albedo of ice	0.34	0.33	0.44	0.43	0.38
α_{slim}	–	Lower limit of snow albedo	0.66	0.66	0.66	0.66	0.66
α_{flim}	–	Lower limit of firn albedo	0.56	0.56	0.56	0.56	0.56
α_{stof}	–	Albedo snow–firn transition	0.03	0.03	0.03	0.03	0.03
α_{ilim}	–	Lower limit of ice albedo	0.16	0.16	0.27	0.27	0.22
Γ_T	K km^{-1}	Temperature lapse rate	-6.0	-6.0	-5.3	-5.3	-5.7
Γ_p	mm km^{-1}	Precipitation lapse rate	2.3	2.3	1.2	1.2	1.8
DEBM							
α_{ter}	–	Albedo of terrain	0.21	0.21	0.21	0.21	0.21
ϵ_{ter}	–	Emissivity of terrain	0.95	0.95	0.95	0.95	0.95
T_{sub}	°C	Min subsurface temperature	-30	-30	-30	-30	-30
h	m	Thickness of subsurface	0.10	0.10	0.10	0.10	0.10
Z_{thr}	m	Snow threshold	0.01	0.01	0.01	0.01	0.01
b_1	mm	Roughness rate constant	0.91	0.91	0.91	0.91	0.91
b_2	°C	Roughness rate constant	1.36	1.36	1.36	1.36	1.36
b_3	°C	Roughness rate constant	0.054	0.054	0.054	0.054	0.054
b_4	mm	Roughness rate constant	2.3	2.3	2.3	2.3	2.3
z_{oi}	mm	Roughness length of ice	0.65	0.65	0.20	0.20	0.42

Specific comments

I have very few specific comments or corrections to make, as the paper is commendably free of grammatical errors or uncertainties. However, although in general I think figures are easier to interpret than tables, in this case I wonder if Figures 3 and 5 could be usefully replaced (or at least supplemented) with a table of the actual RMS error (or As for Figure 5) values. The differences between some model runs (e.g. South Glacier 2008 CTIM vs ETIM, North Glacier 2008 DEBM control, temporal and spatio-temporal runs) are very slight, and not easy to discern from the Figures; given the relatively small set of numbers involved, I think simple tables of the results would be more effective.

Whether to present the data in figures 3 and 5 as tables or as figures presented a dilemma from the first drafting of this manuscript. The figures provide the advantage of visually demonstrating the overall patterns and large differences between some of the model runs, while tables allow examination of the finer differences between the model runs but readers tend to lose the overall sense of magnitude. We have supplemented Figures 3 and 5 with Tables 3 and 4 to allow readers to examine the data in both ways. To reduce redundancy the tables do not display the identical statistic to their respective figures. Table 3 expresses RMSE as a percent instead of the m w.e. used in Figure 3. Table 4 contains simulated glacier-wide surface ablation instead of the difference in this quantity between the control run and the various tests used in Figure 5. References to Tables 3 and 4 are worked into the text along with references to figure 3 and 5 in subsections **3.1 Comparison to ablation stakes** and subsection 3.3 **Comparison of transfer experiments to control runs.**

The new table 3 is below:

Table 3. RMSE for each model, experiment, glacier and year, expressed as a percentage (%). S08 is South Glacier 2008, S09 is South Glacier 2009, N08 is North Glacier 2008 and N09 is North Glacier 2009.

	CTIM	ETIM	SEBM	DEBM
S08				
Control (○)	18	17	28	23
Temporal (×)	34	30	35	23
Spatial (□)	63	57	55	41
Spatial-temporal (◇)	59	54	27	36
Regional (R)	47	22	36	28
S09				
Control (○)	16	15	12	14
Temporal (×)	31	23	31	13
Spatial (□)	40	34	17	22
Spatial-temporal (◇)	51	49	48	25
Regional (R)	26	14	21	20
N08				
Control (○)	30	30	45	30
Temporal (×)	85	77	186	40
Spatial (□)	142	127	141	126
Spatial-temporal (◇)	142	134	166	39
Regional (R)	108	142	85	27
N09				
Control (○)	14	14	10	15
Temporal (×)	14	15	53	20
Spatial (□)	39	37	45	29
Spatial-temporal (◇)	70	66	55	29
Regional (R)	24	33	20	25

The new table 4 is below:

Table 4. Simulated glacier-wide surface ablation (A_s) for each model, experiment, glacier and year, expressed as m w.e. S08 is South Glacier 2008, S09 is South Glacier 2009, N08 is North Glacier 2008 and N09 is North Glacier 2009.

	CTIM	ETIM	SEBM	DEBM
S08				
Control (○)	-0.98	-0.96	-0.92	-0.79
Temporal (×)	-0.76	-0.79	-0.94	-0.77
Spatial (□)	-0.45	-0.49	-0.41	-0.67
Spatial-temporal (◇)	-0.71	-0.73	-0.95	-0.74
Regional (R)	-0.65	-0.86	-0.67	-0.80
S09				
Control (○)	-1.06	-1.00	-0.86	-0.81
Temporal (×)	-1.34	-1.17	-0.71	-0.83
Spatial (□)	-0.81	-0.74	-0.94	-0.82
Spatial-temporal (◇)	-0.61	-0.60	-0.49	-0.76
Regional (R)	-0.85	-0.98	-0.79	-0.82
N08				
Control (○)	-0.48	-0.50	-0.31	-0.46
Temporal (×)	-0.72	-0.72	-0.89	-0.50
Spatial (□)	-1.14	-0.99	-0.71	-0.60
Spatial-temporal (◇)	-1.09	-0.98	-0.89	-0.62
Regional (R)	-0.88	-1.01	-0.55	-0.60
N09				
Control (○)	-0.97	-0.89	-0.96	-0.70
Temporal (×)	-0.62	-0.61	-0.44	-0.66
Spatial (□)	-1.43	-1.20	-0.58	-0.79
Spatial-temporal (◇)	-1.47	-1.20	-0.49	-0.79
Regional (R)	-0.85	-0.92	-0.56	-0.79

My other specific comment concerns the title of the paper; overall, I feel that the paper really sets out to compare model parameter transferability rather than the transferability of the models themselves, as the paper does not comment much on the actual model results themselves. This is fine given the aims of the paper, but I think it should be reflected in the title.

We have changed the title of the paper to:

<u>Assessment of parameter transferability in temperature-index and energy-balance melt models</u>

Overall

To conclude, I would be happy to see this paper published with the additions and corrections suggested. I do not think that these require the paper to be re-reviewed, but if the editors felt this was necessary I would be happy to review the paper again.

Response to reviewer R. Braithwaite

I start my review by quoting the short abstract of Van der Veen (1999):

"Numerical models of the cryosphere cannot be verified and their truthfulness in providing an accurate description of actual physical processes cannot be proven conclusively. At best, models can be confirmed by comparing predictions with independent data that were not used to calibrate model parameters. The more such confirmations are achieved, the greater the confidence that can be placed in the model as a representation of Nature. Most prognostic cryospheric models have not be adequately calibrated and confirmed, and skepticism towards their predictions is therefore warranted." We do not have to completely agree with the above quote to see that the present discussion paper by MacDougall et al. may be a very important contribution to this vexed question of evaluating cryospheric model performance.

The melt model is "calibrated" for one situation, e.g. one melt season on one glacier, by calculating the parameter value(s) needed to make the model predictions "fit" the measured data for the situation in questions. For example, the degree-day factor in a simple degree-day model could be calculated as a regression coefficient if one has a series of simultaneous data for melt and temperature within the period in question. If the regression coefficient is accompanied by a relatively high correlation coefficient between observed melt-temperature data, we might be tempted to say that we have "verified" the model for the dataset that we have. However, we still cannot simply assume that our value of degree-day factor is valid for another situation, e.g. for another melt season on another glacier, and we should follow MacDougall et al. by applying the degree-day factor found for one situation to another situation to "confirm" the model in the terms of Van der Veen (1999). Such confirmation will obviously be limited by the available data to only a few melt seasons on a few glaciers but should help to build general confidence in our model. MacDougall et al. deserve full credit for reinvigorating this important problem but I do have some problems with the present paper, which I outline below.

Although interesting and important, this discussion paper claims far too much in its abstract. Applying the transfer principle to only two melt seasons on two nearby glaciers is rather minimalistic and insufficient to make claims about one model being more transferable than another model. On the second page of the paper, the assessment is correctly described as "an optimistic one". The paper's abstract and conclusions should include this important reservation.

We have attempted to make the limitations of our study more conspicuous in the abstract and conclusion, as it was not our intention to inflate the contributions of this study. The manuscript title has also been changed, in response to comments from the other reviewer, with the new title better reflecting the fact that this study examines the transferability of model parameters, rather than models themselves. The revised abstract reads:

"Transferability of glacier melt models is necessary for reliable projections of melt over large glacierized regions and over long time-scales. The transferability of such models and their parameters has been examined for individual model types, but inter-comparison has been hindered by the diversity of validation statistics used to quantify transferability. We apply four common types of melt models---the classical degree-day model, an enhanced temperature-index model, a simplified energy-balance model and a full energy-balance model---to two glaciers in the same small mountain range. The transferability of key parameters within each model is examined between these two sites and over two melt seasons. Within this limited dataset, we find that the full energy balance model produces the most consistent results; deviations in estimated glacier-wide summer surface ablation are ≤35% when parameters derived from the other glacier and/or melt season are used in this model. The other three models exhibit deviations in glacier-wide summer surface ablation of ≥100% in the same experiments. If model parameters are instead calibrated with all the available data, from both glaciers and years, the full energy balance model again yields the most consistent results. Beyond this, we do not find a simple relationship between model complexity and model-parameter transferability. However, further study involving additional field sites and over more melt seasons would be required to generalize these results."

The revised conclusion reads:

"We have examined parameter transferability in space and time within four commonly used types of glacier melt models. These models were applied to two small glaciers in the St. Elias Mountains of northwestern Canada over two melt seasons. In this study, the physically-based energy-balance model yielded the most consistent results; modelled glacier-wide summer surface ablation varied by ≤35% when using model parameters calibrated for a different melt season and/or the other glacier. In the analogous experiments, the other models produced variations in estimated glacier-wide summer surface ablation exceeding 100%. No simple relationship between model complexity and model transferability was observed within this limited dataset. Deviations in estimated glacier-wide summer surface ablation between the models in the control runs themselves, where parameters are calibrated for a specific site and year, were 24-41%. Using "regional" parameter values derived from both glaciers and both years generally increases error in simulated ablation at the stake locations relative to using locally derived parameter values. There is a need for similar experiments to be conducted in other glacierized regions and over longer time scales for general conclusions about model transferability to be drawn."

I understand the transfer concept of MacDougall et al. for a simple melt-climate model like the degree-day model. For example, Braithwaite (1995) calculated degree-day factors for large data samples at two locations in Greenland. His samples were so big, e.g. daily melt and temperature data for 415 days at Nordbogletscher and 512 days at Qamanârssûp sermia, that sampling errors were relatively small. However, following the approach of MacDougall et al., Braithwaite (1995) should have compared models between smaller samples, e.g. for six or seven melt seasons at each of the two glaciers. I will probably attempt this as soon as I have finished this review. As MacDougall et al. correctly note, Hock (2003) gives a summary of degree-day factors for glaciers in widely ranging conditions and it may be possible to test the transferability concept for some of these data.

The above paragraph refers to the degree-day model, i.e. a model with one parameter and one input variable. I find the transferability concept more difficult to understand when applied to the other three models considered by MacDougall et al., i.e. models with more than one *parameter and more than one input variable. I doubt if many people have tried to transfer a complete energy balance model from one situation to another. For example, the model of Arnold et al (1996) seems to involve transfer of longwave radiation and surface roughness parameters from Braithwaite and Olesen (1990) but these are combined with a much-improved approach to shortwave radiation.*

It is true that most previous work on melt model transferability has focused on temperature-index models. However, there have been several recent studies exploring the transferability of more complex models. For example Carenzo et al. (2009) [Carenzo, M. and Pellicciotti, F. and Rimkus, S. and Burlando, P., 2009, Assessing the transferability and robustness of an enhanced temperature-index glacier-melt model, Journal of Glaciology, 55(190), 258-274.] explored the transferability of an enhanced temperature-index model, while MacDougall and Flowers (2010) explored the transferability of a full energy balance model. We are also aware that Dr. F. Pellicciotti's group at ETH Zürich is exploring the transferability of various glacier melt models (Refer to AGU Fall Meeting 2010 abstracts: "Conceptual melt models: the past or valuable tools for future scenarios?" and "Modeling distributed glacier ablation for climate change simulations: comparison of an energy-balance and an enhanced temperature-index model").

The conclusion that the full energy balance model is the most transferable of the four models almost sounds inevitable as it must be easier to transfer a model with many parameters and many input variables. However, the reason why some workers do not use energy balance models is that there is little evidence that they perform better than a simple temperature-index for the modelling of day-to-day variations in melt at some location. Therefore, in the terms of Van der Veen (1999), I suggest that we need to discuss the "calibration" of our models before we discuss their "comparison". Perhaps we need a single statistic to combine the two.

Based on the comment above, there is a need for us to clarify in the manuscript that we are transferring model parameters, rather than models themselves. This fact is reflected in the new title and the revised text, where we have tried to make it clear that we have transferred only model parameters (rather than input meteorological variables or treatments of subprocesses within the models). The concept of parameter transfer for the energy balance models we use should therefore be analogous to parameter transfer for the temperature-index models, with the exception that we cannot transfer all of the energy-balance model parameters; we lack the data to calibrate each parameter individually for each glacier and year, and therefore only transfer those parameters for which we have sufficient calibration data.

In responding to Neil Arnold's review, we have conducted additional simulations whereby each model was calibrated using all of the available data. This procedure would be more analogous to the calibration procedure of Braithwaite (1995), though clearly with far fewer data. These results are presented along with our original results; they are often worse than those of the control runs, and are only sometimes better than those of the runs where local parameters have been transferred in space and time. This type of experiment serves a different purpose than our original experiments: it focuses on comparing the performance of different models when all of the data are used to calibrate each one. We think this is a useful addition to the study and hope that our original intention of assessing the spatial and temporal transferability of parameters within each model type is now clear.

The result that the energy balance model was the most "transferable" was not unexpected, and we recognize that the performance of such a model can be a strong function of the way subprocesses are treated or parameterized within the model. We also find that when local data are available for calibration, the other models (temperature-index and semi-empirical energy balance) perform as well or better than the full energy balance model, so we would not dispute that these simpler models are preferable for some applications. However, we are interested in understanding how best to model melt for locations and times where parameters cannot be specifically calibrated. Although the conclusions we can draw from our study are limited, we suspect that using constant and uniform degree day factors in studies spanning vast areas (e.g. entire mountain ranges) and long (decadal) timescales, as is often done, may be problematic.

The paper is not always easy to understand as it refers the reader to other papers, included in the reference list, for "further details". A recent search on the ISI Web of Science (12 January) failed to detect either MacDougall (2010) or MacDougall and Flowers (2010). No doubt, these papers are on their way and will appear sometime but, if they revise this discussion paper, MacDougall et al. should add further information to make the paper more self-contained.

MacDougall (2010) is a Master's thesis and is currently available at: https://theses.lib.sfu.ca/sites/all/files/public_copies/etd6270_amacdougall_pdf_72594.pdf and should be available officially through the Simon Fraser University library after June 2011. The essential details of MacDougall (2010) are contained in MacDougall and Flowers (2010), which is available as an advanced on-line copy with a subscription to the *Journal of Climate* (http://journals.ametsoc.org/doi/pdf/10.1175/2010JCLI3821.1). This paper is in the printing queue, and is scheduled to be included in the 1 March 2011 issue of the journal. We agree in principle that manuscripts should be as self-contained as possible, however, to explain the energy balance model in full detail would consume four full journal pages. We think that our abbreviated version will be adequate for most readers, while interested readers will be able to refer to MacDougall and Flowers (2010).

Although MacDougall et al. apply the "enhance temperature-index model (ETIM) as it is described in the literature, I think equation (2) is physically incorrect as the radiation melt factor is effectively multiplied by the air temperature. The simplified energybalance model (SEBM) in equation (3) is physically more correct in splitting the energy balance into separate components for shortwave radiation and temperature.

This argument has been made by others, but our motivation for including this model was its widespread popularity. We initially considered including the enhanced temperature-index model of Pelliciotti and others (2005) [Pellicciotti, F., B. Brock, U. Strasser, P. Burlando, M. Funk and J. Corripio, An enhanced temperature-index glacier melt model including the shortwave radiation balance: development and testing for Haut Glacier d'Arolla, Switzerland, *Journal of Glaciology*, 51(175),573-587,2005], but this model is designed to be calibrated against the results of an energy balance model, and hence would not conform to our study requirements.

As a last point, MacDougall et al. quote the wide range of estimates of 21st century sea-level rise as a motivation for re-examining glacier melt models. However, the large differences between the estimates they quote are probably more down to different estimates of glacier volumes and areas in different regions of the world.

This is a good point. We have changed text to acknowledge this other source of uncertainty in the Introduction after line 23, as follows: "Such studies have produced a wide range of projected contributions of mountain glaciers and ice caps to 21st century sea-level rise, from 4 cm Sea Level Equivalent (SLE) (Raper and Braithwaite, 2006) to 36 cm SLE (Bahr et al., 2009), motivating a reexamination of the assumptions necessary to apply these models over large regions. Uncertainties in the total volume of glaciers and ice caps (e.g. Raper and Braithwaite, 2005; Meier et al., 2007) and variability between GCM output (Randall et al., 2007) must also contribute to the range in projected sea-level rise."

MacDougall et al. certainly raise model transferability as an important issue and I value MacDougall et al. as an item of discussion. However, I wonder if they should revise this paper at all as I suspect that they will do further work that may be more reliable, for example for more seasons on more glaciers. Perhaps they should not embarrass their future selves with a premature conclusion!

We do plan to continue work on this topic, but would also like to submit the results of this preliminary study in a revised manuscript. We hope to avoid embarrassing our future selves by tempering the conclusions made in this paper with an acknowledgment of the limitations of our study in its spatial and temporal scope. Many years and glaciers from now, perhaps we will be able to draw more general conclusions about the transferability of melt models and their parameters.

References

Arnold, N.S., I.C. Willis, M.. Sharp, K.S.Richards and W.J.Lawson. A distributed surface energy-balance model for a small valley glacier. I. Development and testing for Haut Glacier d'Arolla, Valais, Switzerland. Journal of Glaciology, 42(140), 1996, 77-89.

Braithwaite, R.J. and O.B. Olsen. A simple energy-balance model to calculate ice ablation at the margin of the Greenland ice sheet. Journal of Glaciology 36(123), 1990, 222-228.

Braithwaite, R.J. Positive degree-day factors for ablation on the Greenland ice sheet studied by energy-balance modelling. Journal of Glaciology 41(137), 1995, 153-160.

Hock, R. Glacier melt: a review of processes and their modelling. Prog. Phys. Geog.,29(3), 362-391, 2005.

Van der Veen, C.J. Evaluating the performance of cryospheric models. Polar Geography,23 (2), 1999, 83-96.