Environmental Science and Engineering
University of Northern British Columbia
Prince George, BC, Canada
Email: stephen.dery@unbc.ca
Tel: +1 (250) 960-5193
Fax: +1 (250) 960-5845

12 January 2017

Dr. Mark F. P. Bierkens
Special Eric Wood Issue
Hydrology and Earth System Sciences

Dear Dr. Bierkens:

Thank you for your decision on our manuscript titled "Evaluating uncertainties in modelling the snow hydrology of the Fraser River Basin, British Columbia, Canada" by Siraj Ul Islam and Stephen J. Déry (Reference # HESS-2016-469). We fully appreciate the detailed and constructive comments provided by the anonymous Referees #1 and #2. We have carefully revised our manuscript by addressing all concerns and comments of the referees' to the best of our ability. Our point-by-point response (in bold italic lettering) to each of the referees' comment/suggestion is enumerated in the attached document.

Please let us know if additional information or points of clarification are required.

Thank you.

Sincerely yours,

Stephen Déry

Referee # 1:

This manuscript provides an assessment of the uncertainties in hydrological modelling over the FRB, BC, Canada. The authors used four precipitation dataset and a calibration routine to quantify and characterize the uncertainties to related to meteorological and parameters estimates. The manuscript is well written and covers an interesting topic, however, I feel some issues have not been properly addressed. I suggest moderate revisions to the manuscript before publication. In general I think some of the procedures could be implemented more consistent to ensure that the conclusions of the paper are better.

We recognize that there are complexities in our procedures that were not thoroughly addressed. In our revision process, we have paid special attention to ensure our results and discussions are more consistent. This is accomplished by means of additional hydrological model simulations and analyses.

Page 11 Line 19-20 The authors want to study the impact of different forcing dataset and their related uncertainties in hydrological simulations. They mention a couple of times that the errors in mountainous precipitation might cause significant discrepancies between dataset and therefore they use four different datasets to study that impact. Although the datasets all have their own resolution, the authors decide to bilinear interpolate all of them to a common grid of 25km without taking into account elevation corrections. I believe this procedure will add to the forcing uncertainty, especially for the coarser resolution since they are downscaled to a resolution without including the high resolution elevation data to correct for orographic effects. I think it would be good if the authors could provide some estimate for the uncertainty added by the bilinear interpolation without conditioning on the elevation profile.

We agree that the elevation correction should be explicitly accounted for in the interpolation methodology to correct for orographic effects. Such correction is highly important when interpolating from coarser to higher spatial resolutions, especially in mountainous terrain (Dodson and Marks, 1997). However, in our interpolation methodology the NARR (32 km) dataset is interpolated from coarse resolution, curvilinear grids to slightly higher (25 km) resolution, rectilinear grids. On the other hand, both the PCIC (6 km) and ANSUPLIN (10 km) datasets are interpolated to a coarser resolution (25 km). Interpolating the NARR dataset from a 32 km to a 25 km resolution does not induce much elevation dependent uncertainties since the change in orography remains minimal between mean elevations at 25 km and 32 km grid resolutions (Figure A). Thus the relationship of atmospheric variables such as air temperature with elevation remains nearly identical at both resolutions. In the revision, we have nonetheless highlighted this source of uncertainty at Page 11, Lines 15-21 and Page 12, Lines 1-2 along with the details of the interpolation process.

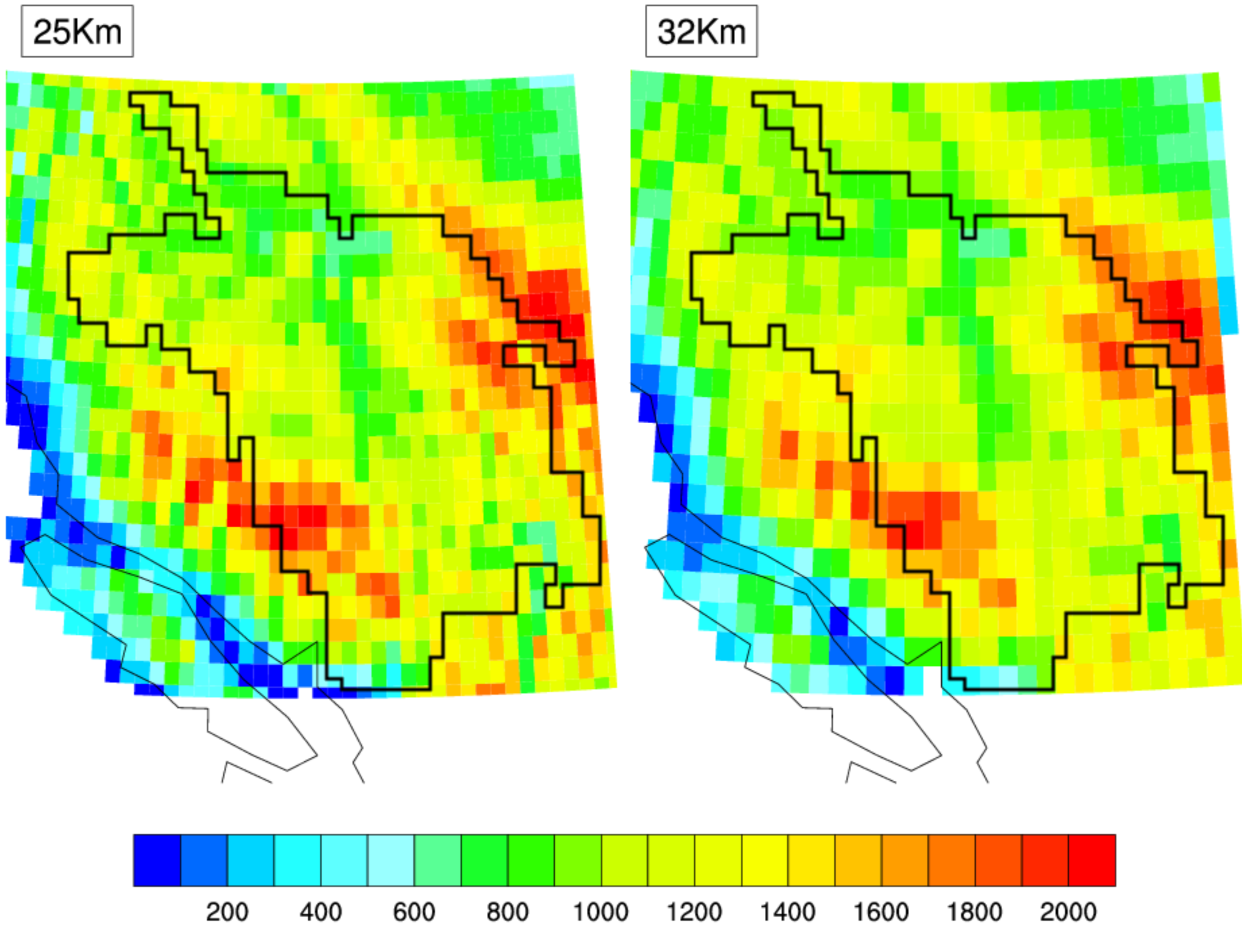


Figure A: Comparison of mean elevation (m) at 25 km and 32 km resolutions over the part of British Columbia covering the Fraser River Basin (black outlines).

Section 2.4.1 Why did the authors select these parameters? Is there sensitivity information that could be used to identify the most sensitive parameters? Maybe the impact of the routing model is more prominent, while it is not calibrated and doesn't account for the reservoirs and lakes present in the FRB. In addition, none of the snow parameters is calibrated, while the authors mention the importance of snow throughout the manuscript. Maybe a calibration on the snow processes (compaction, sublimation or just simple degree day factor), would further benefit the discharge simulation at the outlet and for the sub-catchments.

We selected these parameters for calibration based on the manual calibration experience from previous studies by Nijssen et al. (1997), Su et al. (2005), Shi et al. (2008), Kang et al. (2014, 2016) and Islam et al. (2016), among others. VIC is a physically based hydrologic model that has many (about 20, depending on how the term "parameter" is defined) parameters that must be specified. However, the usual implementation approach involves the calibration of only six soil parameters. Such parameters have the largest effects on the hydrograph shape and are the most sensitive parameters in the water balance components (Nijssen et al. 1997; Su et al. 2005). These parameters must be estimated from observations, via a trial and error procedure that leads to an acceptable match of simulated discharge with observations.

For the snow calibration, we have fixed the value of thresholds for maximum (at which snow can fall) and minimum (at which rain can fall) air temperature as 0.5°C and -0.5°C, respectively. These values were adjusted based on the region's climatology and are kept constant for all the simulations in the global control file. Parameters related to the snow albedo are adjusted using the traditional VIC algorithm based on the US Army Corps of Engineers empirical snow albedo decay curves for transitions from snow accumulation to ablation. We have added this detail in Section 2.3.2 at Page 14, Line 20 and Page 15, Lines 1-13 of the revised paper.

Page 15, why is on the PCIC forcing data used for the optimizer. This almost ensures that the PCIC will have the best performance in the following evaluation sections. It might be more interesting (and more work), to calibrate for every forcing dataset individual and then use these four parameter sets for the validation with every unique forcing dataset leading to sixteen combinations. This also gives you four simulations per forcing dataset as a result of the different calibrations. I know this is some work, but it is feasible. I feel it would lift the quality of the overall uncertainty analysis and thereby better support the conclusions of the paper.

As discussed in our manuscript, the VIC model is quite sensitive to the meteorological forcing data, most notably precipitation. This means that if the forcing data change, the soil parameters that result in reproducing simulated streamflow will change accordingly. Our comparison of hydrological simulations driven by four forcing datasets revealed that PCIC and UW driven VIC simulations are noticeably better with high NSE values and reproduce hydrographs similar to that from observations. We only used PCIC forcing data for VIC model simulations to investigate the uncertainties in the model calibration process. Our primary goal is to evaluate optimizer sensitivity to a unique set of parameter limits. We want to see how the MOCOM optimizer results in different optimized parameters and change the overall simulated hydrograph in the calibration process. Evaluating the same sensitivity for each dataset is beyond the scope of this paper. However, we have repeated our methodology with the UW dataset with new set of parameters limits. Instead of three, we have run five different experiments. As expected, the optimized final values of calibration parameters are different for each set of initial parameter range producing different hydrographs for the main stem Fraser River at Hope, BC. In the revised manuscript, we only used the UW dataset to force VIC model as this dataset along with our VIC model implementation is examined extensively over the FRB in Kang et al, (2014) and (2016). We have updated the text in section 3.3 of our revised manuscript and have modified Figure 8 and Table 3 accordingly. Thank you again for your thoughtful review and for highlighting these difficult but important questions.

Minor comments:

I have discussed the ANUSPLIN acronym with some colleagues over the lunch break. We believe the authors could maybe come up with a better name.

This dataset is based on the statistical models generated by the Australian National University Spline (ANUSPLIN), which are the thin-plate smoothing spline-based climate interpolation algorithms (Hutchinson et al., 2009; Hopkinson et al., 2011). In studies such as Eum et al. (2014), Curry et al. (2016) and Irwin et al. (2017), ANSUPLIN is used as the standard name

for this dataset. We have therefore retained the same name to remain consistent with common usage in the literature.

Page 3 Line 13-14: measurement of the response metric -> objective function and the calibration variable

We have modified this line by replacing "measurement of the response metric" with "objective function and the calibration variable".

Page 7 Line4: Why did the author select the PDO rather than the more influential ENSO signal for Western Canada.

We agree that the FRB is strongly teleconnected to both PDO and ENSO phases. However, we only focused our attention on the PDO teleconnections as the hydrological model calibration is usually performed over many years (minimum of 5 to 10 years) rather than a few individual years. As ENSO teleconnections are short term (a season or two), it is not feasible to evaluate its influence on the total calibration period. We have discussed this issue in detail in response to Referee #2.

Page 11 Line 8: The PGF is not really high temporal resolution, it uses satellite observations, which are corrected with gauges at monthly temporal resolution. I think it would be good to provide the reader with the data sources of the PGF. This is important to understand the performance of the WU dataset.

Thank you for your suggestion. The monthly precipitation data are sourced from the University of Delaware. The gauge undercatch is first applied on the monthly data and are then interpolated to daily values based on the work by Sheffield et al. (2006). We have extended this discussion by including the following detail at Page 10, Lines 20-21 and Page 11, Lines 1-3 in our revised manuscript.

"To improve the precipitation estimates, the monthly data are adjusted to account for gauge undercatch by using the methods outlined by Adam and Lettenmaier (2008). Such adjustment is important since gauge-based precipitation measurements may underestimate solid precipitation in winter by 10%–50% (Adam and Lettenmaier 2003)."

Section 2.2 and 2.4, maybe combine these sections since they both cover VIC and could be easily combined into on VIC section. Otherwise move section 2.3 forward to have the two VIC sections following one another.

We have combined VIC model related discussions as section 2.3 in our revised manuscript.

Page 12 Line 19, maybe remove the number of columns and rows, the domain would be sufficient

We have modified the sentence at Page 14, Line 2 by removing "composed of 34 rows and 42 columns".

Page 13 Line 3 Citation, year could be without the brackets

Correction is made at Page 14, Line 6.

Page 14 Line 18 Why not loop over the year 1979 rather than no spin-up. If the forcing of 1979 were to be recycled for five year and the stabilized ICs could then be used rather than no spin-up. This would ensure that the NARR simulation is more equal to the others and therefore the difference can be really attributed to the difference in forcing rather than a cold model start.

We agree that the VIC model spin up period should be the same for all four forcing datasets. Except for the NARR driven VIC simulations, the PCIC, UW and ANUSPLIN driven simulations are integrated using a five-year spin up period prior to 1979. In our revised manuscript, we have looped recursively the NARR driven simulation for five years using the year 1979 as the forcing data and have re-run the model to ensure our methodology is consistent with the other three sets of simulations. However the results and conclusions of the manuscript do not change as there is little difference in the outcome of the NARR simulations. We have revised the text at Page 16, Lines 16-19 accordingly.

Page 14 Line 18-20 Once calibration. . .(Table 1) -> please clarify. It is not entirely clear what you want to do here.

We have revised the text as follows at Page 16, Lines 19-20.

"After calibration, the model validation runs were initialized with five different state files to produce five ensemble members."

Page 18 Line 9-12 Do you have estimates of the cross-correlation between the precipitation products. Up to what extend are they derived from the same input data.

The cross correlation of the ANUSPLIN, UW and PCIC datasets (being driven by the different number of station observations) could be high but the bias in precipitation magnitude makes these datasets different. This is due to the different methodologies used in their development.

Page 22 Line 15 mentions a VIC sensitivity experiment, it would be great to show some of these results to get a better understanding of the model parameter uncertainty.

We have now included the SWE and hydrograph of the NARR sensitivity experiments as Figure 5 in the supplementary document.

Page 24 Line 13-14 "air temperatures are more crucial for hydrological simulations", I would argue that this is true for the timing, but not so much for the total streamflow volume (at least not for the FRB, where evap is low). Maybe rephrase to "are crucial for the runoff timing in hydrological simulations"

We have modified the text at Page 27, Line 4-5 as suggested.

Figure 2 When does the water year start? October 1st? Please clarify the captions of the figures, they could be more self-explaining

Caption of Figure 2 is revised accordingly.

Figure 7, What is a,b,c,d,e,f? No explanation, also not in the caption of Figure 6

Caption of Figure 7 is revised as:

“Same as Figure 6 but for the FRB’s six major sub-basins (a) Fraser-Shelley (UF), (b) Stuart (SU), (c) Nautley (NA), (d) Quesnel (QU), (e) Chilko (CH) and (f) Thompson-Nicola (TN).

Referee # 2:

The objectives of this study, which focuses on the modelling of snow hydrology of the Fraser River Basin (FRB) of British Columbia (BC), Canada, using the Variable Infiltration Capacity (VIC) model forced with several high-resolution gridded climate datasets (i.e., ANUSPLIN, NARR, UW, and PCIC), were to comprehensively assess uncertainties related to: (i) driving datasets, (ii) optimization of model parameters, and (iii) model calibration during cool and warm phases of the Pacific Decadal Oscillation (PDO).

This is a very well-written paper that deals with relevant subject matters to the hydrological modelling community. The paper delivers sound results with respect to objective (i), however, I believe those associated with objectives (ii) and (iii) are somewhat incomplete. All the results are discussed with respect to mean values. I believe the strength of the paper would be increased if the interannual variability was discussed as well. Moreover, at the end of the paper, the reader is left with an incomplete take-home message; which begs the following question: how could we use the outcomes of the paper if we were asked to run the model to answer questions related to hydroelectric development or other water resources management issue in the FRB? Something is missing here, like a solid recommendation. Here are a few suggestions I can take from the paper: whenever hydrological modelling is performed one should: (i) analyse the stream flow record with respect to previously identified teleconnection correlations? (ii) perform various calibrations in order to find the optimal set of parameter values? (iii) have several snow elevation bands, may be every 100 m? The readers want a clear take-home message from the authors on how to use their findings. Incidentally, the authors made such a reference in the introduction of the paper, but did follow through in their conclusion (last sentence on p. 4, lines 18-20).

We agree with the referee's concerns for objectives (ii) and (iii). We have therefore revised our results and discussion with the set of additional VIC simulations. This is to improve the discussion and the overall take home message of our findings. To address and compare interannual variability of VIC simulations driven by four different datasets, we have analyzed the simulated runoff coefficient of variation (CV) for the Fraser River at Hope, BC for each simulation. We have also expanded the conclusion section of our manuscript with recommendations for applications of hydrological modelling and uncertainties in water resource management.

I have made suggestions in the following list of specific comments below on how to fulfill what I perceive as shortcomings. As a side note, I found a bit difficult the exercise of jumping back and forth between the content of the main manuscript and that of the Supplemental File. Perhaps, the introduction of the content of the Supplement File and the content of Section 3.1 could be transformed into a useful technical note.

To limit number of figures, we have only presented core results in the manuscript and have moved all the extra figures to the supplementary document. We expect that the online version of the article in HESS will provide a direct link to the supplementary document to facilate the reader going back and forth between documents.Further to this, it is important the reader has immediate access to the supplementary results as publishing this as a technical note would require considerable time and effort leading to a separate publication, which would inhibit

understanding of the current effort. As such, we maintain the use of a supplementary document for this paper.

That being said, I strongly encourage the authors to address these comments as I feel the paper could certainly be a significant contribution to the community.

Specific comments:

To my knowledge several researchers have conducted studies related to objective (i) of this paper and, thus, the authors should consult the following references and relate their work accordingly:

- Essou, G.R.C., R. Arsenault, R., F.P. Brissette. 2016. Comparison of climate datasets for lumped hydrological modeling over the continental United States. J.of Hydrology, 537: 334-345.

- Essou, G.R.C., F. Saberly, P. Lucas-Picher, F.P. Brissette, A. Poulin. 2016. Can Precipitation and Temperature from Meteorological Reanalyses Be Used for Hydrological Modeling? J. of Hydrometeorology, DOI: 10.1175/JHM-D-15-0138.1

- Arsenault, R., F.P. Brissette. 2014. Continuous streamflow prediction in ungauged basins: the effects of equifinality and parameter set selection on uncertainty in regionalization approaches. Water Resour. Res., 50 (7): 6135-6153

- Sabarly, F., G. Essou, P. Lucas-Picher, A. Poulin, F. Brissette, 2016: Use of Four Reanalysis Datasets to Assess the Terrestrial Branch of the Water Cycle over Quebec, Canada. Journal of Hydrometeorology, 17, 1447-1466

Thank you for providing a list of useful studies related to our work. In our revised manuscript, we have included them and relate these studies to our results.

To my knowledge, two other studies have dealt with the problem of using calibration and validation series that have different climatological characteristics (objective (iii) of the paper), those of Klemeš (1986. Operational testing of hydrological simulation models. Hydrolog. Sci. Journal 31(1): 13-24) and Seiller et al. (2012. Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions. Hydrology and Earth System Sciences, European Geosciences Union, 16 (4): 1171-1189.).

- In the introductory section of the manuscript, I invite the authors to relate objective (iii) of the study with the content of the aforementioned papers.

All these studies are now cited in the introduction at Page 7, Line 10.

P. 5, line 19: the authors refer to Islam *et al*. (2016), but it is missing from the list of references.

This citation is now included in the list of references.

P. 14, lines 4-6, it is written: « The VIC model calibration is applied to the Fraser River's main stem at Hope, BC and the FRB's major sub-basins, namely the Upper Fraser at Shelley (UF), Stuart (SU), 5 Nautley (NA), Quesnel (QU), Chilko (CH) and Thompson-Nicola (TN) Rivers (Fig. 1a and Supplementary Table 1). »

- Does this mean there is a set of parameter values different for each sub-basin and one for the drainage area between Hope and the outlet of each sub-basin? It is common practice to do so with many distributed models, but I do not see anywhere in the paper the performance of the model with respect to each subbasin outlet. Have I missed something?

In our manuscript, we focused on the Fraser River's main stem at Hope, BC for analysis. However, we have calibrated the VIC model for all the major sub-basins of the FRB using each different forcing datasets and results are already presented in Figure 7 of the manuscript. In response to this comment, we have added a new table (Table A) reporting NSE scores of the model forced with the PCIC data for all the sub-basins in the supplementary document as Supplementary Table 2.

Table A: The Nash–Sutcliffe coefficient of efficiency (NSE) for the PCIC driven VIC calibration (1979–1990) and validation (1991–2006) periods for six major sub-basins of the FRB.

Basins	***1979-1990 Calibration***	***1991-2006 Validation***
	Daily NSE	***Daily NSE***
Fraser-Shelley (UF)	***0.77***	***0.76***
Stuart (SU)	***0.70***	***0.86***
Nautley (NA)	***0.23***	***0.15***
Quesnel (QU)	***0.89***	***0.83***
Chilko (CH)	***0.69***	***0.65***
Thompson- Nicola (TN)	***0.86***	***0.80***

P. 14, lines 16-18: to be consistent, should not the five-year spin up period been applied to all gridded datasets; that is from 1979 to 1985; or loop over at least a couple of years until convergence is achieved and then undertake the calibration exercise?

We agree that the VIC model spin up period should be the same for all four forcing datasets. Except for the NARR driven VIC simulations, the PCIC, UW and ANUSPLIN driven simulations are integrated using a five-year spin up period prior to 1979. In our revised manuscript, we have looped recursively the NARR driven simulation for five years using the

year 1979 as the forcing data and have re-run the model to ensure our methodology is consistent with the other three sets of simulations. We have revised the text at Page 16, Lines 16-19 accordingly.

P. 15, lines 4-11: Why using the PCIC gridded dataset for the optimizer uncertainty runs? Why modifying the parameter space? Why not conducting several optimizer uncertainty runs and look at the various local optima? That is what is usually done. By reducing the parameter space, the authors are restricting, in theory, the number of local optima, yet there could still be several of them in this modified space. At the end we get three sets of local optima, yet there are multiple other sets. This is an interesting twist, but:

- Please provide more details on this calibration strategy. The outcome of this work is incomplete as I believe there should have been multiple calibrations for each experiment in order to fully illustrate the equifinality. Please discuss.

The comparison of hydrological simulations driven by four forcing datasets revealed that the PCIC and UW simulations are noticeably better with high NSE values and reproduce hydrographs similar to that from observations. Therefore, we only used PCIC forcing data for VIC model simulations to investigate objective (ii). We agree that in the literature (e.g., Lindenschmidt et al. 2007; Shen et al. 2008; Sudheer et al. 2011, etc.), different methodologies and several optimizer uncertainty runs are used to address overall uncertainty in calibration parameters without modifying the parameter space. However our primary goal is to evaluate optimizer sensitivity to a unique set of parameters limits. We want to see how the MOCOM optimizer results in different optimized parameters and changes the overall simulated hydrograph during the calibration process. One can see from our analysis that within a broad range of parameter limits, only one particular subset of parameter range is optimally converging and produces more reliable calibration results (Figure 8) with high NSE values (Table 3). In the revision process, we have repeated our methodology with the UW dataset with new set of parameters limits. Instead of three, we have run five different experiments. As expected, the optimized final values of calibration parameters are different for each set of initial parameter range producing different hydrographs for the main stem Fraser River at Hope, BC. In the revised manuscript, we only used the UW dataset to force VIC model as this dataset along with our VIC model implementation is examined extensively over the FRB in Kang et al, (2014) and (2016). We have updated the text in section 3.3 of our revised manuscript and have modified Figure 8 and Table 3 accordingly. The overall conclusion of this analysis is that the automated optimizers used to converge calibration parameters still rely on the hydrologist's experience and some manual adjustment of initial calibration parameter ranges. We have added this discussion in our revised manuscript at Page 28, Lines 9-11.

P. 15, lines 12-21: why not six (6) experiments to fully explore the question (2 cool phases x 3 warm phases)?

- This is an interesting experiment, yet I believe it is incomplete as not all possibilities were explored (see Seiller *et al*., 2012).

- Are not there any other highly-impacting teleconnection that affect the basin hydrology (*e.g.*, ENSO)?

We agree with the referee's suggestion. We have now extended our experiments by adding two additional PDO runs (Table B). As per the availability of forcing data, we can only evaluate PDO phases within the 1950-2006 time period. For each calibration experiment in one particular phase of the PDO, the automated MOCOM optimizer is used to optimize calibration parameters. The NSE is calculated for the calibration and validation periods using the daily observed streamflow data for the Fraser River at Hope, BC. While evaluating additional PDO experiments, our conclusion remains the same, i.e. the calibration is biased toward the cold or warm phase of the PDO and therefore it is necessary to avoid PDO phase shifts in the hydrological model calibration and validation process. In essence, it is better to run multiple calibration experiments within the required time period. We have updated Table 4 and the text at Page 29, Lines 18-20 and Page 30, Lines 1-2 accordingly.

Table B: UW driven VIC PDO runs and their performance metrics (NSE coefficient) in the calibration and validation time period evaluated for Fraser River at Hope, BC.

	Calibration		***Validation***	
	NSE (Time slice)	***PDO Phase (Flows)***	***NSE (Time slice)***	***PDO Phase (Flows)***
PDO1	***0.84 (1981-1990)***	***Warm (low flows)***	***0.84 (1991-2001)***	***Warm (low flows)***
PDO2	***0.84 (1956-1965)***	***Cool (high flows)***	***0.85 (1966-1976)***	***Cool (high flows)***
PDO3	***0.84 (1967-1976)***	***Cool (high flows)***	***0.79 (1977-1987)***	***Warm (low flows)***
PDO4	***0.86 (1977-1987)***	***Warm (low flows)***	***0.80 (1967-1976)***	***Cool (high flows)***
PDO5	***0.89 (1991-2001)***	***Warm (low flows)***	***0.87 (1981-1990)***	***Warm (low flows)***

The FRB is strongly teleconnected to PDO and ENSO phases (Thorne and Woo, 2011). Our VIC implementation has been tested for its response to warm and cool phases of the PDO and ENSO. The model realistically simulates anomalies of runoff, SWE, and snowcover under these climatic conditions. In Figure B (Supplementary Figure 8), the effects of the PDO on annual time series of air temperature, precipitation, SWE_{melt} and runoff are shown for the FRB.

Figures C and D represent the response of VIC simulated snowcover (SC) composite anomalies to warm (Figure C) and cool (Figure D) phases of ENSO. ENSO phases are identified in the observation to estimate composites. Consistent with the findings reported in other studies such as Mantua et al. (1997), Fleming et al. (2007) and Shrestha et al. (2016), the anomalies are below (above) normal in warm (cool) ENSO phases.

In our uncertainty analysis, we only focused our attention on the PDO teleconnections as the hydrological model calibration is usually performed over many years (more than 5 to 10 years)

rather than a few individual years. As ENSO teleconnections are short term (a season or two), it is not feasible to evaluate its influence on the total calibration period.

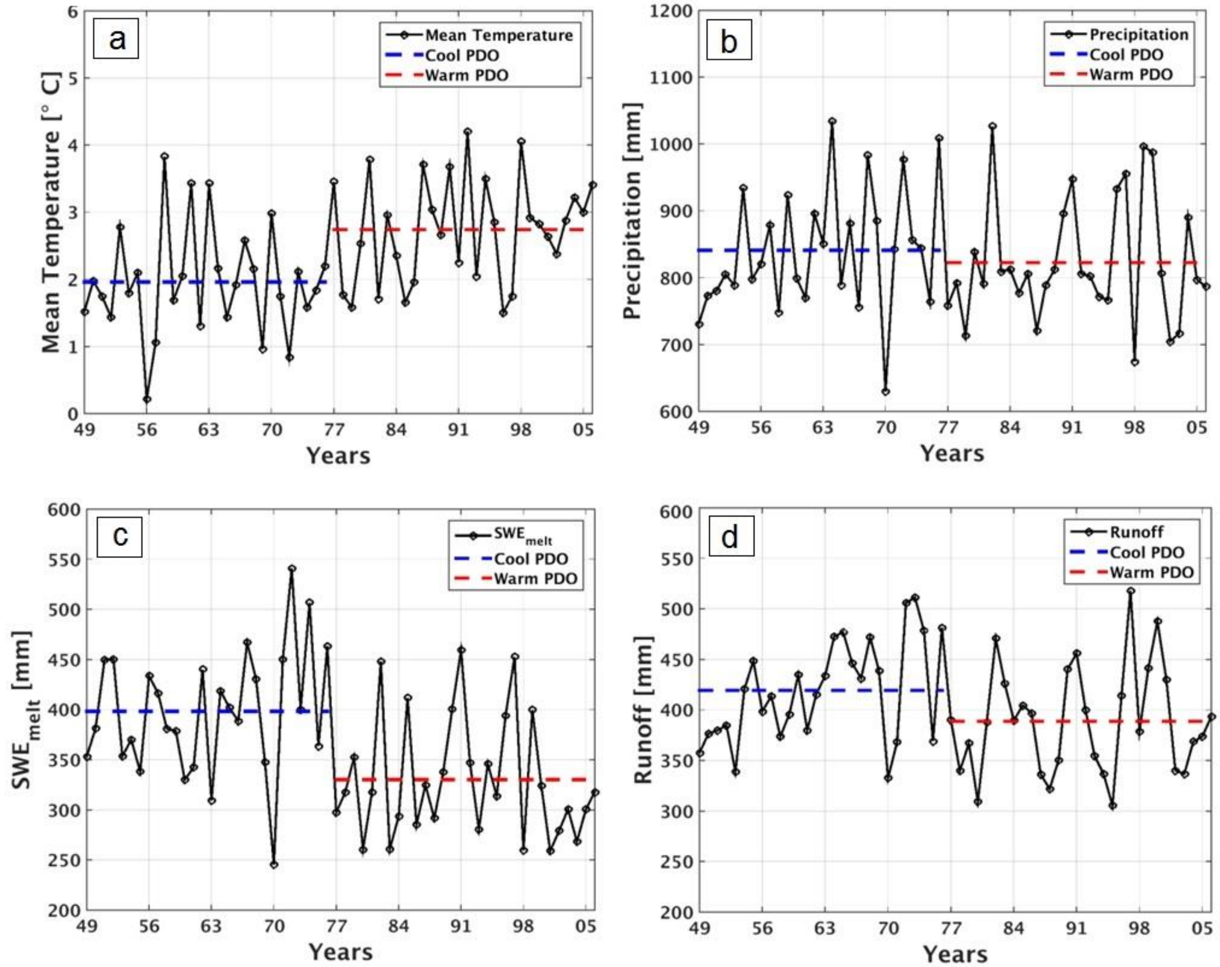


Figure B: Annual variation of mean (a) air temperature, (b) precipitation, (c) SWEmelt and (d) runoff in cool (blue line) and warm (red line) phases of the PDO. Air temperature and precipitation are extracted from the UW forcing dataset whereas SWE$_{melt}$ and runoff are the UW-VIC simulation output. (a), (b) and (c) shows areally-averaged values over all FRB's grid cells whereas runoff is calculated using external routing model for Fraser River at Hope.

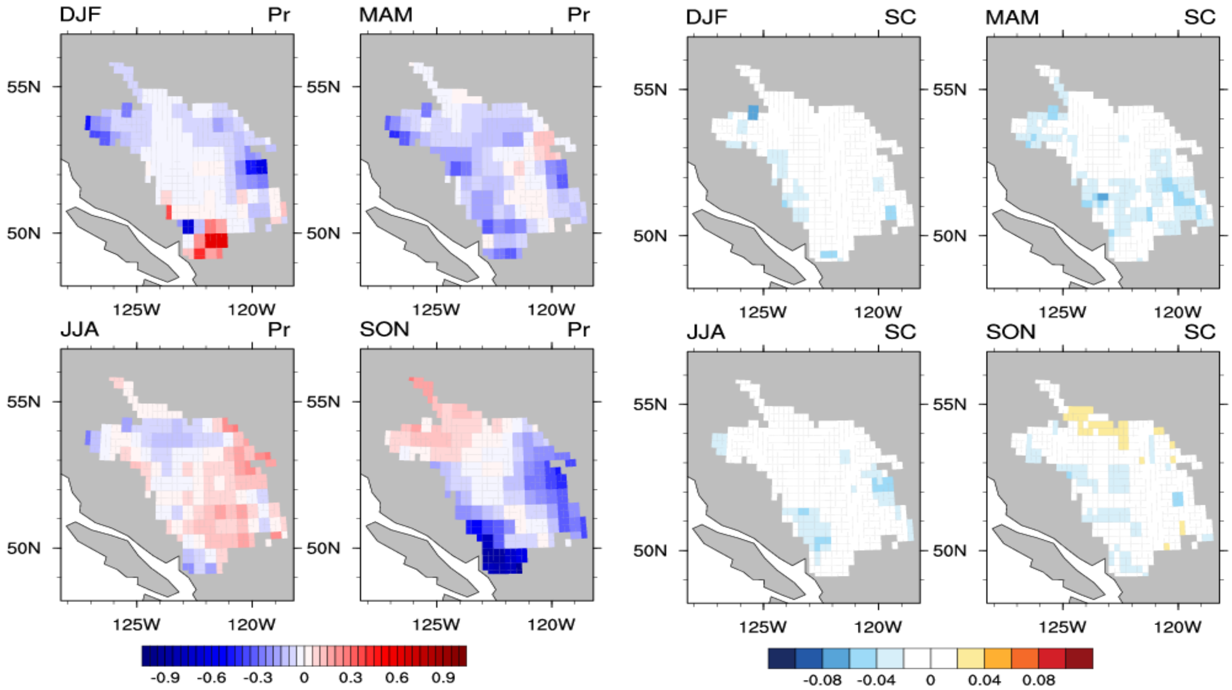


Figure C: Seasonal composite of precipitation (Pr ~ mm day^{-1}) and snowcover (SC ~ fraction) anomalies (1949-2006) in El Niño years. Seasons are based on averaging Dec-Jan-Feb (DJF), Mar-Apr-May (MAM), Jun-Jul-Aug (JJA) and Sep-Oct-Nov (SON) months.

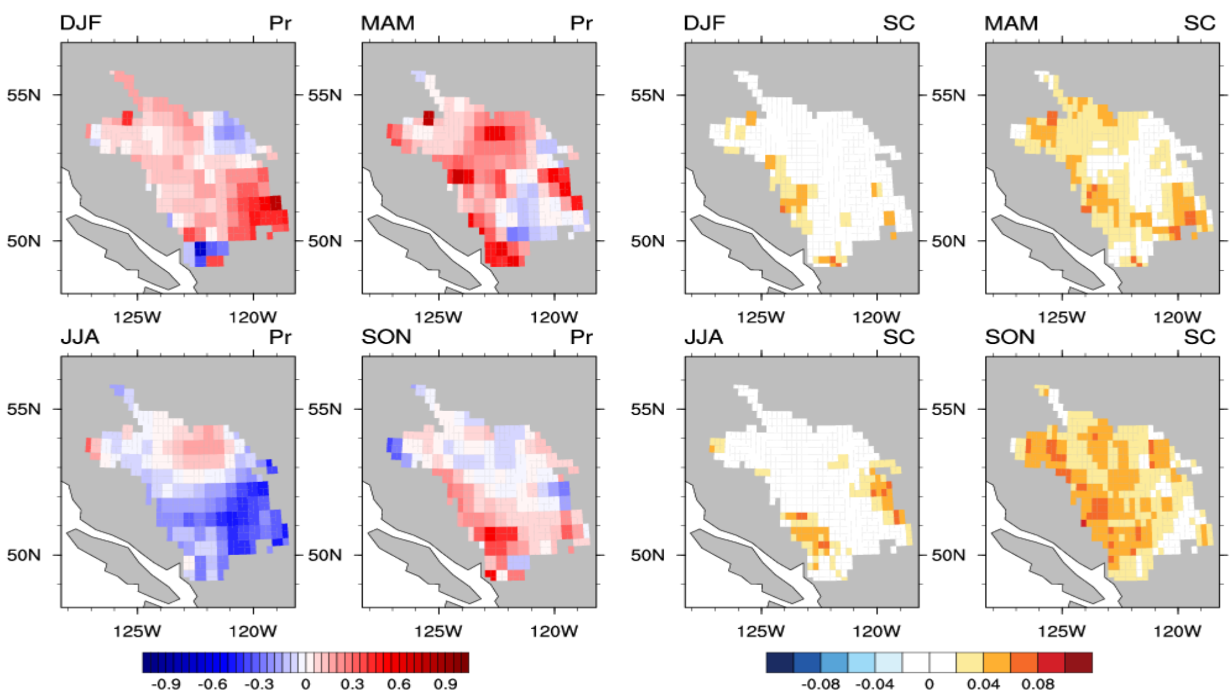


Figure D: Seasonal composite of precipitation (Pr ~ mm day^{-1}) and snowcover (SC ~ fraction) anomalies (1949-2006) in La Niña years. Seasons are based on averaging Dec-Jan-Feb (DJF), Mar-Apr-May (MAM), Jun-Jul-Aug (JJA) and Sep-Oct-Nov (SON) months.

P. 16, lines 2-3: Perhaps the three regions mentioned here should be depicted on the related Figures, that those figures where knowledge of the region is warranted to follow what is presented and discussed in the paper (Figure 4 in the manuscript and Figures 1-3 and 6 in the Supplement File).

The boundaries of three subregions are already shown in Figure 1b representing the model gridcells and mean elevations. Drawing these boundaries on each spatial panel will render the figures much too crowded. For clarity and easier interpretation of the results, we have therefore not drawn these boundaries on each spatial plot.

Given the results introduced in Section 3.1, I think it would have been interesting to have the following calibration parameters: snow-to-rainfall temperature threshold and a vertical temperature gradient.

- Does the VIC model have such parameters? If that is the case, why not calibrating them?
- Is there a publication that has introduced a formal sensitivity analysis of the VIC model?
- Why focusing on soil-related parameters solely and not others as well? In other words, the authors should first introduce to the readers all the model parameters and provide a rationale for their choice.

We selected these parameters for calibration based on the manual calibration experience from previous studies by Nijssen et al. (1997), Su et al. (2005), Shi et al. (2008), Kang et al. (2014, 2016) and Islam et al. (2016), among others. VIC is a physically based hydrologic model that has many (about 20, depending on how the term "parameter" is defined) parameters that must be specified. However, the usual implementation approach involves the calibration of only six soil parameters. Such parameters have the largest effects on the hydrograph shape and are the most sensitive parameters in the water balance components (Nijssen et al. 1997; Su et al. 2005). These parameters must be estimated from observations, via a trial and error procedure that leads to an acceptable match of simulated discharge with observations.

For the snow calibration, we have fixed the value of thresholds for maximum (at which snow can fall) and minimum (at which rain can fall) air temperature as 0.5°C and -0.5°C, respectively. These values were adjusted based on the region's climatology and are kept constant for all the simulations in the global control file. Parameters related to the snow albedo are adjusted using the traditional VIC algorithm based on the US Army Corps of Engineers empirical snow albedo decay curves for transitions from snow accumulation to ablation. We have added this detail in Section 2.3.2 at Page 14, Line 20 and Page 15, Lines 1-13 of the revised paper.

P. 19, lines 12-15: « The lower SWE in the NARR-VIC simulation is probably due to the warmer air temperature during winter and spring (Fig. 2b). Winter temperatures being warmer in the NARR dataset may alter the phase of precipitation partitioning with more rainfall than snowfall, and hence less SWE in the NARR-VIC simulation. »

- This is formulated as a hypothesis, but do not the authors have the input precipitation dataset to validate this hypothesis?

We have validated this result by a VIC model sensitivity experiment where the air temperature was perturbed by 2°C while keeping the precipitation unchanged. The simulated SWE and runoff decreased (a nearly 25% decrease in peak runoff) with 2°C rises in air temperature forcings. The detail of this experiment is already mentioned in the manuscript at Page 25, Lines 1-4. We did not alter the phase of precipitation in this experiment considering that the VIC model automatically partitions the precipitation type (solid versus liquid) during its simulation. We have now included Supplementary Figure 5 to respresent the result of NARR sensitivity run for daily SWE and discharge.

P.20, line 4: what about a sensitivity analysis of the impact of the number of elevation bands on the output variables of interest? Why ten in other words, why not twenty?

The VIC model is mostly configured using 1 to 5 elevation bands in many studies focusing different regions (Haddeland at al. 2002; Shrestha et al. 2012, 2016; Oubeidillah et al. 2014 and many more). Depending on the required research question and the available resources, a suitable number of elevation bands can be generated in the VIC model. In our implementation of the VIC model, we have used 10 elevation bands by utilizing a high resolution digital elevation model. Considering the complex terrain, computational time and 25 km resolution grid, 10 elevation bands are sufficient to investigate elevation dependent changes. Increasing elevation bands to 20 will add substantially more computational time without contributing much to the elevation dependent changes due to model coarse resolution.

P.22, Section 3.2.2: for each gridded dataset – VIC simulation what about calculating a runoff to precipitation coefficient to make inferences about model robustness?

The VIC model robustness is thoroughly evaluated in studies such as Kang et al. (2014, 2016) using the UW forcing data for the period 1949-2006 over the FRB and it subbasins. In this manuscript, we are mainly focusing on the comparison of different datasets. Evaluating the runoff to precipitation coefficient is beyond the scope of this paper. However such analysis is already used for evaluating the UW driven VIC simulation over the FRB. For example, Figure E shows the time series of the ratio of water year runoff to precipitation (R/P). There is clearly a declining trend in R/P over the 1949-2006 period, indicating less runoff productivity as the climate warms across the FRB. This is a result of the transition of precipitation from snowfall to rainfall, and a tendency towards greater evapotranspiration.

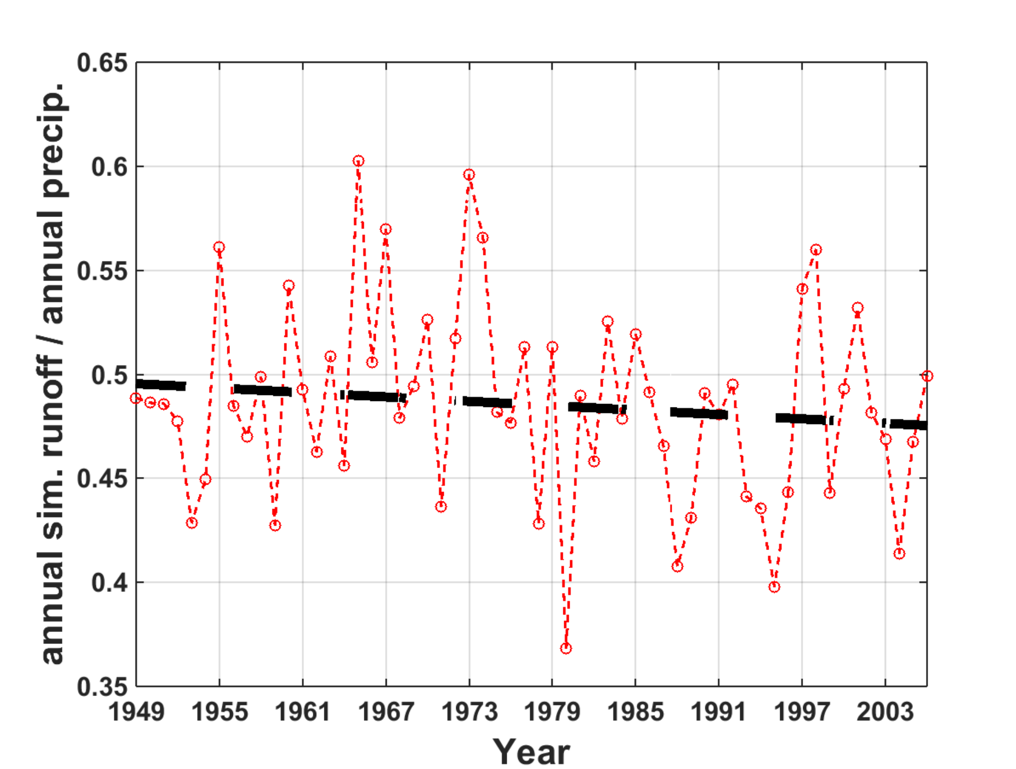


Figure E: Annual ratio of UW driven simulated runoff to precipitation (R/P) in the FRB, water years 1949-2006.

P.24, line 5: « …(Troin *et al*., 2015, 2016)…» Please check as there is only one Troin *et al*. in the list of reference!

Reference for Troin et al. (2016) is now included in the reference list as:

Troin, M., Poulin, A., Baraer, M. and Brissette, F.: Comparing snow models under current and future climates: Uncertainties and implications for hydrological impact studies, J. Hydrol., 540, 588–602, doi:10.1016/j.jhydrol.2016.06.055, 2016.

P.26, lines 17-19: « For each set of calibration experiments, the calibration parameters are different, which affects the air temperature lapse rates and thus evapotranspiration, the formation of the snowpack, and the timing of snowmelt. »

- I thought the temperature lapse rate was not a calibration parameter or is it?

The temperature lapse rate is not a calibration parameter. In the VIC model, mean grid temperature is lapsed to each elevation band. Precipitation falls as snow or rain depending on the lapsed temperature. The model defines the lapse rate based on the elevation and number of elevations bands during its simulation. We have modified this sentence at Page 29, Line 11-12 in our revised manuscript.

P.28, lines 5-6: « The trend analysis further highlighted uncertainties of the NARR Dataset. »

- Uncertainties in terms of what...I am not sure I follow well the authors here. Please provide additional details.

We have revised this sentence at Page 30, Lines 19-20 as "The monthly trend analysis distinguished the NARR dataset by showing decreased trends in air temperature and increased trends in precipitation and its VIC driven runoff."

P29, line 3-4: is not the following sentence « Improving the quality of precipitation data may thus lead to more accurate hydrological responses in the FRB. » contradicting this other sentence « Thus uncertainties in air temperatures are more crucial for hydrological simulations over the FRB rather than those in precipitation» on page , lines 13-14?

- Please help me here!

Thanks for highlighting this issue. We have revised the first sentence at Page 31, Lines 17-19 as:

"While the air temperature plays a dominant role in the hydrological simulations, improving the quality of precipitation data can lead to more accurate hydrological responses in the FRB. Considerable precipitation bias can substantially degrade the model performance."

P.29, line 12: not being familiar with the FRB, I was a bit surprise to learn here that the authors had never mentioned in the description of the basin that there are glaciers in the study area!

In Alberta and British Columbia, glaciers account for an estimated area of 26,700 km^2 (Bolch et al. 2010). There are more than 1000 glaciers in the northern and central Rockies spanning a total area of 838 km^2. However these glaciers contribute relatively little to the annual streamflow to rivers of western Canada. For example, Comeau et al. (2009) studied the glacier contributions to streamflow generation by applying a hydrological model at 9 km resolution to the North and South Saskatchewan Rivers originating in the Canadian Rocky Mountains. They separate the mean annual streamflow contributions from glacier retreat and summer melting of the seasonal snowpack. For the period 1975-1998, they estimate that glacier melt (ice volume losses) made up only 2.0% of mean annual discharge of the Bow River in Calgary.

In our study, the dynamics of glaciers cannot be simulated in the current version of the VIC model as glaciers and their dynamics are not included in the model physics. Nonetheless the effects of glaciers may not change our results significantly as we have used the VIC model on ~25 km grid cell resolution (625 km^2 area per grid cell) in our study. Furthermore, glaciers cover only 1.5% of the FRB (Shrestha et al. 2012) and provide only a modest contribution to streamflow, primarily in late summer (August/early September). Although glacier dynamics are not simulated in the VIC model there are some cells where there is a perennial snowpack. Year after year this water does not melt out. This is a recognized challenge with the VIC model and is dealt with differently in each implementation, ranging from eliminating these cells from analysis (as in our study) to introducing a simple, conceptual representation of glacier mass balance into VIC, modelled using perennial snow in combination with VIC built in snow routines where a portion of VIC grid cells are identified as glacier cells and used to form a glacier mask (Schnorbus et al. 2011). In this study, we compared those cells with perennial snowpack to Baseline Thematic Mapping (BTM) and found that the glaciating cells match the

location of observed glaciers. However there were very few grid cells with a perennial snowpack due to the low resolution and were therefore masked in the analysis. We have included this detail at Page 9, Lines 3-5 and Page 18, Lines 18-20 in the revised manuscript.

P.29, lines 14-15: « Such structural uncertainties need to be evaluated by examining multiple hydrological models over the same region. »

- That is fine to mention that, but this statement does not have anything to do with what was presented in the paper...why bring this issue and others up at this point, on top of it in the conclusion...maybe there should be a subsection untitled recommendations for future work.

We have revised the manuscript by removing these lines.

Figures:

Figures 2, 3, 6 & 7

- What about producing a graph illustrating the daily values of the coefficient of variation so the readers can appreciate the interannual variability for each data set?

As suggested, we have calculated the daily CV of routed runoff for each dataset for the Fraser River at Hope (Figure F). We have added this graph as panel (b) in Figure 6 of the manuscript.

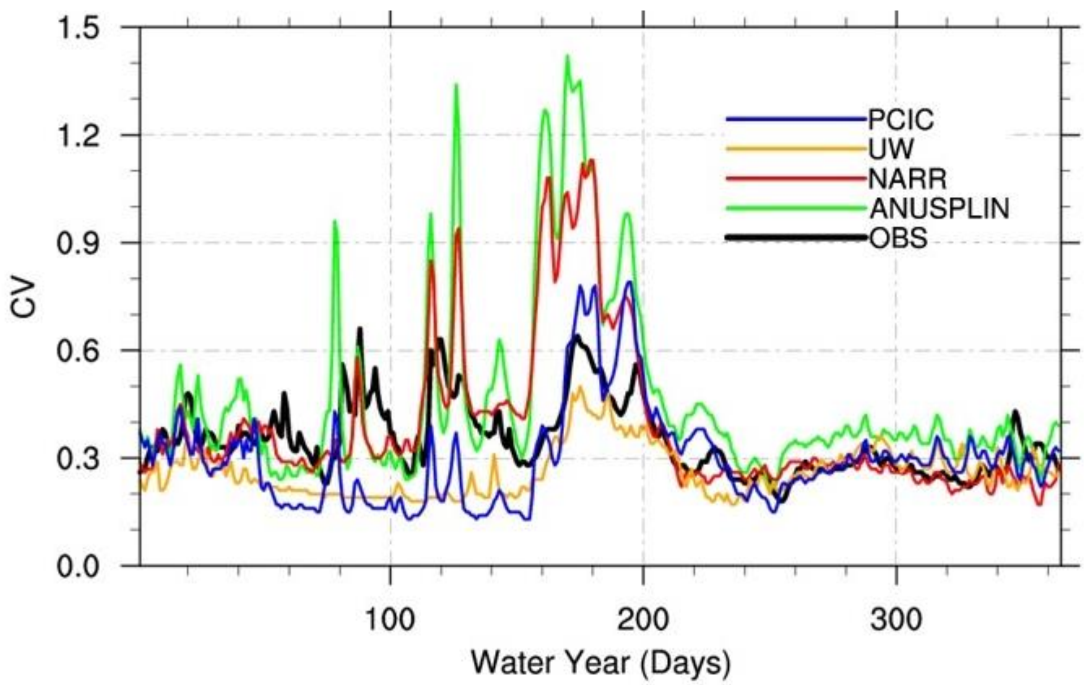


Figure F: The observed and simulated daily runoff coefficient of variation (CV) for Fraser at Hope over water years 1979-2006. An external routing model is used to calculate runoff for the ANUSPLIN-VIC, NARR-VIC, UW-VIC and PCIC-VIC simulations.

Figure 2

- Days with January 1st as Julian day 1? In other words, what are the actual Julian days defining the water Year (1 October to 30 September of the following calendar year)? This comment apply to all relevant graphs introduced in the paper and supplement file.

The water year is defined in the text at Page 18, Line 16. It starts on 1 October and ends on 30 September of the following calendar year. To assist the reader, we have included the start and end dates of the water year in the figure captions (where required).

Figure 5

- Boxplots would have been useful to highlight the interannual variability.

Each panel of Figure 5 represents four different datasets and 10 elevation bands. Including the interannual variability on these plots will make these results more complex by making the different datasets indistinguishable.

Other comments

Other specific comments are all related to verb tense can be found in the annotated manuscript. That is, to improve the reading and understanding I suggest using the past tense instead of the present tense. This may be viewed as an old fashion way of reporting scientific work. Now, I am also open to a strong argument to refute my point of view.

We have revised the manuscript by using past tense as suggested.

References:

Adam, J. C. and Lettenmaier, D. P.: Application of new precipitation and reconstructed streamflow products to streamflow trend attribution in Northern Eurasia, J. Clim., 21(8), 1807–1828, doi:10.1175/2007JCLI1535.1, 2008.

Adam, J. C.: Adjustment of global gridded precipitation for systematic bias, J. Geophys. Res., 108(D9), 1–15, doi:10.1029/2002JD002499, 2003.

Bolch, T., Menounos, B., Wheate. R.: Landsat-based inventory of glaciers in western Canada, 1985-2005. Remote Sensing of Environment 114: 127-137, 2010.

Comeau, L. E. L., Pietroniro, A., Demuth, M. N.: Glacier contribution to the North and South Saskatchewan Rivers. Hydrological Processes 23: 2640-2653, 2009.

Curry, C. L., Tencer, B., Whan, K., Weaver, A. J., Giguère, M. and Wiebe, E.: Searching for Added Value in Simulating Climate Extremes with a High-Resolution Regional Climate Model over Western Canada, Atmosphere-Ocean, 54(4), 364–384, doi:10.1080/07055900.2016.1158146, 2016.

Dodson R., Marks, D.: Daily air temperature interpolated at high spatial resolution over a large mountainous region, Clim. Research, 8:1−20, 1997.

Eum, H. Il, Dibike, Y., Prowse, T. and Bonsal, B.: Inter-comparison of high-resolution gridded climate data sets and their implication on hydrological model simulation over the Athabasca Watershed, Canada, Hydrol. Process., 28(14), 4250–4271, doi:10.1002/hyp.10236, 2014.

Fleming S.W., Whitfield P. H., Moore R. D., Quilty E. J.: Regime dependent streamflow sensitivities to Pacific climate modes cross the Georgia–Puget transboundary ecoregion. Hydrological Processes 21: 3264–3287, 2007.

Haddeland, B., Matheussen, V., Lettenmaier, D. P. :Influence of spatial resolution in a macroscale hydrologic model, Water Resources Research, 38 (7), pp. 1124–1133, 2002.

Hopkinson, R. F., Mckenney, D. W., Milewska, E. J., Hutchinson, M. F., Papadopol, P. and Vincent, L. A.: Impact of aligning climatological day on gridding daily maximum-minimum temperature and precipitation over Canada, J. Appl. Meteorol. Climatol., 50(8), 1654–1665, doi:10.1175/2011JAMC2684.1, 2011.

Hutchinson, M. F., McKenney, D. W., Lawrence, K., Pedlar, J. H., Hopkinson, R. F., Milewska, E. and Papadopol, P.: Development and testing of Canada-wide interpolated spatial models of daily minimum–maximum temperature and precipitation for 1961–2003, J. Appl. Meteorol. Climatol., 48(4), 725–741, doi:10.1175/2008JAMC1979.1, 2009.

Irwin, S., Srivastav, R. K., Simonovic, S. P. and Burn, D. H.: Delineation of precipitation regions using location and atmospheric variables in two Canadian climate regions: the role of attribute selection, Hydrol. Sci. J., 62(2), 191–204, doi:10.1080/02626667.2016.1183776, 2017.

Kang, D. H., Gao, H., Shi, X., Islam, S. U. and Déry, S. J.: Impacts of a Rapidly Declining Mountain Snowpack on Streamflow Timing in Canada's Fraser River Basin, Sci. Rep., 6, 19299, doi:10.1038/srep19299, 2016.

Kang, D. H., Shi, X., Gao, H. and Déry, S. J.: On the changing contribution of snow to the hydrology of the Fraser River Basin, Canada, J. Hydrometeorol., 15(4), 1344–1365, doi:10.1175/JHM-D-13-0120.1, 2014.

Lindenschmidt, K. E., Fleischbein, K., and Baborowski, M.: Structural uncertainty in a river water quality modelling system, Ecol. Model., 204, 289–300, 2007.

Mantua N. J., Hare S. R., Zhang Y., Wallace J. M., Francis R. C.: A Pacific Interdecadal Climate Oscillation with impacts on salmon production. Bulletin of the American Meteorological Society 78: 1069–1079, 1997.

Nijssen, B., Lettenmaier, D. P., Liang, X., Wetzel, S. W., Wood, E. F.: Streamflow simulation for continental-scale river basins. Water Resour. Res., 33, 711–724, 1997.

Oubeidillah, A. A., Kao, S. C., Ashfaq, M., Naz, B. S. and Tootle, G.: A large-scale, high-resolution hydrological model parameter data set for climate change impact assessment for the conterminous US, Hydrol. Earth Syst. Sci., 18(1), 67–84, doi:10.5194/hess-18-67-2014, 2014.

Schnorbus, M., Bennett, K., Werner, A. and Berland, A. J.: Hydrologic Impacts of Climate Change in the Peace, Campbell and Columbia Sub-basins, British Columbia, Canada. Pacific Climate Impacts Consortium, University of Victoria: Victoria, BC, 157 pp., 2011

Shen, Z. Y., Hong, Q., and Yu, H.: Parameter uncertainty analysis of the non-point source pollution in the Daning River watershed of the Three Gorges Reservoir Region, China, Sci. Total Environ., 405, 195–205, 2008.

Shi, X., Wood, A. W. and Lettenmaier, D. P.: How Essential is hydrologic model calibration to seasonal streamflow forecasting?, J. Hydrometeorol., 9(6), 1350–1363, doi:10.1175/2008JHM1001.1, 2008.

Shrestha, R. R., Schnorbus, M. A., Peters, D. L.: Assessment of a hydrologic model's reliability in simulating flow regime alterations in a changing climate, Hydrol. Processes, doi: 10.1002/hyp.10812, 2016.

Shrestha, R. R., Schnorbus, M. A., Werner, A. T., Berland, A. J.: Modeling spatial and temporal variability of hydrologic impacts of climate change in the Fraser River basin, British Columbia, Canada. Hydrol. Processes, 26, 1840–1860, doi:10.1002/hyp.9283, 2012.

Su, F. G., et al.: Streamflow simulations of the terrestrial Arctic domain, J. Geophys. Res.-Atmos., 110(D8), 2005.

Sudheer, K. P., Lakshmi, G., and Chaubey, I.: Application of a pseudo simulator to evaluate the sensitivity of parameters in complex watershed models, Environ. Modell. Softw., 26, 135–143, 2011

Thorne R. and Woo M. K.: Streamflow response to climatic variability in a complex mountainous environment: Fraser River Basin, British Columbia, Canada, Hydrol. Processes 25, 3076–85, 2011.

Evaluating uncertainties in modelling the snow hydrology of the Fraser River Basin, British Columbia, Canada

Siraj Ul Islam and Stephen J. Déry

Environmental Science and Engineering Program,
University of Northern British Columbia,
3333 University Way, Prince George, BC, V2N 4Z9, Canada

Correspondence to: Stephen J. Déry (sdery@unbc.ca)

Abstract. This study evaluates predictive uncertainties in the snow hydrology of the Fraser River Basin (FRB) of British Columbia (BC), Canada, using the Variable Infiltration Capacity (VIC) model forced with several high-resolution gridded climate datasets. These datasets include the Canadian Precipitation Analysis and the thin-plate smoothing splines (ANUSPLIN), the North American Regional Reanalysis (NARR), University of Washington (UW) and Pacific Climate Impacts Consortium (PCIC) gridded products. Uncertainties are evaluated at different stages of the VIC implementation starting with the driving datasets, optimization of model parameters, and model calibration during cool and warm phases of the Pacific Decadal Oscillation (PDO).

The inter-comparison of the forcing datasets (precipitation and air temperature) and their VIC simulations (snow water equivalent (SWE) and runoff) reveal widespread differences over the FRB, especially in mountainous regions. The ANUSPLIN precipitation shows a considerable dry bias in the Rocky Mountains whereas the NARR winter air temperature is 2°C warmer than the other datasets over most of the FRB. In the VIC simulations, the elevation-dependent changes in the maximum SWE (maxSWE) are more prominent at higher elevations of the Rocky Mountains where the PCIC-VIC simulation accumulates too much SWE and ANUSPLIN-VIC yields an underestimation. Additionally, at each elevation range, the day of maxSWE varies 10 to 20 days between the VIC simulations. The snow melting season begins early in the NARR-VIC simulation whereas the PCIC-VIC simulation delays the melting indicating seasonal uncertainty in SWE simulations. When compared with the observed runoff for the Fraser River main stem at Hope, BC, the ANUSPLIN-VIC simulation shows considerable underestimation of runoff throughout the water year owing to reduced precipitation in the ANUSPLIN forcing dataset. The NARR-VIC simulation yields more winter and spring runoff and earlier decline of flows in summer due to a nearly 15-day earlier onset of the FRB springtime snowmelt.

Analysis of the parametric uncertainty in the VIC calibration process shows that the choice of the initial parameter range plays a crucial role in defining the model hydrological response for the FRB. Furthermore, the VIC calibration process is biased toward cool and warm phases of the PDO and the choice of proper calibration and validation time periods is important for the experimental setup. Overall the VIC hydrological response is prominently influenced by the uncertainties involved in the forcing datasets rather than those in its parameters optimization and experimental setups.

Keywords: Uncertainties, hydrological modelling, climate datasets, runoff, calibration, Fraser River

1 Introduction

While advances in computational power and ongoing developments in hydrological modelling have increased the reliability of hydrologic simulations, the issue of adequately addressing the associated uncertainty remains challenging (Liu and Gupta, 2007). There is a growing need for proper estimation of uncertainties associated with hydrological models and the observations required to drive and evaluate their outputs. Hydrological simulations of snow processes and related hydrology depend critically on the input climate forcing datasets, particularly the precipitation and air temperature (Reed et al., 2004; Mote et al., 2005; Tobin et al., 2011). Both of these input forcings regulate the quantity and phase of modelled precipitation and affect the response of simulated snow accumulation and runoff. The model results therefore rely heavily on the quality of these forcings as the uncertainty (measurement errors, etc.) in such data will propagate through all hydrological processes during simulations (Wagener and Gupta, 2005; Anderson et al., 2007; Tapiador et al., 2012). Studies such as Essou et al. (2016a) compared hydrological simulations of different observed datasets over the continental United States (US). They reported that there are significant differences between the datasets, although all the datasets were essentially interpolated from almost the same climate databases. Furthermore ~~in~~ Essou et al. (2016b)~~, they~~ compared the hydrological response of three reanalysis datasets over the US and found precipitation biases in all reanalyses, especially in summer and winter in the southeastern US. The uncertainties in hydrological simulations also arise from the model parameters, its structure and in the objective function and the calibration variable ~~measurement of the response metric~~ that is used for model calibration. Hence the reliability of input forcings along with the capability of the hydrological model and the experimental setup ultimately determine the fate of hydrological variables essential for water resource management.

Several observed gridded climate datasets of precipitation and air temperature (Mesinger et al., 2006; Hopkinson et al., 2011), based on available observational data, post-processing techniques and, in some cases, climate modelling, are currently available over the Canadian landmass to facilitate climate and hydrological simulations. These datasets provide long term gridded precipitation and air temperature records on hourly and daily bases making them especially useful for hydrological simulations, particularly over areas where in situ station densities are low. However, these datasets, being spatially interpolated or assimilated to gridcells, rely mainly on the spatial density of the observational network, which is often quite low in mountainous regions (Rinke et al., 2004). Observational data incorporated into gridded datasets may also contain measurement errors and missing records that translate into the data interpolation and contribute to the overall uncertainty in gridded data products. Such uncertainties are assessed in many studies focusing on the forcing data (Horton et al., 2006; Graham et al., 2007; Kay et al., 2009; Eum et al., 2014).

The quality of hydrological modelling depends on how well a model simulates the regional detail and topographic characteristics of the region, especially in mountainous regions. However, most mountainous regions exhibit higher errors in gridded datasets because they are usually based on an uneven number of stations that are mostly located at lower elevations (Eum et al., 2012). This is true for most large basins in western Canada that exhibit highly variable elevation ranges and strong climatological heterogeneity. One such large basin is British Columbia’s (BC’s) Fraser River Basin (FRB), which is vital for Canada’s environment, economy and cultural identity. Its mountainous snowpack serves as a natural reservoir for cold-season precipitation, providing snowmelt driven flows in summer. Evaluating uncertainties in modelling the FRB's hydrology is crucial for informed decision-making and water resources management. This includes the communication of the uncertainties, propagated into the model predictions, in an appropriate manner to decision makers or stakeholders, thereby allowing confidence in the model results.

Although the currently available gridded datasets (reanalysis and interpolated) over the FRB are derived from observational stations using various interpolation and assimilation techniques, they may still have systematic biases because of their grid resolution, the density of the surface station network used for data assimilation, and the topographic characteristics of the FRB. In the FRB, 23% of the basin exceeds 1500 m in elevation whereas roughly 5% of the in situ meteorological stations surpasses this elevation (Shrestha et al., 2012). Such mismatch between station densities at different elevations makes the precipitation interpolation at higher elevations excessively influenced from the lower elevation stations (Stahl et al., 2006; Rodenhuis et al., 2009; Neilsen et al., 2010). Therefore, despite extensive implementation of hydrologic modelling with single observed forcings (e.g. Shrestha et al., 2012; Kang et al., 2014, 2016, etc.), evaluation of the uncertainties in forcing datasets remains a critical and challenging issue for the FRB. As such, the first step is to evaluate available observation-based forcing datasets for their suitability to be used in hydrological modelling over the FRB.

In Canada, numerous studies have assessed the performance of the hydrologic simulations driven by only one particular driving dataset (Pietroniro et al., 2006; Choi et al., 2009; Bennett et al., 2012; Kang et al., 2014). Sabarly et al. (2016) haveused four reanalysis datasets to assess the terrestrial branch of the water cCycle over Quebec with satisfactory results over 1979-2008. Eum et al. (2014) recently compared hydrological simulations driven by several high-resolution gridded climate datasets over western Canada's Athabasca watershed and found significant differences across the simulations. While BC's snowpacks and hydrology are well studied in the literature (Danard and Murty, 1994; Choi et al., 2010; Thorne and Woo, 2011; Déry et al., 2012; Shrestha et al., 2012; Kang et al., 2014, 2016; Islam et al., ~~2016~~2017; Trubilowicz et al., 2016), detailed inter-comparisons of available observational forcing in terms of their hydrological response is not thoroughly analysed, particularly over the FRB's complex topography. In this study, we therefore investigate the simulated hydrological response of uncertainties associated with air temperature and precipitation forcing on the FRB's mountainous snowpack and runoff. To achieve this, four forcing datasets, namely the Canadian Precipitation Analysis and the thin-plate smoothing splines (ANUSPLIN hereafter; Hopkinson et al., 2011), the North American Regional Reanalysis (NARR hereafter; Mesinger et al., 2006), University of Washington (UW hereafter; Shi et al., 2013) and Pacific Climate Impacts Consortium (PCIC hereafter; Shrestha et al., 2012) gridded products are applied to the FRB. These datasets are explored across three different regions and multiple elevation ranges. ~~Over the FRB region, while~~ T~~t~~he PCIC and UW datasets are ~~individually~~ used by~~in~~ Shrestha et al. (2012) and Kang et al. (2014, 2016), respectively to drive the VIC hydrological model over the FRB whereas~~,~~ t~~t~~he NARR and ANUSPLIN datasets are not yet evaluated over this region. However, the NARR dataset is used in studies focusing on other regions of Canada (Woo and Thorne, 2006; Choi et al, 2009; Ainslie and Jackson, 2010; Eum et al., 2014; Trubilowicz et al., 2016). To our knowledge, this is the first comprehensive study that collectively examines the spatial and elevation dependent hydrological response of these datasets for the FRB.~~these datasets collectively based on the spatial and elevation dependent hydrological response~~.

Along with forcing datasets, many studies have focused their attention either on model structure (Wilby and Harris, 2006; Jiang et al., 2007; Ludwig et al., 2009; Poulin et al., 2011; Velazquez et al., 2013) or on calibration parameters (Teutschbein et al., 2011; Bennett et al., 2012). Arsenault et al. (2014) estimated the uncertainty due to parameter set selection using the hydrological model over theseveral basins in Quebec. They showed that parameter set selection can play an important role in model implementation and predicted flows. For parameter uncertainty, a hydrological model can have many equivalent local optima within a realistic parameter space (Poulin et al., 2011). Therefore, several different parameter sets may be available for the same ''optimal'' measure of efficiency during the optimization process (i.e. parameter non-uniqueness; Beven, 2006). Here we evaluate the parameter uncertainties involved in the model calibration process, i.e. calibration optimizer sensitivity to parameter initial limits. Moreover we focus on another unique aspect of modelling uncertainty related to the selection of time periods for model calibration and validation under changing climatic conditions on decadal time scales. Studies such Klemeš (1986) and Seiller et al. (2012) highlighted the issue of calibration and validation of hydrological modelling under different climatological conditions. In this study, we estimate In this case, the hydrological model sensitivity to different climatological conditions by first analysingfocusing on the FRB's air temperature and precipitation teleconnections with cool and warm phases of the Pacific Decadal Oscillation (PDO).

Overall, the main goals of this study are: (i) to compare and identify the most reliable available gridded forcing datasets for the hydrological simulations over the FRB region; (ii) to evaluate hydrological modelling responses of different driving datasets over a range of FRB elevations; (iii) to assess the uncertainty involved in the model calibration process by focusing on the optimizer used for parameter optimization; and (iv) to evaluate the calibration process under changing climatic conditions.

To ~~analyse the FRB's snowmelt-driven hydrology~~achieve these four objectives, the macroscale Variable Infiltration Capacity (VIC) hydrological model (Liang et al., 1994, 1996) is used as the simulation tool. The VIC model conserves surface water and energy balances for large-scale watersheds such as the FRB (Cherkauer et al., 2003). It has been successfully implemented, calibrated and evaluated over the FRB (Shrestha et al., 2012; Kang et al., 2014; Islam et al., ~~2016~~2017).

Formatted: Don't adjust space between Latin and Asian text, Don't adjust space between Asian text and numbers

The remainder of this paper is structured as follows. Section 2 discusses the FRB, ~~the VIC model,~~ the driving datasets, the VIC model and experimental setup. Section 3 describes the forcings inter-comparison, hydrological simulations, parameter sensitivity and uncertainly related to the PDO. Section 4 summarizes and concludes this study.

2. Study Area, Model and Methodology

2.1 Fraser River Basin (FRB)

The FRB is one of the largest basins of western North America spanning 240,000 km^2 of diverse landscapes with elevations varying from sea level to ~~Mount Robson, its tallest peak at~~ 3954 m above sea level at Mt. Robson, its tallest peak (Benke and Cushing, 2005). It covers the mountainous terrain of the Coast and Rocky Mountains along with dry central plateaus (Fig. 1). The FRB's headwaters are in the Rocky Mountains with its major tributaries the Stuart, Nechako, Quesnel, Chilcotin, Thompson, and Harrison Rivers. The Fraser River runs 1400 km through the ~~Fraser Canyon~~whole basin before reaching

Hope, BC, where it veers westward to drain into the Salish Sea and the Strait of Georgia at Vancouver, BC (Benke and Cushing, 2005; Schnorbus et al., 2010).

In winter, considerable amounts of snow ~~normally~~ usually accumulate at higher elevations, except in coastal areas. In late spring and early summer, snowmelt from higher elevations induce peak flows in the main stem of the Fraser River and its many tributaries (Moore and Wondzell, 2005), which rapidly decline in late summer following the depletion of snowmelt. Owing to its complex mountainous ranges, the FRB's hydrologic response varies considerably across the basin, differentiating it into snow-dominant, hybrid (rain and snow), or rain-dominant regimes (Wade et al., 2001). The glaciers cover only 1.5% of the FRB (Shrestha et al., 2012) and provide only a modest contribution to streamflow, primarily in late summer (August/early September).

~~2.2 Variable Infiltration Capacity (VIC) Model~~

~~The VIC model resolves energy and water balances and therefore requires a large number of parameters, including soil, vegetation, elevation, and daily meteorological forcings, at each gridcell. To evaluate hydrological responses over complex terrain, the model simulates the subgrid variability in topography and precipitation by dividing each gridcell into a number of snow elevation bands (Nijssen et al., 2001b). The model utilizes a mosaic-type representation by partitioning elevation bands into a number of topographic tiles that are based on high-resolution spatial elevations and fractional area. The snow model embedded in the VIC model is then applied to each elevation tile separately (Gao et al., 2009).~~

~~The VIC model is widely used in many hydrological applications including water availability estimation and climate change impacts assessment in North America (Maurer et al., 2002; Christensen and Lettenmaier, 2007; Adam et al., 2009; Cuo et al., 2009; Elsner et al., 2010; Gao et al., 2010; Wen et al., 2011; Oubeidillah et al., 2014;) and around the world (Nijssen et al., 2001a,b; Haddeland et al., 2007; Zhou et al., 2016). It is also commonly used to simulate hydrologic responses in snowmelt-dominated basins (Christensen and Lettenmaier, 2007; Hidalgo et al., 2009; Cherkauer and Sinha, 2010; Schnorbus et al., 2011).~~

2.~~3~~ 2 Datasets

Along with recent developments in hydrological models, several observation-based gridded datasets are now available to drive the models such as ANUSPLIN, NARR, UW and PCIC. These meteorological forcing datasets are developed using high-resolution, state-of-the-art data interpolation and (for NARR only) assimilation techniques. This is to improve the quality of forcing data to analyse a model's hydrological response over ~~of~~ any particular basin.

The ANUSPLIN dataset, developed by Natural Resources Canada (NRCan), contains gridded data of daily maximum and minimum air temperature (°C), and total daily precipitation (mm) for the Canadian landmass south of 60°N at ~10 km resolution (NRCan, 2014). This Canadian dataset uses a trivariate thin-plate smoothing spline technique referred to as ANUSPLIN (Hutchinson et al., 2009) with recent modifications (Hopkinson et al., 2011). Eum et al. (2014) used the ANUSPLIN dataset for hydrological modelling over Alberta's Athabasca watershed and reported underestimations in simulated runoff, owing to a dry bias in ANUSPLIN precipitation.

NARR was developed at 32 km spatial and 3-hourly temporal resolution to improve the National Centers for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) global reanalysis data by employing the Eta Data Assimilation system for the North American domain for the period from 1979 to the current year. The interannual variability of the NARR seasonal precipitation and accuracy of its temperature and winds are found superior to earlier versions of the NCEP/NCAR reanalysis datasets (Mesinger et al., 2006; Nigam and Ruiz-Barradas, 2006). Choi et al. (2009) investigated the applicability of air temperature and precipitation data from NARR for hydrological modelling of selected watersheds in northern Manitoba. They found that NARR air temperature and precipitation data are in much better agreement with observations than the NCEP–NCAR Global Reanalysis-1 dataset (Kalnay et al., 1996; Kistler et al., 2001) dataset. Woo and Thorne (2006) used air temperature precipitation and precipitation air temperature data from two global reanalysis datasets and from NARR as input to a hydrological model for the Liard River Basin in western subarctic Canada and reported significant improvement in its hydrological simulations. NARR output has also been used in regional water budget calculations (Luo et al., 2007; Ruane, 2010; Sheffield et al., 2012). Choi et al. (2009) and Keshta and Elshorbagy (2011) reported that NARR output isare suitable for hydrologic purposesmodelling, especially when other observations are unavailable. However, they focused on the Canadian Prairies, where the topography is not complex.

The UW dataset of daily precipitation, maximum and minimum air temperature, and average wind speed are based on the extended gridded UW dataset (Shi et al., 2013; Adam et al., 2006; Adam and Lettenmaier, 2008). Monthly precipitation originates from the University of Delaware observed land surface precipitation product (Matsuura and Willmott, 2009), which was adjusted to account for gauge undercatch and was converted to daily data using the high temporal precipitation dataset from Sheffield et al. (2006). To improve the precipitation estimates, the monthly data were adjusted to account for gauge undercatch by using the methods outlined by Adam and Lettenmaier (2008). Such adjustment is important since gauge-based precipitation measurements may underestimate solid precipitation in winter by 10%–50% (Adam and Lettenmaier 2003). Daily wind speeds are extracted from the NCEP/NCAR reanalysis datasets (Kalnay et al., 1996).

The PCIC dataset of precipitation, maximum and minimum temperature, and wind speed was derived primarily from Environment and Climate Change Canada (ECCC) climate station observations, with additional inputs from the United States Co-operative Station Network, the BC Ministry of Forests, Lands and Natural Resource Operations, the BC Ministry of Environment's Automated Snow Pillow network, and BC Hydro's climate network (Schnorbus et al., 2011; Shrestha et al., 2012). These data are available at ~6 km resolution and were corrected for point precipitation biases and elevation effects (Schnorbus et al., 2011).

The ANUSPLIN, NARR, UW and PCIC datasets are available at 10 km, 32 km, 25 km and 6 km spatial resolution, respectively, and at a daily time scale. To facilitate comparison, the ANUSPLIN, NARR and PCIC datasets are were regridded to 25 km resolution using bilinear interpolation to match the scale of the current VIC implementation. The NARR (32 km) dataset iswas interpolated from coarse resolution curvilinear grids to slightly higher (25 km) resolution rectilinear grids. On the other hand, both the PCIC (6 km) and ANSUPLIN (10 km) datasets arewere interpolated to a coarser resolution (25 km). The elevation correction, which is important when interpolating from coarser to higher spatial resolutions (Dodson and Marks, 1997), iswas not used to correct the orographic effects for the NARR dataset. Interpolating the NARR dataset from a 32 km to a 25 km resolution does not induce much elevation dependent uncertainties since the change in orography remains minimal between mean elevations at 25 km and 32 km grid resolutions. Thus the relationship of atmospheric variables such as air temperature with elevation remains nearly identical at both resolutions.

~~The d~~Daily wind speeds ~~variable, which is one of the~~a required VIC input variable~~ forcing~~, are~~ is~~ not available for the ANUSPLIN dataset. We therefore used the PCIC based wind speeds in the ANUSPLIN driven VIC simulations. The PCIC wind speeds are sourced from the Environment~~ Canada~~ and Climate Change Canada station product (Schnorbus et al., 2011).

To calibrate and validate the VIC model simulated flows, we used daily streamflow data from ECCC's Hydrometric Dataset (HYDAT; Water Survey of Canada 2014). These data ~~are~~ were extracted and compiled into a comprehensive streamflow dataset for the FRB spanning 1911–2010 (Déry et al., 2012).

In addition, we compared the simulated SWE with observations from the BC River Forecast Centre's network of snow pillow sites (BC Ministry of Forests, Lands and Natural Resource Operations, 2014). The snow pillow stations record the weight of the accumulated snowpack (SWE) on a daily basis. Based on the availability of data, we used SWE observations from four sites located at Yellowhead (ID: 1A01P) and McBride (ID: 1A02P) in the upper Fraser and at Mission Ridge (ID: 1C18P) and Boss Mountain Mine (ID: 1C20P) in the middle Fraser. Due to data~~the~~ availability~~ of data~~, we used the 1996-2006 time period for the Yellowhead, Mission Ridge and Boss Mountain Mine snow pillows and 1980-1986 for the McBride location. Detailed information about these sites is available in Kang et al. (2014) and Déry et al. (2014).

2.~~2~~3 Variable Infiltration Capacity (VIC) Model

The VIC model resolves energy and water balances and therefore requires a large number of parameters, including soil, vegetation, elevation, and daily meteorological forcings, at each gridcell. To evaluate hydrological responses over complex terrain, the model simulates the subgrid variability in topography and precipitation by dividing each gridcell into a number of snow elevation bands (Nijssen et al., 2001a~~b~~). The model utilizes a mosaic-type representation by partitioning elevation bands into a number of topographic tiles that are based on high-resolution spatial elevations and fractional area. The snow model embedded in the VIC model is then applied to each elevation tile separately (Gao et al., 2009).

The VIC model is widely used in many hydrological applications including water availability estimation and climate change impacts assessment in North America (Maurer et al., 2002; Christensen and Lettenmaier, 2007; Adam et al., 2009; Cuo et al., 2009; Elsner et al., 2010; Gao et al., 2010; Wen et al., 2011; Oubeidillah et al., 2014~~;~~) and around the world (Nijssen et al., 2001a,b; Haddeland et al., 2007; Zhou et al., 2016). It is also commonly used to simulate hydrologic responses in snowmelt-dominated basins (Christensen and Lettenmaier, 2007; Hidalgo et al., 2009; Cherkauer and Sinha, 2010; Schnorbus et al., 2011).

2.~~4~~ 3.1 The VIC Implementation

The VIC model, as set up by Kang et al. (2014) and Islam et al. (~~2016~~2017) for the FRB, is employed for evaluating the model's ability to simulate the FRB's hydrological response when driven by different observational forcings. The model was previously applied to the FRB to investigate its observed and projected changes in snowpacks and runoff. In this study, we performed model integrations over the entire FRB using gridcells ~~composed of 34 rows and 42 columns~~ spanning 48°– 55°N and 119°–131°W. The model is configured at 0.25° spatial resolution using a daily time step, three soil layer depths and ten vertical snow elevation bands. Once an individual VIC simulation is completed, the runoff for the basin is extracted at an outlet point of the given sub-basin, using an external routing model that simulated a channel network (adapted from Wu et al., ~~(~~2011~~)~~) with several nodes (Lohmann et al., 1996; 1998a, b). Streamflow is converted to areal runoff by dividing it by the corresponding sub-basin area. Daily runoff at the outlet cell is integrated over time to obtain total water year~~annual~~ runoff for a selected basin. Other than the calibration parameters, the soil and vegetation parameters, leaf area index (LAI) and albedo data are kept identical as per the Kang et al. (2014)~~the~~ VIC model implementation to the FRB ~~reported in Kang et al. (2014)~~.

2.~~4~~3.~~1~~ 2 Calibration

To explore the feasible parameter space, we used the University of Arizona multi-objective complex evolution (MOCOM-UA) optimizer for the VIC calibration process (Yapo et al., 1998; Shi et al., 2008). MOCOM-UA searches a set of VIC input parameters using the population method to maximize the Nash–Sutcliffe efficiency (NSE) coefficient (Nash and Sutcliffe, 1970) between observed and simulated runoff. Six soil parameters are used in the optimization process, i.e. b_infilt (a parameter of the variable infiltration curve), Dsmax (the maximum velocity of base flow for each gridcell), Ws (the fraction of maximum soil moisture where nonlinear base flow occurs), D2 and D3 (the depths of the second and third soil layers), and Ds (the fraction of the Dsmax parameter at which nonlinear base~~-~~flow occurs).

These calibration parameters we~~a~~re selected based on the manual calibration experience from previous studies by Nijssen et al. (1997), Su et al. (2005), Shi et al. (2008), Kang et al. (2014, 2016) and Islam et al. (201~~6~~7). VIC is a physically based hydrologic model that has many (about 20, depending on how the term "parameter" is defined) parameters that must be specified. However, the usual implementation approach involves the calibration of only these six soil parameters. Such parameters have the largest effects on the hydrograph shape and are the most sensitive parameters in the water balance components (Nijssen et al., 1997; Su et al., 2005). These parameters must be estimated from observations, via a trial and error procedure that leads to an acceptable match of simulated discharge with observations.

For the snow calibration, the value of thresholds for maximum (at which snow can fall) and minimum (at which rain can fall) air temperature were fixed as 0.5°C and -0.5°C, respectively. These values were adjusted based on the region's climatology and were kept constant for all simulations in the global control file. Parameters related to the snow albedo were adjusted using the traditional VIC algorithm based on the US Army Corps of Engineers empirical snow albedo decay curves for transitions from snow accumulation to ablation.

Final values of these six calibrated parameters ~~are~~ were estimated for each forcing dataset by a number of simulation iterations minimizing the difference between the simulated and observed monthly flow.

While the MOCOM-UA automated optimization process utilizes monthly streamflow ~~in~~ during calibration, we evaluated the overall model performance on daily time scales using NSE and correlation performance metrics.

The VIC model calibration is applied to the Fraser River's main stem at Hope, BC and the FRB's major sub-basins, namely the Upper Fraser at Shelley (UF), Stuart (SU), Nautley (NA), Quesnel (QU), Chilko (CH) and Thompson-Nicola (TN) ~~rRivers~~basins (Fig. 1a and Supplementary Table 1). These sub-basins contribute 75% of the annual observed Fraser River discharge at Hope, BC with the largest contributions from the TN, UF and QU sub-basins (Déry et al., 2012).

2.~~4~~3.~~2~~ 3 Experiments

A series of different VIC experiments w~~erare~~as performed to (i) compare the VIC model's response when driven by different forcings, (ii) evaluate the uncertainties related to the VIC optimizer, and (iii) investigate the effect of PDO teleconnections on the VIC calibration and validation time periods. For objective (i), we used all the four datasets to run VIC simulations to facilitate detailed comparison of different datasets and their hydrological response. In objectives (ii) and (iii), rather than the intercomparison of datasets, our goal is to evaluate the uncertainties in the model implementation particularly in its calibration process. We therefore only used the UW dataset to force the VIC model as this dataset along with our VIC model implementation is ~~thoroughly~~ examined extensively over the FRB ~~region~~ in Kang et al.~~,~~ (2014) and (2016). The~~se detail of all the~~ experiments ~~are were~~are~~is~~ categorized as ~~below~~ follows:

1) **Inter-comparison runs:** The VIC model was~~is~~ driven by each forcing dataset for 28 years (1979 to 2006) with 1979-1990 as the calibration period and 1991-2006 as the validation period using the MOCOM-UA optimizer (Table 1). The VIC simulations driven by ANUSPLIN, UW and PCIC forcings are initiated five years prior to the year 1979 to allow model spin-up time. Since NARR is

not available until 1979, its VIC simulations were recursively looped for five years using the year 1979 as the forcing datathe model spin up time is not considered. After calibration, the model validation runs were initialized with five different state files to produce five ensemble members (Table 1)Once calibration is completed, validation runs are initiated with five different initial conditions (lagged by 24-hour intervals) to avoid any random uncertainties such as those in initial conditions (Table 1). The ANUSPLIN, NARR, UW and PCIC driven ensemble mean VIC simulations are named referred as to ANUSPLIN-VIC, NARR-VIC, UW-VIC and PCIC-VIC, respectively. These ensemble simulations are were run for the whole FRB and its UF, SU, NA, QU, CH and TN sub-basins.

2) **Optimizer uncertainty runs:** Our comparison of hydrological simulations driven by four forcing datasets revealed that PCIC and UW driven VIC simulations are noticeably better with high NSE values and reproduce hydrographs similar to that from observations. Here we only used the PCICUW forcing data for VIC model simulations to investigate the uncertainties in the model calibration process for the 1979-1990 time period. Our primary goal is to evaluate optimizer sensitivity to a unique set of parameter limits. We want to see how the MOCOM optimizer results in different optimized parameters and change the overall simulated hydrograph in the calibration process. Here the VIC model is driven only by the PCIC dataset for the 1979-1990 calibration time period using the MOCOM-UA optimizer. We performed the optimization of six soil parameters, i.e. b_infilt, Dsmax, Ws, D2, D3 and Ds in three five experimental setups using different initial ranges of parameter limits. The VIC calibration experiments (OPT1, OPT3, OPT4 and OPT53) are were run using two four narrow ranges selected from the maximum limits of calibration parameters. The

same experiment is then run with maximum limits of the calibration parameters (OPT2). Calibration parameters, their initial ranges and final optimized values for all ~~three~~the experiments are given in Table 3. The OPT1, OPT2, OPT3, OPT4 and ~~OPT3~~ OPT5 simulations ~~are~~ were run over the whole FRB only.

3) **PDO uncertainty runs:** ~~Both~~ We used ~~the~~ the UW ~~and PCIC~~ dataset~~s~~ ~~are used~~ to drive long term (1950-2006) VIC simulations. This is to capture the decadal variability of cool and warm phases of the PDO. ~~Along with this, t~~Five~~hree~~ different experiments, namely PDO1, PDO2, PDO3, PDO4 and ~~PDO3~~ PDO5 ~~are~~ were performed with calibration periods of 1981-1990, 1956-1965,~~ and~~ 1967-1976, 1977-1987 and 1991-2001 and with corresponding validation periods of 1991-2001, 1966-1976, ~~and~~ 1977-1987, 1967-1976 and 1981-1990, respectively (Table 4). Each time period ~~is~~ was selected to capture cool or warm PDO phases, i.e. its cool (1956-1965 and 1967-1976) and warm (1981-1990, 1991-2001 and 1977-1987) phases. For each calibration experiment in one particular phase of the PDO, the MOCOM-UA ~~is~~ was used to optimize calibration parameters. The NSE ~~is~~ was calculated for the calibration and validation periods using the daily observed streamflow data for the Fraser River at Hope. All PDO simulations ~~are~~ were run over the whole FRB only.

2.~~5~~ 4 Analysis Strategy

The analyses ~~are~~ were performed for three FRB hydro-climatic regimes: the Interior Plateau, the Rocky Mountains and the Coast Mountains (Moore, 1991). These three regions ~~are~~ were chosen given their distinct physiography and hydro-climatic conditions. The gridcell partitioning of these three regions and their elevations are shown in Fig. 1b. Results in this study mainly focused on the Fraser River main stem at Hope, BC since it covers 94% of the basin's drainage area and has a continuous streamflow record over the study periods. However, the inter-comparison runs are were also compared over the FRB's major sub-basins. The total runoff is was calculated using the sum of baseflow and runoff. Seasonal variations are were assessed by averaging Dec-Jan-Feb (DJF), Mar-Apr-May (MAM), Jun-Jul-Aug (JJA) and Sep-Oct-Nov (SON) months for winter, spring, summer and autumn, respectively.

In the SWE analysis, the snowmelt is was calculated by taking the difference between maximum and minimum SWE over the water year (1 October to 30 September of the following calendar year). The corresponding day of the water year having maximum SWE (maxSWE) is referred to named as maxSWE-day.

Although glacier dynamics are not included in the VIC model physics, the model produce a perennial snowpack in few gridcells in its output. We compared those cells to Baseline Thematic Mapping (BTM) and found that the glaciating cells match the location of observed glaciers. We therefore masked those gridcell in the SWE analysis considering that the effects of glaciers may not change our results significantly due to the ~25 km model grid cell resolution (625 km^2 area per grid cell) used in this study.

The Mann–Kendall test (Mann, 1945; Kendall, 1970) is was used to estimate monotonic trends in the input forcing data and the simulated hydrological variables. This non-parametric trend test has been used in several other studies to detect changing hydrological regimes (Lettenmaier et al., 1994; Ziegler et al., 2003; Déry et al., 2005, 2016; Kang et al., 2014). Trends are were considered to be statistically-significant when $p < 0.05$ with a two-tailed test.

Formatted: Right: 0 cm, Space Before: 9 pt, After: 0 pt, Adjust space between Latin and Asian text

Formatted: Font: Italic

3 Results and Discussion

We first examine the ANUSPLIN, NARR, UW and PCIC datasets to investigate how substantial are the differences in precipitation and air temperature at several temporal and spatial scales across the FRB and its sub-regions. The VIC simulations, driven by these forcing datasets, are then discussed to evaluate uncertainties in simulated SWE and runoff. This is followed by the discussion of uncertainty in the VIC calibration process.

3.1 Forcings Datasets Inter-comparison

The daily mean air temperature of ANUSPLIN, NARR, UW and PCIC datasets remains below 0°C from November to March and rises above 0°C in early spring over all three FRB sub-regions (Fig. 2). While the inter-datasets seasonal variability of air temperature is quite similar, the winter in NARR is ~2°C warmer compared to the remaining datasets. The grid scale seasonal differences (PCIC minus ANUSPLIN, NARR and UW) of mean air temperature (Supplementary Fig. 1) spatially quantify the inter-datasets disagreements (Supplementary Fig. 1). While the PCIC-ANUSPLIN and PCIC-UW differences are within ±1°C, the PCIC-NARR difference exceeds 2°C over most of the FRB in DJF and SON, revealing NARR air temperatures to be quite warmer than in the PCIC dataset.

The magnitudes of daily mean precipitation vary markedly amongst datasets. Winter precipitation, which begins in November and persists until April, shows greater inter-datasets differences, particularly over the Rocky and Coast Mountains. Compared to the PCIC and UW datasets, the ANUSPLIN precipitation is underestimated in all three regions with nearly 2.0 mm day^{-1} to 5.0 mm day^{-1} differences in the Rocky and Coast Mountains, respectively. This underestimation is more evident in the

PCIC-ANUSPLIN spatial difference ~~(Supplementary Fig. 2)~~ revealing up to 5 mm day^{-1} difference over the mountainous regions (Supplementary Fig. 2). The precipitation differences in the Interior Plateau ~~remain nearly~~approach zero for all ~~the~~ datasets. The maximum intraseasonal variability arises in the Coast Mountains ranging from 10.0 mm day^{-1} of precipitation in winter and nearly zero in summer. The range of inter-datasets spread for peak precipitation varies from 5.0 mm day^{-1} to 10.0 mm day^{-1} during winter for the Coast Mountains. Precipitation in the Coast Mountains is more variable due to its proximity to the Pacific Ocean where the interaction between steep elevations and storm track positions is quite complex. In the Coast Mountains, the NARR precipitation is underestimated and is comparable to ANUSPLIN.

The underestimation of the ANUSPLIN mountainous precipitation is probably due the thin plate smoothing spline surface fitting method used in its preparation. For NARR, air temperature and precipitation uncertainties may have been induced by the climate model used to assimilate and produce the reanalysis product.

3.2 Hydrological Simulations

The ANUSPLIN-VIC, NARR-VIC, UW-VIC and PCIC-VIC simulation performance ~~is~~ was evaluated using the NSE and correlation coefficients by calibrating and validating against observed daily streamflow for the Fraser River at Hope (Table 2). The NSE scores are much higher for the PCIC-VIC and UW-VIC simulations compared to the ANUSPLIN-VIC and NARR-VIC. The lower NSE score in the ANUSPLIN-VIC simulation reflects a dry precipitation bias in the ANUSPLIN dataset. As the model configuration, resolution, and soil data ~~are~~ were identical for all VIC simulations, different NSE values reveal uncertainty associated only with each observational forcing dataset. Despite the low NSE score of the ANUSPLIN-VIC simulation, the correlation coefficient is significantly high. The bias in the simulated streamflow is contributing to the lower NSE coefficient whereas the phase of seasonal flow is quite similar to the observed flow in the ANUSPLIN-VIC simulation. There may be additional sources of uncertainty due to the method used to assess simulation accuracy. For example, instead of using NSE, other model evaluation metrics such as the Kling-Gupta Efficiency (KGE) coefficient (Gupta et al., 2009) may produce different levels of model accuracy.

The ANUSPLIN-VIC, NARR-VIC, UW-VIC and PCIC-VIC simulated SWE and snowmelt, areally-averaged over the FRB's three sub-regions, show similar seasonal variability but considerably different magnitudes, especially over mountainous regions. Figure 3a shows these differences for the Rocky Mountains revealing the range of peak SWE from 400 mm for ANUSPLIN to >600 mm for PCIC. The dry bias in ANUSPLIN precipitation forcing induces lower SWE magnitudes in the ANUSPLIN-VIC simulation. The lower SWE in the NARR-VIC simulation is probably due to the warmer air temperature during winter and spring (Fig. 2b). Winter temperatures being warmer in the NARR dataset may alter the phase of precipitation partitioning with more rainfall than snowfall, and hence less SWE in the NARR-VIC simulation. Such differences in SWE are reflected in the associated snowmelt (Fig. 3b) where the NARR-VIC simulation shows earlier snowmelt. This is further investigated by VIC sensitivity experiments and is discussed later in the text. Grid-scale differences in simulated SWE (Fig. 4) and runoff (Supplementary Fig. 3) arise most notably over the mountainous regions. In the interior FRB, the simulation differences between PCIC-VIC and ANUSPLIN-VIC ~~means~~ mean SWE are within a 10 mm range whereas such differences exceed 50 mm to 100 mm~~ range ~~ for the NARR-VIC and UW-VIC simulations.

In the FRB's mountainous regions, the VIC model can lead to inaccurate snowpack estimates if the elevation dependence on snow accumulation and ablation ~~are ~~is not modelled properly. As mentioned in section 2.~~4~~3, we used ten elevation bands in our VIC implementation so that each band's mean elevation ~~is ~~was used to lapse the gridcell average air temperature and precipitation to produce more reliable estimates. We clustered the elevation distribution within 10 bands into different elevation ranges. This alloweds~~s~~ in-depth analysis of the elevation dependent variation of mean SWE that is of particular importance for the Rocky and Coast Mountains regions of the FRB. We examined the magnitude of maxSWE and corresponding maxSWE-day of the water year between all simulations and elevation ranges (Fig. 5). The difference in maxSWE between all VIC simulations increases with elevation~~s~~, particularly the Rocky Mountains where higher elevations (>1400 m) show large disagreement between simulated maxSWE (Fig. 5a). In the Interior Plateau, the NARR-VIC simulated maxSWE exceeds 300 mm whereas all other simulations are within 200 mm. The maxSWE elevation dependent variation is quite complex in the Coast Mountains. However the simulation~~s~~ differences at elevations >1400 m are smaller compared to the lower elevations below 1000 m. Apart from maxSWE magnitude, the maxSWE-day variation differs considerably across the VIC simulations. Generally, the maxSWE-day varies by nearly~~~2 ~~ two months between lower and higher elevations as snow onset occurs later in autumn. While the maxSWE-day variation is quite complex within each elevation range, the NARR-VIC maxSWE-day is earliest whereas PCIC-VIC delays the snow accumulation over the 600-2000 m elevation range in the Rocky Mountains. There is nearly 20 days of simulated~~ions~~ variation in maxSWE-day at the Rocky Mountains highest elevation range. Such variation highlights the uncertainties in seasonality of the VIC simulated snowpacks. For the Interior Plateau and the Coast Mountains, no consistent pattern of maxSWE-day variation exists for any particular simulation.

3.2.1 Comparison of observed versus s~~S~~imulated SWE ~~comparison with observations~~

As mentioned earlier, all gridded climate forcing datasets are based on station observations. The density of stations in the FRB's mountainous regions remains quite low and therefore induces higher uncertainties in the observational gridded products. It is important to quantify the spatial discrepancy between the simulated (0.25° gridcell) and observed (snow pillow station dataset) SWE that may lead to an uncertainty in snow estimations by models (Elder et al., 1991; Tong et al., 2010). We used observed SWE from BC snow pillow sites and the VIC simulated SWE data over the same elevation and overlapping continuous time periods at four different locations in the upper and middle Fraser where a high volume of SWE accumulates seasonally.

The daily time series of VIC simulated SWE (Supplementary Fig. 4) follows the observed interannual variability in snow accumulation but with considerable differences across simulations. The PCIC-VIC simulation accumulates too much SWE compared to observations in the gridcell corresponding to the Yellowhead location. This overestimation is further explored for this site~~Yellowhead~~ by expanding the time series back to 1979 (not shown), which reveals issues with PCIC precipitation data only during 1996-2004 ~~exhibiting~~ with considerable above normal anomalies at Yellowhead. While ANUSPLIN-VIC shows lower SWE amounts, the NARR-VIC and UW-VIC simulations reproduce the observed variation quite reasonably for Yellowhead. For McBride, all simulations are more or less comparable except ANUSPLIN-VIC showing a SWE underestimation compared to observations. In the middle Fraser, the UW-VIC simulation is quite comparable to observations whereas the PCIC-VIC simulation underestimates SWE at Mission Ridge. Both ANUSPLIN-VIC and NARR-VIC underestimate SWE in the middle Fraser locations. The observed SWE values in the lower Fraser locations are not well captured by VIC, perhaps owing to the region's coastal influence and strong sensitivity to air temperatures (not shown). These results highlight the importance of accurate precipitation forcings to simulate SWE. Along with this, even small perturbations in air temperature can change the phase of precipitation, which directly contributes to changes in SWE accumulation.

3.2.2 Comparison of observed versus simulated Simulated rRunoff comparison with observations

The VIC simulated flows are routed to produce hydrographs for the Fraser River at Hope, BC (Fig. 6a). Comparison of simulated runoff with observations shows the highly consistent model performance for PCIC-VIC and UW-VIC simulations whereas the runoff is considerably lower for the ANUSPLIN-VIC simulation. The NARR-VIC hydrograph is comparable in magnitude with observations but the runoff timing is considerably shifted (~15 days) yielding more winter and spring runoff and earlier decline of flows in summer. The shift in the hydrograph is probably caused by the 2°C warmer air temperatures causing earlier snowmelt. This finding is was confirmed by a VIC sensitivity experiment where the air temperature is was perturbed by 2°C while keeping the precipitation unchanged (Supplementary Fig.). Similar to the case of NARR-VIC results, the simulated SWE, snowmelt and runoff decreases with 2°C rises in air temperature forcings (Supplementary Fig. 5). The coefficient of variation for all four datasets reveals that variability in the PCIC-VIC and UW-VIC simulations is similar to observations (Fig. 6b). We -further produced the hydrographs for the FRB's six major sub-basins to compare VIC simulation runs of each basin. Similar to the hydrograph of the Fraser River at Hope, the ANUSPLIN-VIC runoff shows considerable disagreement with the observed hydrograph, especially in the UF, QU and TN basin owing to the dry bias in its precipitation forcing. Moreover, NARR-VIC runoff is overestimated in the SU, NA and CH sub-basins whereas for UF, QU and TN, the simulated runoff underestimates observed flows. Consistent with spatial differences of mean air temperature and runoff (Supplementary Figs. 1 and 3), the warmer NARR air temperatures (compared to PCIC) over the SU, NA and CH sub-basins in winter and spring induce more snowmelt and hence overestimate runoff. In contrast, over the UF, QU and TN, the NARR air temperature is comparatively cooler in winter. This may reduce the snowmelt driven runoff causing underestimation over these sub-basins. The PCIC-VIC hydrographs are better in most of the basins with high NSE scores (Supplementary Table 2).

Differences seen in the FRB's flow magnitude and timing clarify the impact of forcing uncertainties on the simulations. Such variation in simulated runoff especially during the snow-melting period (Apr-Jul) is either due to the uncertain amount of precipitation or the magnitude of air temperature in the forcing datasets.

We further investigated differences in forcings and their VIC simulation based on their climatic trends. The monthly climate trends in air temperature, precipitation and simulated runoff (Supplementary Fig. 56) shows relatively similar that the warm air temperatures warming during winter (up to 3°C in December) and the with the declined of precipitation (mainly snowfall) during winter for all the four is relatively similar between different forcing datasets. The magnitude of trends in the NARR dataset is somewhat lower for air temperature and higher for precipitation compared to the other three datasets. In the simulated runoff, the monthly variation of trends generally agrees among simulations, but the trends are weak in the ANUSPLIN-VIC and UW-VIC simulations whereas the PCIC-VIC and NARR-VIC simulations exhibit strong trends. In the NARR-VIC simulations, runoff trends are affected by lower air temperature and higher precipitation trends yielding increasing runoff. Grid-scale trends show widespread differences in the NARR-VIC runoff, particularly in the interior of the FRB when compared to ANUSPLIN-VIC, UW-VIC and PCIC-VIC monthly trends (Supplementary Fig. 67). All the four simulations exhibit strong positive runoff trends in April followed by declining trends in May in the Rocky Mountains (the UF and TN sub-basins).

The inter-comparison analysis shows that the uncertainties in forcing datasets contribute substantially to the performance of the VIC model. This is consistent with studies reporting that the uncertainties in model structure contribute less to snowpack and runoff simulations (Troin et al., 2015, 2016), whereas the uncertainties in forcing datasets are the predominant sources of uncertainties (Kay et al., 2009; Chen et al., 2011). Using the NARR dataset, the systematic biases in simulations and the substantial effect of lateral boundary conditions on the performance of the regional model have also been identified in many other studies (de Elia et al., 2008; Eum et al., 2012).

While the small differences in precipitation are acceptable, the air temperature uncertainties play an especially important role in the hydrological simulations. In the FRB, air temperature controls summer water availability, making regional snowpacks more vulnerable to temperature-induced effects, rather than precipitation. Thus uncertainties in air temperatures are crucial for the runoff timing in hydrological simulations are more crucial for hydrological simulations over the FRB rather than those in precipitation.

3.3 Uncertainty in Calibration Optimizer

We further investigated the uncertainty in the optimization of parameters during the calibration process. Many studies have evaluated the parameter uncertainties by adding random noise to the calibration parameters. We used a different approach by estimating the uncertainty in the MOCOM-UA optimizer used in the calibration of parameters. This ~~is~~ was to estimate the optimizer uncertainty during the VIC calibration process using different values of initial parameters limits. The optimization process for the OPT1, OPT2, OPT3, OPT4 and OPT5~~3~~ experiments required 39, 89, 61, 52, and 56 ~~82, 115 and 45~~ iterations, respectively, to optimize the b_inf, Ds, Ws, D2, D3 and Dsmax parameters to their final values (~~Fig. 8a-f,~~ Table 3). The corresponding mean monthly (as the optimizer cannot utilize daily data) runoff for the Fraser River at Hope in the OPT1, OPT2, OPT3, OPT4 and OPT5 ~~OPT1, OPT2 and OPT3~~ experiments are quite different when compared to observations (Fig. 8~~g~~). The NSE scores reveal different accuracy for the ~~three~~ five simulations even when the parameters' initial range in the OPT1, OPT3, OPT4 and OPT5~~3~~ experiments is a subset of the OPT2 experiment. The optimization process for parameter calibration would require an expert's experience to set the initial parameter range to converge them to their optimal values. Note that if the initial parameter uncertainty distribution is set as wide as it is physically meaningful, then the optimization will require more computational time to converge toward the Pareto optimum. However, to set the initial parameter limits, the subjective judgement and skill based on experience is needed.

While we performed many sets of experiments with different initial parameters, only OPT1's initial limits produced higher NSE and utilized less computational time. The estimation of hydrologic model parameters depends significantly on the availability and quality of the precipitation and observed streamflow data along with the accuracy of the routing model used. It is therefore important to consider bias correction of forcing datasets as part of automatic calibration. The observed streamflow data used to calibrate the model are often based on water levels that are converted to discharge by the use of a rating curve, which can also induce uncertainty in the observed discharge data. The overall conclusion of this analysis is that the automated optimizers used to converge calibration parameters still rely on the hydrologist's experience and some manual adjustment of initial calibration parameter ranges.

3.4 Uncertainty in calibration due to PDO Phases

The FRB~~'s~~ streamflow varies from year to year as well on decadal timescales depending on the timing and magnitude of precipitation and air temperatures during the preceding winter and spring. Given that the FRB air temperature and precipitation ~~is~~ are influenced by cool and warm phases of the PDO (Mantua et al., 1997; Fleming et al., ~~2007~~2010; Whitfield et al., 2010; Thorne and Woo, 2011), the choice of VIC calibration and validation periods may induce uncertainty in calibration. The influence of PDO phases in the forcing dataset can produce different snowpack and runoff levels in the hydrological simulation~~ (Supplementary Fig. 7)~~. The long term UW-VIC simulations (1949-2006) show higher mean SWE and runoff levels in a cool PDO phase (1949-1976) and lower mean values in a warm PDO phase (1979-2006) (Supplementary Fig. 8). The interannual variations show earlier peak flows characterized by a warm PDO, in response to warmer basin conditions, increased rainfall, and earlier snowmelt. The VIC model calibrations may be biased towards hydrologic conditions of the warm and cold PDO phases and may induce uncertainties in the results. The model performance could be improved by calibrating and validating the model in the same PDO phase (experiments PDO1, PDO2 and PDO5~~2~~), i.e. the NSE coefficient is similar in the calibration and validation periods (Table 34). If the calibration is performed in the cool PDO phase and validation in the warm PDO phase (experiment PDO3), the NSE score decreases to 0.79 for the validation period since the model calibration is biased towards the cool conditions, simulating higher flows for the Fraser River at Hope owing to more snow and later snowmelt. The same is true if the calibration and validation is performed in the warm and cool PDO phases respectively (experiment PDO4). For each set of calibration experiments, the calibration parameters are different, which affects the air temperature lapse rates and thus evapotranspiration, the formation of the snowpack, and the timing of snowmelt. Figure 9 shows observed and simulated runoff for the Fraser River at Hope revealing lower observed peak flows ~2.70 mm day^{-1} in a warm PDO phase (PDO1) and higher peak flows ~3.30 mm day^{-1} in a cool PDO phase (PDO2). Interestingly the UW driven PDO simulations underestimate peak flows in the warm PDO phase and overestimate them in the cool PDO phase whereas the NSE coefficient for both the cool and warm PDO phases is almost equivalent (Table 4). PDO 4 and PDO 5 experiments further support these findings.

This analysis reveals that the hydrological model performance changes considerably changes with different climatic conditions and the choice of the calibration and validation time periods, an important factor in hydrological simulations. The proper implementation of a hydrological model does need a requires a careful calibration strategy to produce reliable hydrological information important for water resource management..

We repeated the PDO experiments using the PCIC forcing dataset to compare UW and PCIC simulations under PDO phases. The PCIC PDO simulation shows better agreement with observed data in both cool and warm phases of the PDO compared to the UW PDO simulations. Consistent with UW

PDO3 simulation, PCIC PDO3 simulation shows poor performance during validation periods if the calibration is performed in a cool PDO phase (Fig. 9 e, f).

4. Conclusions

This study utilized ANUSPLIN, NARR, UW and PCIC observation-based gridded datasets to evaluate systematic inter-datasets uncertainties and their VIC simulated hydrological response over the FRB. The uncertainties involved in the optimization of model parameters and model calibration under cool and warm phases of the PDO are were also examined.

The air temperature in the PCIC and UW datasets were comparable while the PCIC precipitation remains quite high in the Rocky Mountains compared to the UW and NARR datasets. The ANUSPLIN precipitation forcing had a considerable dry bias over mountainous regions of the FRB compared to the NARR, UW and PCIC datasets. The NARR winter air temperature was 2°C warmer than the other datasets over most of the FRB. The PCIC-VIC and UW-VIC simulations had higher NSE values and more reasonable hydrographs compared with observed flows for the Fraser River at Hope. Their performance for many of the FRB's major sub-basins remained satisfactory. The PCIC-VIC simulation revealed higher SWE compared to other datasets, probably due to its higher precipitation amounts. The ANUSPLIN-VIC simulation had considerably lower runoff and NSE values along with less SWE and snowmelt amounts owing to its reduced precipitation. The NARR dataset showed warm winter air temperatures, which influenced its hydrological response by simulating less SWE and decreased snowmelt, and hence lower runoff. The monthly trend analysis distinguished the NARR dataset by showing decreased trends in air temperature and increased trends in precipitation and its VIC driven runoff. The trend analysis further highlighted uncertainties of the NARR dataset.

The elevation dependence of maxSWE showed disagreements over the higher elevations of the Rocky Mountains between simulations where the PCIC-VIC simulation overestimated SWE and ANUSPLIN-VIC resulted in underestimation. Furthermore the elevation dependent variation of the maxSWE-day fluctuates fluctuated considerably between simulations.

The parametric uncertainty in the VIC calibration process revealeds that the choice of the initial parameter range plays a crucial role in defining the model performance. During the PDO phases, choice of the calibration and validation time periods play a crucial role in defining the model hydrological response for the FRB. Model calibration was biased towards hydrologic conditions of the warm and cold PDO phases. The UW-VIC PDO simulations underestimated and overestimated the peak flows in the warm and cool PDO phases, respectively.

This study's inter-comparison revealed spatial and temporal differences amongst the ANUSPLIN, NARR, UW and PCIC datasets over the FRB, which is essential to capture the uncertainties in modelling hydrologic responses. Overall, the PCIC and UW datasets had reliable results for the FRB snow hydrology whereas the ANUSPLIN and NARR datasets had issues with either precipitation or with air temperature. The FRB's snow-dominated hydrology and its complex elevation profile require highly accurate meteorological station densities to increase the reliability of the high resolution gridded datasets. While the air temperature plays a dominant role in the hydrological simulations, improving the quality of precipitation data can lead to more accurate hydrological responses in the FRB. Considerable precipitation bias can substantially degrade the model performance.~~Improving the quality of precipitation data may thus lead to more accurate hydrological responses in the FRB.~~ There is the need for concrete methods to deal with the increasing uncertainty associated with the models themselves, and with the observations required for driving and evaluating the models.

In this study, the FRB hydrological response varied considerably ~~varied~~ under different forcing datasets, modelling parameters and remote teleconnections. However there are other sources of uncertainties not discussed here that may establish a range of possible impacts on hydrological simulations. First, the hydrological model used in this study runs at a daily time step, which can be increased to hourly to refine the model performance. ~~The uncertainties in the VIC model structure include here t~~The lack of the representation of glaciers in the current version of the VIC model~~. The lack of glacier's dynamics in the current version of the VIC model may~~ may ~~also~~ induce uncertainties in model results. Along with these, the VIC simulations are also affected by intrinsic uncertainties in its parameterizations such as, for example, the representation of cold processes (e.g., snowpacks and soil freezing).~~ Such structural uncertainties need to be evaluated by examining multiple hydrological models over the same region. The lack of glacier's dynamics in the current version of the VIC model may also induce uncertainties in model results.~~ The in situ soil moisture observations that are not necessarily representative of the model grid scale may also contribute to the overall uncertainties in the results. Finally, hydrological simulations are mainly validated using comparisons between simulated and observed flows, which depend on routing models that may contain structural uncertainties. Our future work will investigate such uncertainties using high temporal and spatial resolution hydrological models over the FRB.

Acknowledgments

This work ~~is~~ was supported by the NSERC-funded Canadian Sea Ice and Snow Evolution (CanSISE) Network. This manuscript was motivated by the Eric Wood Symposium held at Princeton University on 2-3 June 2016. The authors are grateful to colleagues from the Pacific Climate Impacts Consortium (PCIC) for providing ongoing assistance with this research and to Dennis Lettenmaier ~~and the Land Surface Hydrology Research Group~~ at ~~the University of California~~UCLA for providing ~~model and~~ assistance in ~~its~~ the VIC model implementation. The authors are thankful to Michael Allchin (UNBC) for plotting Fig. 1a, Xiaogang Shi (Xi'an Jiaotong-Liverpool University) for development and improvements of the UW dataset, and Do Hyuk Kang (NASA GSFC) for helping in setting up the VIC model over the FRB. Thanks to the anonymous referees and the handling editor for constructive comments that greatly improved the paper

References

Adam, J. C., Hamlet, A. F. and Lettenmaier, D. P.: Implications of global climate change for snowmelt hydrology in the twenty-first century, Hydrol. Process., 23, 962–972, doi:10.1002/hyp, 2009.

Adam, J. C. and Lettenmaier, D. P.: Application of new precipitation and reconstructed streamflow products to streamflow trend attribution in Northern Eurasia, J. Clim., 21(8), 1807–1828, doi:10.1175/2007JCLI1535.1, 2008.

Adam, J. C., Clark, E. A., Lettenmaier, D. P. and Wood, E. F.: Correction of global precipitation products for orographic effects, J. Clim., 19(1), 15–38, doi:10.1175/JCLI3604.1, 2006.

Adam, J. C. and Lettenmaier, D. P: Adjustment of global gridded precipitation for systematic bias, J. Geophys. Res., 108(D9), 1–15, doi:10.1029/2002JD002499, 2003.

Ainslie, B. and Jackson, P. L.: Downscaling and bias correcting a cold season precipitation climatology over coastal southern british British columbia Columbia using the regional atmospheric modeling system (RAMS), J. Appl. Meteorol. Climatol., 49(5), 937–953, doi:10.1175/2010JAMC2315.1, 2010.

Anderson, J., Chung, F., Anderson, M., Brekke, L., Easton, D., Ejeta, M., Peterson, R. and Snyder, R.: Progress on incorporating climate change into management of California's water resources, Clim. Change, 87(1 SUPPL 1):91, doi:10.1007/s10584-007-9353-1, 2007.

Arsenault, R. and Brissette, F. P.: Continuous streamflow prediction in ungauged basins: The effects of equifinality and parameter set selection on uncertainty in regionalization approaches, Water Resour. Res., 50(7), 6135–6153, doi:10.1002/2013WR014898, 2014.

BC Ministry of Forests, Lands and Natural Resource Operations: Automated Snow Pillow Data. [Available online at http://bcrfc.env.gov.bc.ca/data/asp/.], 2014.

Benke, A. C. and Cushing, C. E.: Rivers of North America, Elsevier Press, New York, 607-732, 2005.

Bennett, K. E., Werner, A. T. and Schnorbus, M.: Uncertainties in hydrologic and climate change impact analyses in headwater basins of British Columbia, J. Clim., 25(17), 5711–5730, doi:10.1175/JCLI-D-11-00417.1, 2012.

Beven, K.: A manifesto for the equifinality thesis, J. Hydro., 320, 18–36, 2006.

Chen, J., Brissette, F. P., Poulin, A. and Leconte, R.: Overall uncertainty study of the hydrological impacts of climate change for a Canadian watershed, Water Resour. Res., 47(12), W12509, doi:10.1029/2011WR010602, 2011.

Formatted: Font color: Auto

Formatted: No widow/orphan control

Cherkauer, K. A. and Sinha, T.: Hydrologic impacts of projected future climate change in the Lake Michigan region, J. Great Lakes Res., 36(SUPPL. 2), 33–50, doi:10.1016/j.jglr.2009.11.012, 2010.

Cherkauer, K. A., Bowling, L. C. and Lettenmaier, D. P.: Variable infiltration capacity cold land process model updates, Glob. Planet. Change, 38(1–2), 151–159, doi:10.1016/S0921-8181(03)00025-0, 2003.

Choi, G., Robinson, D. A. and Kang, S.: Changing northern hemisphere snow seasons, J. Clim., 23(19), 5305–5310, doi:10.1175/2010JCLI3644.1, 2010.

Choi, W., Kim, S. J., Rasmussen, P. F. and Moore, A. R.: Use of the North American Regional Reanalysis for hydrological modelling in Manitoba, Can. Water Resour. J., 34(1), 17–36, doi:10.4296/cwrj3401017, 2009.

Christensen, N. S. and Lettenmaier, D. P.: A multimodal ensemble approach to assessment of climate change impacts on the hydrology and water resources of the Colorado River Basin, Hydrol. Earth Syst. Sci., 11, 1417–1434, 2007.

Cuo, L., Lettenmaier, D. P., Alberti, M. and Richey, J. E.: Effects of a century of land cover and climate change on the hydrology of the Puget Sound basin, Hydrol. Process., 23(6), 907–933, doi:10.1002/hyp.7228, 2009.

Danard, M. and Murty, T. S.: On recent climate trends in selected salmon-hatching areas of British Columbia, J. Clim., 7(11), 1803–1808, doi:10.1175/1520-0442(1994)007<1803:ORCTIS>2.0.CO;2, 1994.

de Elía, R., Caya, D., Côté, H., Frigon, A., Biner, S., Giguère, M., Paquin, D., Harvey, R. and Plummer, D.: Evaluation of uncertainties in the CRCM-simulated North American climate, Clim. Dyn., 30(2), 113–132, doi:10.1007/s00382-007-0288-z, 2008.

Déry S. J., Hernández-Henríquez, M.., Owens P. N., Parkes M. W. and Petticrew E. L.: A century of hydrological variability and trends in the Fraser River Basin, Environ. Res. Lett., 7(2), 24019, doi:10.1088/1748-9326/7/2/024019, 2012.

Déry, S. J., Stieglitz, M., McKenna, E. C. and Wood, E. F.: Characteristics and trends of river discharge into Hudson, James, and Ungava Bays, 1964-2000, J. Clim., 18(14), 2540–2557, doi:10.1175/JCLI3440.1, 2005.

Déry, S. J., Knudsvig, H. K., Hernández-Henríquez, M. A. and Coxson, D. S.: Net snowpack accumulation and ablation characteristics in the Inland Temperate Rainforest of the Upper Fraser River Basin, Canada, Hydrology, 1(1), 1–19, doi:10.3390/hydrology1010001, 2014.

Déry, S. J., Stadnyk, T. A., MacDonald, M. K., and Gauli-Sharma, B.: Recent trends and variability in river discharge across northern Canada, Hydrol. Earth Syst. Sci., 20, 4801-4818, doi: 10.5194/hess-20-1-2016, 2016.

Dodson R., Marks, D.: Daily air temperature interpolated at high spatial resolution over a large mountainous region, Clim. Research, 8:1–20, 1997.

Elder, K., Dozier, J. and Michaelsen, J.: Snow accumulation and distribution in an Alpine Watershed, Water Resour. Res., 27(7), 1541–1552, doi:10.1029/91WR00506, 1991.

Elsner, M. M., Cuo, L., Voisin, N., Deems, J. S., Hamlet, A. F., Vano, J. aA., Mickelson, K. E. B., Lee, S. Y. and Lettenmaier, D. P.: Implications of 21st century climate change for the hydrology of Washington State, Clim. Change, 102(1–2), 225–260, doi:10.1007/s10584-010-9855-0, 2010.

Environment Canada: Climate database. [Available online at http://climate.weather.gc.ca.], cited 2014.

Essou, G. R. C., Arsenault, R. and Brissette, F. P.: Comparison of climate datasets for lumped hydrological modeling over the continental United States, J. Hydrol., 537, 334–345, doi:http://dx.doi.org/10.1016/j.jhydrol.2016.03.063, 2016a.

Essou, G. R. C., Sabarly, F., Lucas-Picher, P., Brissette, F. and Poulin, A.: Can Pprecipitation and Ttemperature from Mmeteorological Rreanalyses Bbe Uused for Hhydrological Mmodeling?, J. Hydrometeorol., 17(7), 1929–1950, doi:10.1175/JHM-D-15-0138.1, 2016b.

Eum, H. Il, Dibike, Y., Prowse, T. and Bonsal, B.: Inter-comparison of high-resolution gridded climate data sets and their implication on hydrological model simulation over the Athabasca Watershed, Canada, Hydrol. Process., 28(14), 4250–4271, doi:10.1002/hyp.10236, 2014.

Eum, H. Il, Gachon, P., Laprise, R. and Ouarda, T.: Evaluation of regional climate model simulations versus gridded observed and regional reanalysis products using a combined weighting scheme, Clim. Dyn., 38(7–8), 1433–1457, doi:10.1007/s00382-011-1149-3, 2012.

Fleming, S. W. and Whitfield, P. H.: Spatiotemporal mapping of ENSO and PDO surface meteorological signals in British Columbia, Yukon, and southeast Alaska, Atmosphere-Ocean, 48(2), 122–131, doi:10.3137/AO1107.2010, 2010.

Foreman, M. G. G., Lee, D. K., Morrison, J., Macdonald, S., Barnes, D. and Williams, I. V.: Simulations and retrospective analyses of Fraser watershed flows and temperatures, Atmosphere-Ocean, 39(2), 89–105, doi:10.1080/07055900.2001.9649668, 2001.

Fowler, H. J., Blenkinsop, S. and Tebaldi, C.: Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling, Int. J. Climatol., 27(12), 1547–1578, doi:10.1002/joc.1556, 2007.

Formatted: Indent: Left: 0 cm, Hanging: 0.75 cm, Space Before: 12 pt, Line spacing: single

Formatted: Not Highlight

Formatted: Font: 12 pt

Gao, H., Tang, Q., Shi, X., Zhu, C., Bohn, T. J., Su, F., Sheffield, J., Pan, M., Lettenmaier, D. P. and Wood, E. F.: Water Budget Record from Variable Infiltration Capacity (VIC) Model. In Algorithm Theoretical Basis Document for Terrestrial Water Cycle Data Records (unpublished), 2009.

Gao, H., Tang, Q., Ferguson, C. R., Wood, E. F. and Lettenmaier, D. P.: Estimating the water budget of major US river basins via remote sensing, Int. J. Remote Sens., 31(14), 3955–3978, doi:10.1080/01431161.2010.483488, 2010.

Graham, L. P., Hagemann, S., Jaun, S. and Beniston, M.: On interpreting hydrological change from regional climate models, Clim. Change, 81, 97–122, doi:10.1007/s10584-006-9217-0, 2007.

Gupta, H. V., Kling, H., Yilmaz, K. K. and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J. Hydrol., 377(1–2), 80–91, doi:10.1016/j.jhydrol.2009.08.003, 2009.

Haddeland, I., Skaugen, T. and Lettenmaier, D. P.: Hydrologic effects of land and water management in North America and Asia: 1700-1992, Hydrol. Earth Syst. Sci., 11(2), 1035–1045, doi:10.5194/hess-11-1035-2007, 2007.

Hidalgo, H. G., Das, T., Dettinger, M. D., Cayan, D. R., Pierce, D. W., Barnett, T. P., Bala, G., Mirin, A., Wood, A. W., Bonfils, C., Santer, B. D. and Nozawa, T.: Detection and attribution of streamflow timing changes to climate change in the western United States, J. Clim., 22(13), 3838–3855, doi:10.1175/2009JCLI2470.1, 2009.

Hopkinson, R. F., Mckenney, D. W., Milewska, E. J., Hutchinson, M. F., Papadopol, P. and Vincent, L. A.: Impact of aligning climatological day on gridding daily maximum-minimum temperature and precipitation over Canada, J. Appl. Meteorol. Climatol., 50(8), 1654–1665, doi:10.1175/2011JAMC2684.1, 2011.

Horton, P., Schaefli, B., Mezghani, A., Hingray, B. and Musy, A.: Assessment of climate-change impacts on alpine discharge regimes with climate model uncertainty, Hydrol. Process., 20, 2091–2109, 2006.

Hutchinson, M. F., McKenney, D. W., Lawrence, K., Pedlar, J. H., Hopkinson, R. F., Milewska, E. and Papadopol, P.: Development and testing of Canada-wide interpolated spatial models of daily minimum–maximum temperature and precipitation for 1961–2003, J. Appl. Meteorol. Climatol., 48(4), 725–741, doi:10.1175/2008JAMC1979.1, 2009.

Islam, S. U., Déry, S. J., Werner, A. T.: Future climate change impacts on snow and water resources of the Fraser River Basin, British Columbia, J. Hydrometeorol., JHM-D-16-0012.1, doi:10.1175/JHM-D-16-0012.1, 2017.

Jiang, T., Chen, Y. D., Xu, C., Chen, X., Chen, X. and Singh, V. P.: Comparison of hydrological impacts of climate change simulated by six hydrological models in the Dongjiang Basin, South China, J. Hydrol., 336(3–4), 316–333, doi:10.1016/j.jhydrol.2007.01.010, 2007.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R. and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, Bull. Am. Meteorol. Soc., 77(3), 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.

Kang, D. H., Gao, H., Shi, X., Islam, S. U. and Déry, S. J.: Impacts of a Rapidly rapidly Declining declining Mountain mountain Snowpack snowpack on Streamflow streamflow tTiming in Canada's Fraser River Basin, Sci. Rep., 6, 19299, doi:10.1038/srep19299, 2016.

Kang, D. H., Shi, X., Gao, H. and Déry, S. J.: On the changing contribution of snow to the hydrology of the Fraser River Basin, Canada, J. Hydrometeorol., 15(4), 1344–1365, doi:10.1175/JHM-D-13-0120.1, 2014.

Kay, A. L., Davies, H. N., Bell, V. A. and Jones, R. G.: Comparison of uncertainty sources for climate change impacts: Flood frequency in England, Clim. Change, 92(1–2), 41–63, doi:10.1007/s10584-008-9471-4, 2009.

Kendall, M. G.: Rank Correlation Methods. 4th ed. Griffin, 202 pp., 1970.

Keshta, N. and Elshorbagy, A.: Utilizing North American Regional Reanalysis for modeling soil moisture and evapotranspiration in reconstructed watersheds, Phys. Chem. Earth, 36(1–4), 31–41, doi:10.1016/j.pce.2010.12.001, 2011.

Klemeš, V.: Operational testing of hydrological simulation models, Hydrol. Sci. J., 31(1), 13–24, doi:10.1080/02626668609491024, 1986.

Kistler, R., Collins, W., Saha, S., White, G., Woollen, J., Kalnay, E., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., van den Dool, H., Jenne, R. and Fiorino, M.: The NCEP–NCAR 50–Year Reanalysis: Monthly Means CD–ROM and Documentation, Bull. Am. Meteorol. Soc., 82(2), 247–267, doi:10.1175/1520-0477(2001)082<0247:TNNYRM>2.3.CO;2, 2001.

Lettenmaier, D. P., Wood, E. F. and Wallis, J. R.: Hydro-climatological trends in the continental United States, 1948-88, J. Clim., 7(4), 586–607, doi:10.1175/1520-0442(1994)007<0586:HCTITC>2.0.CO;2, 1994.

Liang, X., Lettenmaier, D. P., Wood, E. F. and Burges, S. J.: A simple hydrologically based model of land surface water and energy fluxes for general circulation models, J. Geophys. Res. Atmos., 99(D7), 14415–14428, doi:10.1029/94JD00483, 1994.

Liang, X., Wood, E. F. and Lettenmaier, D. P.: Surface soil moisture parameterization of the VIC-2L Model: Evaluation and modifications, Glob. Planet. Chang., 13, 195–206, 1996.

Liu, Y. and Gupta, H. V.: Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, Water Resour. Res., 43(7), W07401, doi:10.1029/2006WR005756, 2007.

Lohmann, D., Nolte-Holube, R. and Raschke, E.: A large-scale horizontal routing model to be coupled to land surface parametrization schemes, Tellus, Ser. A Dyn. Meteorol. Oceanogr., 48(5), 708–721, doi:10.1034/j.1600-0870.1996.t01-3-00009.x, 1996.

Lohmann, D., Raschke, E., Nijssen, B. and Lettenmaier, D. P.: Regional scale hydrology: I. Formulation of the VIC-2L model coupled to a routing model, Hydrol. Sci. J., 43(1), 131–141, doi:10.1080/02626669809492107, 1998a.

Lohmann, D., Raschke, E., Nijssen, B. and Lettenmaier, D. P.: Regional scale hydrology: II. Application of the VIC-2L model to the Weser River, Germany, Hydrol. Sci. J., 43(1), 143–158, doi:10.1080/02626669809492108, 1998b.

Luo, Y., Berbery, E. H., Mitchell, K. E. and Betts, A. K.: Relationships between land surface and near-surface atmospheric variables in the NCEP North American Regional Reanalysis, J. Hydrometeorol., 8, 1184–1203, doi:10.1175/2007JHM844.1, 2007.

Mann, H. B.: Nonparametric tests against trend, Econom., 245–259, doi:10.1017/CBO9781107415324.004, 1945.

Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M. and Francis, R. C.: A Pacific interdecadal climate oscillation with impacts on salmon production, Bull. Am. Meteorol. Soc., 78(6), 1069–1079, doi:10.1175/1520-0477(1997)078<1069:APICOW>2.0.CO;2, 1997.

Matsuura, K., and Willmott C. J.: Terrestrial precipitation: 1900–2008 gridded monthly time series (version 2.01). Center for Climatic Research, Department of Geography, University of Delaware, Newark, DE, digital media. [Available online at http://climate.geog.udel.edu/~climate/.], 2009.

Maurer, E. P., Wood, A. W., Adam, J. C., Lettenmaier, D. P. and Nijssen, B.: A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States, J. Clim., 15(22), 3237–3251, doi:10.1175/1520-0442(2002)015<3237:ALTHBD>2.0.CO;2, 2002.

Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., Jović, D., Woollen, J., Rogers, E., Berbery, E. H., Ek, M. B., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y., Manikin, G., Parrish, D. and Shi, W.: North American regional reanalysis, Bull. Am. Meteorol. Soc., 87(3), 343–360, doi:10.1175/BAMS-87-3-343, 2006.

Moore, R. and Wondzell, S. M.: Physical hydrology and the effects of forest harvesting in the Pacific Northwest: A review, J. Am. Water Resour. Assoc., 41(4), 763–784, doi:10.1111/j.1752-1688.2005.tb03770.x, 2005.

Moore, R. D.: Hydrology and Water Supply in the Fraser River Basin. In Water in Sustainable Development: Exploring Our Common Future in the Fraser River Basin, AHJ Dorcey, JR Griggs (eds). Wastewater Research Centre, The University of British Columbia: Vancouver, British Columbia, Canada; 21–40, 1991.

Mote, P. W., Hamlet, A. F., Clark, M. P. and Lettenmaier, D. P.: Declining mountain snowpack in western North America, Bull. Am. Meteorol. Soc., 86(1), 39–49, doi:10.1175/BAMS-86-1-39, 2005.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I - A discussion of principles, J. Hydrol., 10(3), 282–290, doi:10.1016/0022-1694(70)90255-6, 1970.

Neilsen, D., Duke, G., Taylor, B., Byrne, J., Kienzle, S. and der Gulik, T. Van: Development and verification of daily gridded climate surfaces in the Okanagan Basin of British Columbia, Can. Water Resour. J., 35(2), 131–154, doi:10.4296/cwrj3502131, 2010.

Nigam, S. and Ruiz-Barradas, A.: Seasonal hydroclimate variability over North America in global and regional reanalyses and AMIP simulations: Varied representation, J. Clim., 19(5), 815–837, doi:10.1175/JCLI3635.1, 2006.

Nijssen, B., Schnur, R. and Lettenmaier, D. P.: Global retrospective estimation of soil moisture using the Variable Infiltration Capacity land surface model, 1980–93, J. Clim., 14(8), 1790–1808, doi:10.1175/1520-0442(2001)014<1790:GREOSM>2.0.CO;2, 2001ba.

Nijssen, B., O'Donnell, G. M., Lettenmaier, D. P., Lohmann, D. and Wood, E. F.: Predicting the discharge of global rivers, J. Clim., 14(15), 3307–3323, doi:10.1175/1520-0442(2001)014<3307:PTDOGR>2.0.CO;2, 2001a2001b.

Nijssen, B., Schnur, R. and Lettenmaier, D. P.: Global retrospective estimation of soil moisture using the Variable Infiltration Capacity land surface model, 1980–93, J. Clim., 14(8), 1790–1808, doi:10.1175/1520-0442(2001)014<1790:GREOSM>2.0.CO;2, 2001b.

Nijssen, B., Lettenmaier, D. P., Liang, X., Wetzel, S. W., Wood, E. F.: Streamflow simulation for continental-scale river basins. Water Resour. Res., 33, 711–724, 1997.

NRCan.: Regional, national and international climate modeling. Retrieved from http://cfs.nrcan.gc.ca/projects/3?lang=en_CA, 2014.

Oubeidillah, A. A., Kao, S. C., Ashfaq, M., Naz, B. S. and Tootle, G.: A large-scale, high-resolution hydrological model parameter data set for climate change impact assessment for the conterminous US, Hydrol. Earth Syst. Sci., 18(1), 67–84, doi:10.5194/hess-18-67-2014, 2014.

Pietroniro, A., Leconte, R., Toth, B., Peters, D. L., Kouwen, N., Conly, F. M. and Prowse, T. D.: Modelling climate change impacts in the Peace and Athabasca catchment and delta: III - Integrated model assessment, Hydrol. Process., 20(19), 4231–4245, doi:10.1002/hyp.6428, 2006.

Poulin, A., Brissette, F., Leconte, R., Arsenault, R. and Malo, J. S.: Uncertainty of hydrological modelling in climate change impact studies in a Canadian, snow-dominated river basin, J. Hydrol., 409(3–4), 626–636, doi:10.1016/j.jhydrol.2011.08.057, 2011.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F. and Seo, D. J.: Overall distributed model intercomparison project results, J. Hydrol., 298, 27–60, 2004.

Rinke, A., Marbaix, P. and Dethloff, K.: Internal variability in Arctic regional climate simulations: Case study for the SHEBA year, Clim. Res., 27(3), 197–209, doi:10.3354/cr027197, 2004.

Rodenhuis, D., Bennett, K. and Werner, A., Murdock, Q. and Bronaugh,: Hydro-climatology and future climate impacts in British Columbia. Pacific Climate Impacts Consortium, University of Victoria, Victoria, BC, 132 pp., 2009

Ruane, A. C.: NARR's atmospheric water cycle components. Part I: 20-year mean and annual interactions, J. Hydrometeorol., 11(6), 1205–1219, doi:10.1175/2010JHM1193.1, 2010.

Sabarly, F., Essou, G., Lucas-Picher, P., Poulin, A. and Brissette, F.: Use of f~~F~~our r~~R~~eanalysis d~~D~~atasets to a~~A~~ssess the t~~T~~errestrial b~~B~~ranch of the w~~W~~ater c~~C~~ycle over Quebec, Canada, J. Hydrometeorol., 17(5), 1447–1466, doi:10.1175/JHM-D-15-0093.1, 2016.

Schnorbus, M., Bennett, K. and Werner, A.: Quantifying the water resource impacts of mountain pine beetle and associated salvage harvest operations across a range of watershed scales : Hydrologic modelling of the Fraser River Basin,~~ 2010~~. Information Report: BC-X-423, Natural Resources Canada, Canadian Forestry Service, Pacific Forestry Centre, Victoria, BC, 79 pp, 2010.~~.~~

Schnorbus, M., Bennett, K., Werner, A. and Berland, A. J.: Hydrologic Impacts of Climate Change in the Peace, Campbell and Columbia Sub-basins, British Columbia, Canada. Pacific Climate Impacts Consortium, University of Victoria: Victoria, BC, 157 pp., 2011

Seiller, G., Anctil, F., and Perrin, C.: Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions, Hydrol. Earth Syst. Sci., 16, 1171-1189, doi:10.5194/hess-16-1171-2012, 2012.

Formatted: Not Highlight

Sheffield, J., Goteti, G., and Wood, E. F.: Development of a 50-yr high-resolution global dataset of meteorological forcings for land surface modeling, J. Clim.~~ate~~, 19(13), 3088 – 3111, 2006.

Sheffield, J., Livneh, B. and Wood, E. F.: Representation of terrestrial hydrology and large-scale drought of the continental United States from the North American Regional Reanalysis, J. Hydrometeorol., 13, 856–876, doi:10.1175/JHM-D-11-065.1, 2012.

Shi, X., Déry, S. J., Groisman, P. Y. and Lettenmaier, D. P.: Relationships between recent pan-arctic snow cover and hydroclimate trends, J. Clim., 26(6), 2048–2064, doi:10.1175/JCLI-D-12-00044.1, 2013.

Shi, X., Wood, A. W. and Lettenmaier, D. P.: How e~~E~~ssential is hydrologic model calibration to seasonal streamflow forecasting?, J. Hydrometeorol., 9(6), 1350–1363, doi:10.1175/2008JHM1001.1, 2008.

Shrestha, R. R., Schnorbus, M. A., Werner, A. T. and Berland, A. J.: Modelling spatial and temporal variability of hydrologic impacts of climate change in the Fraser River basin, British Columbia, Canada, Hydrol. Process., 26(12), 1841–1861, doi:10.1002/hyp.9283, 2012.

Stahl, K., Moore, R. D., Floyer, J. A., Asplin, M. G. and McKendry, I. G.: Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density, Agric. For. Meteorol., 139(3–4), 224–236, doi:10.1016/j.agrformet.2006.07.004, 2006.

Su, F. G., et al.: Streamflow simulations of the terrestrial Arctic domain, J. Geophys. Res.-Atmos., 110(D8), 2005.

Tapiador, F. J., Turk, F. J., Petersen, W., Hou, A. Y., García-Ortega, E., Machado, L. A. T., Angelis, C. F., Salio, P., Kidd, C., Huffman, G. J. and de Castro, M.: Global precipitation measurement: Methods, datasets and applications, Atmos. Res., 104–105, 70–97, doi:10.1016/j.atmosres.2011.10.021, 2012.

Teutschbein, C., Wetterhall, F. and Seibert, J.: Evaluation of different downscaling techniques for hydrological climate-change impact studies at the catchment scale, Clim. Dyn., 37(9), 2087–2105, doi:10.1007/s00382-010-0979-8, 2011.

Thorne, R. and Woo, M. K.: Streamflow response to climatic variability in a complex mountainous environment: Fraser River Basin, British Columbia, Canada, Hydrol. Process., 25(19), 3076–3085, doi:10.1002/hyp.8225, 2011.

Tobin, C., Nicotina, L., Parlange, M. B., Berne, A. and Rinaldo, A.: Improved interpolation of meteorological forcings for hydrologic applications in a Swiss Alpine region, J. Hydrol., 401(1–2), 77–89, doi:10.1016/j.jhydrol.2011.02.010, 2011.

Formatted: Font: Not Bold, Not Italic

Tong, J., Déry, S. J., Jackson, P. L. and Derksen, C.: Testing snow water equivalent retrieval algorithms for passive microwave remote sensing in an alpine watershed of western Canada, Can. J. Remote Sens., 36, S74-S86, doi:10.5589/m10-009, 2010.

Troin, M., Arsenault, R. and Brissette, F.: Performance and uncertainty evaluation of snow models on snowmelt flow simulations over a Nordic catchment (Mistassibi, Canada), Hydrology, 2(4), 289–317, doi:10.3390/hydrology2040289, 2015.

Troin, M., Poulin, A., Baraer, M. and Brissette, F.: Comparing snow models under current and future climates: Uncertainties and implications for hydrological impact studies, J. Hydrol., 540, 588–602, doi:10.1016/j.jhydrol.2016.06.055, 2016.

Trubilowicz, J. W., Shea, J. M., Jost, G. and Moore, R. D.: Suitability of North American Regional Reanalysis (NARR) output for hydrologic modelling and analysis in mountainous terrain, Hydrol. Process., 30(13), 2332–2347, doi:10.1002/hyp.10795, 2016.

Velazquez, J. A., Schmid, J., Ricard, S., Muerth, M. J., Gauvin St-Denis, B., Minville, M., Chaumont, D., Caya, D., Ludwig, R. and Turcotte, R.: An ensemble approach to assess hydrological models' contribution to uncertainties in the analysis of climate change impact on water resources, Hydrol. Earth Syst. Sci., 17(2), 565–578, doi:10.5194/hess-17-565-2013, 2013.

Wade, N. L., Martin, J. and Whitfield, P. H.: Hydrologic and climatic zonation of Georgia Basin, British Columbia, Can. Water Resour. J., 26(1), 43–70, doi:10.4296/cwrj2601043, 2001.

Wagener, T. and Gupta, H. V.: Model identification for hydrological forecasting under uncertainty, Stoch. Environ. Res. Risk Assess., 19(6), 378–387, doi:10.1007/s00477-005-0006-5, 2005.

Water Survey of Canada: HYDAT database. [Available online at http://www.ec.gc.ca/rhc-wsc/.], 2014.

Wen, L., Lin, C. A., Wu, Z., Lu, G., Pomeroy, J. and Zhu, Y.: Reconstructing sixty year (1950-2009) daily soil moisture over the Canadian Prairies using the Variable Infiltration Capacity model, Can. Water Resour. J., 36(1), 83–102, doi:10.4296/cwrj3601083, 2011.

Whitfield, P. H., Moore, R. D., Fleming, S. W. and Zawadzki, A.: Pacific Decadal Oscillation and the hydroclimatology of Western Canada—Review and prospects, Can. Water Resour. J., 35(1), 1–28, doi:10.4296/cwrj3501001, 2010.

Wilby, R. L. and Harris, I.: A framework for assessing uncertainties in climate change impacts: Low-flow scenarios for the River Thames, UK, Water Resour. Res., 42(2), W02419, doi:10.1029/2005WR004065, 2006.

Woo, M. K. and Thorne, R.: Snowmelt contribution to discharge from a large mountainous catchment in subarctic Canada, Hydrol. Process., 20, 2129–2139, 2006.

Wu, H., Kimball, J. S., Mantua, N. and Stanford, J.: Automated upscaling of river networks for macroscale hydrological modeling, Water Resour. Res., 47(3), W03517, doi:10.1029/2009WR008871, 2011.

Yapo, P. O., Gupta, H. V. and Sorooshian, S.: Multi-objective global optimization for hydrologic models, J. Hydrol., 204(1–4), 83–97, doi:10.1016/S0022-1694(97)00107-8, 1998.

Zhou, T., Nijssen, B., Gao, H. and Lettenmaier, D. P.: The contribution of reservoirs to global land surface water storage variations, J. Hydrometeorol., 309–325, doi:10.1175/JHM-D-15-0002.1, 2016.

Ziegler, A. D., Sheffield, J., Maurer, E. P., Nijssen, B., Wood, E. F. and Lettenmaier, D. P.: Detection of intensification in global- and continental-scale hydrological cycles: Temporal scale of evaluation, J. Clim., 16(3), 535–547, doi:10.1175/1520-0442(2003)016<0535:DOIIGA>2.0.CO;2, 2003.

Table 1: Description of VIC inter-comparison experiments performed using observational forcings.

VIC model Driving data	Time Span	Ensemble Runs	VIC configuration	Description
The Canadian Precipitation Analysis and the thin-plate smoothing splines (**ANUSPLIN**, Hopkinson et al., 2011)	1979-1990 (Cal.) 1991-2006 (Val.)	5	Domain = 48°– 55°N and 119°–131°W Resolution = 25 km × 25 km, Time Step = Daily Soil Layers = 3 Vertical elevation bands = 10	Validation runs are initiated 5 times with different initial conditions.
North American Regional Reanalysis (**NARR**, Mesinger et al., 2006)				
Pacific Climate Impacts Consortium (**PCIC**, Shrestha et al., 2012)				
University of Washington (**UW**, Shi et al., 2013)				

Table 2: Daily pPerformance metrics for the VIC inter-comparison runs. Calibration (1979–1990) and validation (1991–2006) for Fraser River main stem at Hope, BC is evaluated using the Nash–Sutcliffe Efficiency (NSE) coefficient and correlation coefficient (r, all statistically-significant at $p < 0.05$).

Experiments	**1979-1990 Daily Calibration**		**1991-2006 Daily Validation**	
	NSE	**r**	**NSE**	**r**
ANUSPLIN-VIC	0.54	0.91	0.55	0.94
NARR-VIC	0.67	0.85	0.81	0.90
PCIC-VIC	0.90	0.96	0.90	0.95
UW-VIC	0.82	0.94	0.80	0.92

Table 3: Parameters used to optimize during the calibration process for mean daily runoff for the Fraser River at Hope. OPT1, OPT2, OPT3, OPT4 and OPT5~~3~~ are different experiments using same forcing data but with different initial range for each calibration parameter.

VIC model Calibration Parameters *(units)*	**Description**	**Initial Range (Final Optimized Parameters)**				
		Experiment OPT1	**Experiment OPT2**	**Experiment OPT3**	**Experiment OPT4**	**Experiment OPT5**
b_inf	Controls the partitioning of precipitation (or snowmelt) into surface runoff or infiltration	0.2-0.00001 **(0.07)**	0.3-0.00001 **(0.16)**	0.25-0. 10 **(0.10)**	0.1-0.0001 **(0.08)**	0.16-0.12 **(0.12)**
Ds	Fraction of maximum baseflow velocity	0.1-0.000001 **(0.05)**	0.9-0.00001 **(0.09)**	0.30-0.04 **(0.05)**	0.6-0.0001 **(0.19)**	0.09-0.03 **(0.05)**
Ws	Fraction of maximum soil moisture content of the third soil layer at which nonlinear baseflow occurs	0.6-0.20 **(0.33)**	1.0-0.1 **(0.49)**	0.65-0.20 **(0.50)**	0.5-0.3 **(0.42)**	0.35-0.20 **(0.31)**
D2 (m)	The second soil layer thicknesses, which affect the water available for transpiration	1.0-0.7 **(0.82)**	3.0-0.7 **(1.02)**	0.80-0.70 **(0.76)**	2.8-1.0 **(1.07)**	0.80-0.70 **(0.78)**
D3 (m)	The third soil layer thicknesses, which affect the water available for baseflow	2.5-0. 7 **(1.66)**	5.5-0.7 **(2.70)**	2.00-1.00 **(1.82)**	3.0-1.0 **(1.38)**	1.8-1.2 **(1.76)**
Dsmax (mm day-1)	Maximum baseflow velocity	18.0-12.0 **(16.0)**	30.0-12.0 **(22.71)**	23.0-12.0 **(14.28)**	18-12 **(16.22)**	16-13 **(14.11)**
Monthly NSE	-	**0.93**	**0.84**	**0.92**	**0.89**	**0.91**

Table 4: Daily Pperformance metrics for the UW forcing driven PDO runs. Calibration and validation for Fraser River main stem at Hope, BC is evaluated using the NSE coefficient using the dataset. See text for the detail of PDO experiments.

Experiment Name	**Calibration**		**Validation**	
	NSE (Time period)	PDO Phase (Flows)	NSE (Time period)	PDO Phase (Flows)
PDO1	0.84 (1981-1990)	Warm (low flows)	0.84 (1991-2001)	Warm (low flows)
PDO2	0.84 (1957-1966)	Cool (high flows)	0.85 (1967-1976)	Cool (high flows)
PDO3	0.84 (1967-1976)	Cool (high flows)	0.79 (1977-1987)	Warm (low flows)
PDO4	0.86 (1977-1987)	Warm (low flows)	0.80 (1967-1976)	Cool (high flows)
PDO5	0.89 (1991-2001)	Warm (low flows)	0.87 (1981-1990)	Warm (low flows)

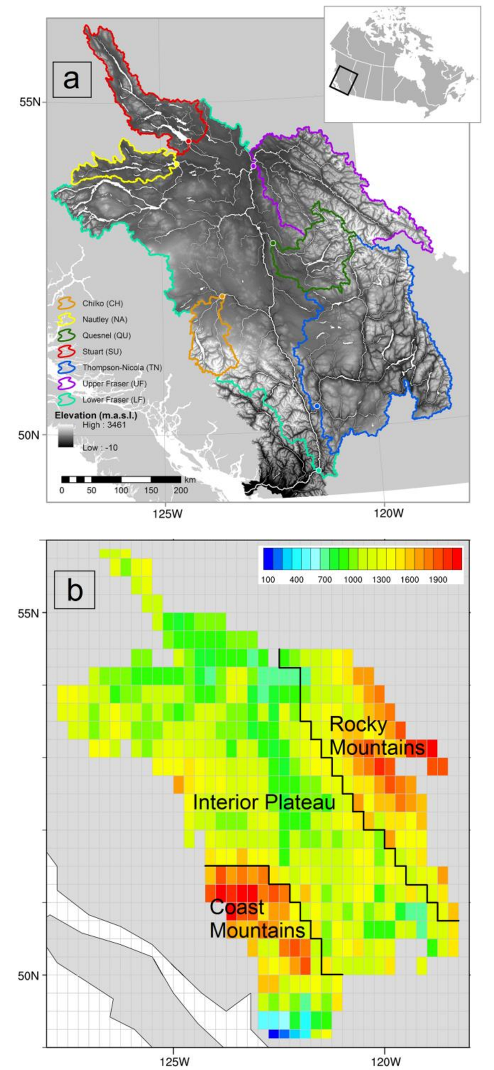


Figure 1: (a) High-resolution digital elevation map of the FRB with identification of major sub-basins including the Fraser River main stem at Hope, BC. (b) FRB mean elevation (m) per VIC model gridcell. The location of the hydrometric gauge on the Fraser Rivers' main stem at Hope is highlighted in both plots with ~~as~~ black circles.

Formatted: Centered

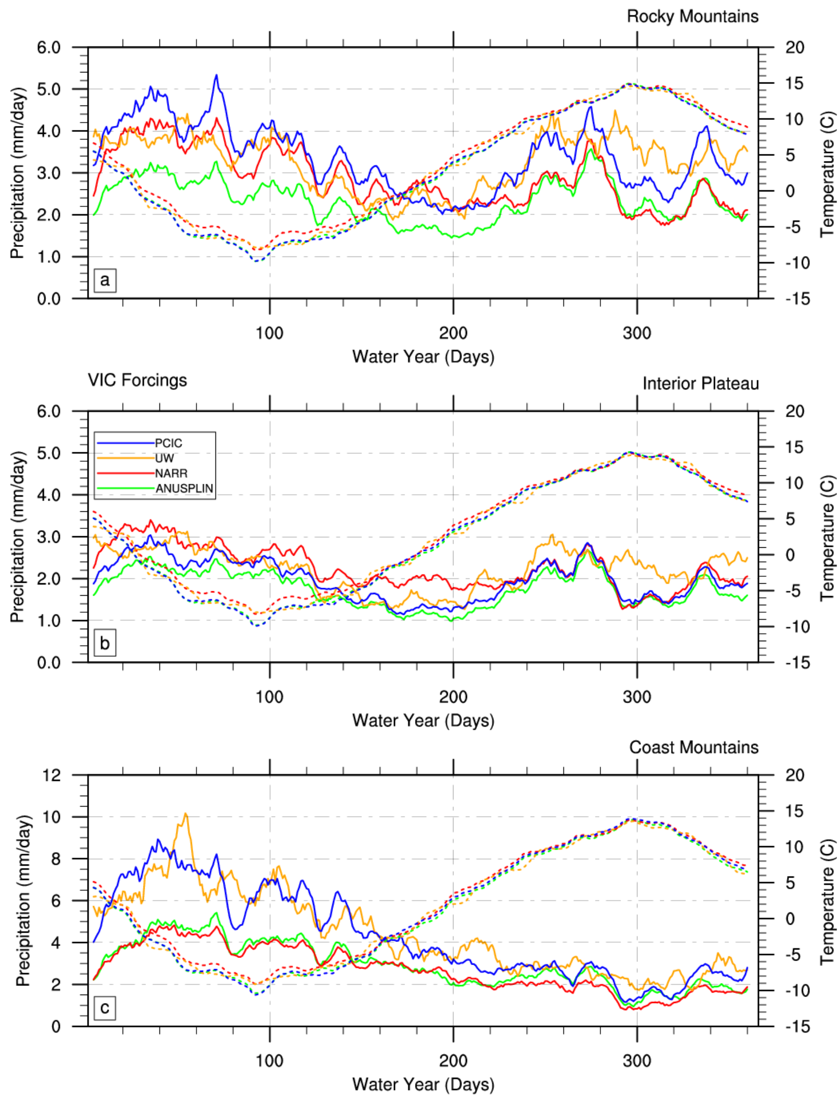


Figure 2: Area-averaged time series of mean daily air temperature (dotted lines) and daily precipitation (solid lines) over the (a) Rocky Mountains, (b) Interior Plateau, and (c) Coast Mountains for the ANUSPLIN, NARR, UW and PCIC forcing datasets, water years 1979-2006. **Water year starts on 1 October and ends on 30 September of the following calendar year.**

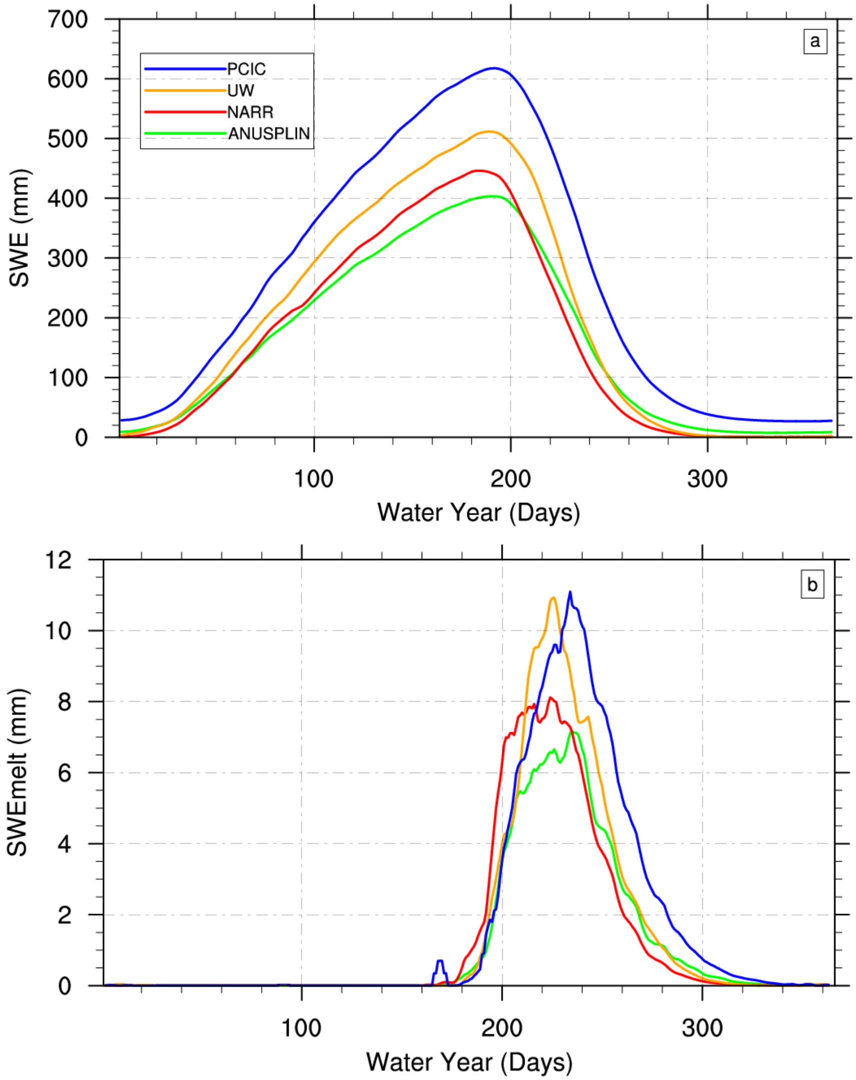


Figure 3: Area-averaged time series of daily mean (a) SWE and (b) SWE_{melt} for the ANUSPLIN-VIC, NARR-VIC, UW-VIC and PCIC-VIC simulations averaged over the Rocky Mountains, water years 1979-2006. Water year starts on 1 October and ends on 30 September of the following calendar year.

Formatted: Subscript

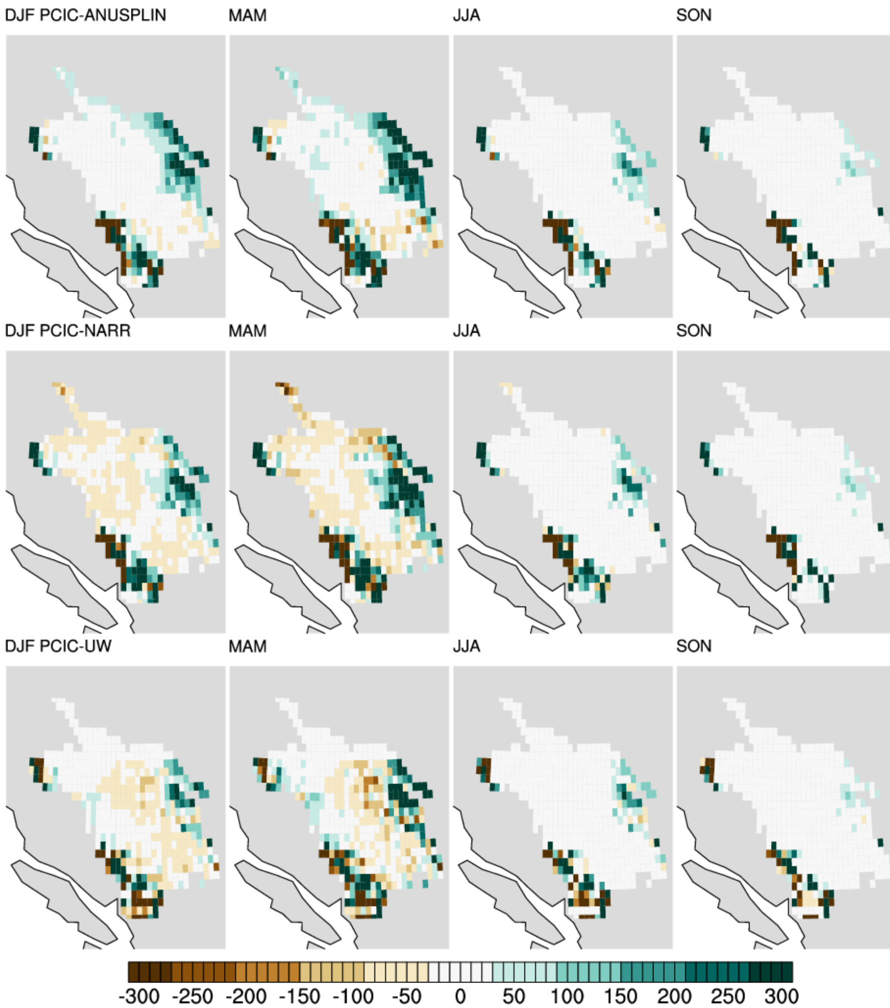


Figure 4: Spatial differences of mean seasonal SWE (mm) based on PCIC-VIC minus ANUSPLIN-VIC (1st row), NARR -VIC (2nd row) and UW (3rd row) simulations, water years 1979-2006. DFJ, MAM, JJA and SON correspond to winter, spring, summer and autumn, respectively.

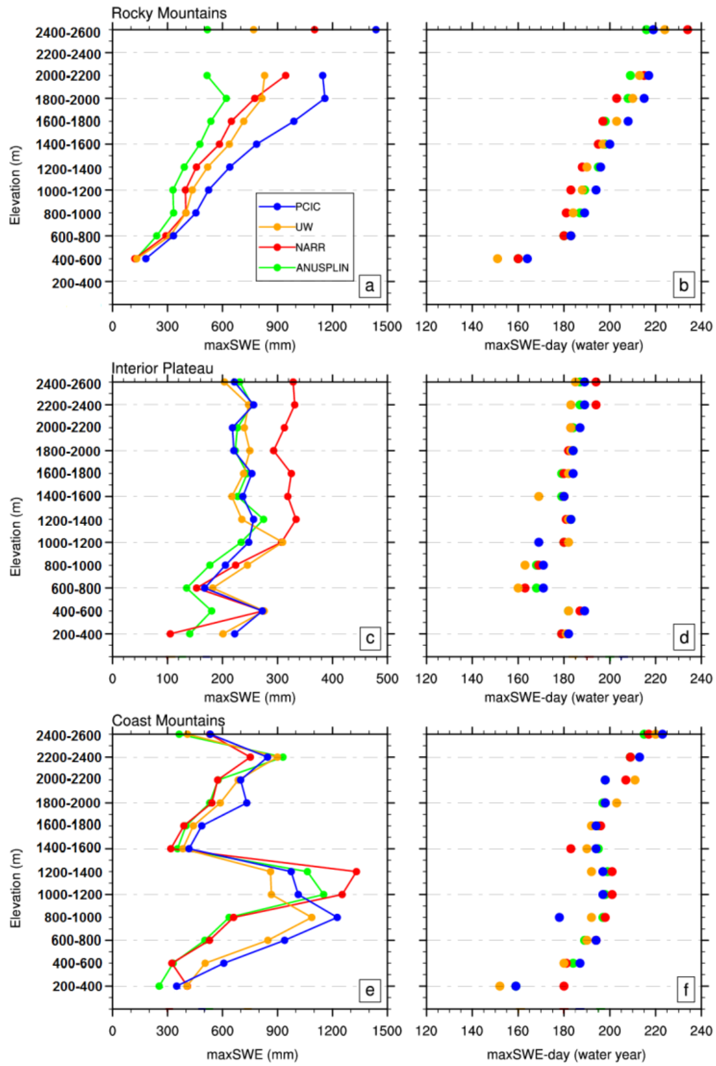
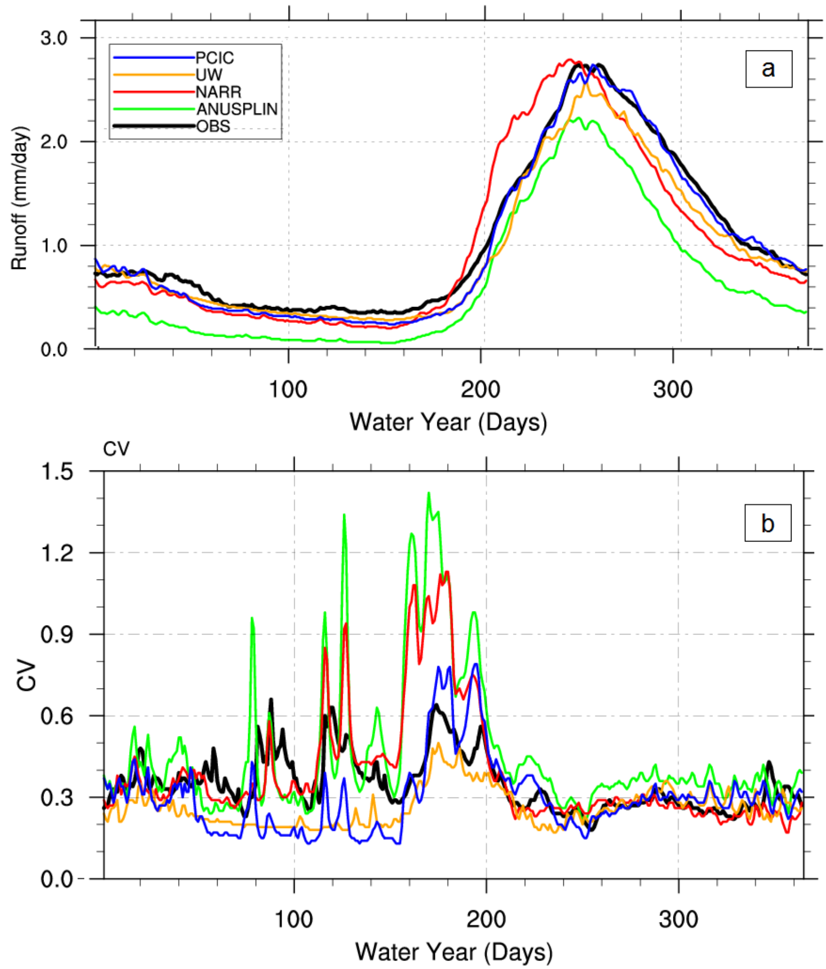


Figure 5: Variation of (a, c, e) maxSWE and corresponding (b, d, e) maxSWE-day for the ANUSPLIN-VIC, NARR-VIC, UW-VIC and PCIC-VIC simulations averaged over the (a, b) Rocky Mountains, (c, d) Interior Plateau and (e, f) Coast Mountains, water years 1979-2006.



Formatted: Font: Italic

Figure 6: The simulated and observed daily (a) runoff and (b) coefficient of variation (CV) for the Fraser River at Hope averaged over water years 1979-2006. An external routing model is used to calculate runoff for the ANUSPLIN-VIC, NARR-VIC, UW-VIC and PCIC-VIC simulations. Water year starts on 1 October and ends on 30 September of the following calendar year.

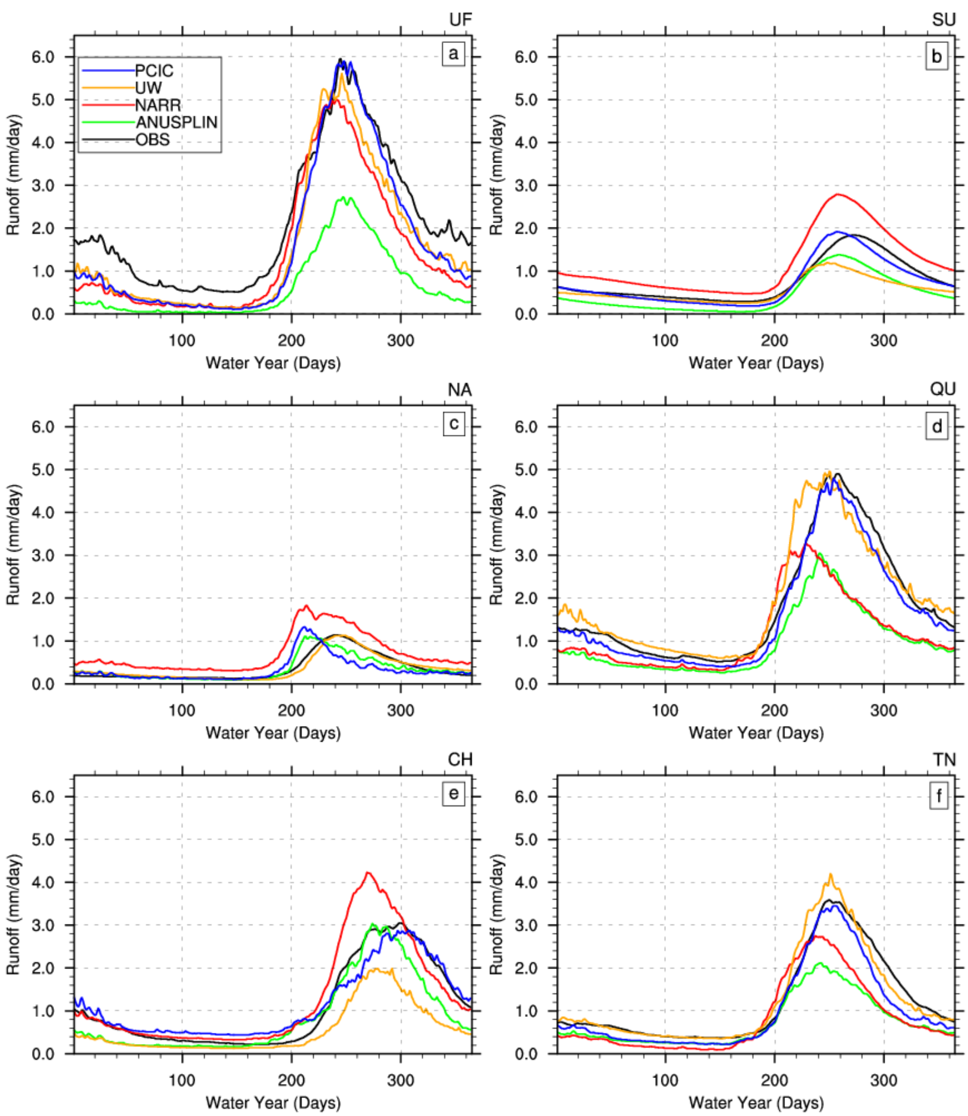


Figure 7: Same as Figure 6a but for the FRB's six major sub-basins (a) Fraser-Shelley (UF), (b) Stuart (SU), (c) Nautley (NA), (d) Quesnel (QU), (e) Chilko (CH) and (f) Thompson-Nicola (TN).

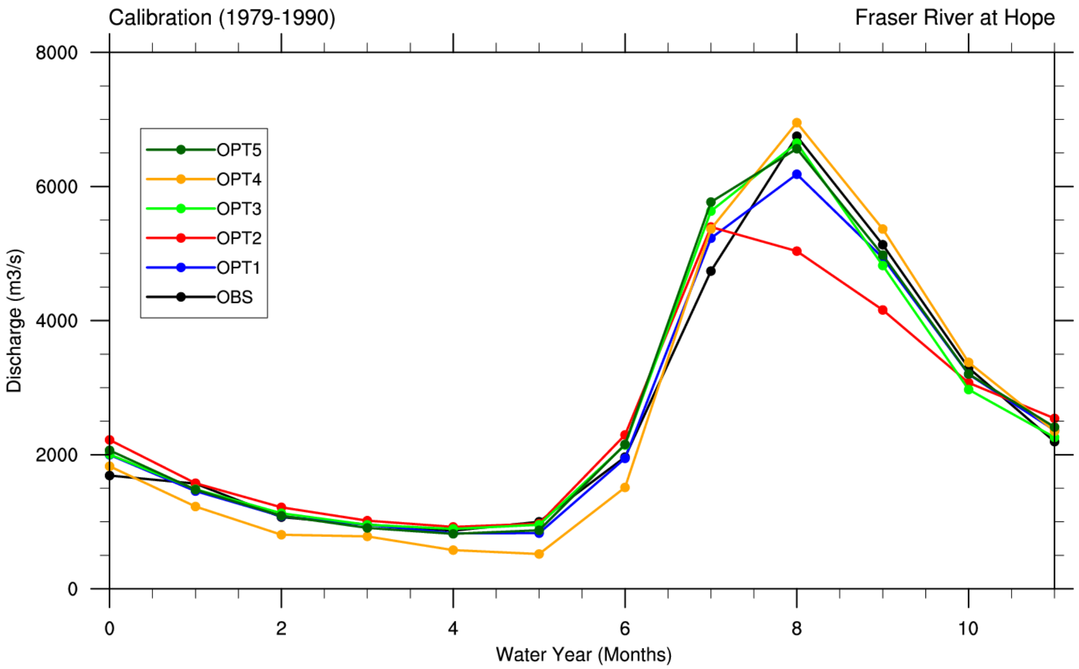


Figure 8: UW-VIC simulations using five different parameter sets (labelled as OPT1, OPT2, OPT3, OPT4 and OPT5, see text and Table 3 for details) are compared for mean monthly discharge for the Fraser River at Hope during calibration period 1979-1990. The black curve represents observed monthly discharge.Figure 8: UW-VIC simulations using five different parameter sets (labelled as OPT1, OPT2, OPT3, OPT4 and OPT5, see text and Table 3 for details) are compared for mean monthly runoff for the Fraser River at Hope.

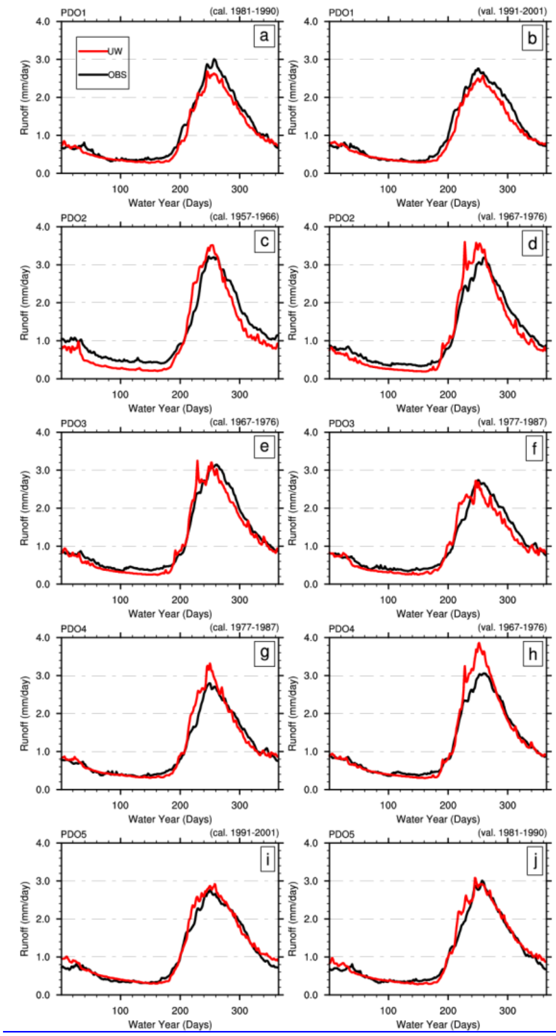


Figure 9: UW-VIC simulated daily runoff during calibration (cal.) and validation (val.) for the Fraser River at Hope, BC. PDO1, PDO2, PDO3, PDO4 and PDO5 refer to the VIC experiments performed during different experimental setups (see text and Table 4 for details). Water year starts on 1 October and ends on 30 September of the following calendar year.

Formatted: Centered

Figure 9: UW-VIC simulated daily runoff during calibration (cal.) and validation (val.) under different PDO phases for the Fraser River main stem at Hope, BC. PDO1 (cal. 1981-1990 val.1991-2001), PDO2 (cal. 1956-1965 val.1966-1976) and PDO3 (cal. 1967-1976 val.1977-1987) refer to VIC experiments performed during warm, cool and cool/warm PDO phases (see text and Table 4 for details). Water year starts on 1 October and ends on 30 September of the following calendar year.