spin-up period (Figure 3.5). The first step, for the period 1985-1992, was based on 1985 glacier areas. The second step, for the period 1993-1999, was based on glacier extents from 2000. This approach explicitly incorporates the hydrologic effects of glacier retreat during the calibration period, helping to avoid parameter bias that could arise from using only one glacier coverage. For example, using only the 1985 glacier cover would result in glacier cover being overestimated through most of the calibration period, possibly distorting the values of parameters that govern the intensity of glacier melt; that is, the model would have to simulate an artificially reduced melt intensity to compensate for the overestimate of glacier area. Glacier net mass balance (b_n) for the entire basin was derived from net mass balances for each elevation band. Net mass balances for elevation bands were calculated as sums of winter and summer balance. October 1st was taken as the beginning of the accumulation season but the ablation period was allowed to be extended beyond that date in case of some ice melt in the next accumulation season.

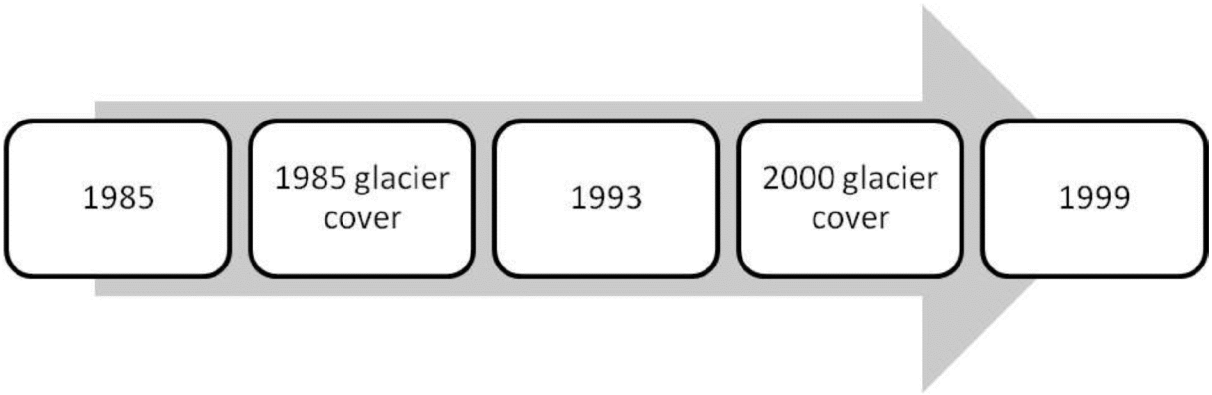

Figure 3.5. Time periods and glacier coverages used for model calibration.

The calibration procedure, outlined in Figure 3.6, starts with finding a benchmark parameter set by maximizing the Nash-Sutcliffe efficiency (E) or, in terms of GLUE, the generalized likelihood measure. This was done with *genoud* (Mebane & Sekhon 2009), an optimization algorithm in R (R Development Core Team 2009) that combines evolutionary algorithm methods with a steepest gradient descent algorithm. A large negative number was returned for parameter sets that did not fulfill the multiple criteria listed in Figure 3.6, to ensure that the optimization algorithms not only maximize E but also search for solutions that meet the additional criteria. In a second step, a Latin Hypercube Search (LHS) with 10,000 model runs was performed. Latin hypercube designs are most often used in highly

dimensional problems, where it is important to sample efficiently from distributions of input variables. Parameter sets from the 10,000 model runs were constrained by criteria given in Figure 3.6. If no parameter sets with Nash-Sutcliffe efficiencies greater than the benchmark efficiencies minus a threshold were found, all parameter sets were rejected and the second step was repeated with adjusted (narrowed) parameter ranges until enough parameter sets (~30) were found that fulfilled all criteria (i.e., behavioural parameter sets).

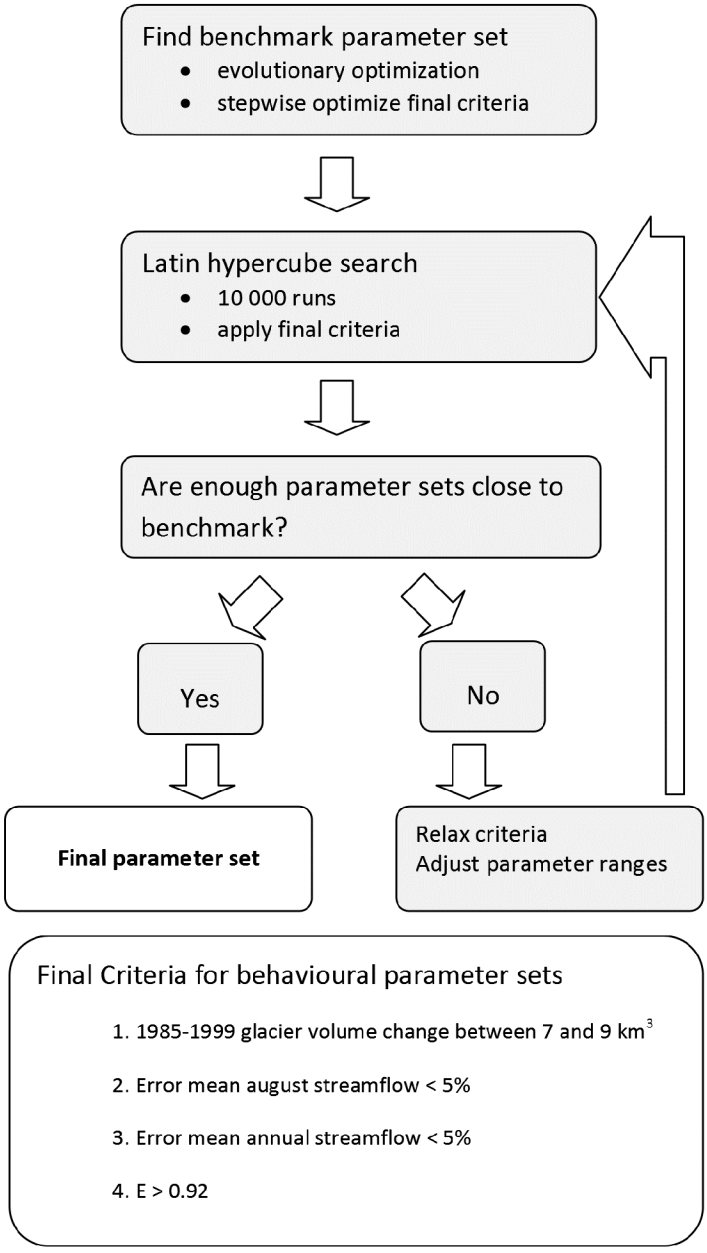

Figure 3.6. Flow chart illlustrating model calibration procedure.

There are many examples in the hydrologic literature where acceptable parameter sets are obtained by random sampling of the usually high-dimensional parameter space and by subsequently picking the best performing parameter sets according to one or multiple criteria (Konz and Seibert 2010; Stahl *et al.* 2008). In comparison to this common procedure, the procedure described here differs in the sense that it ensures that the final ensemble parameter set contains solutions that are within a given range of the potentially "best" solution in a given parameter space. In a highly dimensional parameter space, random sampling with even thousands of model runs does not guarantee that the best parameter combinations are found. Without prior knowledge of how well the "best" possible solution performs, the modeller will usually relax criteria in order to obtain enough acceptable parameter sets, with the possible result that criteria for acceptable parameter sets are more relaxed than necessary. Theoretically, there are two ways to proceed when no parameter set found by the LHS provides solutions close to the benchmark parameter set: either increase the sample size or adjust the prior parameter distributions (decrease the range). Increasing the sample size is the favorable solution because it likely leads to a more diverse set of parameters. However, the number of model runs is limited by computational power (even with multiple CPUs it would take months for the Mica basin). Given the time constraints in this study, adjusting the prior parameter distributions was the preferred option.

The parameter ranges for model calibration and uncertainty analysis were based on default values provided in the HBV-EC manual (Canadian Hydraulics Centre 2006), values reported in previous studies (Stahl *et al.* 2008; Hamilton *et al.* 2000), our own experience with applying HBV-EC on other catchments, and by visually testing the influence of parameters on the simulated hydrograph: with the modest glacier cover in Mica, a visual inspection of simulated hydrographs provides more information on the sensitivity of modelled streamflow to the various glacier parameters than a single goodness of fit measure such as E.

HBV-EC routes water with only two reservoirs (slow and fast). For Mica basin, the slow reservoir effectively models the low flow period during winter. As a consequence, runoff routing during the melt period is modelled primarily with the single fast reservoir, which makes it difficult for the model to simulate the occasional intense summer (convective) rainfalls, for which the spatial pattern and nonlinearity of runoff generation differs from that associated with snowmelt. To minimize the effect of intense rainfalls on the parameterization of meltwater generation and runoff, we excluded the five highest rainfall events from model calibration and uncertainty analysis.

3.4.2 Results

Prior parameter distributions for LHS were assumed uniform at all stages. As a generalized likelihood measure we used the Nash-Sutcliffe efficiency (E). Because the slow reservoir is used to model low flows only, KS (Table 3.2) can be constrained to parameter ranges that match just the winter flows.

The benchmark parameter set obtained by the combined evolutionary-steepest gradient optimization matched observed streamflow with a Nash-Sutcliffe efficiency of 0.93 for the calibration period and 0.95 for the test period. A first 10,000 run LHS within the initial parameter ranges (parameter range step 1 in Table 3.2) found no acceptable parameter sets that met all criteria. Although 28 parameter sets had $E >$ 0.91, all of these parameter sets had to be rejected because none fulfilled all of the additional criteria. Without prior knowledge of the benchmark E, the common procedure would now have been to find acceptable solutions with relaxed criteria. However, from the benchmark parameter set we knew that better solutions are possible and that 10,000 runs are just too few to sample the parameter space for acceptable solutions. Therefore, a second LHS with adjusted parameter ranges was performed. This search found 17 acceptable parameter sets, but histograms indicated that two parameters in the acceptable parameter sets were predominantly sampled near a range boundary. We therefore performed a third LHS with slightly refined parameter ranges. A detailed description of the calibration process is provided below.

Figure 3.7 shows the wide range of modelled glacier volume changes that can still lead to E close to the benchmark. Results from the first LHS suggest that equifinal parameter solutions are possible with glacier volume loss ranging from 5-40 km^3. Unlike the findings reported by Stahl *et al.* (2008), E does not peak at the observed glacier volume loss, which is likely due to the lower glacier coverage in Mica and lower sensitivity of streamflow simulations to glacier parameters. This result underlines the importance of using observed glacier volume changes to constrain model parameters, particularly in basins with modest glacier coverage like Mica. Note that the blue dots in Figure 3.7, the second LHS, represent higher maximum E because of the greater sampling density within the restricted parameter space and not necessarily because the glacier volume loss is close to the observed. More intense sampling within the parameter space that leads to higher glacier volume losses would possibly lead to higher E at higher glacier volume losses as well. Figure 3.7 also shows that, although there is a wide range in glacier volume loss that can lead to high E, no acceptably high E can be found with parameter combinations that result in volume gain by glaciers.

Table 3.2. Description of calibrated model parameters, benchmark parameter values, parameter ranges for LHS, and correlation between each parameter and glacier net mass balance (b_n) and Nash-Sutcliffe efficiency (E).

Model routine	Parameter	Description	Benchmark Optimization	Parameter range Step 1		Correlation Step 1		Parameter range Step 2		Correlation Step 2		Parameter range Step 3		Correlation Step 3	
				from	to	b	E	from	to	b	E	from	to	b	E
Climate	TLAPSE	Temperature lapse rate (°C m^{-1})	0.00760	0.006	0.009	0.51	0.25	0.0075	0.009	0.48	-0.37	0.0075	0.009	0.58	-0.42
	PLAPSE	Fractional precipitation increase with elevation (m^{-1})	0.00003	0.00002	0.0005	0.22	-0.31	0.00002	0.0002	0.23	-0.10	5E-06	0.0001	0.28	-0.03
Snow	AM	Influence of aspect/slope on melt factor	0.12481	0	0.7	0.00	0.08	0.1	0.6	-0.01	-0.31	0	0.6	-0.11	-0.32
	CMIN	Melt factor for winter solstice in open areas (mm $°C^{-1}$ day^{-1})	2.96739	2	4	-0.60	-0.07	2.2	3.4	-0.69	0.55	2.6	3.4	-0.50	0.33
	DC	Increase of melt factor between winter and summer solstice (mm $°C^{-1}$ day^{-1})	0.18447	1	3	-0.44	-0.28	0.5	2	-0.42	-0.05	0	1.2	-0.45	0.10
	MRF-low	Ratio between melt factor in forest to melt factor in open areas below 1200 m	0.71696	0.4	0.9	0.00	0.00	0.6	0.8	0.00	-0.01	0.6	0.8	0.01	0.00
	MRF-high	Ratio between melt factor in forest to melt factor in open areas above 1200 m	0.72330	0.4	0.9	-0.02	0.00	0.6	0.8	-0.01	0.01	0.6	0.8	0.00	0.01
Glacier	MRG	Ratio of melt of glacier ice to melt of seasonal snow	1.04472	1	2	-0.33	-0.23	1	1.4	-0.24	-0.04	1	1.4	-0.33	0.00
	DKG	Difference between minimum and maximum outflow coefficients for glacier water storage (day^{-1})	0.02942	0.005	0.1		-0.07	0.01	0.05		0.01	0.01	0.05		0.01

Model routine	Parameter	Description	Benchmark Optimization	Parameter range Step 1		Correlation Step 1		Parameter range Step 2		Correlation Step 2		Parameter range Step 3		Correlation Step 3	
Glacier (cont.)	KGMIN	Minimum outflow coefficient for glacier water (day^{-1})	0.01652	0.005	0.1		-0.01	0.01	0.05		0.09	0.005	0.03		0.09
Runoff	KF	Proportion of fast reserviour release (day^{-1})	0.14108	0.01	0.4		0.23	0.1	0.35		-0.19	0.05	0.3		0.18
	ALPHA	Exponent to adjust release rate of fast reservoir	0.13276	0.01	0.3		0.04	0.05	0.2		-0.11	0.05	0.2		0.05
	KS	Proportion of slow reservoir release (day^{-1})	0.01442	0.0005	0.05		-0.04	0.001	0.03		0.15	0.0005	0.015		0.07
	FRAC	Fraction of runoff directed to fast reservoir	0.89401	0.5	0.9		0.11	0.7	0.9		0.35	0.7	0.9		0.49

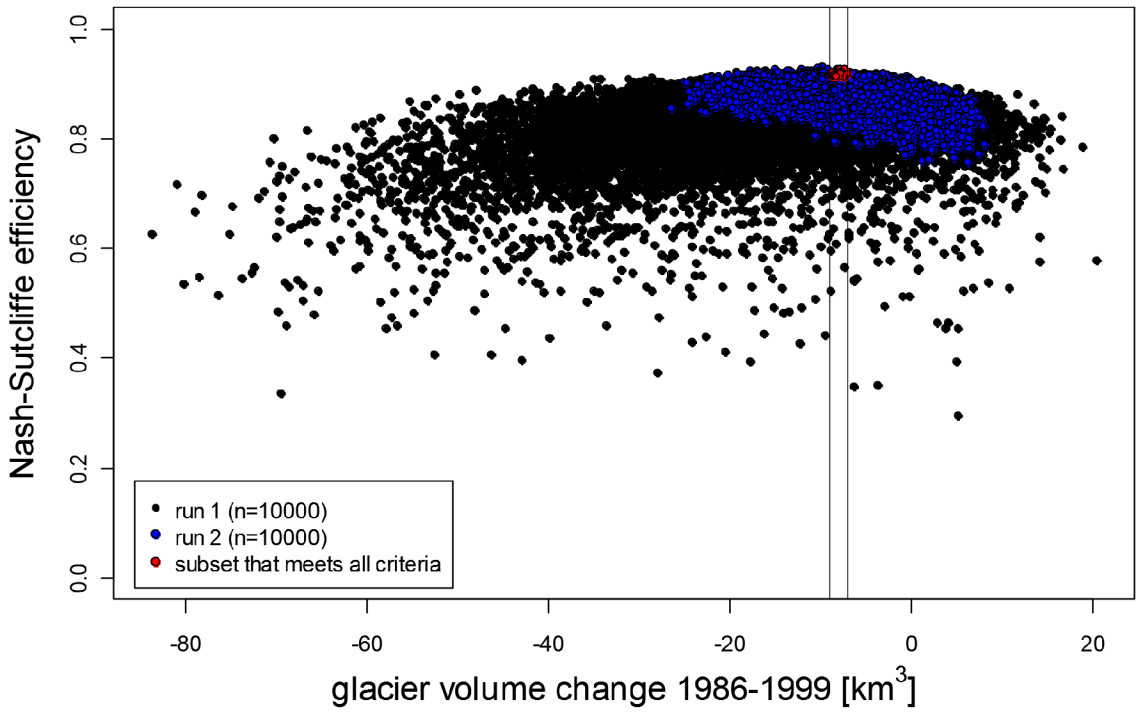

Figure 3.7. Nash-Sutcliffe efficiency (*E)* plotted against simulated glacier volume change for 10,000 model runs in the initial Latin Hypercube Search (black) and for 10,000 model runs in a Latin Hypercube Search with adjusted *prior* parameter distributions (blue). Red dots indicate acceptable parameter combinations.

The temperature lapse rate (*TLAPSE*), the melt factor at winter solstice (*CMIN*), and the precipitation lapse rate (*PLAPSE*) are the calibrated parameters that have the highest correlations with both the glacier net mass balance (b_n) and *E* (Table 3.2). Because they had winter mass balance measurements, Stahl *et al.* (2008) were able to fix the climate parameters *TLAPSE* and *PLAPSE* at an initial step during model calibration, separately from the calibration using streamflow data. This approach was not possible in this study as we have no winter balance measurements, so a greater amount of uncertainty in these parameters (wider parameter ranges) is propagated all the way through to streamflow predictions. Other important parameters are the ratio of melt of glacier ice to melt of seasonal snow (*MRG*) and the increase of melt factor between winter and summer solstice (*DC*). *MRG* and *DC* are both correlated with b_n at all steps during the uncertainty analysis, but are correlated with *E* only in the first LHS with wide parameter ranges. The routing parameters with the highest correlation with *E* are the fast reservoir release coefficient

(*KF*) and the fraction of runoff directed to the slow reservoir (*FRAC*). *ALPHA*, the exponent to adjust the release rate of the fast reservoir, has little influence on *E*.

As mentioned above, the slow reservoir coefficient (*KS*) was constrained to model winter flow and hence had a small parameter range, which is why *KS* shows low correlation to *E* (it would likely have a higher correlation if a wider parameter range were considered). Glacier reservoir coefficients and the melt reduction factor under forest (*MRF*) show weak correlation to both glacier volume change and *E*.

Besides *TLAPSE*, the most important model parameters are *CMIN* and *DC*, both of which comprise the base melt factor C_0 (Eq. 3.1). We will use C_0 as an example to illustrate how the use of both glacier volume loss and model efficiency for streamflow (*E*) help to constrain the calibration. For non-zero values of *DC*, C_0 varies sinusoidally between a minimum value (*CMIN*) at the winter solstice to a maximum value (*CMIN* + *DC*) at the summer solstice. A *DC* of zero would mean that C_0 is constant in time, whereas a *CMIN* of zero would result in C_0 with the highest possible seasonal variation. In the first LHS, we chose the parameter ranges to be wide enough to generate solutions between those maxima.

Figure 3.8, which is based on results from the first LHS, shows that melt factors greater than 5.4 (blue) lead to solutions with efficiencies lower than 0.85 and over-predicted glacier volume losses. Melt factors lower than 2.7 lead to glacier volume losses that are smaller than observed. Plotting simulated and observed hydrographs with different combinations of *CMIN* + *DC* (not shown here) revealed that combinations with low *CMIN* and high *DC*, which generate a high seasonal variation of C_0, produce a much-too-delayed onset of runoff compared to observations. Hence, in the second LHS, the prior parameter ranges of *CMIN* and *DC* can be adjusted to combinations of *CMIN* and *DC* that are between 5.4 and 2.7 with a minimum *DC* of zero. Other parameter ranges were adjusted according to parameter distributions that meet relaxed criteria applied to the first LHS (Figure 3.9).

Figure 3.10 shows that within the adjusted parameter ranges, with random variation of all other parameters (also *MRG*), any melt factor from 2.6 to 4.6 can lead to model efficiencies that are close to the benchmark value of *E*. Measured glacier volume loss is the only option to further constrain the melt parameters and hence to narrow the error bars over future streamflow predictions related to glacier changes.

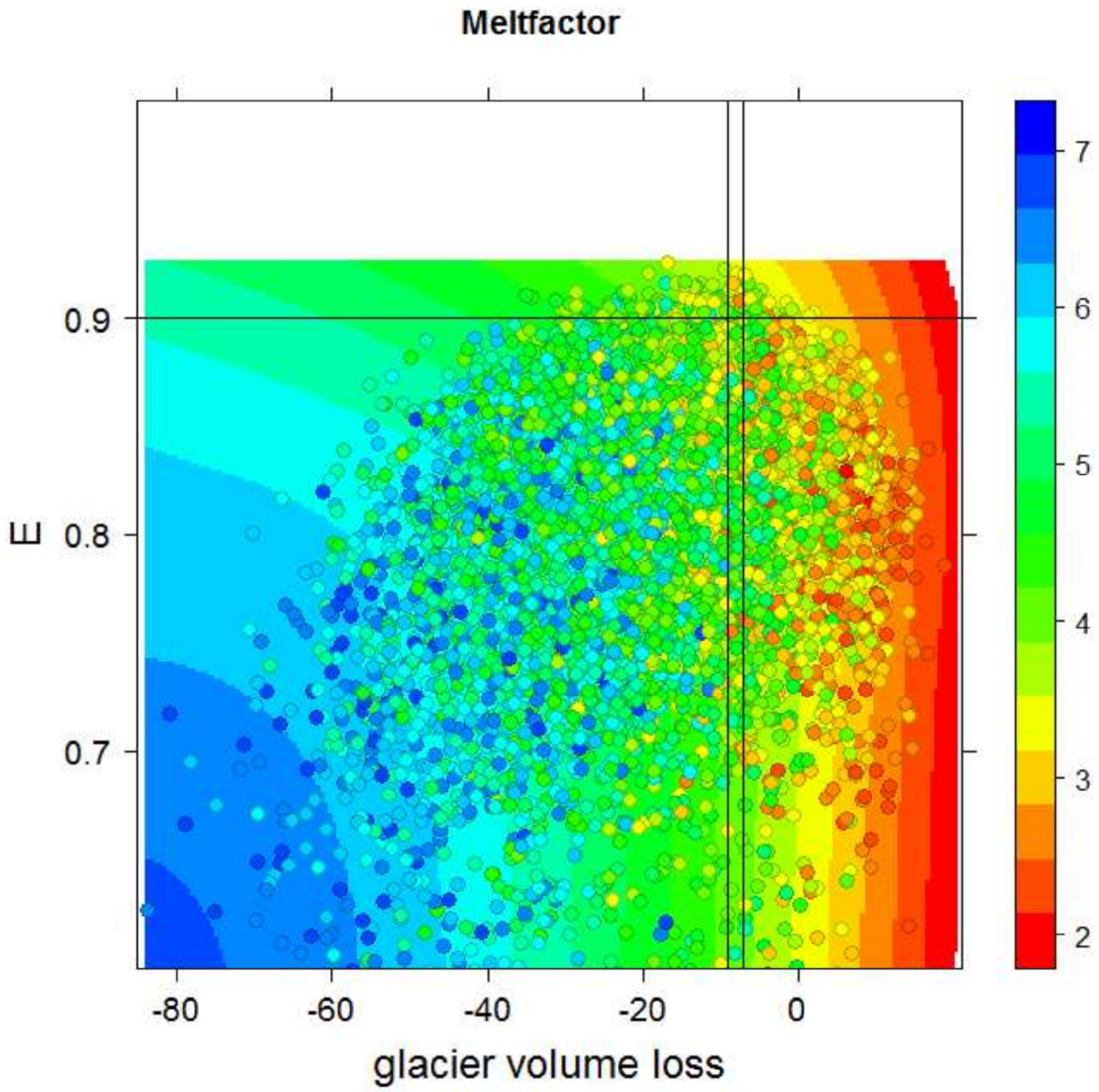

Figure 3.8. Variation of the parameter C_0 interpolated over the resulting values of model efficiency (E) and glacier volume change for the first Latin Hypercube Search. Each plotted point represents a model run. All parameters besides C_0 were varied randomly.

.

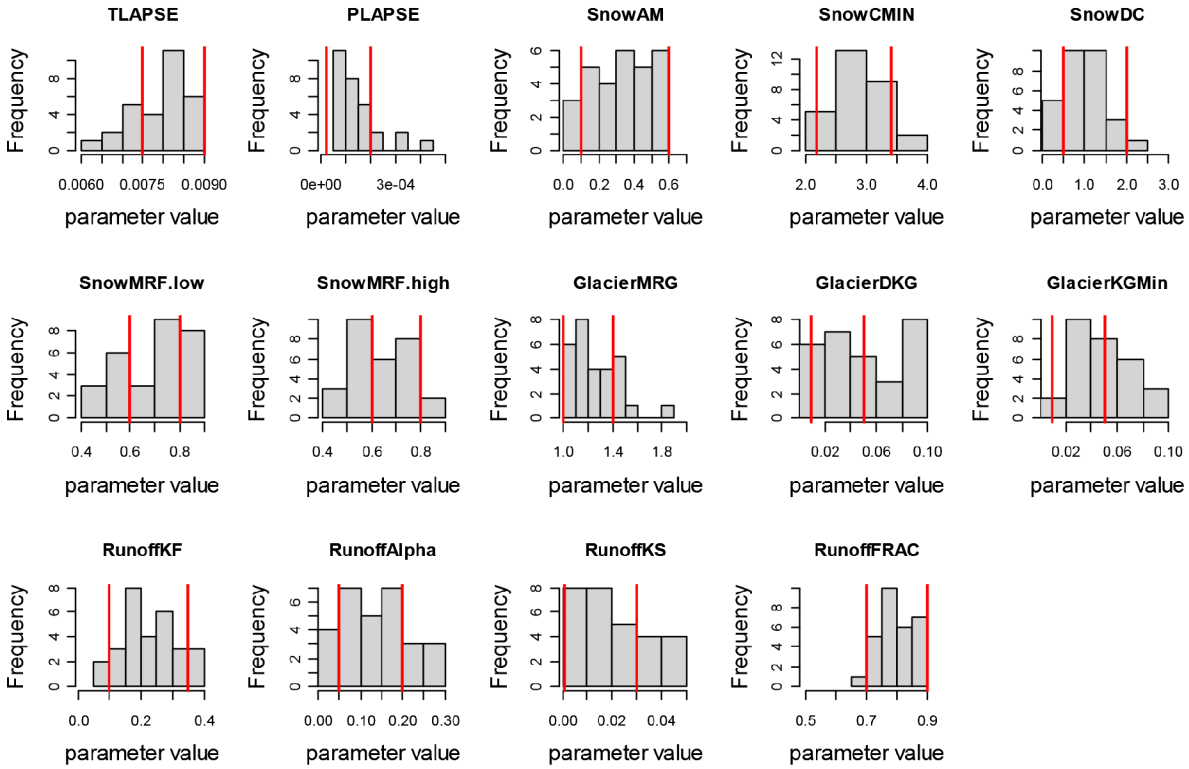

Figure 3.9. Histograms of parameters that meet relaxed criteria applied to the first Latin Hypercube Search. The error for mean August streamflow was relaxed to 10 % (from 5 %) and the constraint for E was relaxed to > 0.87 (from E > 0.91). Red lines indicate adjusted parameter ranges.

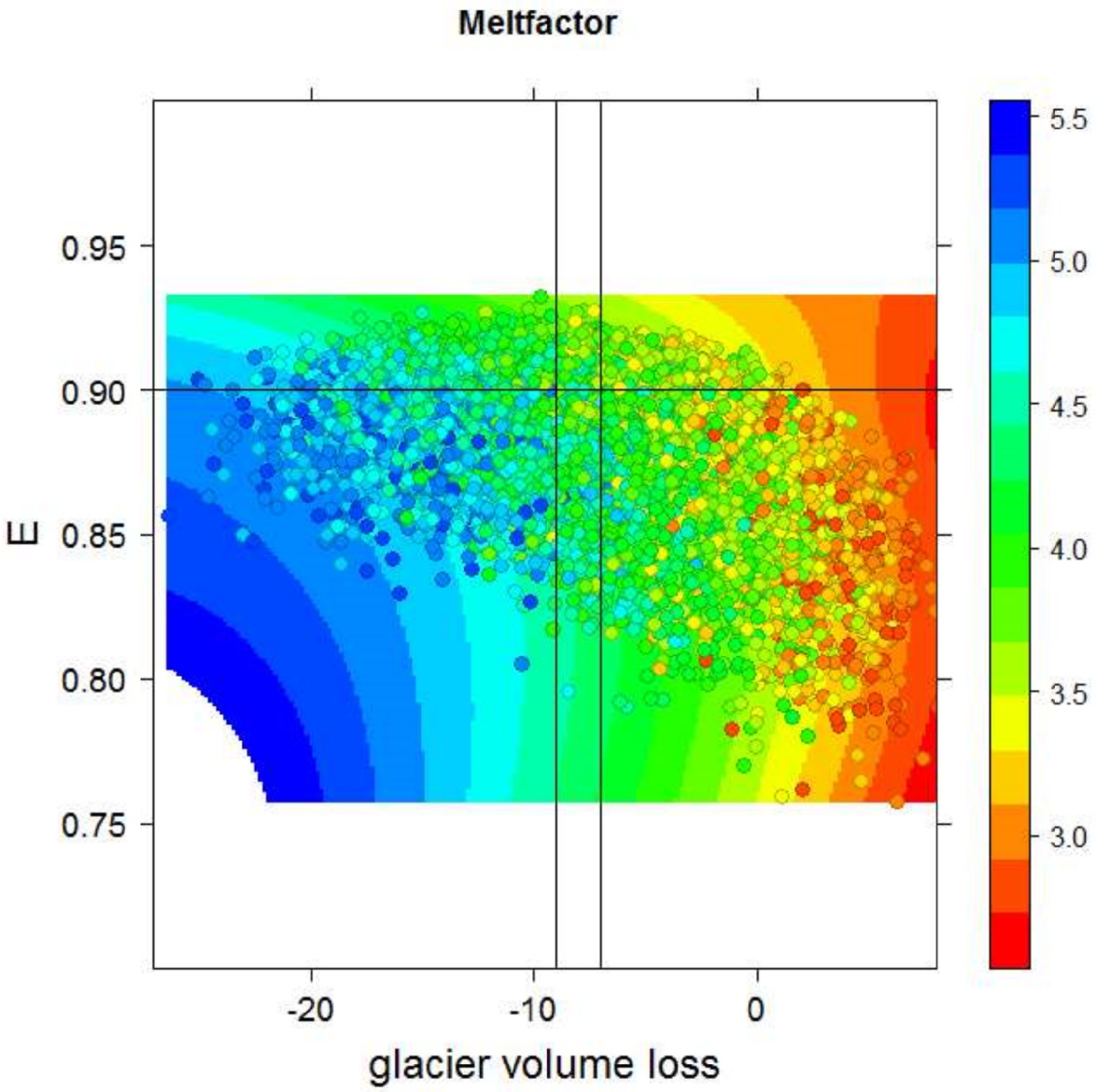

Figure 3.10. Variation of the parameter C_0 interpolated over the resulting values of model efficiency (E) and glacier volume change for the second Latin Hypercube Search. Each plotted point represents a model run. All parameters besides C_0 were varied randomly.

From the 10,000 model runs in the third LHS, 705 parameter sets met the final criterion of $E > 0.915$, but only 23 of these also met the additional criteria given in Figure 3.8. This again underlines the importance of the use of glacier mass balance or volume change data to constrain model parameters. All of the acceptable 23 parameter sets are considered equally representative for Mica basin. All parameter sets

reproduce the seasonal peak flows as well as low flows, but have difficulty with modelling intense rainfall events, especially during autumn (Figure 3.11). This is not surprising, since, as mentioned before, one of the two reservoirs (the slow reservoir) is used to model the low flows during winter, and the single fast reservoir cannot simultaneously represent runoff generation due to melt and rainfall given the differences in their spatial patterns and nonlinearity. Since this model weakness only affects daily peak flows, particularly during autumn, it should not undermine the model's ability to make projections of the effects of climate change and glacier response on monthly or seasonal patterns of melt-generated runoff.

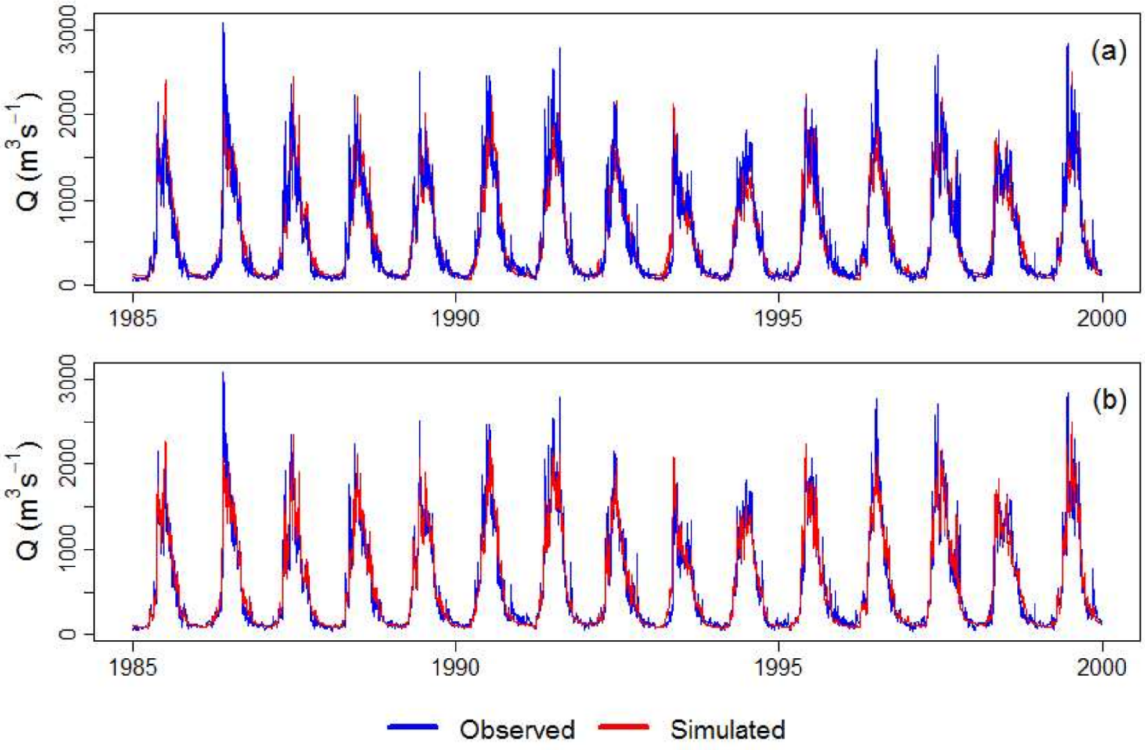

Figure 3.11. Observed and simulated discharge for the calibration period (1985-1999). (a) observed and simulated discharge predicted with the best model; (b) observed and the ensemble of simulated discharge.

3.4.3 Discussion

There is disagreement in the hydrologic literature as to whether the treatment of uncertainty in hydrologic model predictions should be founded in formal Bayesian statistics or on informal Bayesian approaches

such as the GLUE method (there is also debate as to whether the GLUE approach should be seen as an informal Bayesian method, as a special case of a formal Bayesian method, or as a non-Bayesian method). The main advantage of a formal Bayesian method is that individual error sources, e.g., model structural errors and parameter uncertainty, can be separated. Since our main interest was the assessment of the overall uncertainty in predicting the hydrological impacts of glacier change, we chose a GLUE type approach that we modified to fit the specific needs of this study. In the traditional GLUE approach, parameter ranges should be widened so that the envelope of ensemble parameter sets cover all or part of the observed discharge (symmetrically). However, in a model that does not incorporate strong nonlinearity, such as HBV-EC, this can lead to parameter combinations that do not represent accurately the dominant runoff generation processes. For example, for Mica basin, HBV-EC simply does not appear to be able to model a heavy rain event on top of a seasonal melt peak with high accuracy. Criteria can be relaxed and parameter ranges widened so that some parameter sets could cover such events, but this would heavily distort conclusions regarding the relative roles of snow and glacier melt drawn from such an ensemble. One also has to bear in mind that despite the fact that parameter sets in our ensemble converge at the final criterion, E, they still cover the whole range of all other criteria including uncertainties associated with the observed glacier volume loss. Predictions made outside the observed range of climate forcings are more likely to diverge.

3.5 Model testing

Model testing on streamflow for the period 2000-2007 using the observed glacier extents from 2005 yielded an efficiency of 0.95 for the best model, a slightly better performance than during the calibration period. One possible reason for the improved model performance is that observed rather than estimated data were available for the Floe Lake climate station during the test period. As for the calibration period, HBV-EC performed well on the seasonal melt peak as well as on low flows during winter, but has some weakness capturing intense rainfall events.

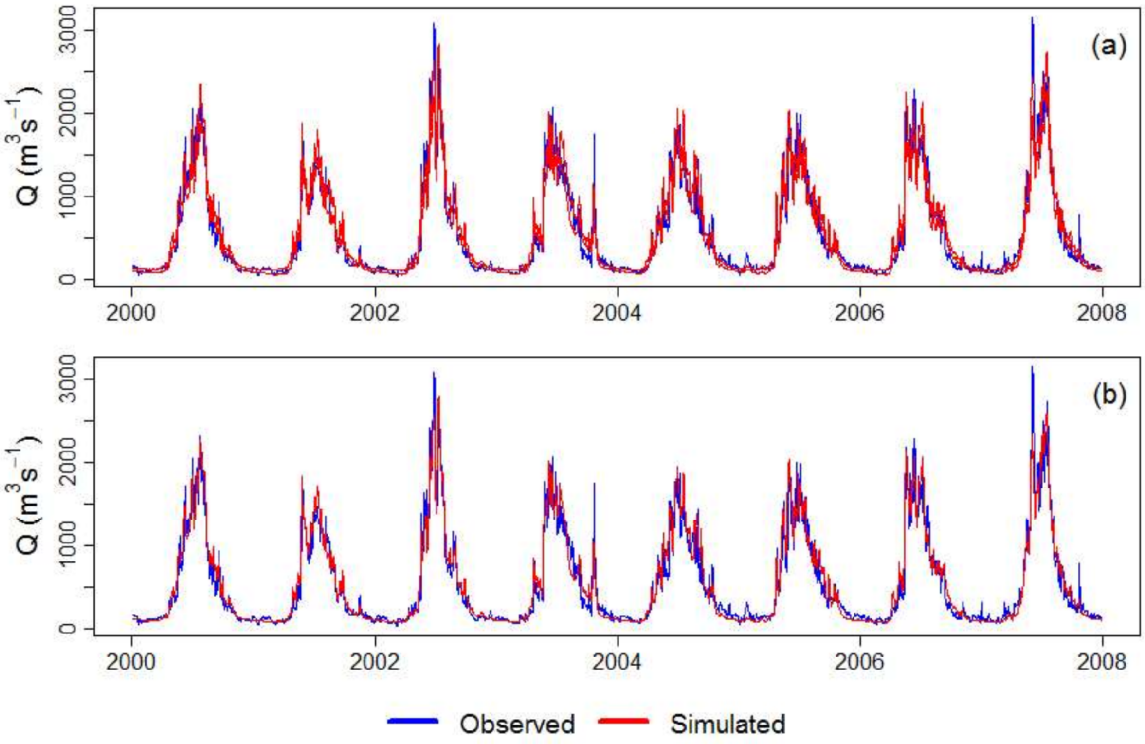

Figure 3.12. . Observed and simulated discharge for the test period (2000-2007). (a) observed and simulated discharge predicted with the best model; (b) observed and the ensemble of simulated discharge.

Snow water equivalent data for three snow pillows were available from 1995 onwards. Since these data were not used for model calibration, we used them to test the ability of the calibrated HBV-EC to simulate snow processes for the period from 1995-2007. Since HBV-EC is a semi-distributed model based on the concept of GRUs, the model does not predict state variables for a specific location but only for each GRU. This needs to be considered when comparing simulated and observed snow water equivalent (*SWE*). Despite the difference in spatial scales associated with the modelled and observed values, *SWE* predicted by HBV-EC shows encouraging agreement with observations, with regression fits between predicted and observed *SWE* having R^2 of 0.82, 0.77, and 0.86 for the Molson Creek, Floe Lake, and Mount Revelstoke snowpillows, respectively (Figure 3.13). The model accurately predicted the timing of the onset of snowmelt as well as the rate of decrease of *SWE* during the ablation period at all three snow pillow sites. The model tends to underpredict the peak *SWE*. For some years this underprediction is within the error bars of *SWE* measurements (±5% according to Gray and Male, 1981). This underestimation may be caused, in part, by the fact that snow pillows tend to overestimate *SWE* due to creep (downslope deformation of a snowpack), which puts additional load on the pillows (Gray and Male 1981). However, there are some station-years in which the underestimation is too large to be simply

attributed to measurement errors (e.g., 1996-1997 at Floe Lake). Interestingly, the timing of these types of errors is not consistent among stations. For example, in the water year 1996-1997, peak SWE was reproduced reasonably accurately at Molson Creek and Mount Revelstoke. This pattern of errors likely reflects, in part, the inherent variability in precipitation patterns from year to year, which are not properly represented through the use of fixed lapse rates in each climate zone.

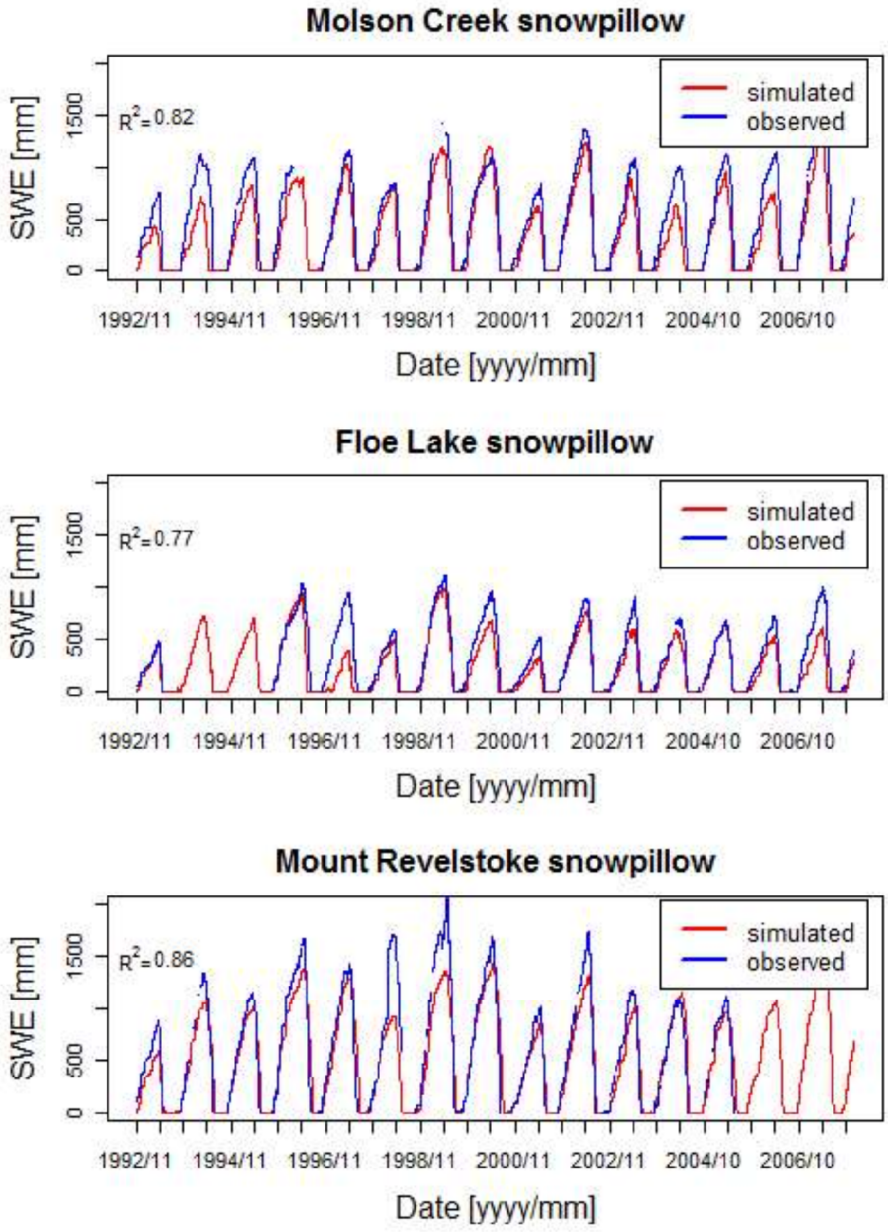

Figure 3.13. Simulated (using the best model) and observed snow water equivalents for three snow pillow sites. Simulations are for the GRU that corresponded to the snow pillow sites.

3.6 Historic contributions of glacier melt inferred from HBV-EC

3.6.1 Historic trends in climate data

Four of the five climate stations in Mica (MCA, RGR, MOL and RAD) have records that are long enough to explore historic temperature and precipitation trends. RGR, RAD and MOL have records starting in 1967, 1969 and 1984, respectively, and both temperature and precipitation trends are calculated based on observations only. All stations have been backfilled to 1960 to model historic trends in streamflow.

Mean annual air temperature shows increasing trends for all four climate stations (Figure 3.14). Mean annual air temperature from the early 1960s up until the 1976-77 shift in Pacific climate (e.g., Mantua et al., 1997) is more or less steady. Following 1976, mean annual temperature increases at all climate stations. The highest elevation climate station, MOL (1935 m asl.), shows the highest increase in mean annual temperature amongst all climate stations, with a slope of 0.051 °C/year. During the post-1976 period, mean annual air temperature steadily increased at all climate stations (0.037 °C/year at MCA, 0.033 °C/year at RAD, 0.029 °C/year at RGR).

Less clear trends are found in the precipitation data (not shown), with a few extreme years dominating the overall trend at each station. While precipitation at the higher elevation climate stations MOL and RAD is slightly increasing, a weak decreasing trend can be observed at the lower elevation climate station MCA.

3.6.2 Historic variations in streamflow and inferred contributions from glacier melt

Mean annual discharge shows high inter-annual variation but no trend; the slope of a fitted linear model was not significantly different from zero based on a two-tailed t test (Figure 3.15). Mean July discharge also does not show a significant trend with time. However, both August and September mean discharges significantly decrease with time ($p < 0.05$). The trend, estimated by linear regression, is -7.1 $m^3s^{-1}year^{-1}$ for mean August discharge and -4.3 $m^3s^{-1}year^{-1}$ for mean September discharge.

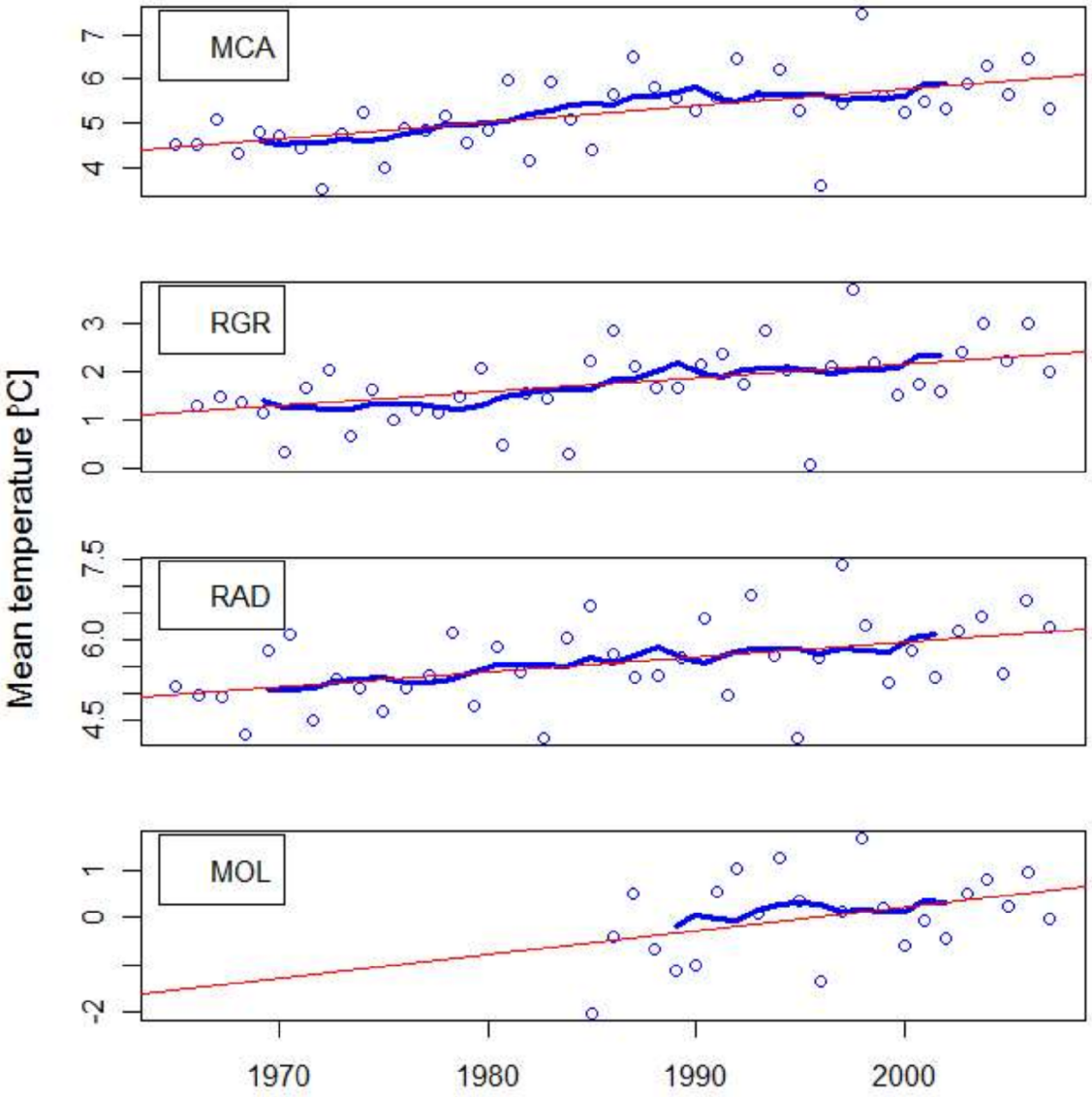

Figure 3.14. Mean annual temperatures for MCA, RGR, RAD, and MOL and their linear trend (red line) for the period 1965-2007. The bold blue line is a 10-year moving average.

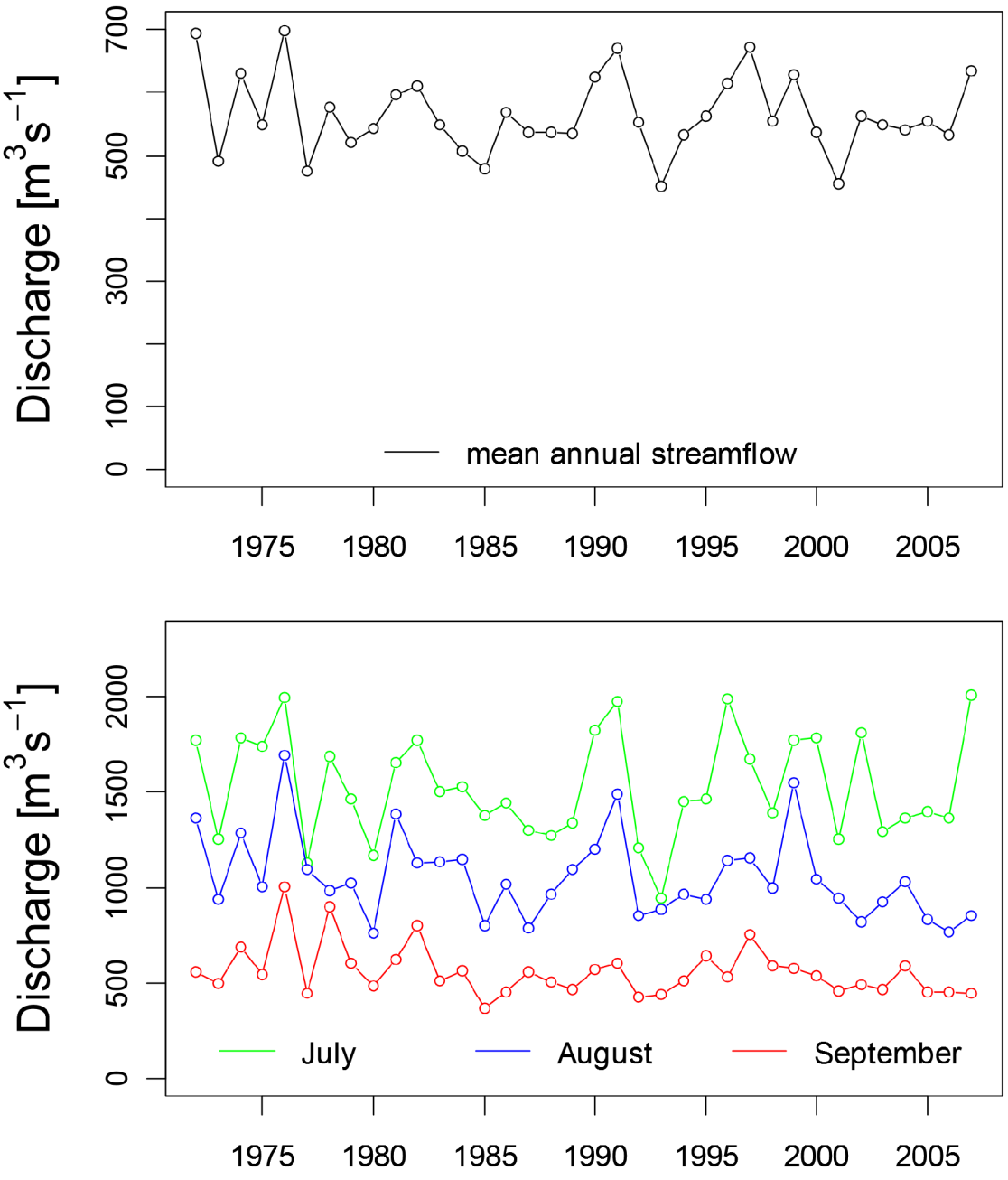

Figure 3.15. Time series of mean annual discharge (top) and mean discharge for July, August and September (bottom) for the period 1972-2007.

Historic contributions of glacier melt can be modelled based on either observed or simulated glacier areas. Simulations of basin wide glacier area from the UBC Regional Glaciation Model (UBC-RGM) show

good agreement with observations (Figure 3.16). After a period with increasing glacier area prior to the Pacific climate shift in 1976-77, which was associated with a shift to more negative net mass balance in western Canada (e.g., Moore and Demuth, 2001), glacier areas steadily decreased up to the end of the data records in 2007 (Figure 3.16). This decrease in glacier area is confirmed by observations. Hydrographs from simulations based on simulated and observed glacier masks are visually indistinguishable.

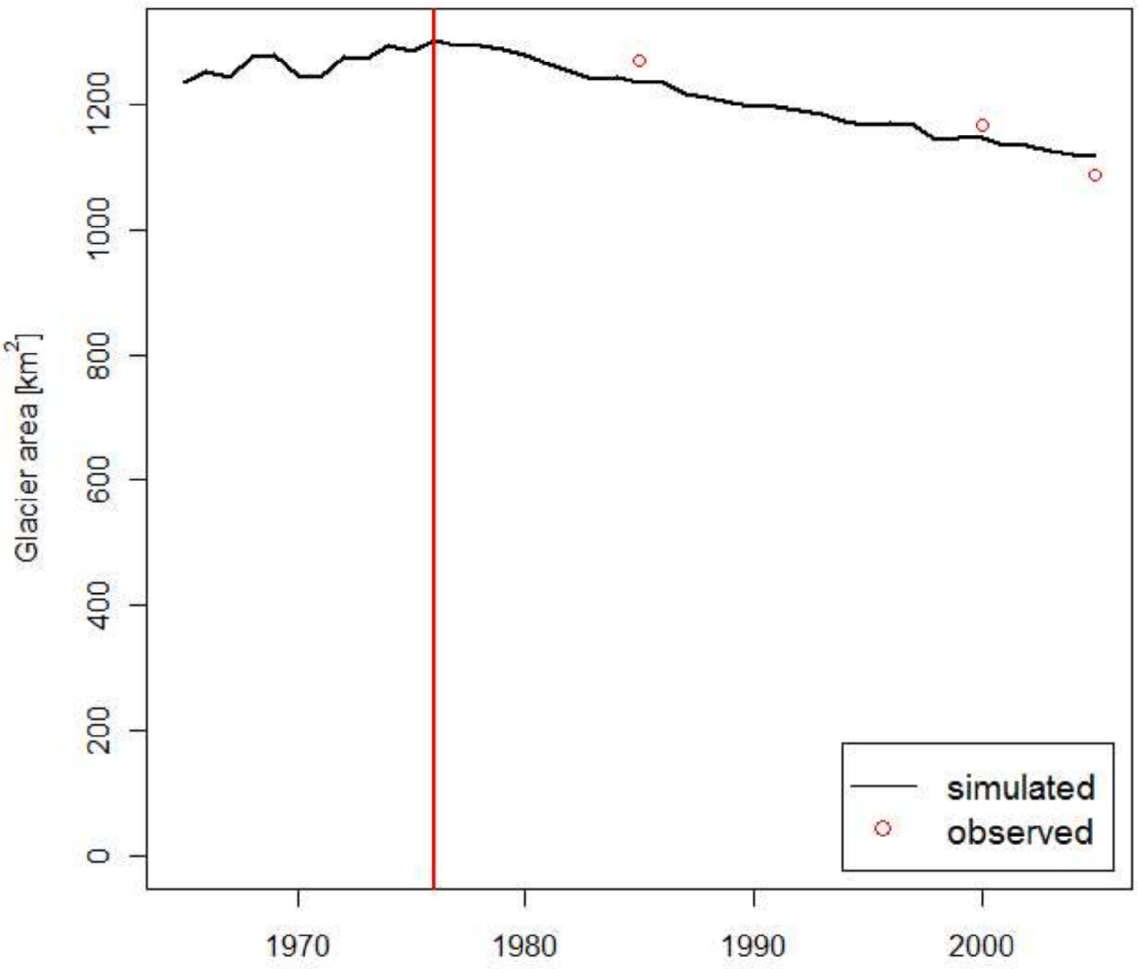

Figure 3.16. Historic simulated and observed glacier area. The vertical red line indicates the 1976-77 Pacific climate shift.

The difference between streamflow simulations with and without glaciers provides an estimate of the contribution of glacier ice melt to discharge, taking into account the fact that snowmelt and rainfall runoff from the areas currently covered by glaciers would occur even if the glaciers completely disappeared. In the no-glacier runs, all glacier cover is converted to open land cover. The differences are calculated from the ensemble means for glacier and no-glacier scenarios (23 runs, each based on a different "behavioural" parameter set). Figure 3.17 shows that the annual contribution of ice melt to total streamflow varies between 3% and 9%, and is on average 6%.

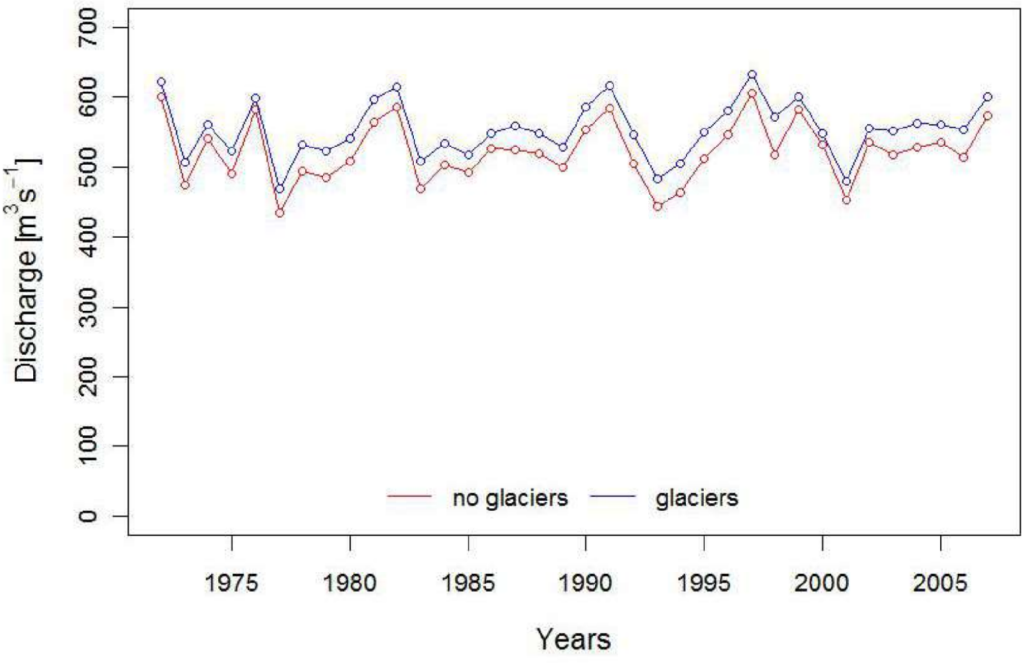

Figure 3.17. Historical mean annual streamflow simulated with and without glaciers (simulations based on observed glacier areas).

Mean monthly discharges in July are more or less identical between glacier and no-glacier scenarios. In July, some years have a higher discharge in the no-glacier scenario. This occurs because the routing in the presence of glaciers differs from routing with no glaciers: the glacier reservoir can lag flows from a few days up to several weeks, depending on the parameter values. This behaviour has been documented empirically (e.g., Stenborg, 1970). Mean August streamflow would be up to 25% lower if there were no glaciers, though the variation of contributions is high (coefficient of variation = 7%). The relative

contribution of glaciers is highest in September, when ice melt can provide up to 35% of the discharge. September contributions of ice melt are also less variable over time, with a coefficient of variation of 5%.

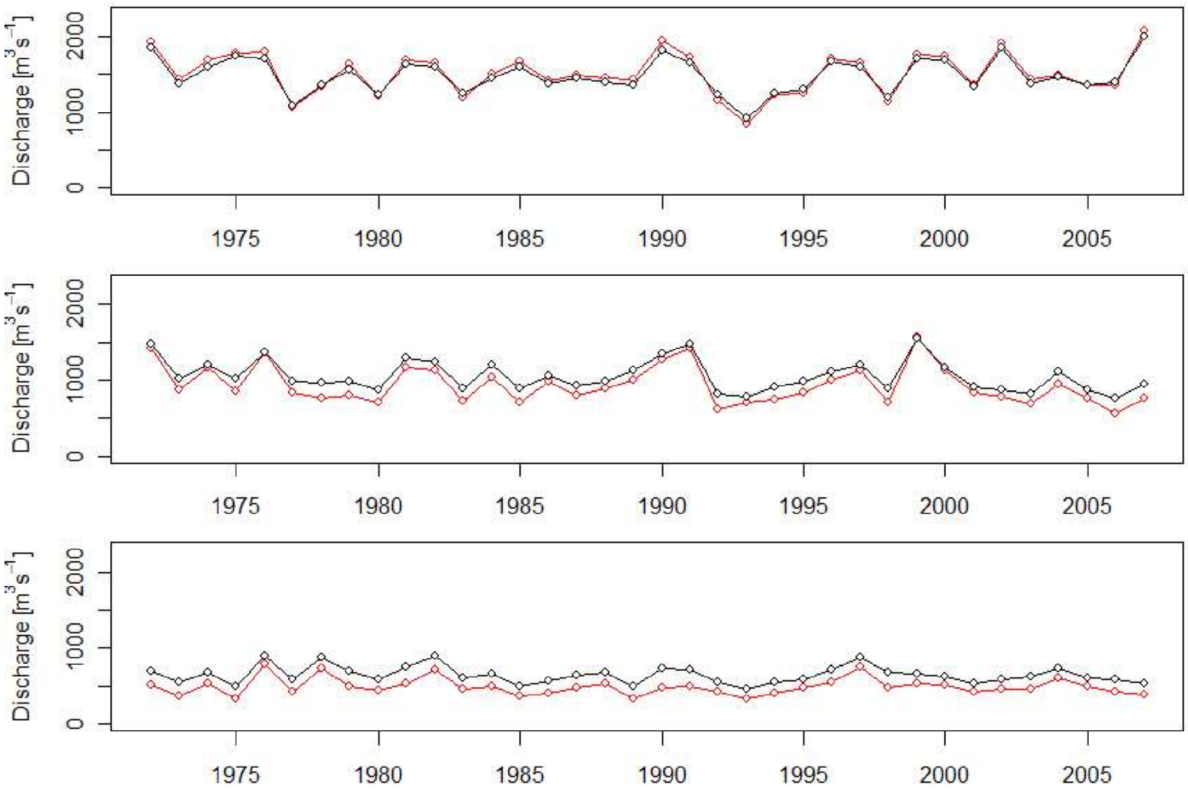

Figure 3.17. From top to bottom: mean July, August and September streamflow simulated with and without glaciers.

Figure 3.18 compares the mean and range of ensemble predictions for two hydroclimatically contrasting years for simulations with glaciers to simulations where glaciers have been removed and replaced by open land cover. Glaciers and glacier ice melt are particularly important in years with early snowmelt such as 1998, the year with the highest modelled ice melt (Figure 3.18), where glaciers can contribute to more than 20% of the flow for periods of more than two months. In years with late snow melt, for example the year 2000, glaciers have a minor effect on discharge.

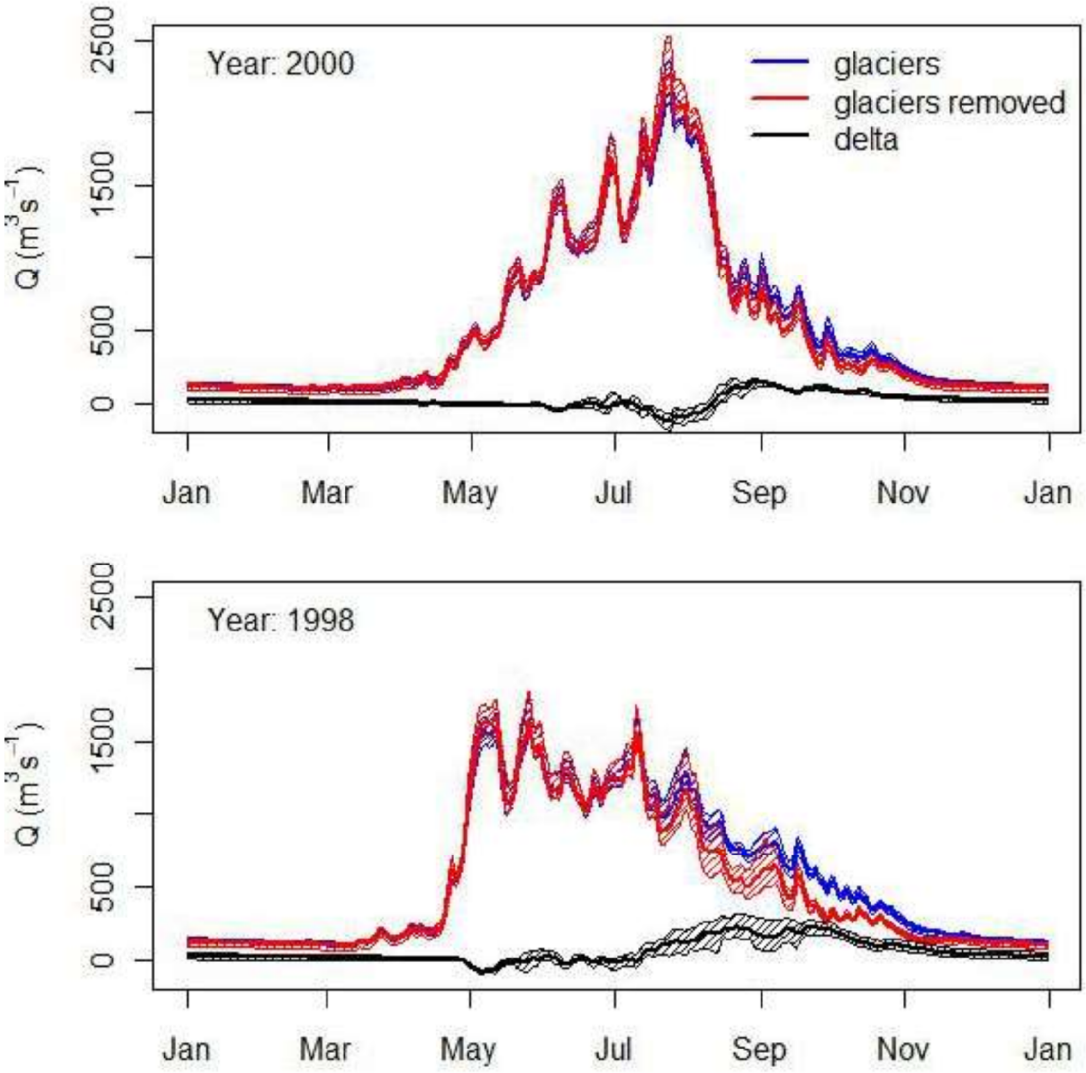

Figure 3.18. The effect of glaciers on discharge shown by comparing simulations with and without glaciers for the year with the highest modelled ice melt (1998) and the year with the lowest ice melt (2000).

3.7 Summary

The calibrated HBV-EC model accurately predicts the overall seasonal hydrograph, including the seasonal peak and the low flows. Nash-Sutcliffe efficiencies exceeding 0.93 were achieved for the independent test period. The calibrated model was able to reproduce the onset of snowmelt as recorded at snow pillows, as well as the rate of snowmelt. Constraining the model calibration using observed changes

in glacier volume increases confidence in the model's predictions of the relative roles of snow and glacier melt. Running the model with and without glaciers suggests that glacier ice melt contributes from 3 to 9% of annual runoff, depending on the year. The contribution in August is up to 25% of discharge in that month, while the contribution in September is up to 35% of discharge. In all cases, the contribution of glacier melt is highest in years with low snow accumulation and early snow disappearance.

4 Modelling glacier dynamics

4.1 Overview

There is a substantial literature devoted to numerical ice dynamics modelling applied to large ice sheets such as the present-day Antarctica and Greenland ice sheets or the now-vanished Laurentide and Fennoscandian ice sheets. At a much smaller scale, the flow of individual "research" glaciers has been successfully modelled. In both cases, the problem geometry, in particular the topography of the subglacial bed $B(x,y)$, and the glacier mass balance forcing are usually well known, well constrained or spatially smooth. In contrast, the problem of simultaneously modelling the flow of many glaciers within a region of complex topography, which we call "regional glaciation modelling,"is comparatively new and can introduce special challenges. There are two particularly difficult challenges. First, the subglacial topography is unknown, although it is subject to physical and plausibility constraints. Second, the climate forcing is certain to have a high degree of spatial variability owing to elevation and orographic influences on temperature and precipitation, while climate fields obtained from observation networks, climate reanalysis products and GCM output are all coarsely resolved. Thus downscaling from these fields to the scale of individual glaciers presents a significant challenge. Much of the art and science of regional glaciation modelling focuses on developing the best possible glacier mass balance model from these downscaled climate fields. Figure 4.1 summarizes the main components of a regional glaciation model.

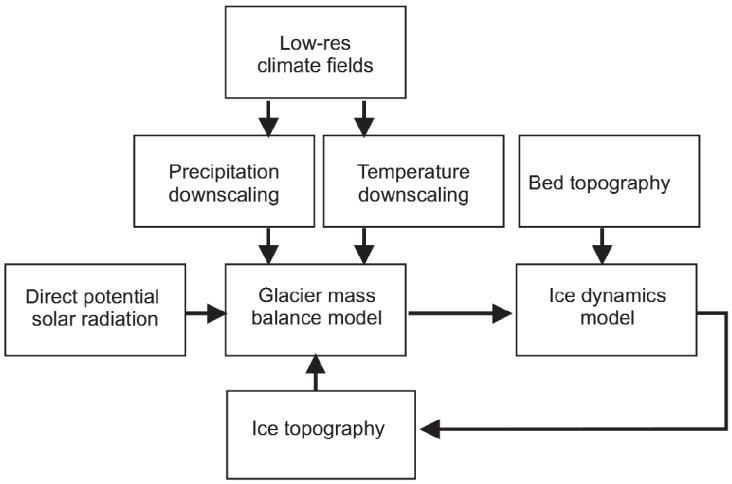

Figure 4.1. Flow chart showing major components of the UBC regional glaciation model.

4.2 Landscape representation

All aspects of the modelling presented in this section require a digital elevation model (DEM). The derivation of the subglacial topography additionally requires maps of glacier coverage that temporally correspond to the time of acquisition of the DEM. The baseline DEM that was used for the generation of the downscaling and surface mass balance is the Shuttle Radar Tomography Mission version 4 with 90 m spatial resolution described by Farr *et al.* 2007 (hereafter SRTM; accessed via http://srtm.csi.cgiar.org/) . For the ice dynamics modelling, this DEM was supplemented by data from the British Columbia Terrain Resource Information Management (TRIM) program acquired from BC Hydro (Scott Weston, BC Hydro, personal communication, 2010). These two DEMs are nested for the purpose of ice dynamics modelling, with the TRIM data used within the Mica basin and SRTM used without. Both DEMs were resampled from their native resolution to 200 m. There is a slight offset apparent between the SRTM and the TRIM DEMs at the watershed boundary where the data meet. Glacier coverages, or "ice masks" I_{ij} were derived from Landsat ETM data by Bolsch *et al.* (2010) and are based on scenes captured in year 2000. We take $I_{i,j}=1$ for ice-covered cells and $I_{i,j}=0$ for ice-free cells. From these two pieces of information, the area of every glacier in the study region is computed and ice thickness $H_{i,j}$ is estimated for the time t_0 at which ice surface topography S and I were observed. The algorithm (Clarke, unpublished) used to estimate $H_{i,j}$ assumes that individual glaciers have a basal shear stress τ that depends on glacier area A. Applying the scaling analysis of Bahr (1997) we use the relation $\tau=cA^{0.25}$ where the constant of proportionality c is assigned to yield ice thicknesses that either match observations or, lacking these, align with experience elsewhere. This approach is judged to be superior to an artificial neural network method previously described by Clarke *et al.* (2009), but optimal estimation of ice thickness from surface data remains a challenging and important research question that merits additional attention. All data have been projected into the so-called NARR projection, which is a Lambert Conformal Conic projection whose parameters are given as supplementary information. All modelling takes place within this projection and final results are reprojected to the BC Albers projection used by BC Hydro. This approach preserves the accuracy of the forcing data by avoiding reprojection of these data and of the associated interpolation of these fields.

4.3 Climate representation

4.3.1 North American Regional Reanalysis

A major part of our strategy for glacier modelling was to develop a "most accurate" set of both driving meteorological variables and glacier mass balance for the Mica catchment. For this purpose, we have chosen the North American Regional Reanalysis (NARR), a high-resolution atmospheric reanalysis data

set which is fully described in Mesinger *et al*. (2006). This reanalysis comprises a full suite of atmospheric data (>200 variables) on a ~32 km grid at 29 pressure levels and at 3-hourly intervals for the period 1979 through the present. These data are derived from short term Global Climate Model simulations with observational data assimilated 8 times per model day. As such, the NARR represents a self-consistent set of atmospheric data. We use several of the NARR variables to apply our process-based downscaling, which is described in Section 4.4.

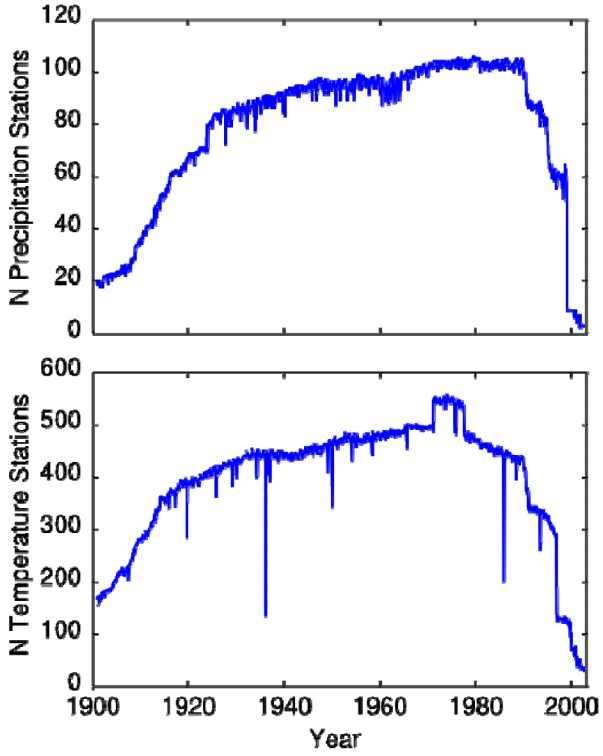

Figure 4.2. Time series showing the number of stations contributing to the central CRUTS2.1 grid cell that was used to simulate glacier growth during the 20th century. (a) Precipitation. (b) Temperature. For both records, there are substantially more stations contributing to the grid cell in the middle and late part of the 20th century than earlier and later.

4.3.2 CRUTS2.1

Although simulation of the past fluctuations of glaciers was not an explicit objective of our project, this step is essential to the glacier dynamics modelling that will be described below. To meet this need for 20th century glacier forcings, two approaches were taken. The first was to utilize the 20th century simulations available for the GCMs that we used in this study. This approach was followed with the hope that the time series of forcings from the individual GCMs would yield a consistent picture of the glaciological past and future, rather than the patchwork that can result from using multiple data sources.

The second approach was to use a gridded data set containing interpolated temperature and precipitation observations throughout the 20th century. The TS2.1 data from the Climate Research Unit of the University of East Anglia (hereafter CRUTS2.1; New *et al.* 2000, Mitchell and Jones, 2005) was chosen for the 20th century simulations. The 20th century data from the GCMs differed from one another and from CRUTS2.1 and it was considered unlikely that output from any of the GCMs could be successfully used to drive a glacier mass balance model for the 20th century. Thus we decided against using GCM-based mass balance models to initialize our simulations of glacier changes in the 21st century.

The CRUTS2.1 data are derived from station measurements of several variables including 2 m air temperature, precipitation, cloud cover, and humidity. These observational data are aggregated to monthly anomalies and then interpolated to a 0.5° resolution geographic grid using an inverse distance weighting (IDW) scheme. Several quality concerns arise in the production of these data. The IDW interpolation allows distant stations to contribute information on climate fields (as far away as 1200 km for temperature and 450 km for precipitation). For periods later in the century when there are numerous climate stations, this presents little problem due to the high station density relative to the grid spacing. For early in the 20th century, however, temperature records from the Canadian prairies could have contributed to the temperature reconstruction over the BC interior and the Canadian Rocky Mountains. Mitchell and Jones (2005) do not quantify the error associated with these changes, so it is impossible to assess the uncertainty that this contributes to the modelling. However, it can be assumed that there is a general decrease in model uncertainty from the onset of simulations to present. For the purposes of this study, with its focus on future glacier changes, errors incurred in spin-up early in the 20th century should have minimal effect on 21st century simulations. Furthermore, in this study we only use the temperature and precipitation variables, which have the greatest number of contributing stations.

4.3.3 GCMs

To represent the 21st century climate for Mica region we use six GCMs from the ensemble of 22 models originally produced for the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4). These six GCMs, listed in Table 4.1, are chosen according to their ability to simulate various aspects of recent climate (1979-1999) over this region. Details of the evaluation methods and criteria for model selection are presented in Chapter 5.

We use the projected temperature (air temperature at 850 hPa) and precipitation fields from the six selected GCMs for emission scenarios A2, A1B and B1 (see Chapter 5, Box 5.2 for a short description of these scenarios). Our general approach is to produce a high resolution reference downscaling at daily

temporal resolution, and then use statistics calculated from these data to generate the monthly input fields for our mass balance model. For the GCM scenarios for future climate and the CRUTS2.1-forced mass balance simulations, we apply the monthly anomalies in temperature and precipitation to a monthly climatology derived from our high resolution downscale – the so-called *delta approach*. The delta approach uses the anomaly fields (anomalies in the forcing variables) from the parent GCM (or in CRU) to adjust the high spatial resolution mean-of-month temperature and precipitation fields. This downscaling is fully described in section 4.4. The formulation for the delta approach is given by the equations in Box 4.1. For temperature, the anomaly is additive to the NARR field while for precipitation the anomaly is multiplicative. For larger modelling domains, the delta approach retains the spatial structure of anomalies in the parent GCM, but for the scale of the Mica catchment, the anomalies are largely driven by, at most, several GCM grid nodes. Thus the bulk of the spatial structure in the delta-derived temperature and precipitation fields is due to the spatial structure of the mean state of the high resolution, NARR-derived downscale. This is not seen to be a major problem because glaciers typically respond in unison (although with varying magnitudes) over the O(100 km) spatial scales within the Mica catchment (Huybers and Roe, 2009). Thus, it is most important to accurately characterize the local vertical gradients in temperature and precipitation rather than the meso-scale lateral anomalies. The projected climate fields from GCMs and the historical data obtained from the Climate Research Unit (CRU) of the University of East Anglia are downscaled by applying the GCM anomaly for a given month to the high resolution climatology developed from the NARR downscaling. The equations used in this transfer step of our downscaling method are listed in Box 4.1. We consider the NARR-based downscaling as the reference mean climate for the period 1979–1999. All GCM data are initially interpolated to 10×NARR grid, which is approximately 320×320 km in conical conformal Lambert map projection.

Figure 4.3 shows the time series of temperature and precipitation change (2001–2100), projected from the six GCMs and three emission scenarios, for the grid cell covering the Mica region. As illustrated, the projected temperature increase for this region from various emission scenarios is in the range of 2 to 5 K depending on the choice of GCM and emission scenario. For the annual sums of precipitation, the projections range from precipitation with no significant increasing trend (e.g. GFDL-CM2.0) to precipitation increasing by up to 20% (e.g. CGCM3.1 models) by the end of the 21st century.

Table 4.1. List of GCMs used in glacier modelling. Resolutions are given in degrees longitude by degrees latitude.

Model	Governing body	Country	Resolution
CGCM3.1(T47)	Canadian Centre for Climate Modeling and Analysis	Canada	2.8 x 2.8
CGCM3.1(T63)			1.85 x 1.85
CSIRO-Mk3.0	CSIRO Atmospheric Research	Australia	1.85 x 1.85
GFDL-CM2.0	U.S. Department of Commence/NOAA/ Gophysical Fluid Dynamics Laboratory	USA	2.5 x 2.0
MIRO3.2(hires)	Center for Climate System Research (The University of Tokyo), National Institute for Environmental Studies, and Frontier Research Canter for Global Change (JAMSTEC)	Japan	1.12 x 1.12
ECHAM/MPI-OM	Max Planck Institute for Meteorology	Germany	1.85 x 1.85

Box 4.1. Method used to obtain downscaled climate fields from low-resolution CRU and GCM fields.

$$T^*(x,y,t) - T_{\mathrm{R}}(x,y,\bar{t}_k) + \left[T_{\mathrm{M}}(x,y,t) - T_{\mathrm{M}}(x,y,\bar{t}_k)\right] \tag{4.1a}$$

$$P^*(x,y,t) = P_{\mathrm{R}}(x,y,\bar{t}_k) \times \left\{ \frac{P_{\mathrm{M}}(x,y,t)}{P_{\mathrm{M}}(x,y,\bar{t}_k)} \right\} \tag{4.1b}$$

$T^*(x,y,t)$ = projected downscaled temperature field at (x,y)
$T_{\mathrm{R}}(x,y,\bar{t}_k)$ = high-resolution downscaled monthly mean temperature at (x,y) from NARR (1979 1999)
$T_{\mathrm{M}}(x,y,\bar{t}_k)$ = monthly mean temperature at (x,y) from GCM or CRU (1979 1999)
$P^*(x,y,t)$ = projected downscaled precipitation field at (x,y)
$P_{\mathrm{R}}(x,y,\bar{t}_k)$ = high-resolution downscaled monthly mean precipitation at (x,y) from NARR (1979 1999)
$P_{\mathrm{M}}(x,y,\bar{t}_k)$ = monthly mean precipitation at (x,y) from GCM or CRU (1979 1999)

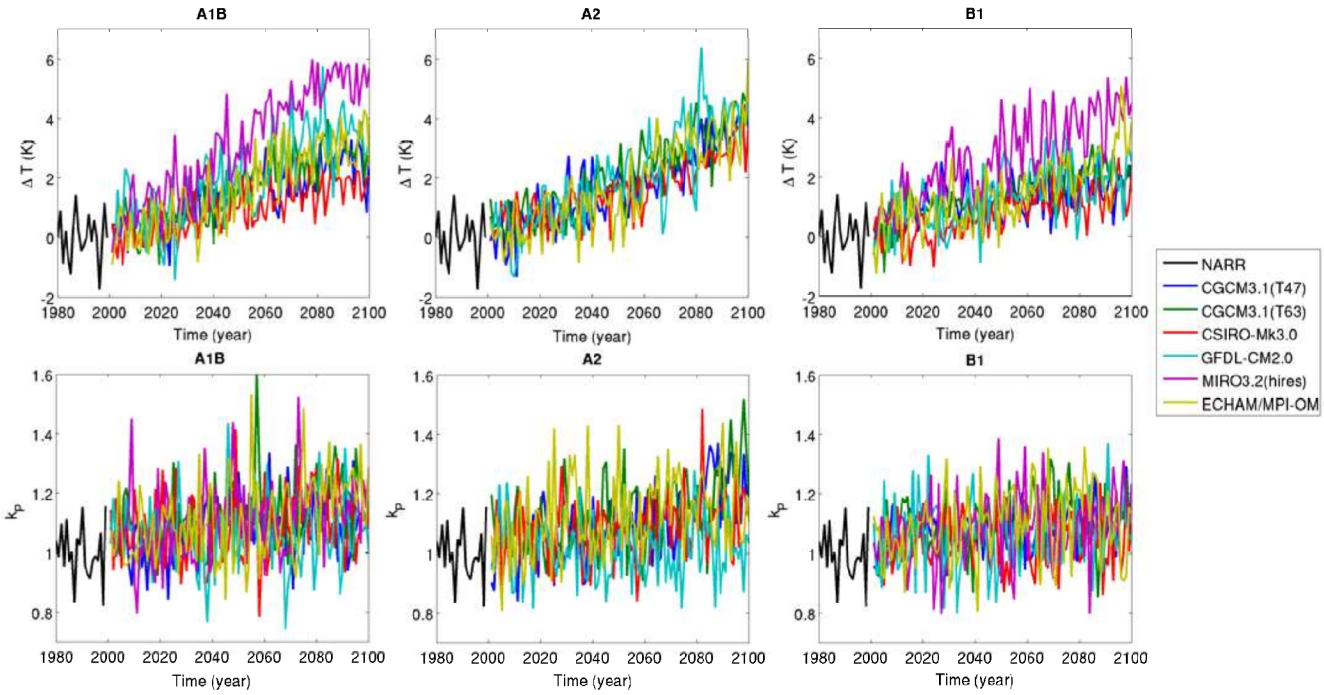

Figure 4.3. Time series of annual temperature and precipitation change for Mica region (Columbia Basin, BC) from NARR and six GCMs for the scenarios A1B (left), A2 (middle) and B1(right). Temperature change ΔT corresponds to annually averaged ΔT_M in Box 4.1, while precipitation change (k_P) corresponds to annually averaged ratio $P_M(x,y,t)/P_M(x,t,t_k)$ in Box 4.1.

4.4 Temperature and precipitation downscaling of NARR

Our modelling efforts require high spatial resolution input fields of temperature and precipitation. The ice dynamics model is set to operate at a nominal resolution of 200 m in order to resolve the small scale topography in which glaciers both initially form and finally depart. Smaller grid spacing would add little in terms of extra information and larger spacing would begin to exclude important topographic features. Generating the high resolution climatology from NARR over a large region requires downscaling that is physics-based and without the requirement of extensive calibration. These characteristics lend confidence in the transferability of the method to future and past climatic conditions. This section provides a brief description of the downscaling of NARR data only (as opposed to downscaling of CRUTS2.1 and GCM data), and the full description is available in Jarosch *et al*. (2010).

4.4.1 Precipitation downscaling

As mentioned above, our downscaling aim was to apply a physics-based model without using a full dynamical downscaling which, at present, is beyond the scope of this project. To meet this need, a

downscaling that applies the physics of orographic precipitation generation was used. This is the so-called linear orographic precipitation model (LOP model) developed by Smith and Barstad (2004) and Barstad and Smith (2005). A schematic illustration of the model is shown in Figure 4.4. This figure shows how advection of moist air facilitates orographic precipitation on the windward side of orographic barriers. On the lee side, if sufficient atmospheric stability exists, a train of gravity waves will be simulated with suppressed precipitation where subsidence occurs and enhanced precipitation where ascent occurs. The mathematical essence of the LOP is summarized in Box 4.2. Because the model involves only a directional derivative and simplified gravity wave dynamics, it can readily be solved in the Fourier domain. Thus, this approach meets our needs for an accurate representation of the governing physics with modest computational demands. The model simulates orographic uplift as well as lee-side mountain wave dynamics which are responsible for precipitation generation and suppression.

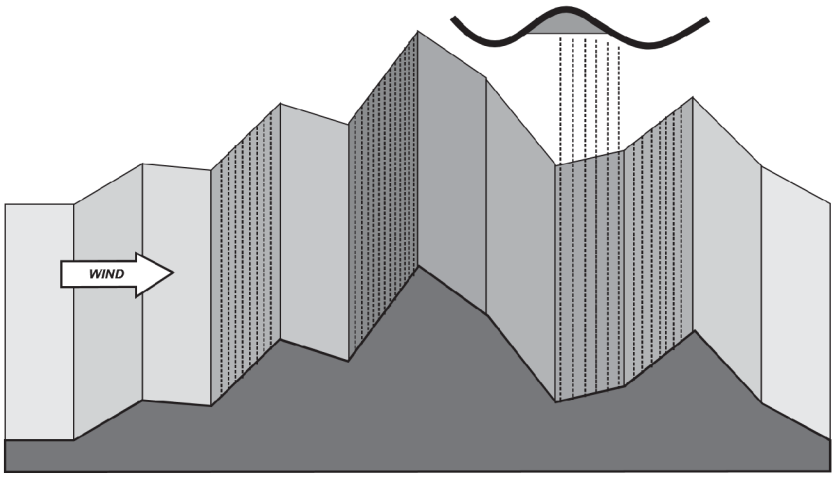

Figure 4.4. Schematic showing the basic ideas of the linear theory of orographic precipitation. When a column of moist air passes over mountainous terrain, the column cools when it is uplifted and relative humidity increases. When the column becomes water saturated precipitation occurs. Under special circumstances waves can form on the lee-side of mountains.

Box 4.2. The linear theory of orographic precipitation.

$$\hat{P}(k,l) = \frac{C_{\mathrm{w}} i\sigma \hat{h}(k,l)}{\left(1 - imH_{\mathrm{w}}\right)\left(1 + i\sigma\tau_{\mathrm{c}}\right)\left(1 + i\sigma\tau_{\mathrm{f}}\right)} \tag{4.2a}$$

$$m(k,l) = \left[\left(\frac{N_{\mathrm{m}}^2 - \sigma^2}{\sigma^2 - f^2}\right)\left(k^2 + l^2\right)\right]^{\frac{1}{2}} \tag{4.2b}$$

$\hat{P}(k,l)$ = two-dimensional Fourier transform of precipitation field $p(x,y)$
k = wavenumber in the x (West-East) direction
l = wavenumber in the y (South-North) direction
C_{w} = uplift sensitivity factor
$\sigma = Uk + Vl$
U = wind velocity in x direction
V = wind velocity in y direction
m = vertical wavenumber
τ_{c} = condensation timescale
τ_{f} = fallout timescale
N_{m} = effective moist static stability
f = Coriolis parameter

Comparison between the NARR-based downscaling and the precipitation data from ClimateBC (Hamman and Wang, 2005; Wang *et al.*, 2006) show very similar performance for the two gridded data sets relative to station data (results not shown). Because we require statistics on the daily variability of temperature and precipitation, our in-house downscaling is preferred. Overall, the NARR-based downscale is dry over the Rocky Mountains as is the ClimateBC product. The NARR-based downscale performs substantially better than a simple spatial interpolation of the NARR precipitation data, which is consistently too-dry over all regions. The dryness in the downscaling arises because the precipitation downscaling is additive to the NARR, so it carries some of the deficits in precipitation that are present in the NARR data set. For instance, the NARR precipitation field is overly dry in the interior of British Columbia especially for mountain stations. Although the orographic model is able to correct for some of this deficiency, the downscaling remains somewhat dry over interior mountains. On average, over the Rocky Mountains the NARR-based downscaling is about 25% deficient in precipitation. This argues for the necessity of a correction factor of roughly 1.33.

Accordingly, we have developed a spatial bias correction by comparing station precipitation amounts to downscaled precipitation amounts. Because we are most interested in wintertime precipitation, we have

developed a bias correction based only on the months November through April. This correction is presented in Figure 4.5 and shows that the largest correction was needed on the western edge of the Mica basin amounting to an enhancement of ~1.7 times the downscaled amount. For much of the Rocky Mountains an enhancement of ~1.2 to ~1.5 is applied. On average, this bias correction is comparable to that needed to correct the 25% deficit noted above. The mass balance results, which will be discussed below, show that the correction was capable of bringing the mean wintertime mass balance into agreement with measurements with some deficiencies with respect to mass balance gradients still present. Closer analysis revealed that there was no consistent bias in vertical mass balance gradient, so no effort was made to correct the precipitation gradients.

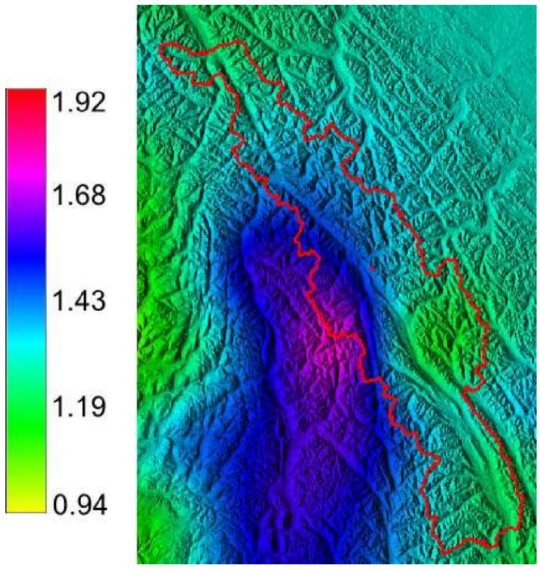

Figure 4.5. Map of precipitation bias correction applied to the downscaled precipitation and to the delta approach for the purpose of simulating glacier mass balance. Overall, the correction acts as an enhancement everywhere with the greatest amounts in the western sector of the Mica catchment. The red line denotes the Mica basin.

4.4.2 Temperature downscaling

The temperature downscaling that we apply makes use of the temperature structure of the reanalysis troposphere. The greatest temperature accuracy in the NARR dataset occurs at mid-troposphere (~400 mbar) with an RMSE there of ~0.7 °C, whereas for surface temperature the RMSE is more than twice that (Mesinger *et al.*, 2006). To downscale temperature we use the NARR temperature and geopotential

height at pressure levels below 500 mbar. A two-component, piecewise-linear model is fit to the temperature structure at daily intervals using least squares regression. The two-component model allows for the possibility of a temperature inversion and other departures from a uniform lapse rate in the vertical column. A schematic representation of this is shown in Figure 4.6. The equation for the piecewise linear fit includes five parameters and is given in Box 4.3. These parameters are derived at all NARR grid locations in the domain and then interpolated to the high resolution of the target downscale topography. The interpolated parameters are then used to calculate temperature at arbitrary altitudes in the study area. For glacier models these altitudes correspond to the DEM surface.

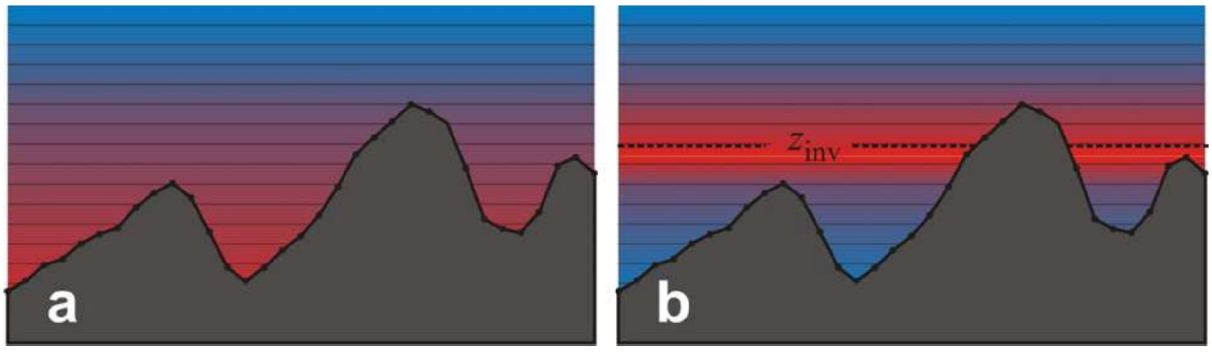

Figure 4.6. Basis of temperature downscaling. The atmosphere is assumed to be vertically stratified and the time-varying thermal lapse rate is determined from the NARR. (a) Simply stratified atmosphere. (b) Atmosphere having a thermal inversion and hence differing lapse rates above and below the inversion level.

Comparisons between the temperature downscale described here and the ClimateBC data show that the NARR-based downscaling has less error during winter and has similar or slightly greater error during summer. A simple downscaling using surface temperature from NARR combined with a fixed temperature lapse rate was also applied and showed worse performance than the other two methods. Both the fixed lapse rate method and the ClimateBC data show seasonal changes in the error indicating a need for at least a seasonally varying lapse rate. For locations with high topographic relief in the Rocky Mountains, the NARR downscaling performs better than the ClimateBC with lower absolute error and lower standard deviation of error between stations. There was no consistent spatial bias in the downscaled temperature relative to station data, so no correction was developed or applied for the purposes of mass balance modelling.

Box 4.3. Method of temperature downscaling.

$$T(x,y) = \begin{cases} h(x,y)\Gamma_{\text{low}} + T_{\text{low}}(z=0) & h(x,y) < z_{\text{inv}} \\ h(x,y)\Gamma_{\text{high}} + T_{\text{high}}(z=0) & h(x,y) \geq z_{\text{inv}} \end{cases} \quad (4.3a)$$

$$z_{\text{inv}} = \frac{T_{\text{low}}(z=0) - T_{\text{high}}(z=0)}{\Gamma_{\text{high}} - \Gamma_{\text{low}}} \quad (4.3b)$$

T = downscaled surface temperature
T_{low} = fitted sea-level temperature for lower segment of curve
T_{high} = fitted sea-level temperature for upper segment of curve
h = surface elevation
Γ_{low} = thermal lapse rate within inversion layer (if present)
Γ_{high} = thermal lapse rate above inversion layer (if present)
z = elevation above sea level
z_{inv} = elevation of inversion

4.5 Mass balance model

This subsection describes how a glacier mass balance model is constructed from downscaled climate fields.

4.5.1 Temperature index ablation model

Mass balance models for alpine glaciers which are described in the published literature range from fairly sophisticated treatments of the surface energy balance (Arnold *et al.*, 1996; Klok and Oerlemans 2002; Hoffman *et al.*, 2008) to simplified temperature index models (Braithwaite, 1981). The needs of the present study are to capture mass balance variability in space and time with minimal input data and also minimal model tuning. Because we are simultaneously modelling all of the glaciers in the Mica catchment, model tuning cannot be performed for individual glaciers. Rather, the mass balance model must be general enough to apply to all of the glaciers in the catchment. Furthermore, there are only five glaciers in western Canada with measurements suitable for model tuning. Surface energy balance models, which are attractive because they are physics-based, do not meet our requirements because they have a number of free parameters to tune. Once tuned, however, they are potentially the most accurate. These models also require comprehensive sets of input data which are not available for the past and future climate scenarios that this study must address. Because the data available from the GCMs is fairly limited in terms of the selection of variables and their temporal resolution, we have chosen the simple

temperature index model that includes solar radiation developed by Hock *et al.* (1999). The formulation for this model is presented in Box 4.4. This melt model contains three tunable parameters and only requires temperature and precipitation as input. Potential, clear-sky solar radiation is also needed, but this can be calculated directly from the surface topography and assumptions about atmospheric transmissivity.

Box 4.4. Basic equations of the radiation-indexed degree-day melt model of Hock *et al.* (1999).

$$M = \begin{cases} \left(\frac{1}{n_\mathrm{d}}\mathrm{MF} + a_\mathrm{snow/ice} I\right) & T > 0 \\ 0 & T \leq 0 \end{cases} \tag{4.4a}$$

$$I = I_0 \left(\frac{R_\mathrm{m}}{R}\right)^2 \psi_\mathrm{a}^{P_\mathrm{a}/P_0 \cos Z} \cos\theta \tag{4.4b}$$

M = surface melt rate ($\mathrm{mm\,h^{-1}}$)
n_d = number of time steps per day
MF = melt factor ($\mathrm{mm\,d^{-1}\,C^{-1}}$)
$a_\mathrm{snow/ice}$ = radiation coefficient (different for snow and for ice)
I = potential clear-sky direct solar radiation at snow/ice surface ($\mathrm{W\,m^{-2}}$)
T = daily mean air temperature (C)
I_0 = solar constant ($1386\,\mathrm{W\,m^{-2}}$)
m = vertical wavenumber
R = instantaneous Sun–Earth distance
R_m = mean Sun–Earth distance
ψ_a = mean atmospheric clear-sky transmissivity
P_a = atmospheric pressure
P_0 = mean atmospheric pressure at sea level
Z = local zenith angle
θ = angle of incidence between normal to grid cell surface slope and solar beam

Model operation is straightforward. Temperature is used to calculate positive degree days. If the daily temperature is zero or below, no melt is computed. For temperatures above zero, melt is the product of positive degree days and the melt factor. The melt factor is a function of potential solar radiation and whether the surface is snow or glacier ice. Three parameters govern the calculated melt factor: the snow and ice radiation coefficients, a_snow and a_ice, and the bulk melt factor *MF*. The mass balance model was developed using daily time steps, so modifications to the input data were required to use a monthly time step. For this, we use average monthly temperatures and monthly total precipitation. When using month-long time steps, a degree-day model must incorporate the inter-daily variability that can generate

substantial melt intensity (or snowfall) even when the mean monthly temperature is low (high). Our approach to incorporating this is given in Section 4.5.3. We have tuned the model using daily and monthly input data and found that the parameters that yield the best model are similar between the two. Accumulation is treated very simply in the model. We apply a temperature threshold for snow of 2°C and assume that all precipitation that falls at temperatures at or below this threshold will contribute to accumulation. Redistribution processes are modelled subsequent to the deposition of snowfall in a model time step. Methods for treatment of these redistribution processes are described in the following section.

The mass balance model requires calculation of potential solar radiation. For this we used the algorithm that is a part of the GRASS GIS package (Hofierka and Suri, 2002). The effects of shading, slope, and aspect are all incorporated. Solar radiation is integrated at a half hourly time step to generate daily totals. For the monthly time step model, these are summed to monthly totals. This computation incorporates a parameter describing the atmospheric turbidity, which influences the intensity of clear-sky solar radiation. We choose a Linke Turbidity coefficient of 3, in keeping with similar values used for mid latitude sites.

4.5.2 Avalanching and wind redistribution

The processes of avalanching and wind redistribution of dry snow are responsible for very strong spatial variability in accumulation on scales ranging from tens of metres to kilometres (Machguth *et al.* 2009). These processes are observationally apparent in the presence of wind-scoured valley walls and low elevation snow accumulation due to wet and dry snow avalanches. A very simple scheme has been developed to begin to address these two processes in the study areas.

Avalanching is incorporated by making a few assumptions about the occurrence and prevalence of large avalanches capable of moving a sufficiently substantial mass of snow to affect glacier mass balance. Redistribution is required to prevent snow accumulation on very steep slopes, such as those present on the eastern edge of the Columbia Icefield, where substantial snow accumulation is unlikely. First, we only attempt to simulate avalanches in terrain above a threshold slope angle of 50°. We assume that all accumulated mass at angles below these remains in place. Second, avalanching is only allowed to take place at the end of the accumulation season on 1 May with the assumption that large scale, wet snow avalanches are most important for mass balance on glaciers. Thus the simulated avalanching is accomplished in one time step per year near the end of the accumulation season. Snow is avalanched and then moved according to several rules: (i) Any snow in excess of 0.5 m water equivalent on slopes above 50° is mobilized; (ii) The material follows a trajectory in the direction of its initial slope azimuth; (iii) Material moves downslope until the slope angle decreases to below 20°; (iv) Deposition obeys an

exponential decay that was modelled to fit avalanche aprons on South Cascade Glacier as derived from a DEM of that glacier.

For the fairly coarse DEM used in this study, avalanching is restricted to higher elevation cirques and ridges. Figure 4.7 shows a sample of water equivalent mass transfer through the avalanching algorithm near the Columbia Icefields. This algorithm is computationally lightweight yet similar to an algorithm developed by Gruber and Bartlet (2007).

Wind redistribution is implemented to simulate the large-scale observed patterns of wind redistribution in the alpine. For example, specific ridges will have a windward face which is consistently scoured and a leeward side that receives enhanced deposition. Thus, the redistribution is performed in a mean, statistical sense rather than by appraising the influence of individual storms and simulating the depositional characteristics of them. The model uses the directional components of the 850 mbar wind velocity vector, and slope, azimuth, and curvature parameters derived from the surface topography. The wind distribution model was driven by the 1979–2008 wintertime (November through April) 850 mbar mean wind vector. As for the delta approach described below, 850 mbar is taken as the pressure level that is closest to the elevation of glaciers in the Mica basin. This method follows that developed by Liston and Elder (2006) and the equations are given in Box 4.5. The redistribution is allowed to be a function of topographic curvature (snow is preferentially deposited in hollows and scoured in convex areas) and the angle of the slope aspect relative to the wind direction (slopes facing into the wind are scoured while those facing away from the wind receive windblown snow). A departure from Liston and Elder's (2006) approach is that we take the inverse of the wind weighting factor directly as the snow loading factor and allow for model tuning to determine the weight of the governing parameters. Thus, when the wind vector is oriented toward topographic slopes and when terrain is convex, L_w becomes large, and snow is effectively scoured. For flat topography where Ω_c and Ω_s are zero, L_w is 1 and snow is neither preferentially deposited nor removed. This wind redistribution model introduces three tunable parameters (γ_s, γ_c, and η). In our tuning we allow γ_s and γ_c, to span the range 0 to +2, which is a greater range than allowed by Liston and Elder (2006).

The tuning for the wind distribution model resulted in weighting the slope orientation parameters most heavily, with the terrain curvature parameter nearly unweighted. This will be discussed further in Section 4.5.4. However, the result is that scouring and deposition occur slope-wide rather than in smaller sub-catchments. This is likely an effect of the tuning data, which are measurements made on the open terrain of glacier centrelines.

Box 4.5. Basic equations of the snow redistribution model

$$\Omega(i,j) = \frac{1}{4}\left\{ \frac{z - \frac{1}{2}\left(z_{i+1,j} + z_{i-1,j} + z_{i,j+1} + z_{i,j-1}\right)}{2\eta} + \frac{z - \frac{1}{2}\left(z_{i+1,j+1} + z_{i+1,j-1} + z_{i-1,j+1} + z_{i-1,j-1}\right)}{2\sqrt{2}\eta} \right\} \tag{4.5a}$$

$$\Omega_{\mathrm{s}} = \beta \cos(\theta - \xi) \tag{4.5b}$$

$$L_{\mathrm{w}} = 1 + \gamma_{\mathrm{c}}\Omega_{\mathrm{c}} + \gamma_{\mathrm{s}}\Omega_{\mathrm{s}} \tag{4.5c}$$

Ω_{c} = topographic curvature factor
Ω_{s} = wind alignment factor
η = topographic curvature length scale
γ_{c} = topographic curvature weighting
γ_{s} = wind alignment weighting
β = topographic gradient
ξ = terrain aspect
θ = wind direction

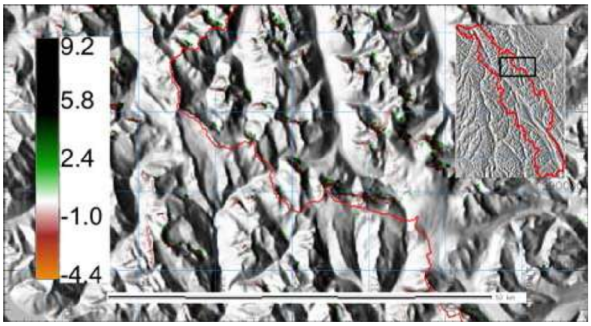

Figure 4.7. Example of snow redistribution by simulated avalanching for a sample region near Columbia Icefields with shaded relief map to provide geographical context. Inset shows the entire Mica watershed. The Columbia Icefield is located near the centre of the map and the red line in both panels is the watershed boundary. The reference grid is at 10 km intervals.

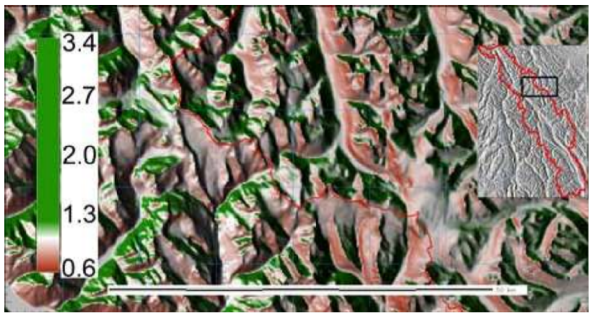

Figure 4.8. Wind redistribution field for the Mica catchment and surroundings. This is derived from climatological mean 850 mbar winds taken from the NARR. South- and west-facing slopes are typically scoured with reductions in snowfall of up to 60%, while some lee-side slopes receive substantially more snow due to wind loading. Inset shows the entire Mica watershed with the zoomed area outlined in black. Red line in both panels is the watershed boundary. The reference grid is at 10 km intervals.

4.5.3 Future and past mass balance via monthly means and the delta approach

The mass balance model is nominally set up to run with a daily time step, which allows the model to respond to daily variability in temperature and precipitation. It is well known that a single snowy day in summer is capable of reducing ablation for several days afterward due to the increased albedo of the snowcover. Although the energy budget is not directly modelled in our approach, a similar effect is simulated with changes in the degree day factor for snow covered ice. Another drawback of forcing a model with longer term mean data is the potential failure to capture periods of ablation when the mean temperature is slightly below the melt threshold. A month long period with a mean temperature of –2 ºC will still likely contain substantial glacier melt. Any model that works on longer-term means would be unable to capture such processes despite their importance for surface mass balance. The meteorological forcing from CRUTS2.1 is available as monthly mean temperature and monthly total precipitation and from the GCMs, upper air data is not available at a daily time step. Using the methods described in this section, the mass balance model integrated on a monthly time scale performs as well as a daily integrated model, runs substantially faster and allows continuity between the monthly available CRU and the GCMs.

To bridge the gap between daily forcing requirements, and monthly forcing availability, we take a stochastic approach to temperature variability. This approach allows accumulation of snow when monthly temperatures are above the threshold for snow, and which allows for ablation when monthly mean temperatures are otherwise below the melting threshold. We use the approach of Marshall *et al.* (2000) for which the equations are given in Box 4.6 and we assume that the monthly temperature distribution is Gaussian. To calculate total positive degree days the Gaussian probability distribution is

integrated from $T = 0$ ºC to infinity following Equation 4.6b. The calculation of precipitation fraction which falls as snow is similar, but the integration is from the snowfall threshold (set at 2 ºC) to negative infinity as given in Equation 4.6a.

Box 4.6. Basis of partitioning precipitation into rain- and snowfall components given the monthly mean temperature and monthly standard deviation of temperature.

$$f_{\text{snow}} = \frac{1}{\sigma_{\text{m}}\sqrt{2\pi}} \int_{-\infty}^{T_{\text{snow}}} \exp\left[-\frac{\left(T - T_{\text{m}}^{\text{A}}\right)^2}{2\sigma_{\text{m}}^2}\right] \mathrm{d}T \tag{4.6a}$$

$$\text{PDD} = \frac{\tau_{\text{m}}}{\sigma_{\text{m}}\sqrt{2\pi}} \int_{0}^{\infty} T \exp\left[-\frac{\left(T - T_{\text{m}}^{\text{A}}\right)^2}{2\sigma_{\text{m}}^2}\right] \mathrm{d}T \tag{4.6b}$$

$$b(x, y, t) = \frac{\rho_{\text{w}}}{\rho_{\text{i}}} f_{\text{snow}} P(x, y, t) \tag{4.6c}$$

f_{snow} = fraction of precipitation to fall as snow
T = temperature
T_{m}^{A} = mean temperature for month m (assumed to be normally distributed)
σ_{m} = standard deviation of monthly temperature about mean value for month m
T_{snow} = threshold temperate at which precipitation falls as snow
PDD = positive degree days
τ_{m} = length of the month m in days
b = ice-equivalent mass balance rate
ρ_{w} = density of water
ρ_{i} = density of ice

To correctly implement this approach, parameters defining the probability distribution are required. These were calculated at each pixel in the 200 m resolution domain using the high resolution, daily NARR downscale. Mean temperatures for the 12 months were calculated by finding the monthly mean and then averaging those means. The standard deviations were calculated by finding the standard deviation in temperature for each day in the calendar year among all like days that occurred in the NARR period (i.e. the standard deviation amongst all 1st of January for the 29 years of the downscaled NARR). These standard deviations were then averaged by month yielding a representative standard deviation for the given month. The integrated probability distribution yields positive degree days directly. For accumulation, the monthly total precipitation is multiplied by the snow fraction calculated as described above to yield total accumulation (see Equation 4.6c).

The anomalies were calculated in a fairly straightforward manner. First, the output for each GCM was interpolated to a spatial grid of resolution 320 km (ten times the grid spacing of NARR) for the purpose of creating a unified product from GCMs of various spatial resolutions. For CRUTS2.1, the 0.5° data were interpolated to a resolution of two times the NARR grid spacing (~64 km) which roughly corresponds to an arc length of 0.5° at the latitude of the Mica catchment. Next, a climatology was established for the period 1979 through 1999 which has temporal overlap with the NARR. From this climatology, the anomaly was calculated by evaluating the difference (ratio for precipitation) between a given month in the future GCM scenario and the GCM climatological state, as well as the difference between a given month in the historical CRUTS2.1 data and its climatological state. Two anomaly maps (one for precipitation and one for temperature) were created for the CRUTS2.1 data and each of the GCM/SRES combinations. These anomaly maps were applied to the NARR downscale climatology by interpolating the anomaly to the 200 m resolution of the downscale and adding the anomaly onto the climatology for temperature and multiplying for precipitation. This derived field was then used to force the mass balance model as described above. Because glaciers have been shown to be most responsive to conditions in the lower, free troposphere (Rasmussen and Conway, 2001; Rasmussen and Conway, 2003), we used temperature anomalies at 850 mbar rather than the surface temperature anomalies in the GCM scenarios.

4.5.4 Model tuning

The mass balance model described above has six parameters that require tuning. Three of these are associated with the ablation part of the mass balance model and the other three are part of the snow redistribution model. Model tuning requires a tuning target for which there are minimal measurements relevant to glacier modelling and none within the Mica basin study area. The glaciers that have appropriate measurements of winter accumulation and summer ablation are the Place, Helm, Tiedemann, and Bridge glaciers in the Coast Mountains of British Columbia and Peyto Glacier, which is situated in Alberta several kilometres east of the Mica catchment. Mass balance measurements were gathered by the Geological Survey of Canada and were quality controlled and made available by Joe Shea (University of Northern British Columbia, personal communication, 2010). These data cover the time span from roughly 1964 through 1995. Of this span, we utilize only the portion that overlaps with the NARR, namely 1979 through 1995. The mass balance data are provided as summer, winter, and net balance by elevation band. To generate comparable data from the mass balance model, grid cells at appropriate elevation ranges on the glaciers of interest were averaged.

The model was tuned in a three-step process. The first step was to tune the model to the parameters relevant for ablation (namely R_f, a_s, a_i) with the accumulation parameters held fixed. The model was next tuned to the parameters important for accumulation (namely η, γ_c, γ_s) this time with the summer parameters held fixed. In the third step, the summer tuning was repeated to confirm those initial parameters. The parameters were tuned by season to limit possible combinations of parameters and speed model tuning. Because the accumulation and summer ablation processes are fairly independent, it is unlikely that relevant parameter combinations were overlooked. Model performance was gauged using the root mean square error (RMSE) fit to the mass balance measurements. In generating the RMSE score, the glaciers of the Coast Mountains were given a 2/3 total weighting while Peyto Glacier results were given a 1/3 weighting. This was done to give leverage to Peyto Glacier, where the climate regime is markedly different than that of most of the Coast Mountain glaciers, and is likely more representative of conditions in Mica basin. Tuning was accomplished via a simplex minimization search across the three-parameter space for each seasonal tuning. Scatter plots showing the relationship between the measured and modelled seasonal mass balance components in the best fitting model are given in Figure 4.9. Overall, the RMSE is lower for winter than summer owing to the lesser mass flux at individual stakes (i.e. summer ablation can reach 10 m w. e. while accumulation rarely exceeds a few metres). For comparison with the group tuning outlined here, the model was tuned to individual glaciers and the statistics compared with the results from group tuning (Table 4.2). All values have units of metres water equivalent. The parameters derived from the tuning are presented in Table 4.3.

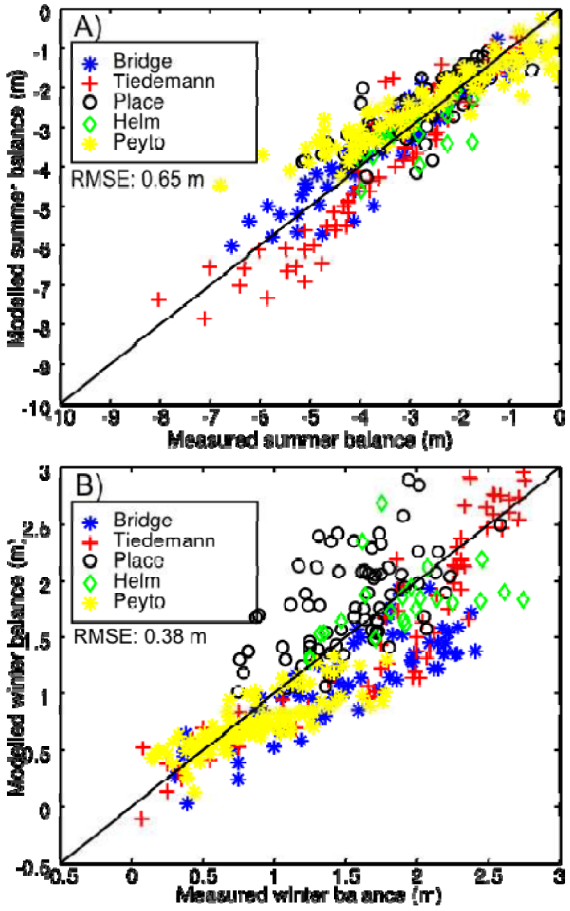

Figure 4.9. Scatterplots comparing modelled summer and winter mass balance to measurements. The RMSE for each seasonal balance is shown. The seemingly larger scatter in winter mass balance is due to the scaling of the axes. For the various glaciers, it can be seen that the tuning was successful in bringing the mean mass balance close to measurement, but often the slopes are different than one. This is especially apparent in the results from Peyto Glacier. (a) Summer mass balance comparison. (b) Winter mass balance comparison.

Table 4.2. RMSE results from mass balance model tuning with units of m w. e. Individual tuning targets are in the left hand column and results for each glacier are given across the table. Clearly the lowest weighted mean RMSE is obtained by tuning to all glaciers simultaneously rather than any individual glacier.

Summer Tuning Target ↓	**Bridge**	**Helm**	**Place**	**Tiedemann**	**Peyto**	**Weighted Mean**
All	0.50	0.50	0.62	0.75	0.77	0.65
Bridge	0.50	0.48	0.63	0.79	0.78	0.66
Helm	0.57	0.43	0.70	0.62	0.87	0.68
Place	0.54	0.58	0.59	0.67	0.85	0.68
Tiedemann	0.81	0.50	0.67	0.56	0.92	0.73
Peyto	0.89	0.45	0.75	1.62	0.63	0.83
Winter Tuning Target ↓	–	–	–	–	–	–
All	0.47	0.43	0.45	0.36	0.26	0.37
Bridge	0.40	1.44	0.87	3.07	0.33	1.07
Helm	0.47	0.41	0.57	0.65	0.23	0.43
Place	0.51	0.49	0.36	0.44	0.33	0.41
Tiedemann	0.47	0.56	0.49	0.28	0.40	0.44
Peyto	0.47	0.41	0.59	1.97	0.22	0.65

From the table, it can be seen that tuning to all glaciers exacts a performance cost when the RMSE for the tuning toward a given glacier is compared with the bulk-tuned RMSE. For the summer tuning, the error incurred in tuning to a single glacier is smaller than the case for winter mass balance. An exception is the case of Tiedemann Glacier when parameters tuned to Peyto Glacier are applied. These two glaciers are situated in disparate climates, so the parameters appropriate for Peyto are unlikely to be appropriate here. Specifically, the tuning for Tiedemann Glacier yielded relatively low values for the snow and ice radiation factors, yet yielded a very high value for the bulk radiation factor (table 4.3). This suggests that albedo is of relatively less importance for Tiedemann and a high bulk radiation factor makes the model of Hock (1999) behave more like a standard degree day model. On Peyto Glacier, the continental setting

likely makes solar radiation a very important contributor to the radiation budget, so these parameters are strongly controlled in the tuning.

For winter tuning, some of the RMSE values are strikingly high, such as the RMSE for Tiedemann Glacier when parameters tuned for the Bridge Glacier are used. Interestingly, the opposite case does not prove true: the RMSE value for the Bridge Glacier is low when the model is driven by parameters tuned for the Tiedemann Glacier. The results for all glaciers are best when driven by parameters tuned for Place Glacier.

These results are not surprising because the tuning parameters are likely somewhat region specific, so this should be viewed as a rough estimate of the uncertainty range associated with using bulk parameters in the model. An attempt was made to determine relationships between model parameters and other variables (such as geographic location, elevation of ELA, total precipitation, mean annual temperature at the ELA). However, the small size of the tuning data set prevented establishing relationships with any kind of statistical confidence, or which had any physical meaning.

Table 4.3. Tuned parameter values for the individual glaciers and for the multi-glacier tuning described in the text.

Tuning Target	R_f (m deg.$^{-1}$ dy^{-1})	a_i (m^3 W^{-1} deg.$^{-1}$ dy^{-1})	a_s (m^3 W^{-1} deg.$^{-1}$ dy^{-1})	η (M)	γ_c	γ_s
Bridge	0.64	0.017	0.011	300	3.00	1.54
Helm	0.58	0.015	0.010	1038	0.00	2.90
Place	0.70	0.014	0.012	1298	0.11	0.00
Tiedemann	2.84	0.005	0.002	2098	1.14	0.00
Peyto	0.53	0.025	0.009	2328	0.00	3.00
Multi Glacier Tuning	1.08	0.015	0.009	255	0.00	1.89

4.5.5 *Adjusting mass balance to dynamic topography*

Computing mass balance on a domain the size of Mica basin is a computationally intensive process, which would be magnified if the calculations were made at annual time steps as the glacier surface evolves. So, rather than coupling the mass balance fully with the ice dynamics model and then downscaling and calculating solar flux on the new topography at an annual model time step, we calculate a mass balance on the static year 2000 DEM for all mass balance years in both past and future. To arrive at a change in mass balance for a change in topography, the mass balance sensitivity to temperature is used combined with a fixed temperature lapse rate. This assumes that the effects of a change in temperature associated with a changing ice surface elevation will outweigh the associated changes in precipitation and solar radiation. Because precipitation is downscaled on a 1 km DEM with its associated smooth topography, our assumptions regarding precipitation are probably accurate for the purposes of these simulations. For the solar component, little change is expected. The change in solar radiation transmitted by the slightly thinner or thicker slice of atmosphere for thicker and thinner ice is probably very small. The only major issue that could arise would be changes in the aspect or shading characteristics of a given glacier pixel. This should have the largest effect on the last remnants of glaciers in high cirques and within the coarse topography of the present-day glacier bed. We do not quantify this effect and assume it is small.

The topographic adjustment requires calculation of the temperature sensitivity of glaciers in the Mica basin. This was accomplished by simulating the glacier mass balance with temperature uniformly increased and decreased by 0.1° C. The sensitivity is calculated on all areas of the domain to account for the possibility of glacier advance or changes in glacier location. The results are presented in Figure 4.10 and are in line with mass balance sensitivity published in the literature for both surface energy balance and temperature index mass balance models. For the most part, sensitivity is highest at the lowest elevations largely because this is most often classified as "ice" in the mass balance model. At high elevations, the sensitivity is low due to both the less frequent occurrence of melting temperatures and the prevalence of snow cover, which has a lower melt factor than glacier ice. With warming in the future, the mass balance sensitivity for a given point is likely to increase, so this method is likely to underestimate melting.

During both the transient and the static mass balance simulations, the evolving mass balance is calculated domain-wide using the mass balance sensitivity db/dT and the change in elevation Δz between the reference DEM and the evolving surface. The calculation follows:

$$\Delta b_n(x,y) = \frac{db_n}{dT}(x,y)\frac{dT}{dz}\Delta z(x,y)$$

$$b_n(x,y) = \bar{b}_n(x,y) + \Delta b_n(x,y)$$

where Δb_n is net mass balance, $\bar{b}_n$ is the mass balance from reference topography, T is temperature, and Δz is change in surface topography from the year 2000 reference. The quantity dT/dz is the atmospheric temperature lapse rate for which a value of 6.5° C km^{-1} is used. A fixed lapse rate is acceptable here because the change in elevation is relatively small, so variable lapse rates will have a fairly small effect. The mass balance forcing calculated on the static topography is adjusted on an annual basis with the topographic change for the present time step. Incorporating this effect is important for the large changes in glacier thickness that have occurred since the dawn of the 20th century and which will occur in the next 100 years. As a sample calculation shows, with a mass balance sensitivity of 0.6 m w. e. °C^{-1} (based on our chosen temperature lapse rate) and Δz of 50 m, the surface mass balance changes by almost 20 cm w. e. Some glaciers have experienced substantially greater surface height change with annual rates of 9 m yr^{-1} over decades (Schiefer *et al.,* 2007) so substantial changes in the surface mass balance associated with this elevation change can be expected.

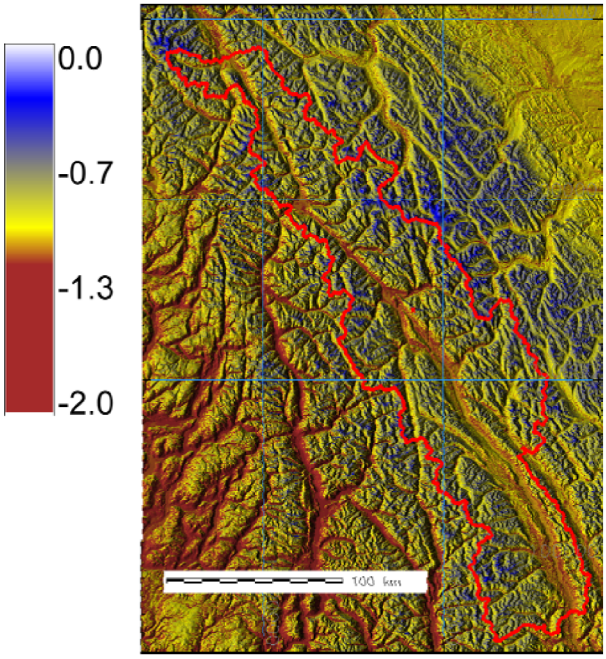

Figure 4.10. Map of mass balance sensitivity to change in temperature of 1°C. Red line denotes the boundary of the Mica catchment. Sensitivity is greatest in valley bottoms and lowest near the summits of the Columbia Icefields. These data are used to correct the mass balance to account for changes in topography associated with evolution of the ice surface in the ice dynamics model.

4.6 Glacier dynamics modelling

4.6.1 Model formulation

The equations governing glacier dynamics are summarized in Box 4.7. Equations 4.7a and 4.7b represent components of the vertically integrated ice creep velocity and Equation 4.7c governs the evolution of ice thickness H in response to ice flow and surface mass balance b. The sliding velocity components are given by V_x and V_y and are governed by a Weertman sliding law. Equations are solved using finite differences on the 200 m spatial resolution grid for the region. The flow law coefficient A is taken as $A=1.0\times10^{-15}$ Pa^{-3} yr^{-1} and the flow law exponent has been fixed at $n = 3$, the conventional value. For all calculations in the present study the sliding coefficient C was set to zero so that the sliding process was completely suppressed. If sliding were activated then the projected rates of deglaciation would increase.

Box 4.7. Basic equations of the glacier dynamics model.

$$Q_x = -\frac{2\mathcal{A}(\rho g)^n}{n+2}\left[(\partial_x S)^2 + (\partial_y S)^2\right]^{\frac{n-1}{2}} H^{n+2}\partial_x S + V_x H \tag{4.7a}$$

$$Q_y = -\frac{2\mathcal{A}(\rho g)^n}{n+2}\left[(\partial_x S)^2 + (\partial_y S)^2\right]^{\frac{n-1}{2}} H^{n+2}\partial_y S + V_y H \tag{4.7b}$$

$$\frac{\partial H}{\partial t} = -\frac{\partial Q_x}{\partial x} - \frac{\partial Q_y}{\partial y} + b \tag{4.7c}$$

$$V_x = \mathcal{C}(\rho g)^m H^m\left[(\partial_x S)^2 + (\partial_y S)^2\right]^{\frac{m-1}{2}}\partial_x S \tag{4.7d}$$

$$V_y = \mathcal{C}(\rho g)^m H^m\left[(\partial_x S)^2 + (\partial_y S)^2\right]^{\frac{m-1}{2}}\partial_y S. \tag{4.7e}$$

H = ice thickness = $S - B$
S = ice surface elevation
B = bed surface elevation
b = ice-equivalent glacier mass balance rate
Q_x = volume flux of ice in the x direction
Q_y = volume flux of ice in the y direction
$\mathcal{A}$ = coefficient for Glen's flow law for ice
ρ = ice density
g = gravity acceleration
n = exponent for Glen's flow law for ice
∂_x = partial derivative in x direction
∂_y = partial derivation in y direction
V_x = sliding velocity in the x direction
V_y = sliding velocity in the y direction
$\mathcal{C}$ = coefficient for glacier sliding law
m = exponent for glacier sliding law

4.6.2 Model spin-up

Although the response time of glaciers is fairly slow and the behavior of glaciers is non-chaotic, initializing the dynamics model is not straightforward. Uncertainties regarding bed topography, mass balance, and ice rheology prevent a "snap shot" start from the year of the most recently available glacier mask and derived bed topography. Doing so generates transient kinematic waves that require nearly a century of simulation to propagate to the termini of the slowest responding glaciers. This transient effect interferes with simulation of the terminus position, ice volume, and ice covered area which are all of critical interest for the hydrological modelling at the heart of this project. To address this problem we take the approach of growing ice from an ice-free topography to a steady state from a glacier mass balance derived from early 20th century climate data. Using this initial state, we launch transient

simulations through the 20th century with the aim of reaching a reasonably accurate ice volume and ice covered area by the onset of the climate projections at 2000 AD. This approach will not provide an exact match between modelled and actual glaciers at any time, but should yield a reasonable glaciation state from which to simulate ice volume loss through the 21st century.

The steady state simulations are initialized with zero ice cover and permitted to run until there is no change in ice volume. The amount of time for steady state to be achieved depends on surface mass balance. A more positive regional mass balance causes the growth of larger glaciers, but also forces a greater mass flux resulting in more rapid advance of glacier termini. The initial state was achieved using the first 30 years of mass balance derived from the CRUTS2.1 (mass balance years 1902–1931) with the mass balance adjusted to the changing ice topography. Initially, the elevation/mass balance feedback acts as a negative mass balance forcing because the deglaciated surface topography is everywhere lower than at the reference state. As the steady state simulation evolves, the elevation feedback leads to positive mass balance forcing. Any initial state will lead to the same steady state ice cover because there is no hysteresis in the ice volume/mass balance relationship for the topographically controlled glaciers of the Mica catchment. In the case of growing ice from zero ice volume, using the CRUTS2.1-derived mass balance, the slowest responding glaciers in the Mica basin take roughly 500 years to reach a steady state. An open question in this approach is whether allowing the glaciers to reach steady state is even appropriate. It is likely that the moraines recording the Little Ice Age advance were deposited by glaciers responding to a much shorter-lived climate forcing than our 500 years of constant climate. Visual inspection of terminus position in the CRUTS2.1-derived steady state compared with Little Ice Age (LIA) trimlines suggests that the initial state is reasonably consistent with the maximum LIA ice extent, although this is difficult to quantify precisely. There are regional differences, however, which are likely due to errors in the surface mass balance as well as inaccuracies in the inferred bed topography. It is assumed that some of the inaccuracies in the initial state remedy themselves during the 100 years of transient simulation that precede implementation of the mass balance derived from GCM climate change scenarios. This assumption is confirmed by the close correspondence between measured glacier area and glacier volume change in the years 1985, 2000 and 2005. Once steady state glaciers have been established, the model is run in transient mode beginning with the first year of CRU-derived mass balance (1902). Although part of the model spin-up process, this step will be described in the next section.

4.6.3 Transient simulations

The transient simulations are operated in a similar fashion to the steady-state runs except the mass balance changes annually. The mass balance time series were constructed from the CRUTS2.1 forcing for 1902–1978, from NARR monthly forcing for 1979–2006, followed by GCM climate change scenarios through year 2100. No efforts were made to prevent major jumps in mean mass balance across the transitions from one mass balance source to the next (such as CRUTS2.1 to NARR). For several of the GCM scenarios, there is a substantial change in mean temperature beginning at the onset of the climate change scenario. Namely, this shift is present within the CSIRO GCM at the interface between the 20th century emissions simulation and the emissions scenario forcing for B1 and A2 forcings.

5 Generation of future climate scenarios

5.1 Selection of GCM and emissions scenarios

Our objective is to select an ensemble of Global Climate Models (GCMs) whose climate projections will be used as drivers for glacier and hydrologic models in BC's Columbia Basin. Our initial set of GCMs consists of 22 models which have been run for the Intergovernmental Panel on Climate Change (IPCC) fourth assessment report (AR4). Data from these 22 GCMs have been collected in the Coupled Model Inter-comparison Project phase 3 (CMIP3), which is archived at the Program for Climate Model Diagnosis and Intercomparison at Lawrence Livermore National Laboratory (LLNL). The CMIP3 data are increasingly being downscaled and used to address regional and local issues in impact studies of climate change (e.g. Coquard *et al.*, 2004; Brekke *et al.*, 2008). Some common questions such studies face are: How should one select the global models to use in the regional studies? If different GCMs give different results, what strategy should be used for selecting the global models? Addressing these questions requires evaluation of GCM performance using a range of different performance measures/metrics (e.g. Gleckler *et al.*, 2008; Pincus *et al.*, 2008; Pierce *et al.*, 2009). These evaluations usually compare modelled versus observed climatologies in the form of root-mean square differences, ranking the models according to their demonstrated ability to reproduce observed features of recent climate. Two recent studies on performance and selection of GCMs (Gleckler *et al.*, 2008; Pierce *et al.*, 2009) provide a set of guidelines for the model selection, and these guidelines can be summarized as follows: (1) enough model realizations must be chosen to account for the effects of the models' natural internal variability; (2) stable hindcasts (and forecasts) can be obtained by including a manageably small group of models – e.g., Pierce *et al.* (2009) found that model skill tends to converge after including approximately 5 different models; and (3) the selection of GCMs depends on the intended application – i.e., there is no universal set of performance metrics that are optimally suited for all applications.

For the purpose of this project we evaluate performance of 22 GCMs from CMIP3, listed in Table 5.1, using a variety of performance metrics. Since our application is in hydrologic and glacier modelling, the evaluation of models is focused on climate variables important for the processes on the climate-watershed interface. Variables that are needed for our downscaling methods and glacier and hydrologic models are air temperature, precipitation and sea-level pressure. Beside these variables we evaluate GCMs based on a few additional variables that are shown to be linked to glacier mass balance (e.g. Matulla *et al.*, 2008): specific humidity at 850 hPa and geopotential heights at 500 and 850 hPa. Some GCMs have multiple runs for each emission scenario, while some only have one run. Therefore, to make our evaluation

analysis consistent, we use only the first run for the GCMs that have multiple runs. In the following sections we describe our evaluation analysis in greater detail.

Table 5.1. Model identification, originating group, and atmospheric resolution

IPCC I.D.	Center and Location	Atmosphere Resolution
BCCR-BCM2.0	Bjerknes Centre for Climate Research (Norway)	T63 L31
CGCM3.1(T47)	Canadian Centre for Climate Modelling and Analysis (Canada)	T47 L31
CGCM3.1(T63)		T63 L31
CSIRO-Mk3.0	CSIRO Atmospheric Research (Australia)	T63 L18
CNRM-CM3	Météo-France, Centre National de Recherches Météorologiques (France)	T42 L45
ECHO-G	Meteorological Institute of the University of Bonn, Meteorological Research Institute of KMA, and Model and Data group (Germany and Korea)	T30 L19
GFDL-CM2.0	US Dept. of Commerce, NOAA	N45 L24
GFDL-CM2.1	Geophysical Fluid Dynamics Laboratory (USA)	N45 L24
GISS-AOM	NASA/Goddard Institute for Space Studies (USA)	90 × 60 L12
GISS-EH		72 × 46 L17
GISS-ER		72 × 46 L17
FGOALS-g1.0	LASG/Institute of Atmospheric Physics (China)	128 × 60 L26
INM-CM3.0	Institute for Numerical Mathematics (Russia)	72 × 45 L21
IPSL-CM4	Institut Pierre Simon Laplace (France)	96 × 72 L19
MIROC3.2(medres)	Center for Climate System Research (The University of Tokyo), National Institute for Environmental Studies, and Frontier Research Center for Global Change (JAMSTEC) (Japan)	T42 L20
MIROC3.2(hires)		T106 L56
MRI-CGCM2.3.2	Meteorological Research Institute (Japan)	T42 L30
ECHAM5/MPI-OM	Max Planck Institute for Meteorology (Germany)	T63 L32
CCSM3	National Center for Atmospheric Research (USA)	T85 L26
PCM		T42 L18
UKMO-HadCM3	Hadley Centre for Climate Prediction and Research, Met Office (UK)	96 × 72 L19
UKMO-HadGEM1		N96 L38

5.2 Evaluation of GCM performance using statistical measures

First, models were screened according to their performance at the global scale based on results from Gleckler *et al.* (2008). As GCMs are designed to replicate the global climate system, poor performance of the model at global scale is taken as an indicator of poorly specified parameters that represent sub-grid-scale processes (often referred to as parameter uncertainty) or parameterization that is missing key process or feedbacks or forcings, such as poorly initialized oceans or land surface processes, etc., (otherwise known as structural uncertainty). Thus, we assume that poor model performance at the global scale that is related to these types of errors might lead to erroneous results at the regional scale. Here we review statistical measures outlined in Gleckler *et al.* (2008). One statistical measure of model fidelity is the root mean square (RMS) difference between a simulated field F and a corresponding reference data set (observations) R. For monthly mean climatological data, the most comprehensive RMS error statistic (E) accounts for errors in both the spatial pattern and the annual cycle, and it is calculated as follows:

$$E^2 = \frac{1}{W}\sum_i \sum_j \sum_t w_{ijt}\left(F_{ijt} - R_{ijt}\right)^2. \tag{5.1}$$

The indices *i, j* and *t* correspond to the longitude, latitude and time dimensions, and *W* is the sum of the weights (w_{ijt}), which for the spatial dimensions are proportional to grid-cell area and for time are proportional to the length of each month. In most cases considered here the sums are accumulated over all 12 months and over one of the spatial domains of interest. The RMS error can be calculated for each climate variable/field of interest. Furthermore, for each climate field Gleckler *et al.* (2008) defined a typical model error as the median of all RMS error calculations. Thus, in an ensemble of 22 GCMs there are 22 RMS error calculations for one climate field using one reference data set. Relative model performance is then defined, for a given model and a given climate field, as a difference between the RMS error and the 'typical' error, normalized by the 'typical' error. Normalizing the RMS calculations in this way yields a measure of how well a given model (with respect to a particular data set) compares with the typical model error. For example, if the relative error has a value of –0.2, then the model's RMS error is 20% smaller than the typical model. Gleckler *et al.* (2008) calculated relative errors for all 22 GCMs from CMIP3 and 26 climate fields, using various observational datasets, and presented their results for global and subglobal domains (North and South hemisphere extra tropics and tropics). GCMs were compared to two gridded observational datasets: a 'primary' and 'alternative' set in each case. Datasets differed by variable, but were primarily sourced from re-analysis (ERA40, European Centre for Medium-Range Weather Forecasts and NCEP/NCAR, National Center for Atmospheric Research) or Earth Radiation Budget Experiment (ERBE)/Clouds and the Earth's Radiant Energy System (CERES) datasets (Gleckler et al., 2008). These products represent the best observationally constrained estimates of the free atmosphere available for the variables evaluated in this work. For a full list of 'primary' and 'alternative' datasets and an explanation of all variables investigated see Table 2 in Gleckler et al. (2008).

In an attempt to define an optimal overall index of model performance for simulating mean annual cycle climatology, Gleckler et al. (2008) introduced a Model Climate Performance Index (MCPI). This index averages each model's relative errors across all of the fields in the study.

Other measures of model performance consider model simulation of inter-annual variability. In Gleckler *et al.* (2008) this is done by examining variances of monthly mean anomalies, computed relative to the monthly climatology for the period 1980–1999. For each GCM and each climate field of interest, the ratio of simulated to observed variances for global and all subglobal domains is calculated. An exploratory Model Variability Index (MVI) is defined as:

$$MVI = \sum_{f=1}^{F} \left[\beta_f - \frac{1}{\beta_f} \right]^2 \tag{5.2}$$

where β_f^2 is the ratio of simulated to observed variance and F is the total number of variables. Defined this way, the MVI is positive definite, with smaller values indicating better agreement with the reference data.

We proceed to evaluate all 22 GCMs at the regional and sub-regional scale to investigate the strength of the CMIP3 models over our area of interest. In our evaluation of GCMs, we use the North American Regional Reanalysis (NARR) as the reference data. Climate fields from all 22 GCMs are interpolated to 10×NARR resolution, which is approximately 320×320 km in conical conformal Lambert map projection. We use monthly data from the modelled 20th century simulations and from NARR, and focus on the overlapping period between the two data sets: 1980–1999. Model performance is analyzed on two spatial domains: 'large,'equivalent to the original NARR domain, and 'small,' which roughly covers the NW corner of the large domain (Figure 5.1). For these two domains and our climate fields of interest (temperature and specific humidity at 850 hPa, geopotential height at 500 hPa and 850 hPa, sea level pressure and precipitation) we calculate relative model errors, variance ratios, and model performance indexes (MCPI and MVI).

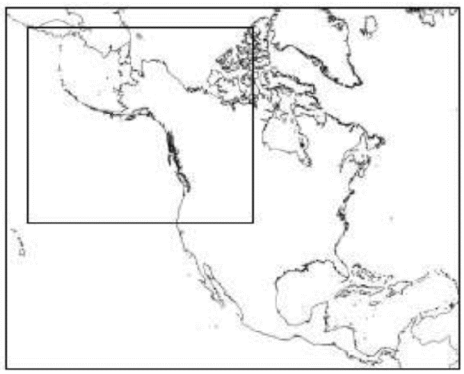

Figure 5.1. Analysis domains. The 'large' domain corresponds to the entire North American Regional Reanalysis (NARR) domain and the 'small' domain to the rectangle within the large one.

The results for relative model errors and MCPI are presented in Figure 5.2, negative numbers indicating that the model's RMS error is smaller than the 'typical' model. Our results echo those of Gleckler *et al.* (2008) that both the models' mean and median (out of all 22 models) are in better agreement with the reference data (NARR) than the typical model. Some models (e.g. ECHAM/MPI-OM, UKMO-HadGEM1, UKMO-HadCM3, MIROC3.2(medres), CGCM3.1(T47)) clearly score better than others,

although no model scores consistently above or below average in all respects. For example, ECHAM/MPI-OM scores high over both domains and all the variables except for precipitation over the 'large' domain. The IPSL-CM4 model, on the other hand, scores low over both domains and all the variables except for specific humidity at 850 hPa, where it scores better than the 'typical' model. According to the relative model errors calculated for the mean annual cycle over 1980–1999, the skill of many of the models differs considerably by variable.

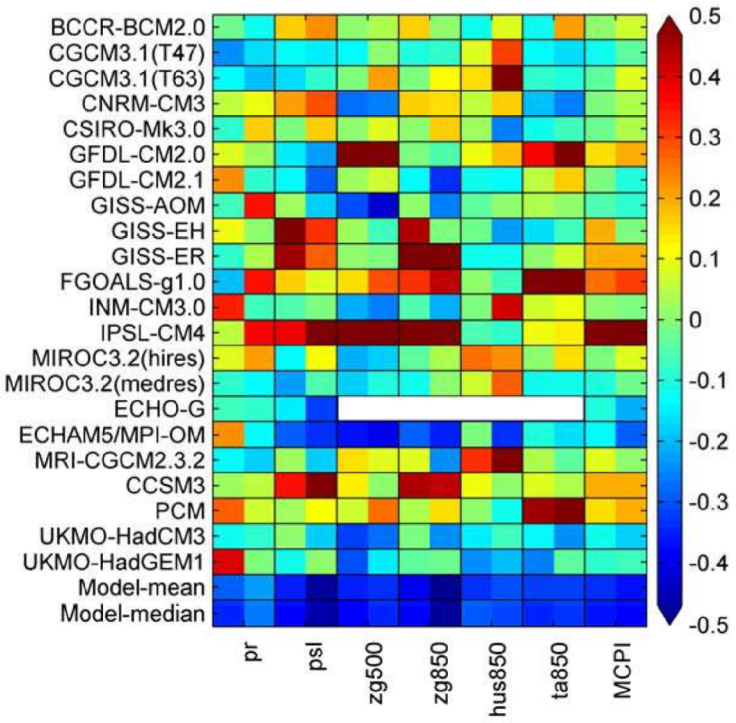

Figure 5.2. Portrait diagram display of relative error metrics for 20th century CMIP3 annual cycle climatology (1980–1990). Each combination of climate variable (*x*-axes) and GCM (*y*-axes) is represented by two grid squares, which correspond to relative errors calculated over the large domain (left square) and the small domain (right square). Variables are: precipitation (pr), sea level pressure (psl), geopotential height at 500 hPa and 850 hPa (zg500, zg850), specific humidity at 850hPa (hus850), and temperature at 850hP (ta850). MCPI is 'Model Climate Performance Index,' calculated as an average of relative errors across the displayed variables.

The results of model evaluation according to their simulation of inter-annual variability (variance ratios and MVI following Gleckler *et al.* (2008) above) are presented in Figure 5.3. Variance ratios (GCM vs. NARR) close to unity indicate that the variance of simulated monthly anomalies (for a given climate variable) compare well with NARR, whereas lower ratios suggest there is too little simulated variability and higher ratios imply too much. The models whose variance compares well with NARR for most climate variables are BCCR-BCM2.0, CNRM-CM3, UKMO-HadCM3, UKMO-HadGEM1. On the other hand, the models whose variance differs the most from NARR for most of the variables are IPSL-CM4, CCSM3 and PCM. Similarly to the previous metric, the skill of many models changes considerably from one variable to the next. It is worthy of note that the variance of precipitation monthly anomalies over the 'large' domain is underestimated from each of the 22 GCMs.

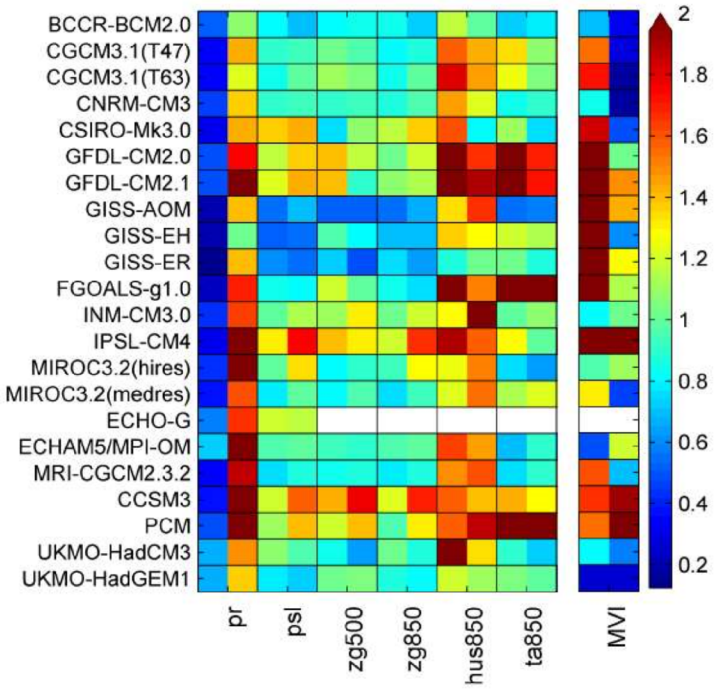

Figure 5.3. Portrait diagram display of variance ratios (GCM/NARR) for 1980–1999 monthly anomalies in the large domain (left square of each column) and small domain (right square). Last column is Model Variability Index (MVI). Climate variables are the same as in Figure 5.2.

According to all applied statistical measures (relative error, MCPI, variance ratios and MVI), the skill of many models appears to depend more on the variable than on the domain. For both MCPI and MVI, the smaller the values, the better the skill. Furthermore, similar to the results in Gleckler *et al.* (2008), most of the models that score high in simulating mean climate (MCPI) also score high in simulating inter-annual

variability (MVI). The models for which this is not the case in our analysis (i.e., the skill of the models is high for MCPI but not for MVI) are GFDL-CM2.1, INM-CM3.0, and ECHAM/MPI-OM.

To summarize the results of models' performance we rank the GCMs according to their scores derived from relative error metrics for annual cycle climatology (1980–1999), MCPI and MVI. For these metrics, models are ranked as being in the top 1 to 5, 6 to 10, 11 to 15, 16 to 20, and 21 to 22, the last being the bottom place (Figure 5.4). As illustrated, the ranking is highly sensitive to the choice of climate variable and the spatial domain. Considering the large spread of variable-specific model performance, Gleckler *et al*. (2008) noted that the overall model performance indices (MCPI and MVI), calculated as average error or average variance ratio across all variables, can hide substantial model errors. Thus, the complexity of the models and the characteristics of their simulated fields cannot be adequately captured by a single measure of performance. Also, it might be fruitful to explore a wide range of metrics, rather than striving for a single index of overall skill. Therefore, in the next section, we apply another performance metric which differs from the metrics based on RMS errors, and which evaluates models on finer temporal scales than the monthly climatologies.

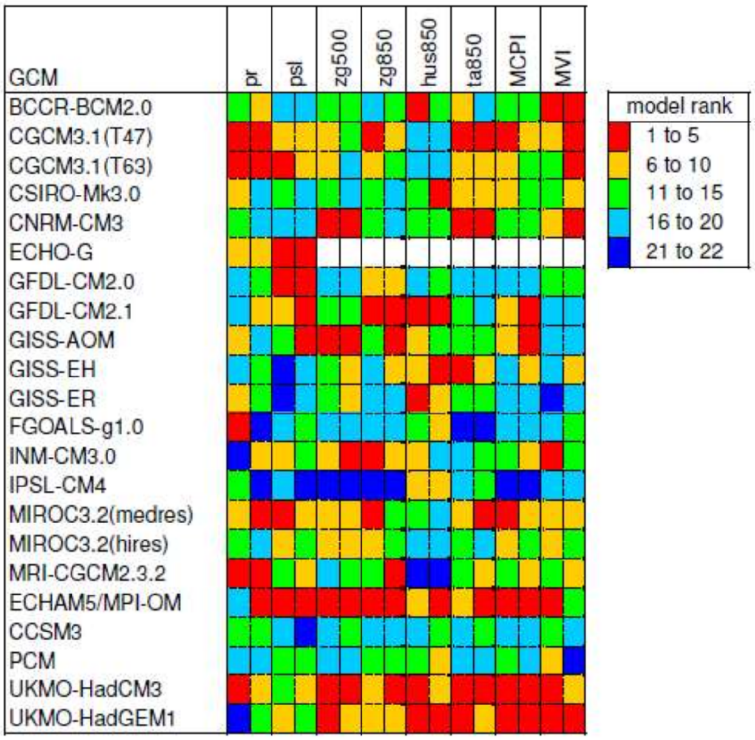

Figure 5.4. Portrait diagram display of model ranking according to the relative errors for six climate variables (pr, psl, zg500, zg850, hus850, and ta850), and performance indexes MCPI and MVI, over 'large' domain (left square in each column) and 'small' domain (right square). Full names of variables are listed in the caption of Figure 5.2.

5.3 Evaluation of GCM performance using self-organizing maps

Because precipitation is the key input variable for modeling snow accumulation, it is important to validate how a GCM reproduces the processes by which precipitation is delivered to glaciers in the study region. Additionally, multiple streamflow regimes are prevalent across BC, each of which is characterized primarily by its climatic drivers, such as rainfall or snowmelt (Eaton and Moore, 2010). Dominant climatic regimes in the province have been linked to synoptic patterns (Stahl *et al.*, 2006). Some synoptic types that characterize patterns of Arctic outflows or Pineapple Express events have been tied to large-scale hydrologic events (Fleming *et al.*, 2007; Eaton and Moore, 2010). Here we evaluate how well the GCMs simulate the frequency of the daily synoptic patterns of sea level pressure in the region. We accomplish this by using a clustering algorithm known as Self-Organizing Maps (SOMs) to identify and classify the characteristic synoptic patterns. SOMs are shown to be a powerful tool for model evaluation, allowing a detailed examination of the differences between simulated and observed atmospheric circulation (e.g. Finnis *et al.*, 2008; Schuenemann and Cassano, 2009). We follow the methodology from

Finnis *et al*. (2008) and Schuenemann and Cassano (2009) and compare sea level pressure (SLP) patterns in 22 GCMs to those in NARR. We aim to find the models that best reproduce the occurrences of NARR synoptic-scale systems (patterns) over our large and small domains.

We use daily SLP from NARR and 21 GCMs (UKMO-HadGEM1 was not included because daily output are not available) for the period 1980–1999, interpolated to 10×NARR grid (approximately 320×320 km in conical conformal Lambert map projection). Prior to the SOM analysis described below, daily SLP anomalies are calculated by subtracting the daily averaged SLP over 1980–1999 from daily SLP at each grid point.

SOM is a common type of unsupervised Artificial Neural Network particularly adept at pattern recognition and classification, and in many respects is analogous to more traditional forms of cluster analysis. Kohonen (2000) offers an explanation of the development and details of the SOM algorithm while Hewitson and Crane (2002) describe the use of SOMs in classifying synoptic patterns in climate data. Our SOM analysis uses the SOM-PAK software, which is available for downloading at http://www.cis.hut.fi/research/som-research (Kohonen *et al*., 1996).

A brief description of SOM algorithm follows: The starting state is an input data set consisting of input vectors, while the final state is a 2-D map (SOM) which consists of nodes that represent characteristic vectors (patterns) of the input data. Initially, a SOM consists of random nodes, where each node has an assigned weight vector (same dimension as input vectors) and an assigned position in the 2-D map space. The procedure for placing a vector onto the map is to find the node with the closest weight vector to the input vector and to assign the map coordinates of this node to the vector. Thus, the steps for creating an SOM are as follows:

1. Randomize the map node weight vectors, W.
2. Select an input vector.
3. Traverse each node in the map and track the node that produces the smallest Euclidean distance between the input vector and the map node weight vector. This node is then called the 'best matching unit'.
4. Update the nodes in the neighbourhood of the best matching unit by shifting them closer to the input vector, for example:

$$W(t + 1) = W(t) + \Theta(t)\alpha(t)(D(t) - W(t)), \quad (5.3)$$

where t is current iteration, D is target input, Θ is restraint due to distance from the best matching unit (usually called the neighbourhood function) and α is learning restraint due to time.

5. Increment t and repeat steps 2 to 5 until t exceeds the chosen limit for time iteration. This is known as the training process.

After the training process, which is repeated several thousand times, individual SOM nodes (or patterns) represent archetypal patterns in the original data, while the full SOM array represents a map of the training data space. The amount of original information retained depends primarily on the size of the SOM, i.e. the number of nodes, with smaller size producing broad generalizations of the input data set, and larger size capturing increasingly fine details. The important feature of an SOM is that the neighbouring nodes (patterns) on the SOM represent similar states, while those that are placed further apart on the map are the most dissimilar.

In our analysis, the input data consist of daily SLP anomalies from the NARR and each GCM separately. Thus, each day's SLP anomaly data represents one input vector and is placed in one row of the input matrix. The first rows have NARR data after which follow the model data in all subsequent rows. A schematic diagram of the SOM training is shown in Figure 5.5. We apply this training on a seasonal basis, producing SLP anomaly patterns that are characteristic for each season (DJF, MAM, JJA, SON). The training is performed independently for the large and small spatial domains. We experimented with different SOM sizes and finally chose three (4×3, 4×4 and 5×4) that provide a reasonable compromise between detail and interpretability. Figure 5.6 illustrates the resulting 4×4 SOM trained with daily SLP anomalies from NARR and one GCM (CCSM3) for winter season over the large and small domains.

Having created SOMs of characteristic SLP anomaly patterns for NARR and each GCM, the next step is to evaluate the model performance. An ideal model would recreate the same synoptic patterns that take place in the real atmosphere, here represented by NARR, and the same frequency of occurrences of each of the SOM nodes. We calculate the node frequency (in %) as the total number of days with a certain seasonal pattern (node) divided by the total number of days for that particular season over the period 1980–1999. Figure 5.7 shows the plots with frequency of occurrences for each of the nodes from the 4x4 SOMs. These plots should be read as follows: each number in Figures 5.7a and 5.7b represents the frequency of the node in the same position as in the SOM in Figure 5.6a. Similarly, each number in Figures 5.7c and 5.7d represents the frequency of the node in the same position as in the SOM in Figure 5.6b. Figures 5.7a and 5.7c show the node frequencies from NARR, for the large and small domains, respectively. Similarly, Figures 5.7b and 5.7d show the node frequencies from the model (CCSM), for the large and small domains, respectively.

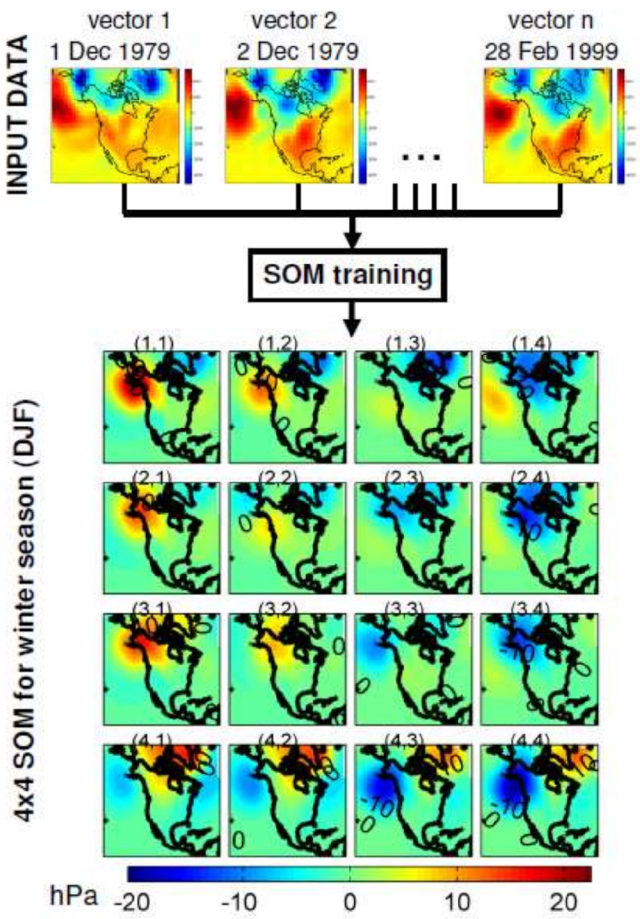

Figure 5.5. Schematic diagram illustrating the SOM training where the input data are winter daily sea level pressure anomalies (from NARR and one GCM) for 1979-1999. 4x4 SOM represents characteristic patterns of sea level pressure anomalies for winter.

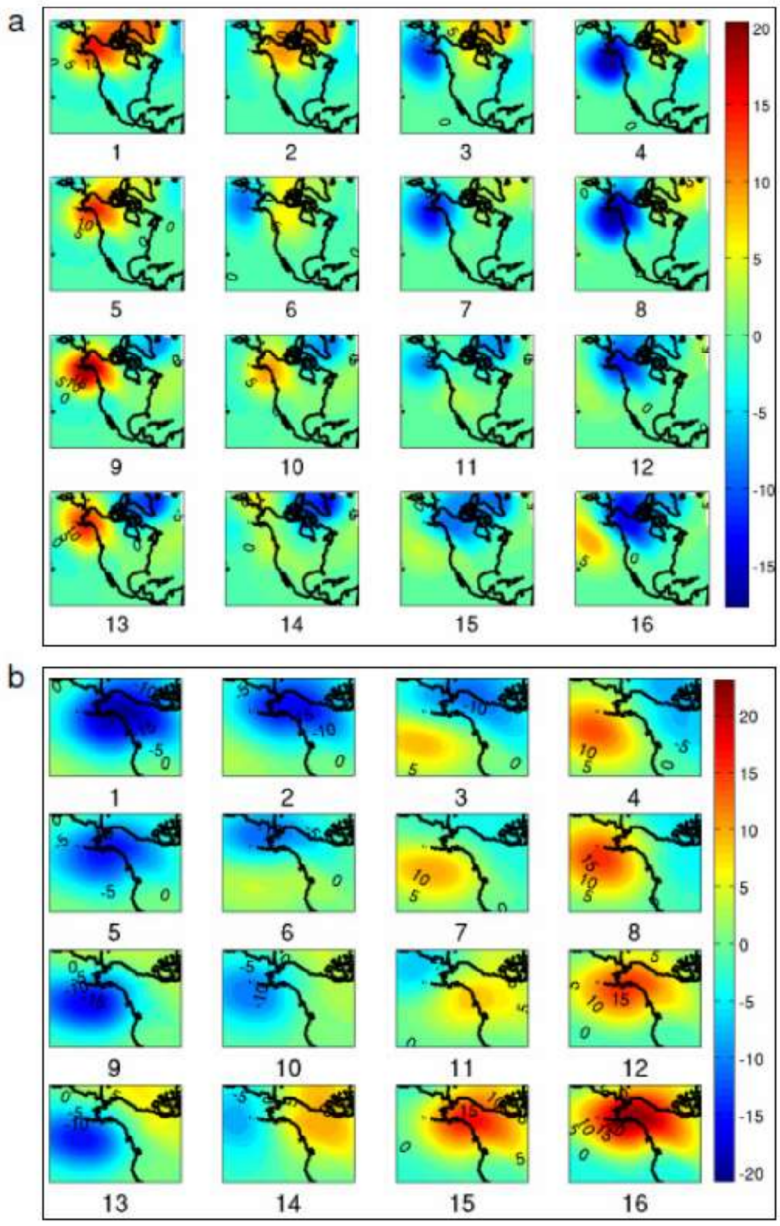

Figure 5.6. 4×4 SOM of winter SLP anomalies (hPa) for the period 1980–1999 trained with NARR and one GCM (CCSM3). (a) Results for the large domain. (b) Results for the small domain. Each numbered node represents one characteristic pattern of daily SLP anomalies.

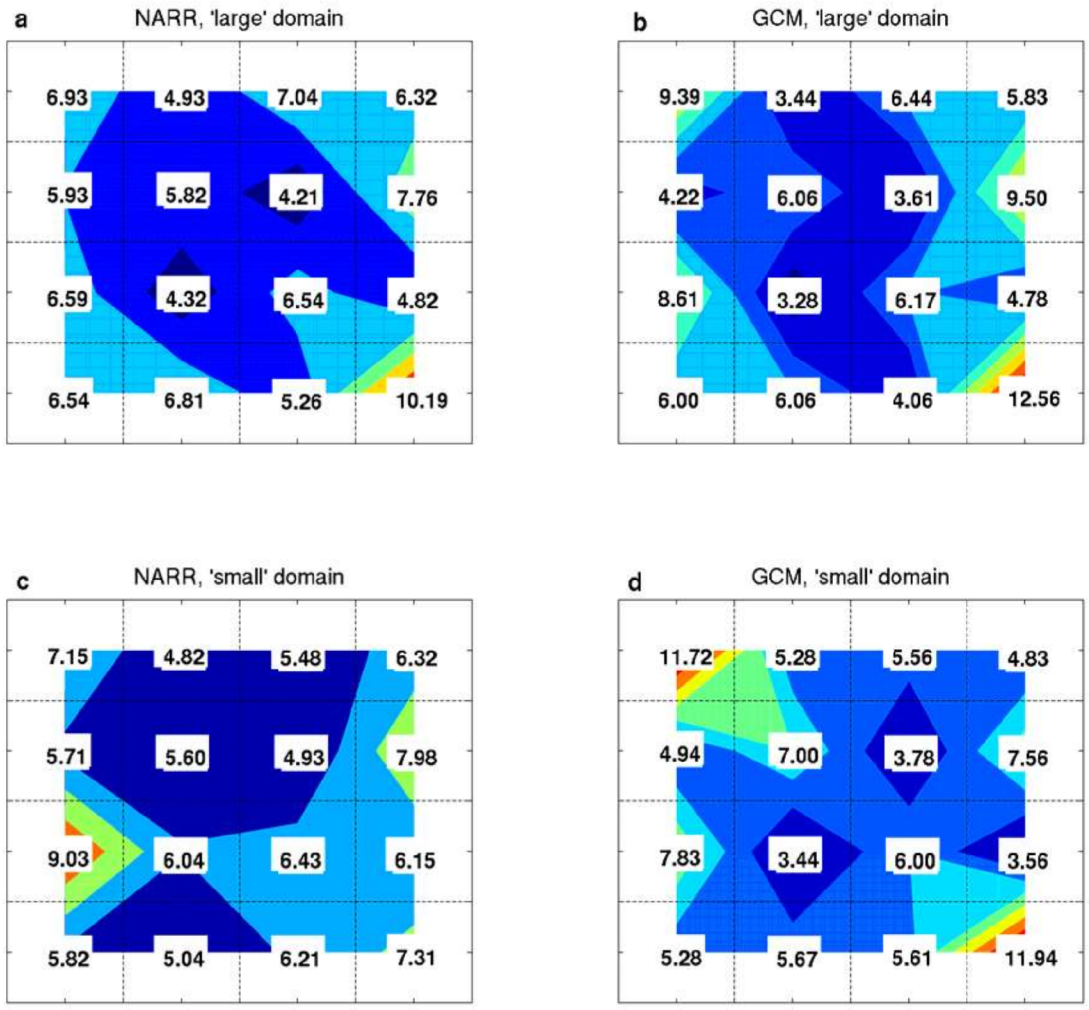

Figure 5.7. Frequency of occurrences (%) of each node in 4×4 SOM for winter season over the period 1980–1999. Darker blue shades indicate smaller values while lighter yellow and red shades indicate larger values. (a) Node frequency of NARR over the large domain. (b) Node frequency of model (CCSM3) over the large domain. (c) Node frequency of NARR over the small domain. (d) Node frequency of model (CCSM3) over the small domain. The corresponding 4×4 SOMs over the large and small domains are shown in Figure 5.6.

The performance of each model can be tested by Pearson correlation coefficient between the reference (NARR) node frequency and each model's node frequency over the same time period (Cassano *et al.*, 2007). We perform this correlation analysis for each season, each SOM size and both spatial domains. The results (r-values) calculated for 4×4 SOM, for all seasons and over both domains are shown in Table 5.2. Values of r that are significantly different from zero (at the 95% confidence level, assessed by a t-

test) are in bold print. These illustrated correlations for 4×4 SOM agree with the results from other SOM sizes (4×3 and 5×4) where the correlation varies widely from one model and season to another, ranging from near perfect correlation to weak inverse relationships. Correlation analysis of frequencies for 4×4 SOM solutions over the large domain shows that the models with correlations significantly different from zero (at the 95% confidence level) in all seasons are CGCM3.1(T63), GFDL-CM2.1, MRI-CGCM2.3.2, and ECHAM5/MPI-OM. The same analysis performed on the small domain reveals that only MIROC3.2(medres) has correlations significantly different from zero in all seasons.

If we measure the success of a model by its ability to achieve correlations that differ significantly from zero, we note that the success of models in reproducing the synoptic patterns depends on the choice of spatial domain. This result is confirmed when the correlation analysis is performed on both domains for the other two SOM sizes (4×3 and 5×4). The correlation analysis also reveals that the success of most models vary with the size of SOM. For example, correlations from BCCR-BCM2.0 calculated over the small domain are significantly different from zero for all seasons in 4×3 SOM, three seasons in 4×4 SOM and two seasons in 5×4 SOM. As expected, the models are most successful for 4×3 SOM because the smallest size of SOM produces the broadest generalizations (patterns) of the input data set. Finally, winter (DJF) is the season with the largest number of significant correlations, i.e. most of the models reproduce well the occurrences of daily synoptic patterns in winter. The season with the least number of model successes is autumn (SON). In Figure 5.8 we illustrate GCM successes by linking color to each model, where each color corresponds to the number of seasons having a correlation significantly different from zero.

Table 5.2. Correlation coefficients, *r*, between node frequencies of 4×4 SOM in NARR and each GCM, on a seasonal basis. Bold type marks correlations significantly different from zero (at the 95% confidence level).

		large domain				small domain			
#	**Model**	**DJF**	**MAM**	**JJA**	**SON**	**DJF**	**MAM**	**JJA**	**SON**
1	BCCR-BCM2.0	**0.61**	*0.49*	**0.68**	*0.44*	**0.54**	0.31	**0.65**	**0.54**
2	CGCM3.1(T47)	**0.60**	*0.35*	**0.77**	**0.68**	**0.55**	**0.58**	**0.56**	*0.49*
3	CGCM3.1(T63)	**0.63**	**0.54**	**0.78**	**0.50**	**0.63**	**0.64**	**0.73**	*0.43*
4	CNRM-CM3	*0.42*	*0.44*	**0.73**	**0.51**	**0.64**	*0.11*	*0.16*	**0.60**
5	CSIRO-Mk3.0	**0.58**	*0.25*	**0.61**	*-0.06*	**0.75**	*0.46*	*0.26*	*-0.19*
6	GFDL-CM2.0	**0.78**	*0.48*	**0.52**	*0.25*	**0.74**	*0.38*	**0.80**	*0.01*
7	GFDL-CM2.1	**0.81**	**0.54**	**0.81**	**0.53**	**0.65**	*0.08*	*0.32*	*0.34*
8	GISS-AOM	*-0.12*	*0.48*	**0.77**	*0.12*	*0.21*	*0.44*	*0.08*	*0.42*
9	GISS-EH	**0.69**	*0.43*	**0.54**	*0.38*	**0.76**	*0.39*	*0.00*	**0.52**
10	GISS-ER	**0.58**	**0.67**	*0.47*	**0.54**	*0.37*	**0.64**	*-0.05*	*0.48*
11	FGOALS-g1.0	**0.50**	**0.76**	*0.25*	*0.35*	**0.51**	*0.40*	*-0.30*	*0.15*
12	INM-CM3.0	**0.68**	*0.39*	**0.67**	*0.26*	**0.80**	**0.61**	*0.10*	*0.20*
13	IPSL-CM4	*0.46*	*0.15*	**0.65**	*-0.31*	**0.66**	*-0.31*	**0.66**	*-0.23*
14	MIROC3.2(hires)	**0.86**	*0.27*	**0.80**	**0.54**	**0.91**	**0.81**	*0.47*	*0.40*
15	MIROC3.2(medres)	*0.40*	**0.79**	**0.72**	**0.51**	**0.63**	**0.78**	**0.72**	**0.62**
16	ECHO-G	**0.75**	*0.41*	**0.67**	**0.74**	**0.72**	*0.03*	**0.61**	**0.59**
17	ECHAM5/MPI-OM	**0.72**	**0.61**	**0.65**	**0.63**	**0.68**	*0.42*	**0.58**	**0.60**
18	MRI-CGCM2.3.2	**0.85**	**0.57**	**0.54**	**0.64**	**0.59**	**0.74**	**0.67**	*0.38*
19	CCSM3	**0.92**	*0.33*	*0.43*	*0.22*	**0.58**	*-0.18*	*0.18*	*-0.02*
20	PCM	*0.29*	**0.72**	*0.24*	**0.63**	**0.52**	**0.57**	*0.23*	*0.17*
21	UKMO-HadCM3	**0.63**	*0.48*	**0.62**	**0.62**	*0.40*	*0.35*	**0.63**	*0.44*