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 to estimate prior parameter ranges, some separation of land use type and the addition of a lake-reservoir routine to the wflow hbv model as routing through lakes is an important process in the case study area. 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 of available data, and more limited time span of available data 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 include representative spring peak events suitable for training the model.

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

The calibration period as shown in Table 1 was selected to correspond with available discharge data in each catchment, include representative peak flow events and 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:

We thank the author for this important comment. As this is a valid consideration, re-analysis of the data was performed to see if inclusion of calibration period data indeed gives a clear advantage to the wflow-hbv model.

The period used for the 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 (depending on the sub-catchment).

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 notably 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 it does not hold up well to validation.

However if we look at the calibration period only, while the skill of wflow hbv is improved, all the MSWEP forced models are improved even more so. This is shown 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		Excluding 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 -	Change in Spearman -	Change in Spearman -	Change in Spearman -

		Excluding calibration period	Calibration period only	Excluding calibration period	Calibration period only
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 models forced with MSWEP are particularly predictive over the calibration period. While the wflow hbv Spearman correlation is improved for the calibration period, the incremental gain is on the lower range when compared to all MSWEP forced models (0.24 compared to 0.29, 0.32 and 0.20 respectively).

This additional analysis suggests 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 that 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:

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

The rank correlation analysis period was 1985-2012 and the data used is provided in the supplemental material. This means that the WFLOW-HBV model was calibrated over 18.5% (5 years) to 37.0% (10 years) of the time period considered in the rank correlation analysis, depending on the sub-catchment. The higher Spearman coefficient of the WFLOW-HBV model found in the rank correlation analysis may therefore be partly attributed to improved process representation of snow accumulation and removal processes including interception and precipitation biases. However, our analysis shows that the performance improvement over the calibration period for the models using the MSWEP forcing is similar, suggesting that the dominant driver of the rank correlation analysis is the choice of forcing meteorological data.

We propose the following additions to be 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.

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:

Peak discharge and snowmelt volume were determined from measured discharge data in the Snare Watershed. Global and local model discharges were not used for either calculation. This should be made clear.

Resolution 3:

The following text will be used in Results – 3.3 Snowmelt Volume based on RC1 and RC2 comments.

Snowmelt volume and peak discharge were calculated and extracted from the measured discharge data at the Catchment 3 outlet. No local or global model data was used in these calculations. Snowmelt volume was approximated using the local minimum method from the hydrograph stream flow separation program (HYSEP) implemented in MATLAB (Burkley, 2012). This is a mathematical technique that mimics manual methods for stream flow separation as opposed to an explicit representation of the physical processes (Sloto and Crouse, 1996). Secondary hydrograph peaks which occurred after the freshet peak and are driven by late-season rainfall events were removed in the snowmelt volume calculation. A simple exponential regression was used to estimate the recession curve from the snowmelt peak (Toebes et al., 1969).

Burkley, J. (2012). Hydrograph Separation using HYDSEP. Retrieved from http://nl.mathworks.com/matlabcentral/fileexchange/36387-hydrograph-separation-usinghydsep/content/f_hysep.m

Sloto, R.A., and Crouse, M.Y., 1996, HYSEP: A Computer Program for Streamflow Hydrograph Separation and Analysis: U.S. Geological Survey Water-Resources Investigations Report 1996–4040, 46 p., https://pubs.er.usgs.gov/publication/wri964040.

Toebes, C., Morrissey, W. B., Shorter, R., & Hendy, M. (1969). Base Flow Recession Curves Handbook of Hydrological Procedures. Wellington, New Zealand.

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

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; this does require further clarification. The take home message is that neither method is ideal, and that it is better to consider both in concert rather than to rely on one exclusively. Data assimilation can provide a means to optimally merge field observations and model states, with 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 to reliably and accurately model streamflow based on calibration, validation and testing statistical results. The WFLOW-HBV 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 as the Spearman coefficient is higher and p-value is lower ($p<0.05$).

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 to anticipate peak discharges. The results of this study demonstrate, however, 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. Improved field measurement techniques that exploit snow distribution across local topography could help further improveme the quality, frequency and predictive ability of ground *measurement data. This data could be 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 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, helping to improve the 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 in the revised manuscript.

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 in the revised manuscript.

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).

Panel names will be added. Additional scatter plots may be added to supplemental material however on basis of

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 weighting. 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 weighting.

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.

Supplementary Data 1:
Rank Correlation Analysis Data Summary

Date	Ground Data (mm)	GlobSnow (mm)	PCR-Glob (mm)	HTESSEL (mm)	WaterGAP (mm)	W3RA (mm)	wflow hbv (mm)	Peak Discharge (m^3/s)	Snowmelt Contribution to Streamflow (mm)
1985	100	120	128	121	93	134	106	180	112
1986	134	131	140	138	103	143	106	224	118
1987	107	122	117	109	85	114	91	136	81
1988	105	135	137	136	100	139	110	170	111
1989	88	125	89	87	66	91	101	101	78
1990	110	149	103	98	74	107	112	144	90
1991	111	190	137	138	100	144	113	255	114
1992	119	217	149	147	105	149	135	178	81
1993	118	149	127	119	92	118	161	200	119
1994	107	125	105	99	76	107	92	67	39
1995	108	155	146	148	108	149	113	76	53
1996	104	118	120	112	85	117	96	237	127
1997	95	108	89	92	67	83	90	309	117
1998	87	114	131	134	97	127	113	80	40
1999	110	103	110	104	80	109	117	335	122
2000	135	109	92	86	68	86	94	136	76
2001	143	130	109	102	78	106	142	306	139
2002	123	99	112	92	82	116	86	80	69
2003	84	93	80	74	59	76	82	58	40
2004	145	108	112	84	81	112	103	85	60
2005	121	126	97	81	69	96	130	210	109
2006	133	132	103	97	74	105	137	287	143
2007	127	127	113	100	81	112	112	113	66
2008	101	142	107	98	74	103	101	142	51
2009	99	113	111	100	82	109	98	141	82
2010	98	134	124	122	92	123	82	88	62
2011	84	125	92	84	61	81	76	64	49
2012	134	124	107	103	78	109	131	149	91
Mean	112	129	114	107	83	113	107	114	91
SD	17.3	25.5	17.9	20.7	13.3	19.8	20.0	17.9	15.1

Supplementary Data 2:
Calibration Ranges and Values for WFLOW-HBV model

Parameter	Description	Lower Bound	Upper Bound	Catchment 1	Catchment 2	Catchment 3
TT	Limit temperature for rain/snow precipitation	-2	2	-0.16	1.04	1.16
TTI	temperature threshold for linear mix of snow/rain precipitation	0	3	1.68	0.010	0.33
CFMAX	Degree day factor	0.4	4	1.60	1.89	2.14
FC	Field Capacity	40	300	125.7	117	50.7
ECORR	Evapotranspiration corrector factor	0.8	1.2	1.13	1.10	1.18
LP	Soil moisture value where soil moisture reaches maximum potential evapotranspiration	0.05	1	0.90	0.15	0.48
KHQ	Upper zone response coefficient	0.0001	0.05	0.028	0.024	0.022
K4	Lower zone response coefficient	0.0001	0.05	0.042	0.033	0.0041
ALPHA	Upper zone runoff coefficient	0.05	2	1.14	0.78	1.29
BETA	Contribution of the soil moisture to the response function	0.4	1	0.75	0.52	0.65
Beta Seepage	Exponent in soil runoff generation equation	0.4	2	1.63	1.87	1.31
WHC	Maximum amount of water that can be stored in snow pack	0.0001	0.2	0.032	0.041	0.181
CFR	Refreezing factor	0.01	0.3	0.14	0.17	0.18
CFLUX	Capillary Rise Rate	0.01	2	1.54	0.83	0.13
PERC	Percolation Rate	0.1	1	0.24	0.46	0.94
RFCF	Rainfall Correction Factor	0.8	1.2	0.80	0.83	0.86
SFCF	Snowfall Correction Factor	0.8	1.2	0.112	1.18	0.89
ICF	Interception (Water)	0	0	0	0	0
	Interception(Tree)	0	0.75	0.03	0.32	0.19
	Interception (Shrub/Tundra)	0	0.75	0.70	0.07	0.64
CEVPF	PET Correction (Water)	1.15	6	2.09	1.76	1.87
	PET Correction (Tree)	1.15	1.15	1.15	1.15	1.15
	PET Correction (Shrub/Tundra)	0.4	1	0.48	0.55	0.55