RC1 Responses

Thank you to the referee for their constructive review and comments. No major objections to comments and review. Manuscript will be updated with the following revisions.

Pg 1. Line 22. Will add sentence:

Study results were unable to demonstrate statistically significant correlations ($p<0.05$) between measured, global hydrological model and GlobSnow-SWE to snowmelt runoff volume or peak discharge.

Pg 1. Line 26: Will update closing sentence to:

This study demonstrates the operational and scientific utility of the global re-analysis datasets in the Sub-Arctic, although knowledge gaps remain in global satellite based datasets for snowpack representation, for example the relationship between passive microwave measured SWE to snowmelt runoff volume.

Pg 2. Line 19: Modify text to:

Precipitation gauge measurements to quantify snowfall at high latitudes have high uncertainty due to the scarcity of meteorological stations, short duration of meteorological measurement records and systematic measurement error (Devine & Mekis, 2008; Mekis & Vincent, 2011; Sugiura et al., 2006).

Pg 3. Line 9: Modify text to:

Research into the reliability of re-analysis products at high latitudes is, however, limited due to a lack of reliable precipitation and SWE data (Mudryk et al. 2015; Wong et al., 2016).

Pg 3. Line 10: The intent here is local (not locally as suggested) in contrast to global models. Will not modify.

Pg 3. Line 15: Correct! Thank you. Modify text from 'but' to 'and'

Pg 3. Line 19: Noted. Modify from 'real and practical' to 'practical'

Pg 3. Line 24: The intent here is local (not locally) as suggested in contrast to global models. Will not modify.

Pg 4. Line 18: Modify text to:

Sublimation, the direct conversion of snow particles to vapour, is a major factor in removing snow from tundra areas (Marsh et al., 1995) and along with wind redistribution is a key driver of spatial variability and quantity of SWE.

Pg 4. Line 27: Modify text to:

Current approaches for hydropower operations in the Snare Watershed use ground SWE measurements and matching with historical discharge records with similar flow characteristics to anticipate discharge.

Pg 5, Line 17: Modify text to:

Historical discharge data was separated into calibration, validation and testing periods.

Pg 6, Line 8: No conflict, Pg 4. Line 27 has been modified

Pg 6, Line 11: Reference added for Penman Monteith (Allen et al, 1998)

Allen, R. G., Pereira, L. S., Raes, D., and Smith, M.: Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56, FAO, Rome, 300, 6541, 1998.

Pg 7, Line 11: Good catch, is unclear. Modify text to:

Secondary hydrograph peaks which occurred after the freshet peak and are driven by late-season rainfall events were removed in the snowmelt volume calculation. The separation of rainfall driven flow increases was performed using a simple exponential regression to estimate the regression curve from the spring melt hydrograph (Toebes et al., 1969) .

Pg 8, Line 14: Due to consistent period of record, can clarify with addition of text as below.

The period of record for all rank correlation analysis was 1985 to 2012.

Pg 8, Line 15: Add text 'or acceptable'

Pg 8, Line 15: Add text and references

Nash-Sutcliffe Efficiency (NSE) values can be on the range of $-\infty$ to 1 where 1 indicates the ideal with no difference between simulated and observed values. (Nash & Sutcliffe, 1970). Percent Bias (PBIAS) gives a measure of the tendency of the simulated results to be larger or less than the observed values. RMSE-observations standard deviation ratio (RSR) has the benefit of with a normalization and scaling factor which facilities comparison (Moriasi et al., 2007). Evaluation using KGE is similar to NSE with an ideal optimized value of 1. (Gupta et al, 2009)

Gupta, H. V.,Kling, H., Koray, Y., & Martinez, G. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology, 377, 80–91.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I —A discussion of principles. Journal of Hydrology, 10(3), 282-290. doi: http://dx.doi.org/10.1016/0022-1694(70)90255-6

Pg 12, Line 4 Will move to recommendations. Text updated as below.

Data products available in Near-Real Time such as MSWEP-NRT, which is a variant of the historic MSWEP dataset, can be similarly applied to as model forcing in remote regions (. Using Delft-FEWS, scheduled model runs can be used to keep model states current and generate regularly scheduled hydrological forecasts (Beck et al, 2017).

Reference added.

Beck, H., van Dijk, A., Leviizzani, V., Schellekens, J., Miralles, G., Martrens, B., de Roo, A., Pappenberger, F., Huffman, G., Wood, E. (2017) MSWEP: 3-hourly 0.1◦ fully global precipitation (1979–present) by merging gauge, satellite, and weather model data [Abstract] Geophysical Research Abstracts, Vol. 19, EGU2017-18289

Pg 12, Line 24 Remove ‘real and’

Pg 12, Line 30 Remove Canadian

Pg 17, Figure 1 Noted, will increase contrast

Pg 18, Figure 4 Noted, will update figure according to recommendation

RC2 Responses

Thank you to the referee for their constructive review and comments. No major objections to comments and review. Manuscript will be updated with the following revisions.

Referee Comment 1: Calibration, validation, testing: (page 5). Please clarify the observations used over the different periods “calibration, validation and testing”. The calibration period is fully automatic? the validation include some human intervention? How was the length of each period decided ?

Author Response 1:

Discharge observation calibration, validation and testing periods are presented in Table 1.

Table 1: Calibration, Validation and Testing Periods

Catchment	**Calibration**	**Validation**	**Testing**
Catchment 1: Indin River Above Chalco Lake	2000-2009	1978-1999	2010-2014
Catchment 2: Snare River Above Indin Lake	2000-2004	1998-1999, 2005-2010	2010-2014
Catchment 3: Snare River Above Ghost River	2000-2009	1984-1999	2010-2014

The calibration period is fully automated using the ALSHO optimization algorithm for single objective optimization of NSE. The model calibration process was somewhat iterative, including initial calibration of a lumped HBV model, some separation of land use type and the addition of a lake-reservoir to the wflow hbv model. Calibration and validation results were available to the modeller throughout this process, and while they were not optimized, were available and known as an implicit indication of model performance.

The length of each period was decided based on the overlapping period and more limited time in Catchment 2: Snare River Above Indin Lake. The calibration period may be considered relatively short, but was found in the model calibration process to produce good calibration and validation results. This shorter period was also a trade-off due to the computationally intensive calibration of the distributed wflow hbv model with a global optimization algorithm. The calibration period was also considered to have representative spring peak events suitable for training the model.

Resolution 1: The above table will be added with the following text:

Calibration period as shown in Table 1 was selected to correspond with available discharge data in each catchment, representative peak flow events and to allow sufficient additional discharge data for validation and testing of the model.

Referee Comment 2: Results in table 3: Which period was considered for the scores calculation in table 3? If the full period was considered, WFLOW-HBV has a clear advantage since most of the period was used in the calibration. If this is the case, please clarify indicating clearly the period used in the validation.

Author Response 2:

As this is a valid consideration, re-analysis of the data was performed to see if inclusion of calibration period data gave a clear advantage to the wflow-hbv model.

The period used for rank correlation analysis was 1985-2012. The WFLOW-HBV model was calibrated over 18.5% (5 years) to 37.0% (10 years) of the rank correlation analysis coverage.

It seems intuitive that rank correlation of the WFLOW-HBV model is increased by calibration of model parameters which affect the maximum annual modelled SWE. The calibration factors affecting SWE accumulation include but are not limited to Interception (ICF), Snowfall Correction Factor (SFCF) and the limit temperature for rain/snow (TT). However, rank correlation analysis is based on monotonicity, not on the magnitude. Calibration directly results in the Improve matching of spring melt volume to measured data, but has less influence on the inter-annual variability of SWE.

When the calibration period (2000-2009) is removed from the rank correlation analysis, the skill of the wflow hbv is notable reduced, as are the other MSWEP forced models as indicated in Table 1 below. This on its own would indicate the calibration was responsible for the stronger rank correlation and that is does not hold up well to validation.

However if we look at the calibration period only, while the wflow hbv skill is improved, all the MSWEP forced models are improved even more so. This is displayed in Table 2 below.

Table 2: Rank Correlation Analysis of Streamflow Contribution to Snowmelt dependent on inclusion of calibration period

Model	Forcing Data	Full Record		No Calibration Period		Only Calibration Period	
		Spearman	p	Spearman	p	Spearman	p
wflow	MSWEP	0.52	0.004	0.29	0.216	0.66	0.004
Ground	Field	0.37	0.056	0.41	0.111	0.35	0.489
HTESSEL	MSWEP	0.47	0.076	0.28	0.291	0.79	0.187
JULES	MSWEP	0.47	0.382	0.25	0.553	0.76	0.803
WaterGap	MSWEP	0.34	0.451	0.23	0.499	0.54	0.676
WaterGAP	WFDEI	0.17	0.465	0.10	0.581	0.15	0.676
W3RA	WFDEI	0.15	0.012	0.12	0.229	0.07	0.060
PCRGLOB	WFDEI	0.14	0.243	0.11	0.565	-0.04	0.162
JULES	WFDEI	0.23	0.011	0.10	0.176	0.24	0.033
HTESSEL	WFDEI	0.25	0.193	0.09	0.559	0.35	0.150
GlobSnow	Data	0.14	0.484	-0.18	0.438	0.42	0.003

Table 3: Change in Spearman Correlation due to exclusion or isolation of calibration period (2000-2009)

Model	Forcing Data	Snowmelt Contribution to Streamflow		Peak Discharge	
		Spearman Change - Excluding calibration	Change in Spearman - Calibration period only	Change in Spearman - Excluding calibration period	Change in Spearman - Calibration period only

		period			
wflow	MSWEP	-0.24	0.14	-0.23	0.28
Ground	Field	0.04	-0.02	0.06	-0.08
HTESSEL	MSWEP	-0.22	0.29	-0.18	0.13
JULES	MSWEP	-0.19	0.32	-0.14	0.19
WaterGap	MSWEP	-0.11	0.20	-0.09	0.10
WaterGAP	WFDEI	-0.07	-0.02	0.02	-0.22
W3RA	WFDEI	-0.03	-0.08	0.07	-0.25
PCRGLOB	WFDEI	-0.04	-0.19	0.02	-0.27
JULES	WFDEI	-0.12	0.01	-0.08	0.25
HTESSEL	WFDEI	-0.16	0.09	-0.10	0.24
GlobSnow	Data	-0.32	0.28	-0.37	0.65

These results indicate that MSWEP is particularly predictive over the calibration period. The wflow hbv Spearman correlation is improved by the calibration period, but the incremental gain is on the lower range for all MSWEP forced models (0.24 compared to 0.29, 0.32 and 0.20 respectively).

This additional analysis suggest that the wflow-hbv is not given a distinct advantage by inclusion of data used to calibrate the model in the rank correlation dataset This is not to say calibrating the wflow-hbv does not improve the rank correlation performance. It can improve the physical representation of some physical processes (interception, rain/snow interfaces) or measurement biases (Snowfall Correction Factor). This incremental improvement from calibration has not been quantified.

Resolution 2:

Additions to made to supplemental data:

1) Rank Correlation Analysis Data Summary containing all annual maximum SWE, snowmelt contribution to streamflow and peak discharge
2) Calibration parameter ranges and results will also be added to supplementary material.

The following text will be added to Results – Section 4.3 Prediction of Snowmelt Volume and Peak Discharge.:

The period used for rank correlation analysis was 1985-2012 meaning the WFLOW-HBV model was calibrated over 18.5% (5 years) to 37.0% (10 years) of the rank correlation analysis time period. The higher Spearman coefficient performance of the wflow hbv model in rank correlation analysis may be partly attributed to improved process representation of snow accumulation and removal processes including interception and precipitation biases. The quantification of the improvement on inter-annual variability and rank correlation due to correlation has not been investigated in this study. The dominant driver of the rank correlation analysis is the choice of forcing meteorological data.

The following text will be added to 5.1 Global re-analysis datasets for predicting streamflow, snowpack accumulation and melt, pg 10, line 13:

Calibration of the local watershed model

Referee Comment 3: How was the snowmelt volume and peak discharge calculated from the global models ? Was it also estimated using the survey SWE observations ? If yes, which methodology was used ?

Author Response 3:

The snowmelt volume and peak discharge are from measured discharge data only. This should be made clear.

Resolution 3:

Rank Correlation Analysis Data Summary containing all maximum annual SWE, snowmelt contribution to streamflow and peak discharge will be added to supplementary material.

The following text will be added to Results – 3.3 Prediction of Snowmelt Volume and Peak Discharge from Maximum Annual SWE.

The snowmelt volume and peak discharge are calculated from measured discharge at downstream Catchment 3 outlet.

Referee Comment 4: It is not clear from the results the affirmation in the discussion (Pag 10 L11:) "Local model maximum annual SWE was found to be a better predictor of snowmelt volume and peak discharge than snowpack survey data." Please clarify this point. Also on this point, later another sentence suggests a similar results (pag 10, L28): "Study results demonstrate that SWE . . . in the Snare Watershed". In both cases, it is suggested that SWE observations are not as good predictors of the local model. Then latter in the discussion: (pag 12 L 16) "as the study shows that ground data is a . . . and peak discharge". Please clarify this point, as this is crucial in this study: Is there an added value of the local model when compared with the survey SWE data in predicting snowmelt volume and peak discharge? (see also the previous comments on clarifying the methodology used)

Author Response 4:

Thanks for this comment, it does require further clarification. The take home message is that neither methods are ideal, and that it is better to consider both in concert than to rely on one exclusively. Data assimilation can provide a means to optimally merge field observations and model states, which knowledge of their corresponding uncertainties.

Resolution 4:

The paragraph at Pag 10 L11 will be changed to:

The local watershed model in this study, forced with global re-analysis datasets and calibrated to available streamflow records is able reliably and accurately model streamflow based on calibration, *validation and testing statistical results. The wflow model is conceptual and has limited representation of physical snow processes, however the modelled maximum annual SWE was found to be a better predictor of snowmelt volume and peak discharge than snowpack survey data.*

Assimilation of snowpack survey data for model state update has the potential to improve SWE estimate and optimally use available information. Data assimilation requires estimates of both model state and observational uncertainty, quantification of which would improve understanding to the relative reliability and applicability of data sources (Liu et al.,2012).

Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Restrepo, P. (2012). Advancingdata assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities, .Hydrol. Earth Syst. Sci.,, 16, 3863-3887. doi: doi:10.5194/hess-16-3863-2012

The paragraph at Pag 10 L11 will be changed to:

SWE is used by operational water managers to predict the inflow volumes from snowmelt and anticipate peak discharge rates. Study results demonstrate that SWE measurement for application in hydrological forecasting is still problematic in the Snare Watershed. Consideration of multiple data sources and methodological improvement of data collection can be used to update model states.

The paragraph at Pag 12 L 16 will be changed to:

The manual collection of end of-winter snowpack survey data is justified, as the study shows that ground data is a comparatively reliable predictor of snowmelt contribution to streamflow and peak discharge. Field measurement improvements that exploit snow distribution across local topography can improvement the quality, frequency and predictive ability of ground measurement data. This data is optimally merged with model data using data assimilation methods (Sun et al., 2016).

Sun, L., Seidou, O., Nistor, I., & Liu, K. (2016). Review of the Kalman-type hydrological data assimilation. Hydrological Sciences Journal, 61(13), 2348-2366. doi: 10.1080/02626667.2015.1127376

In the conclusions, the final paragraph (pg 13) following will be changed to :

This study has demonstrated the utility of global re-analysis datasets for hydrological assessment in the data sparse Canadian Sub-Arctic. In the operational context of the Snare Hydro System, the length and breadth of hydrological assessment as presented here is much greater than could be achieved with local meteorological data. Further research can focus on the optimally merging of observed and modelled snow data to improve predictability of snowmelt volume and peak discharge. The continued development of these datasets and modelling frameworks is promising to help improve understanding of water resources in data sparse Northern regions in the face of climate change.

***A*dditional Comments:**

page 5, L6: “is a based” : “is based”

Noted. Will correct.

page 5, L16-17: page 5, L21: “are conceptual rainfall-runoff models”: The models listed in table 1 are not conceptual rainfall-runoff models. I suggest to change the sentence to: “A set of global hydrological and land-surface models were considered in this study and presented in table 1”

Noted. Will correct.

pag 7, L20:25: Despite a different region in Canada, Snauffer et al (2016) also evaluated several reanalysis and GLOBSNOW. It is worth to cite this paper that also higlithed some limitations of GLOBSNOW.

Thanks for the paper, reference to be added.

Table 2: Please define KGE, PBIAS, RSR

Definitions will be added.

Figure 3: Please add panel names (e.g. a, b c). In the lower panels the blue line refers to GlobSnow or WaterGap ?

Noted, panel names will be added. Lower panel blue line refers to WaterGap.

Figure 4: Also add panel names (e.g. a, b). In addition to the scores in table 3, the scatter plots comparing observed SWE vs. model SWE would be also informative. I suggest to also include these plots (if too much in the main article, at least as supplementary information).

Recommended figures will be added to supplementary materials.

Figure 5 and pag 9 L16: How were the observations interpolated to the 25km grid ?

Observations were interpolated to the 25 km grid using inverse distance to a power interpolation algorithm. More specifically, using the gdal_grid interpolation algorithm invdist with settings weighting power = 2.0, smoothing parameter = 1.0.

The following sentence above will be added to the manuscript.

Observations were interpolated to the 25 km grid using inverse distance to a power interpolation algorithm.

Pag 9, L27: As mentioned above, please avoid the term “conceptual model” here and in other locations. In several places replace please check the usage of “Study results” E.g. pag 10 L28 “Study results” should be “This study demonstrates. . .” or “Our results suggest that..” Also in pag 11 L20 “Study results”

Noted. The text will be reviewed for these terms and updated.