

Article

# Time Variant Sensitivity Analysis of Hydrological Model Parameters in a Cold Region Using Flow Signatures

Ajay Bajracharya <sup>1,2,\*</sup> , Hervé Awoye <sup>2</sup> , Tricia Stadnyk <sup>1,2</sup>  and Masoud Asadzadeh <sup>1</sup>

<sup>1</sup> Department of Civil Engineering, University of Manitoba, MB R3T 5V6, Canada; tricia.stadnyk@ucalgary.ca (T.S.); Masoud.Asadzadeh@umanitoba.ca (M.A.)

<sup>2</sup> Department of Geography, University of Calgary, Calgary, AB T2N 1N4, Canada; oyemonbade.awoye@ucalgary.ca

\* Correspondence: [ajay.bajracharya@ucalgary.ca](mailto:ajay.bajracharya@ucalgary.ca)

Received: 24 February 2020; Accepted: 26 March 2020; Published: 28 March 2020



**Abstract:** The complex terrain, seasonality, and cold region hydrology of the Nelson Churchill River Basin (NCRB) presents a formidable challenge for hydrological modeling, which complicates the calibration of model parameters. Seasonality leads to different hydrological processes dominating at different times of the year, which translates to time variant sensitivity in model parameters. In this study, Hydrological Predictions for the Environment model (HYPE) is set up in the NCRB to analyze the time variant sensitivity analysis (TVSA) of model parameters using a Global Sensitivity Analysis technique known as Variogram Analysis of Response Surfaces (VARS). TVSA can identify parameters that are highly influential in a short period but relatively uninfluential over the whole simulation period. TVSA is generally effective in identifying model's sensitivity to event-based parameters related to cold region processes such as snowmelt and frozen soil. This can guide event-based calibration, useful for operational flood forecasting. In contrast to residual based metrics, flow signatures, specifically the slope of the mid-segment of the flow duration curve, allows VARS to detect the influential parameters throughout the timescale of analysis. The results are beneficial for the calibration process in complex and multi-dimensional models by targeting the informative parameters, which are associated with the cold region hydrological processes.

**Keywords:** hydrological modeling; cold region processes; non-stationarity; variogram analysis of response surfaces; time variant sensitivity analysis

## 1. Introduction

Hydrological models are imprecise representations of real-world hydrological processes governed by mathematical equations and assumptions. Such imperfect representation of real-world hydrological processes often leads to uncertainty in model outputs, which are attributed to different sources of error including measurement error, model structural error and the model parameter uncertainty [1,2]. The model parameter uncertainty is attributed to a general lack of understanding required to accurately determine model parameters, and the fact that parameters compensate for errors in input data and model structure [1–3]. Hydrological models often have a large set of parameters linked to specific processes that dominate at different times of the year, leading to an increase in parameter uncertainty and an inherent seasonality effect of parameter sensitivity and uncertainty [4]. In general, most of the parameters associated with the hydrological processes are unknown and often require calibration to match the model output with the measured dataset(s).

Sensitivity analysis (SA) can assist the modeler in identifying the relative influence each model parameter has on the variation in model output. This information is typically used by the modeler

to limit the number of parameters in calibration, thereby increasing the efficiency and computation speed of the calibration process [5,6]. In practice, SA is useful for determining the most informative parameters prior to model calibration. The least influential parameters can be set to their default values and thus will not require calibration; therefore, SA is recommended before the calibration of a hydrological model is initiated [7].

SA can be broadly classified into two types: (i) Local Sensitivity Analysis (LSA), and (ii) Global Sensitivity Analysis (GSA) [8]. LSA measures the parameter sensitivity at a particular location in the parameter space. However, for complex hydrological models, GSA methods such as Sobol [9] and Morris [10] are recommended because they measure the parameter sensitivity over the entire parameter space, making them robust and reliable [11,12]. These methods, especially the Sobol method, require a large computational budget to reach convergence [13]. Alternatively, response surface or surrogate models can be used to reduce the computational burden of complex models, while maintaining the reliability and robustness of a GSA technique [14–17].

Variogram Analysis of Response Surfaces (VARS) is a recently developed GSA technique [13] that claims to be more computationally efficient and reliable than classical GSA approaches, for continental-scale hydrological modeling. Some recent studies have applied VARS for a detailed SA of model parameters in complex models. For example, Lihare et al. [18] chose Kling Gupta Efficiency (KGE), Nash Sutcliffe Efficiency (*NSE*) and Percent Bias (*PBIAS*) criterion in VARS to identify the most important parameters for the Variable Infiltration Capacity (VIC) model [19] applied to the Lower Nelson River Basin (LNRB). Sensitivity of the model to such parameters is generally dependent upon the evaluation criterion used, and hence the aforementioned study strongly recommends the use of multiple criteria for SA to gain a better understanding of the dynamics among and between model processes.

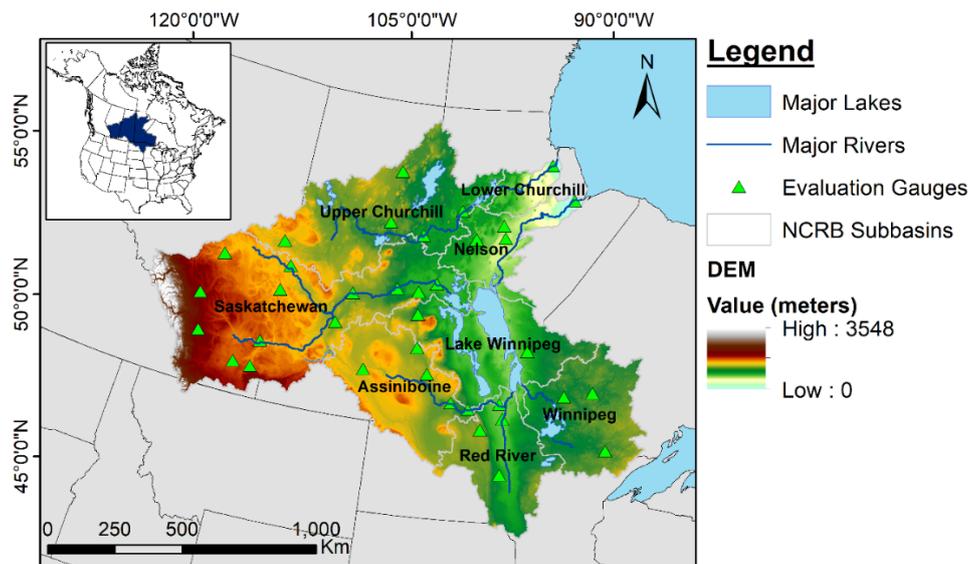
There are several studies in the cold region catchments within Canada to identify the sensitivity of the cold region processes. The catchments under study are diverse as the studies are conducted from the mountainous region of Yukon [20] to the Prairies of Saskatchewan [21]. These studies identify snowmelt runoff and frozen soil infiltration to be highly sensitive to warming temperature and rainfall intensity. Most previous SA studies, including those studies in the cold region, do not account for time varying sensitivity factor in the model parameters [13,22–24]. There are few recent studies in which SA has been implemented by dividing the entire time period into different time slices or window periods [5,25]. The studies that currently exist are based on single, conventional metrics such as *NSE* or Root Mean Square Error (RMSE). Recently, Razavi and Gupta [26] applied a time varying GSA approach based on VARS using different indices of simulated streamflow, evapotranspiration (ET) and soil moisture storage. They recommend the use of TVSA to better control the calibration of the parameters and uncertainty that might arise during the model prediction. In addition to conventional performance evaluation metrics, flow signatures can inform model calibration and uncertainty analysis [27–30]. The use of flow signatures provides feedback to the calibration on different aspects of a hydrograph (e.g., rising or recession limbs versus peak flow), and the overall variability of flow characteristics [27]. Nonetheless, the application of flow signatures has been mostly limited to calibration of hydrological models and has not been presented simultaneously with SA.

Our study employs flow signatures, in addition to conventional error metrics, as evaluation criteria to analyze the seasonal and time variant sensitivity of model parameters for a cold region catchment (NCRB) using Hydrological Predictions for the Environment (HYPE). HYPE is a process-based semi-distributed hydrological model developed by the Swedish Meteorological and Hydrological Institute (SMHI) [31]. Our research addresses two specific objectives: (1) analyze the effect of time resolution on determining the sensitivity of the model parameters, and (2) application of TVSA and flow signatures to identify event-based sensitivity of model parameters in a cold region. Our goal is to determine parameter prioritization during the calibration process for the purposes of reducing the dimensionality of the optimization problem for hydrological modeling in high-latitude, seasonal and cold region environment.

## 2. Materials and Methods

### 2.1. Study Area

NCRB is the third largest catchment in North America [32] with an approximate drainage area of 1.4 million km<sup>2</sup> (Figure 1). It extends from the Rocky Mountains in the west to Lake Superior in the east, with elevation ranging from the sea level (at the outlets draining to Hudson Bay) to 3548 m above mean sea level in the Rocky Mountains. The Nelson and Churchill Rivers are the two major river systems in the basin, with respective catchment areas of 1.07 million km<sup>2</sup> and 0.28 million km<sup>2</sup>, draining the entire catchment into Hudson Bay. The NCRB region can be broadly classified into six freshwater basins: (i) Nelson River, (ii) Churchill River, (iii) Saskatchewan River, (iv) Red River, (v) Assiniboine River, and (vi) Winnipeg River.



**Figure 1.** Map of the Nelson Churchill River Basin with major sub-basins, rivers, lakes and evaluation gauges used for sensitivity analysis.

The Prairie region extending across the southern region of Alberta, Saskatchewan, and Manitoba is the driest part of NCRB, with an average annual precipitation of less than 320 mm. The Rocky Mountains and a small region in Ontario near Lake Superior are the wettest regions of the basin, receiving average annual precipitation of more than 750 mm. The annual average temperature of the catchment varies from +6 °C in the South to −6 °C in the North. During winter, the temperature varies from −4 °C in the South to −22 °C in the North, depending on the latitude. According to climate norms, the winter typically lasts from November to March in the southern latitude. The winter season is longer in the northern latitude of the basin, which generally lasts from October to May.

Precipitation is in the form of snowfall during winter ranging from 29 mm to 120 mm and is stored in the snowpack, which contributes to snowmelt during spring [33]. The water supply within NCRB varies greatly depending on the season, climatic conditions, latitude and topography. According to Manitoba Hydro [33], the Assiniboine River contributes the least to the NCRB water supply with an annual average streamflow of about ~45 m<sup>3</sup>/s between 1981 and 2010. Nelson River is the largest contributor with an annual average streamflow of ~3200 m<sup>3</sup>/s at the Hudson Bay. Lake Winnipeg, the largest lake system within NCRB, contributes an annual average streamflow of ~2180 m<sup>3</sup>/s through its outlet into the Nelson River. The Winnipeg River System, which drains into Lake Winnipeg, contributes an annual average streamflow of ~950 m<sup>3</sup>/s to NCRB. The Red river originating from the northern states, draining into Lake Winnipeg, averages the annual streamflow of ~250 m<sup>3</sup>/s.

NCRB is characterized by numerous small and large lakes. The land cover of the region is mostly forest and wetland, accounting for nearly 70% of the catchment [32]. Great Plains and Prairies distinguish the southern region, whereas the northern region is predominately Boreal Shield. Cold region processes such as snowmelt, frozen soil and permafrost play an integral role in influencing the hydrology of the NCRB. The snowmelt from this region contributes to runoff supplying snowmelt water to lakes, reservoirs, rivers and wetlands [34]. The northern region of the catchment is mostly dominated by permafrost and frozen soil, which greatly influences the runoff with the formation of wetland due to continuous ponding [35–37]. Frozen soil has low infiltration capacity, which allows for the retention of snowmelt water on the surface. However, during the thawing period, the water rapidly infiltrates the unfrozen soil [38].

## 2.2. Hydrological Modeling

The HYPE model of NCRB is adapted from the Hudson Bay HYPE (H-HYPE) configuration described by Tefs et al. [39]. The HYPE model for the entire Arctic region (A-HYPE) is also available in Macdonald et al. [40]. NCRB is discretized into 2693 sub-basins with an average area of 535 km<sup>2</sup> and median area of 330 km<sup>2</sup> (Figure S1). Sub-basins are further classified into homogeneous units called CLASSES in HYPE, which are formed by a unique combination of land cover, soil type and elevation. From the data source (Table 1), eight land cover types (crops, forest, open vegetation, bare soil, open water, glacier, wetland and urban) and seven soil types (coarse, medium, fine, organic, shallow/rock, glacial and urban) are identified for use in HYPE.

**Table 1.** Type of Data used for Hydrological modeling in Hydrological Predictions for the Environment (HYPE) and their sources.

Characteristic/Data Type	Information/Product	Source
Topography (routing and watershed delineation)	Hydro1K	United States Geological Survey (USGS)
Soil characteristics	Harmonized World Soil Database (HWSD) V1.2	Food and Agricultural Organization (FAO) [32]
Land use characteristics	ESA CCI Land Cover 2010 V1.4	ESA Climate Change Initiative—Land cover project 2014
Lakes and wetlands	Global Lakes and Wetlands Database	WWF and the Center for Environmental Systems Research
Reservoirs	Global reservoir and Dam database (GRanD) V1.1	Socioeconomic Data and Applications Center (SEDAC) [33]
Discharge	(i) HYDAT (ii) USGS (iii) Dery et al [41]	(i) Environment Canada (ii) <a href="https://waterdata.usgs.gov/nwis">https://waterdata.usgs.gov/nwis</a> (iii) Nelson and Churchill Outlets only
Precipitation and near-surface air temperature	(i) ERA-Interim (ii) NARR (iii) Hydro-GFD	(i) European Centre for Medium Range Weather Forecasts (ECMWF) [34] (ii) National Centers for Environmental Prediction (NCEP) [35] (iii) Swedish Meteorological and Hydrological Institute (SMHI) [36]
Snow	GlobSnow	Finnish Meteorological Institute (FMI)
Glacier fluctuations	Fluctuations of Glaciers (FoG)	World Glacier Monitoring Service (WGMS) [37]
Evapotranspiration	FLUXNET	<a href="http://fluxnet.ornl.gov">fluxnet.ornl.gov</a>

There are two types of lakes used in HYPE, referred to as ilake (local or inlet lakes) and olake (outlet lakes). The water level and threshold used in ilake and olake control the runoff in the local streams and the main river, respectively. HYPE uses ilake and olake parameters to control the outflow response of lakes, which is parameterized using rating curves. For the Hudson Bay Drainage Basin embedding the NCRB, Macdonald et al. [40] grouped ilake and olake in HYPE into four and three clusters respectively, based on physiographic characteristics using a k-means cluster analysis combined with a silhouette analysis for the selection of the optimal number of clusters.

Table 1 documents the data products used for setting up HYPE for NCRB. Spatial maps such as the Digital Elevation Model (DEM) from United States Geological Survey (USGS), the land use map

from the European Space Agency (ESA), and the soil map from the Harmonized World Soil Database (HWSD) are used to define the spatial characteristics of the NCRB in HYPE. The model is set up with an ensemble of three meteorological reanalysis products: (i) ECMWF Re-Analysis-Interim (ERA-I) (ii) North American Regional Reanalysis (NARR) and (iii) Hydro-GFD. The ensemble dataset is created by taking average of input variables from the three reanalysis products for individual sub-basins. Reanalysis data are the results of data assimilation from different sources such as ground observations, radar and satellite products into a numerical weather prediction model [42].

### 2.3. Evaluation Criteria

The sensitivity of the model parameters in VARS is determined by the model response values associated with all the sampled points. Streamflow generated by the HYPE model using ensemble precipitation and temperature input is evaluated against measured streamflow at multiple gauging stations across the NCRB using three standard error metrics: *NSE* [43], *PBIAS* [44], and Normalized Root Mean Square Error (*NRMSE*).

*NSE* (Equation (1)) represents the model performance in estimating the timing and magnitude of peak flow periods, because it is more sensitive to large simulation errors that usually occur during these periods. *NSE* can range from negative infinity (the worst model performance) to a maximum value of one (the perfect fit between simulated and measured values). A naïve model that results in the long-term average of the measured data has  $NSE = 0$ ; therefore, a positive value for *NSE* is expected for a model that is adequately simulating the system performance.

$$NSE = 1 - \frac{\sum_{t=1}^T (Q_t^m - Q_t^s)^2}{\sum_{t=1}^T (Q_t^m - \overline{Q^m})^2} \quad (1)$$

where,  $Q_t^m$  and  $Q_t^s$  are, respectively, the measured and simulated streamflow ( $m^3/s$ ) at time step  $t$ , and  $\overline{Q^m}$  is the average of the measured streamflow over the time period  $T$ .

*PBIAS* measures the percentage difference between the simulated and measured data, putting an emphasis on the volumetric error in streamflow simulation (Equation (2)). An ideal model would have *PBIAS* of zero as both the simulated and measured streamflow would have the same mean values. However, a *PBIAS* value of zero does not necessary imply that the model is perfect because it is a measure of the volumetric error only. The absolute value of *PBIAS* increases as the difference between the simulated ( $Q_t^s$ ) and measured ( $Q_t^m$ ) streamflow widens:

$$PBIAS = \frac{\sum_{t=1}^T (Q_t^m - Q_t^s)}{\sum_{t=1}^T Q_t^m} \times 100 \quad (2)$$

The *NRMSE* is normalized with respect to the difference between the maximum ( $Q_{max}^m$ ) and minimum ( $Q_{min}^m$ ) values of the measured streamflow (Equation (3)). This promotes fair comparison between different time series dataset with varying magnitude:

$$NRMSE = \frac{\sqrt{\sum_{t=1}^T (Q_t^m - Q_t^s)^2}}{T(Q_{max}^m - Q_{min}^m)} \quad (3)$$

In addition to the aforementioned model evaluation criteria, we have chosen three flow signatures to analyze the influence of the choice of error metrics on the sensitivity of the model parameters. The model performance in simulating the high flows is represented by the percentage difference between the measured and simulated 3-day averaged  $Q_5$  (the value of daily streamflow with a 5% exceedance probability) (Equation (4)). The 3-day averaged  $Q_{95}$  (the value of daily discharge with a 95% exceedance probability) is used to evaluate the model performance in estimating low flows (Equation (5)). The 3-day

averaged streamflow and 7-day averaged streamflow are selected to characterize the low flows and high flows in order to reduce the impact of measured flow uncertainty [45,46]:

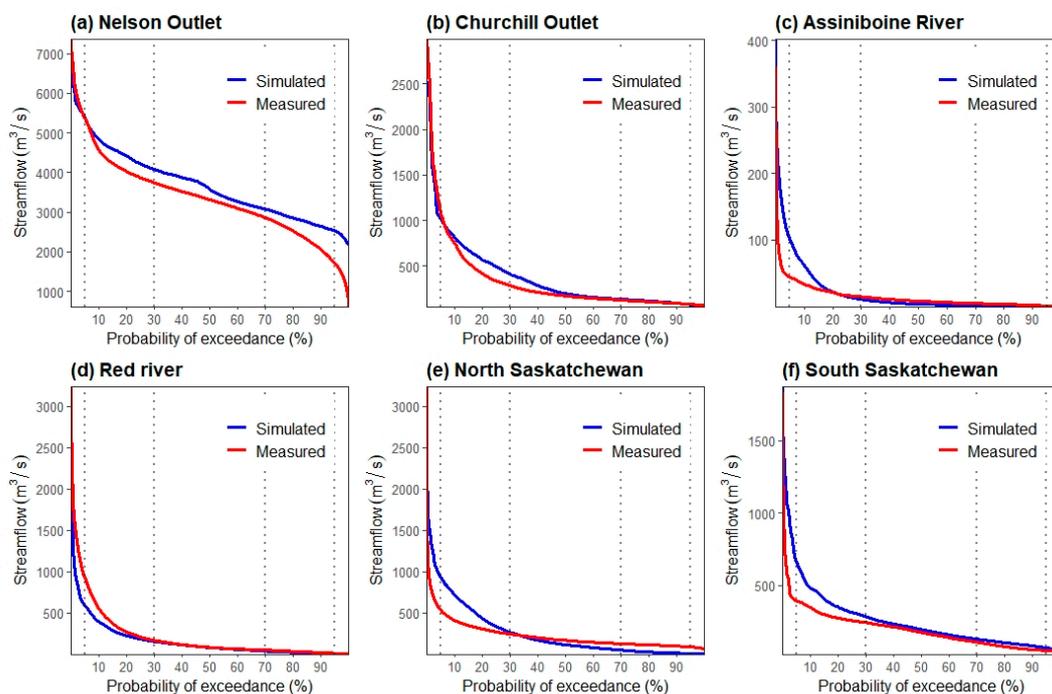
$$\frac{(3day\ avg\ Q_5)_{measured} - (3day\ avg\ Q_5)_{simulated}}{(3day\ avg\ Q_5)_{measured}} \times 100 \tag{4}$$

$$\frac{(3day\ avg\ Q_{95})_{measured} - (3day\ avg\ Q_{95})_{simulated}}{(3day\ avg\ Q_5)_{measured}} \times 100 \tag{5}$$

Finally, we have used the Slope of the Flow duration Curve (SFDC) to evaluate the response of the rising and falling limb of the hydrograph (indicated by the central part of the flow duration curve) to major hydrological events, compared to the real watershed system. SFDC is obtained as in Equation (6) [27] for the region bounded by the 30% and 70% probabilities of exceedance, as represented by the two center dotted lines in Figure 2. It is calculated for both the measured and simulated streamflow and their difference is used as an evaluation metric:

$$SFDC = 100 \times \frac{Q_{30} - Q_{70}}{40 \times Q_d} \tag{6}$$

where,  $Q_{30}$  is the value of daily discharge with a 30% exceedance probability;  $Q_{70}$  is the value of daily discharge with a 70% exceedance probability;  $Q_d = \frac{\sum_{t=1}^T Q(t)}{T}$  is the mean of daily streamflow ( $m^3/s$ );  $T$  is the length of the streamflow data record (days);  $Q(t)$  is the streamflow ( $m^3/s$ ) at time  $t$ .



**Figure 2.** Flow duration curves of simulated streamflow simulated by HYPE using ensemble input data, and measured daily streamflow over 1981–2010 for the major rivers within the Nelson Churchill River Basin (NCRB).

#### 2.4. Sensitivity Analysis Using VARS

In this study, the sensitivity of 34 HYPE parameters is identified using a GSA technique called VARS (Table 2). The parameter range used for SA has been set based on recommendations by Macdonald et al. [40] and other studies related to the HYPE model [47–49]. VARS is based on the directional variogram and covariogram of pairs of points in the response surface. The variogram is a

function quantifying the variance of change in the response surface as a function of distance between samples in the parameter space. The covariogram function quantifies the covariance structure in the response surface as a function of a vector  $h$  that is the distance vector between two points in the decision space. Details on the methodology and application of VARS can be found in Razavi and Gupta [13].

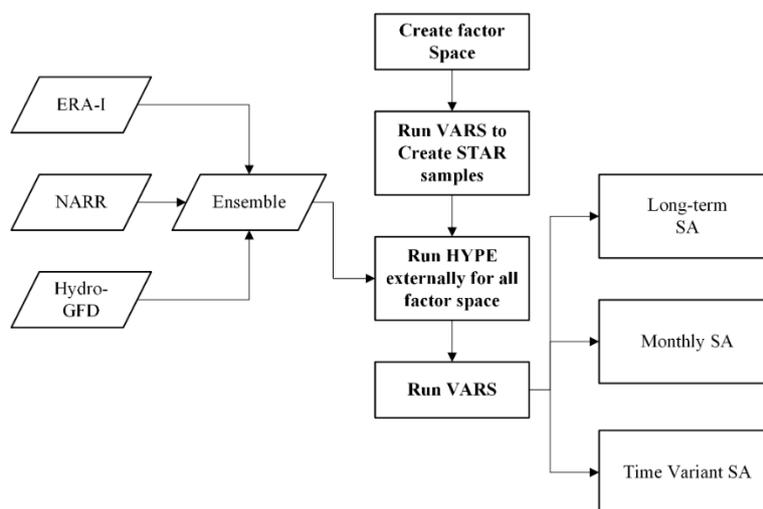
**Table 2.** Description of parameters in HYPE used for sensitivity analysis with lower and upper limit bounds.

S.N.	Parameters	Description	Dependency	Lower Limit	Upper Limit
1	rrc_corr	Correction factor for recession coefficients	Soil type	0.9	1.2
2	kc_corr	Correction factor for Crop Coefficient for PET model	Land use	0.9	1.4
3	fc_corr	Correction factor for fraction of soil available for ET	Soil type	0.8	1.3
4	wp_corr	Correction factor for wilting point	Soil type	0.8	1.1
5	deprl_corr	Correction factor for depth relation for soil temperature memory	Soil type	0.6	1.6
6	fpsno_corr	Correction factor for snow sublimation	Land use	0.8	1.2
7	kc_lake	Crop coefficient factor for lake type land use	Land use	0.7	1.3
8	kc_wetland	Crop coefficient factor for wetland type land use	Land use	0.4	0.9
9	kc_crops	Crop coefficient factor for crop type land use	Land use	0.7	1.3
10	kc_forest	Crop coefficient factor for forest type land use	Land use	0.4	0.9
11	kc_open	Crop coefficient factor for open type land use	Land use	0.7	1.3
12	wfc_coarse	Fraction of soil (coarse) available for ET	Soil type	0.05	0.25
13	wfc_medium	Fraction of soil (medium) available for ET	Soil type	0.1	0.3
14	wfc_fine	Fraction of soil (fine) available for ET	Soil type	0.1	0.3
15	wfc_organic	Fraction of soil (organic) available for ET	Soil type	0.3	0.5
16	wfc_shallow	Fraction of soil (shallow) available for ET	Soil type	0.05	0.15
17	ilrratp(3)	Parameter of rating curve for ilake (cluster 3) outflow (exponent)	Ilake region	1	2
18	ilrratk(1)	Parameter of rating curve for ilake (cluster 1) outflow (rate)	Ilake region	50	70
19	ilrratk(3)	Parameter of rating curve for ilake (cluster 3) outflow (rate)	Ilake region	2	30
20	olrratp(1)	Parameter of rating curve for outlet lake (cluster 1) outflow (exponent)	Olake region	3	4
21	olrratp(4)	Parameter of rating curve for outlet lake (cluster 4) outflow (exponent)	Olake region	1	2
22	olrratk(3)	Parameter of rating curve for outlet lake (cluster 3) outflow (rate)	Olake region	60	100
23	damp	Fraction of delay in the watercourse	Routing	0.4	0.7
24	rivvel	Celerity of flood in watercourse	Routing	0.5	1.5
25–29	bfrozn_(Soil-Type)	Soil dependent infiltration parameter	Frozen soil	1	4
30–34	bcosby_(Soil-Type)	Shape coefficient of soil water potential moisture curve	Frozen soil	2	15

The HYPE model of the NCRB is run externally for all sample solutions generated by VARS based on STAR sampling. Ranking of the sensitivity parameters and indices are based on the Integrated VARS (IVARS). STAR sampling is carried out for each star center defined at certain interval along each dimensions of the parameter space. IVARS is the integration of the directional variogram with different parameter perturbation sizes ranging from 0.1 to 0.5, with an incremental value of 0.1. For example, IVARS<sub>10</sub> refers to the integrated variogram with a small perturbation size of 10% of the parameter range, while IVARS<sub>50</sub> is based on perturbation size equal to 50% of the range of each parameter. Although the investigation of the sensitivity is recommended with different factor ranges, Razavi and Gupta [50] recommend the use of IVARS<sub>50</sub> as a global sensitivity metric if only a single metric is to be chosen. Furthermore, our study also shows that the rankings of parameters are not significantly influenced when comparing parameter rankings amongst IVARS<sub>10</sub>, IVARS<sub>30</sub> and IVARS<sub>50</sub> across six evaluation metrics (Table 3). Therefore, we have opted to use IVARS<sub>50</sub> as the sensitivity metric for all our analysis.

In this study, SA is performed using three approaches corresponding to different time scale resolution: (i) long-term SA, (ii) monthly SA, and (iii) TVSA using a moving window analysis (Figure 3). The long-term SA is based on sensitivity indices calculated from the error metrics computed from simulated and measured streamflow over the entire time period of analysis. In the case of monthly SA, sensitivity indices are determined separately for individual months, to identify the sensitivity based

on seasonality. Compared to the previous two approaches, the TVSA calculates sensitivity indices for each parameter across each window period. For example, for a 30-day window, the sensitivity of the parameters is analyzed based on a moving window of every 30 days using the times series from model simulation. It should be noted that the majority of previous studies applying SA used method (i) [23,51,52], and in this study we compare the more traditional approach to methods (ii) and (iii) in a high-latitude, seasonal, and cold region environment.



**Figure 3.** Methodological framework for time-variant sensitivity analysis of model parameters in HYPE.

For all three SA approaches, we have used four sets of parameters (34 parameters) among which land use and soil type dependent parameters comprise the majority of parameters. Two other sets represent lake discharge and routing parameters. For the application of VARS for HYPE, we have used 20 STAR samples (m) with a resolution of 0.1 (h) on the 34 parameters space of the response surface. The STAR samples are selected using the Latin Hypercube sampling technique. For each star sample (m), a cross-section is generated at equal spacing of resolution specified (h). This results in  $m \times [n \times ((1/h) - 1) + 1] = 6140$  model runs, where n is the number of parameters subjected to SA.

#### 2.4.1. Long-Term SA

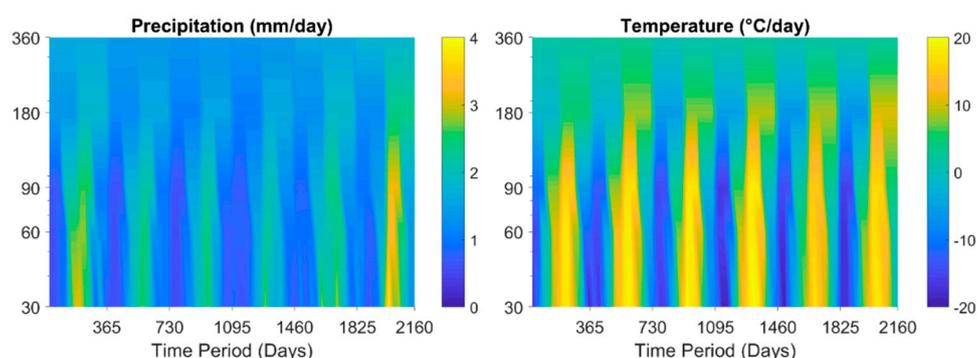
The long-term SA is based on aggregation of sensitivity information from the entire time-period of analysis. Here, we have analyzed the long-term SA for a period of 30 years from 1981–2010 to coincide with the Canadian climate normal period. The aggregated model response values in the form of evaluation metrics are calculated between simulated and measured streamflow for each set of parameters sampled. Parameter rankings are determined based on the ratio of factor sensitivity (RFS) of the parameters. RFS of a parameter is computed as the ratio of the IVARS for that parameter to the sum of IVARS over all parameters considered for SA. The parameter with the largest RFS value is assigned the highest rank (highly dominant), whereas the parameter with the smallest RFS values is assigned the lowest rank (least dominant). The reliability of the ranking of parameters for different evaluation metrics are also computed by VARS. Reliability is generally defined as “the ability of a system or components to function under stated condition for a specified period of time” [53]. In VARS, reliability is the probability of success to determine the true ranking of the parameters based on specific sensitivity indices (IVARS). The reliabilities of parameter rankings are higher in simple models, whereas complex models with a large number of parameters require a large number of STAR samples to attain the state of true ranking [54].

### 2.4.2. Monthly SA

The monthly SA is implemented for a period of 30 years similar to the long-term SA, except that the sensitivity information is aggregated at a monthly timescale. The evaluation metrics are calculated between measured and simulated streamflow for each aggregated month separately. The monthly SA has advantage over long-term SA because it is expected to identify parameters that are strongly influential during particular seasons but less influential during other seasons (e.g., snowmelt parameters in the summer). The importance of some parameters, if not all, during model calibration may therefore change based on the seasonality in the study region and the time-period of analysis. In such cases, calibration should include parameters that are more influential in particular seasons but do not have a high sensitivity index over the whole simulation period. Zhao et al. [55] calibrated a hydrological model for an alpine basin using the stratified calibration approach (SCA). In the SCA method, the calibration of model parameters is implemented by stratifying seasons, which refers to dividing the entire period into dry and wet seasons. Calibration of a hydrological model based on seasonality improves simulation of ET and surface flow during rainy periods [55].

### 2.4.3. Time Variant Sensitivity Analysis

Compared to the previous two methods in which the aggregation of sensitivity information is over a month or long term, the TVSA method is implemented using a certain window period of fixed duration to minimize the loss of sensitivity information. Sensitivity metrics are calculated using VARS for different moving window periods (30, 60, 90, 180 and 360 days) to account for the inter-annual dry and wet cycles in the basin (Figure 4) across a six year time period spanning from 2000 to 2005. The selection of window size is dependent on factors such as quality of data, time period of analysis, and smoothness of the plot desired [56]. Hui et al. [25] used TVSA with window sizes of up to 4 days for hydrologic and sediment parameters. Smaller window sizes are recommended as it reduces the effect of time series aggregation and is therefore able to capture the temporal dynamics of the influential parameters [25,57]. Although some hydrologic processes could occur at shorter durations (<30 days), this analysis is restricted to a minimum 30-day window period to enable the calculation of quantiles for the flow duration curves. The restriction is further compounded by the computational demand of smaller window size for a continental scale catchment. The use of 30 days and 60 days for window size is useful in exploring the seasonal sensitivity, whereas the window size larger than 90 days helps to explore the sensitivity of the model to different parameters at a coarser time scale. The time period of analysis is restricted to only six years due to the very large computational demand required to implement TVSA for every moving window period over a long term for such a large basin. The runtime for a 180-day moving-window is ~3 days for a single evaluation metric (i7-6700 processor, 2.60 GHz). It has increased to 12 days for a single evaluation metric when the window size is reduced to 30 days, and 72 cumulative days for all six-evaluation metrics (for method (iii) only).



**Figure 4.** Daily averaged precipitation (mm/day) and air temperature ( $^{\circ}\text{C}/\text{day}$ ) of the NCRB using the Ensemble dataset for moving windows of 30 days, 60 days, 90 days, 180 days and 360 days from 1/1/2000 to 29/11/2005.

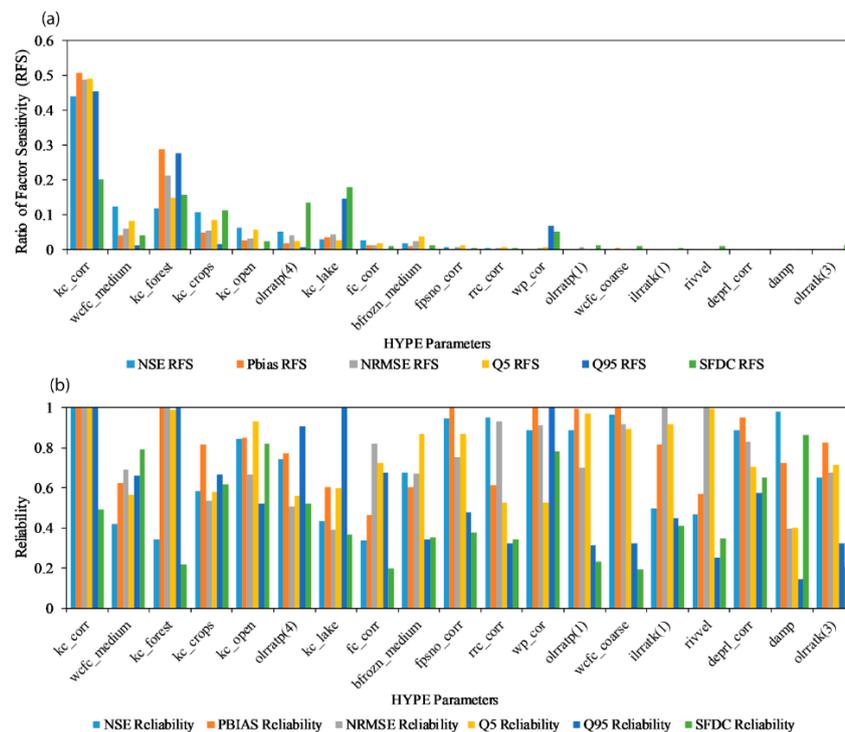
The `pcolor` function of MATLAB is utilized to interpolate the color transition between consecutive time periods and different window sizes for each sensitivity metric. The color transition obtained from interpolation can be jagged, particularly for larger window sizes. Smooth transitions can be achieved by using smaller window sizes [25], which are not implemented in this study due to their large computational demand.

### 3. Results and Discussion

#### 3.1. Long-Term SA

We first analyze the ranking and reliability of HYPE model parameters from the long-term SA based on the sensitivity metrics obtained from  $IVAR_{50}$  for the six evaluation criteria used in this study (Figure 5). Results show that the parameter rankings and their reliability change as the model evaluation metric changes. For all evaluation metrics, the parameters related to land use and soil characteristics are most influential followed by the `ilake` and `olake` parameters.

In terms of relative ranking of the parameters, the crop coefficient correction factor (`kc_corr`) associated with the calculation of ET is the most influential parameter. This is an obvious result, because the `kc_corr` parameter is applied as a correction factor to all land use types. The rankings of the `kc` parameters associated with individual land use type is closely followed by that of the global `kc_corr` parameter. However, these rankings are not stable across all evaluation metrics. The crop coefficient associated with forest (`kc_forest`) is more influential than the crop coefficient associated with open vegetation (`kc_open`) such as grassland and open shrub land. Several studies have found that ET losses from forested areas are higher than those from cropland and open vegetation [58,59]. Most of the frozen soil parameters do not exhibit significant sensitivity and are considered among the least influential parameters. The model shows strong sensitivity to the frozen soil parameter related with medium soil type (`bfrozn_medium`). Therefore, it is ranked among the top ten most influential parameters. The medium soil type alone covers approximately 49% within NRCB, which is almost equal to the coverage of all the other soil types combined. In terms of reliability of the rankings, for *NSE*, the reliability varies from 26% to 100%, with a median reliability of 69%. The median reliability values of the parameter rankings for both *PBIAS* and *NRMSE* criteria remains at 70%. For the 3-day averaged  $Q_5$ ,  $Q_{95}$ , and *SFDC*, the median reliability values degrade to 64%, 65% and 44% respectively. The higher reliability values are more reliable and accurate, and the chances of reproducing the same result is higher under the same condition [60]. Therefore, measurements with higher reliability are preferred to the one with the lower reliability.

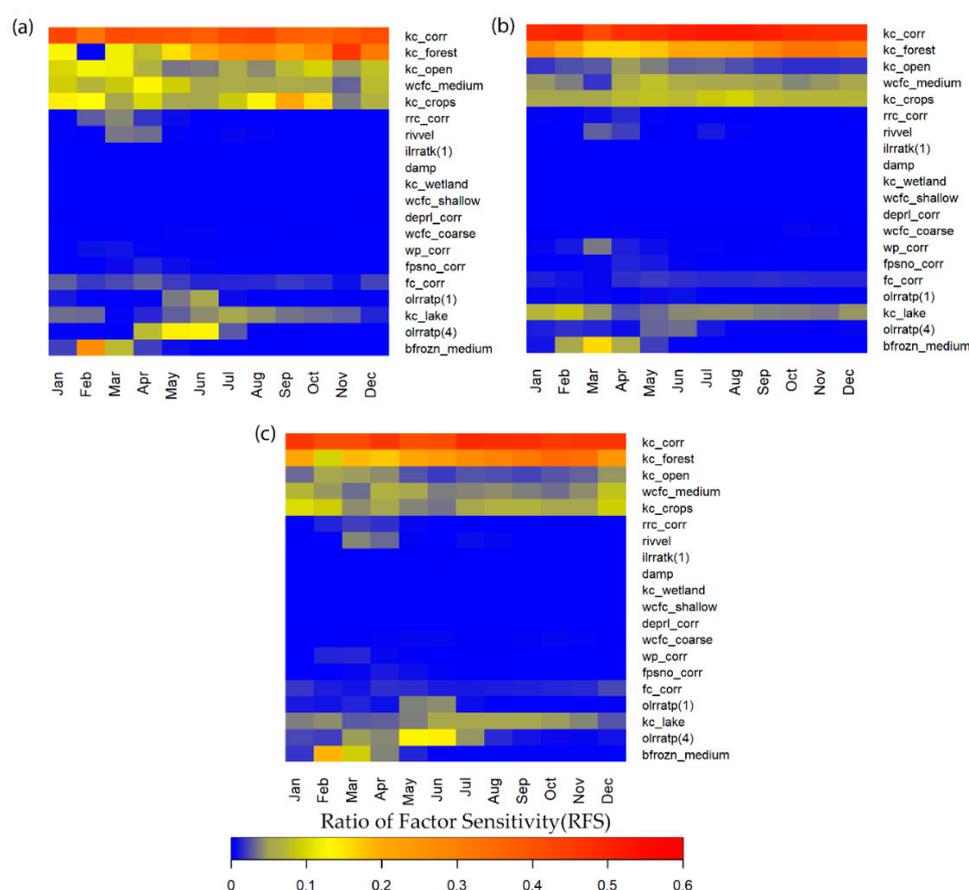


**Figure 5.** (a) Ratio of Factor Sensitivity and the (b) reliability of parameters based on a long-term sensitivity analysis (1981–2010) using Variogram Analysis of Response Surfaces (VARS) and different evaluation criteria for NCRB.

In general, the conventional approach for long-term SA is informative in ranking the parameters based on the sensitivity metrics, and the screening of the important parameters is essential for further calibration of hydrological models. For conventional error metrics such as *NSE*, the least influential parameters which have negligible RFS (e.g., frozen soil parameters related to certain soil types, and some lake parameters) can be eliminated from model calibration, thus reducing the parameter space from the 34 initially selected parameters to a total of 23 parameters (i.e., ~32% reduction). Nevertheless, in this method, the ranking of parameters is based on the sensitivity information aggregated over the entire time period of analysis and is therefore unable to detect the seasonal sensitivity of the parameters, or the time varying nature of the parameters and subsequent hydrologic processes they are controlling. The accumulation of model residuals across the entire period of analysis causes a loss of information content for the model calibration. The model residuals are even larger if the residuals are squared prior to aggregation as in RMSE and *NSE* metrics [5]. To address this issue, it is important to disaggregate the time series before calculating the sensitivity metrics.

### 3.2. Monthly SA

The seasonal nature of some hydrological processes makes the corresponding parameters more influential in particular seasons. For example, the ET parameters (kc group) are expected to be more influential in warmer seasons (Figures 6 and 7). Moreover, the monthly SA gives insight into the model’s sensitivity to seasonal parameters, which appears to be uninfluential from long-term SA results.

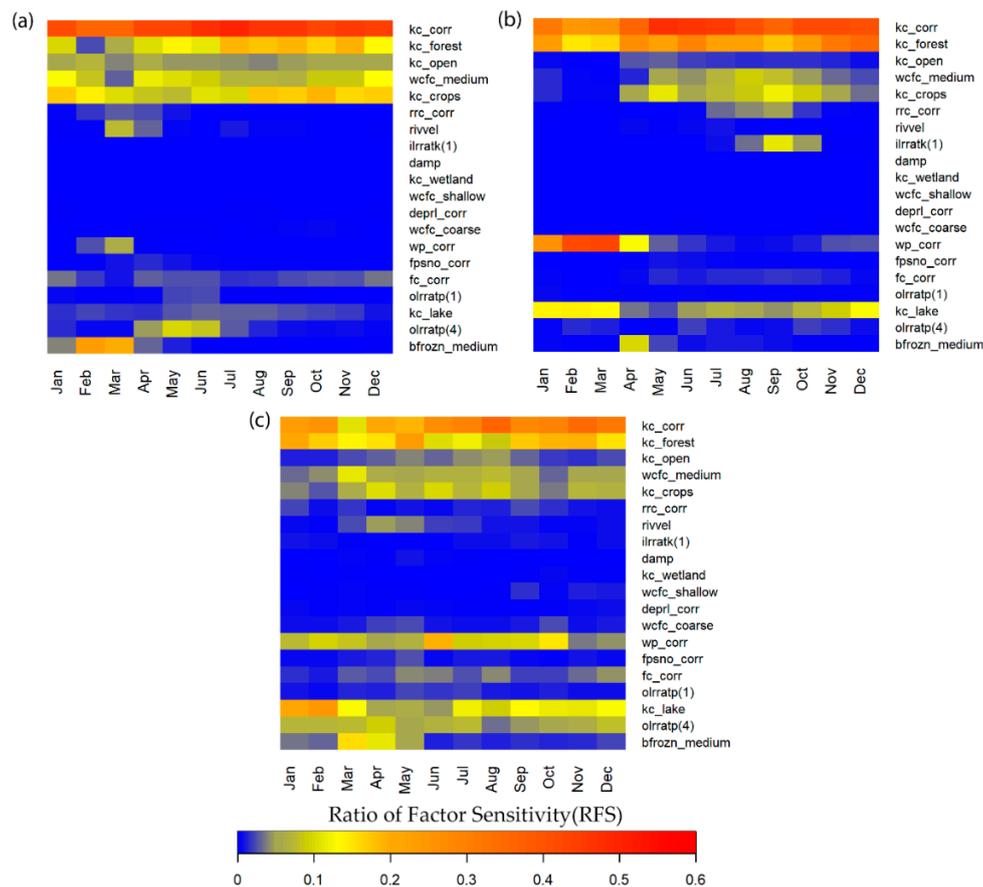


**Figure 6.** Heat map representing the sensitivity of model parameters expressed as a ratio of factor sensitivity based on (a) Nash Sutcliffe Efficiency (*NSE*), (b) Percent Bias (*PBIAS*), and (c) Normalized Root Mean Square Error (*NRMSE*) for the evaluation period of 1981–2010.

The model is least sensitive to the field capacity parameters that are related to organic (*wfcf\_organic*) and fine (*wfcf\_fine*) soil types irrespective of the seasonality. With finer textures, clay soils and wetland peats have a high water retention capacity and significantly low horizontal and vertical hydraulic conductivity [61]. Therefore, *wfcf\_organic* and *wfcf\_fine* likely have an inherent time lag (greater than one month) associated with the hydrologic response time and result in a low sensitivity at a monthly scale. In contrast, across all evaluation metrics the field capacity parameter associated with medium soil textures (*wfcf\_medium*) is highly influential and reveals varying sensitivity on a seasonal basis. The variation in the sensitivity of *wfcf\_medium* parameter depends on the error metrics used but in general, the sensitivity is higher during the summer and spring seasons compared to the winter season. Since the sensitivity of the soil parameter is not only reflected in the spring season when the snowmelt process is activated, the sensitivity is a reflection of the rainfall runoff process more than the snowmelt seasonality.

A comparison between Figures 6 and 7 shows that signature-based metrics better identify parameters that are influential in a particular season (month). For example, the *wp\_corr* parameter related to the wilting point is highly influential only during the months of February and March while using  $Q_{95}$  as an error metric. This suggests the *wp\_corr* parameter is strongly affected by the snowmelt process, which is not necessarily intuitive. Likewise, the *rivvel* parameter controlling the velocity in river reaches and discharge is most influential during spring when snowmelt generates a sharp increase, and high variability, in streamflow. Therefore, *rivvel* based on *NSE* (Figure 6a) and 3-day averaged  $Q_5$  (Figure 7a) shows maximum sensitivity during spring compared to the other metrics. Snowmelt and frozen soil parameters, snow sublimation (*fpsno\_corr*) and frozen soil (*brozrn\_medium*),

are most influential during the spring freshet period. Frozen soil parameters are dependent on soil type and strongly influenced by the soil infiltration capacity. The infiltration capacity of a soil is significantly reduced during the frozen state. During spring, the voids between the soil particles occupied by frozen water start thawing, resulting in an acceleration of the infiltration rate in the soil [62,63].

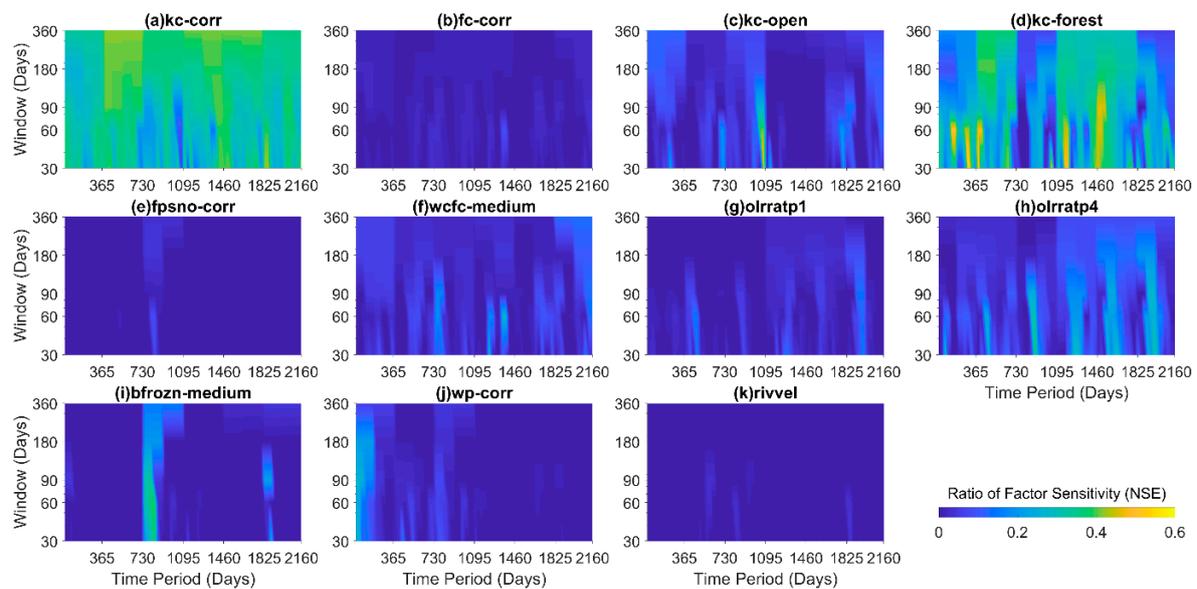


**Figure 7.** Heat map representing the sensitivity of model parameters expressed as a ratio of factor sensitivity based on (a) 3-day averaged  $Q_5$ , (b) 3-day averaged  $Q_{95}$ , and (c) slope of Flow Duration Curve for the evaluation period of 1981–2010.

The monthly SA implemented on 34 hydrological parameters shows variation in the sensitivity indices based on seasonality, unlike the long-term SA in which the sensitivity indices are aggregated over the entire time period. Due to the aggregation of sensitivity indices, the sensitivity values of the parameters in the long-term SA are assumed constant over the entire time period of analysis. In contrast, the monthly SA demonstrates that some parameters such as rivvel and fpsno\_corr are highly influential during a particular season, which corresponds to the snowmelt period (Figure 6). However, their sensitivity remains undetectable during other seasons. The ranking of the parameters also changes seasonally, and neglecting such parameters (and parameter sensitivity changes) can lead to an unrealistic representation of the model processes. Though the monthly SA is able to address the seasonality effect of the parameters, this method is still insufficient to address the issue of temporal aggregation. The conclusion drawn from the monthly SA relative to the long-term SA warrants further investigation of the model’s sensitivity to influential parameters at a finer time scale (i.e. TVSA) to investigate parameters that may not exhibit seasonality but are important for a more realistic, physically representative model and simulation.

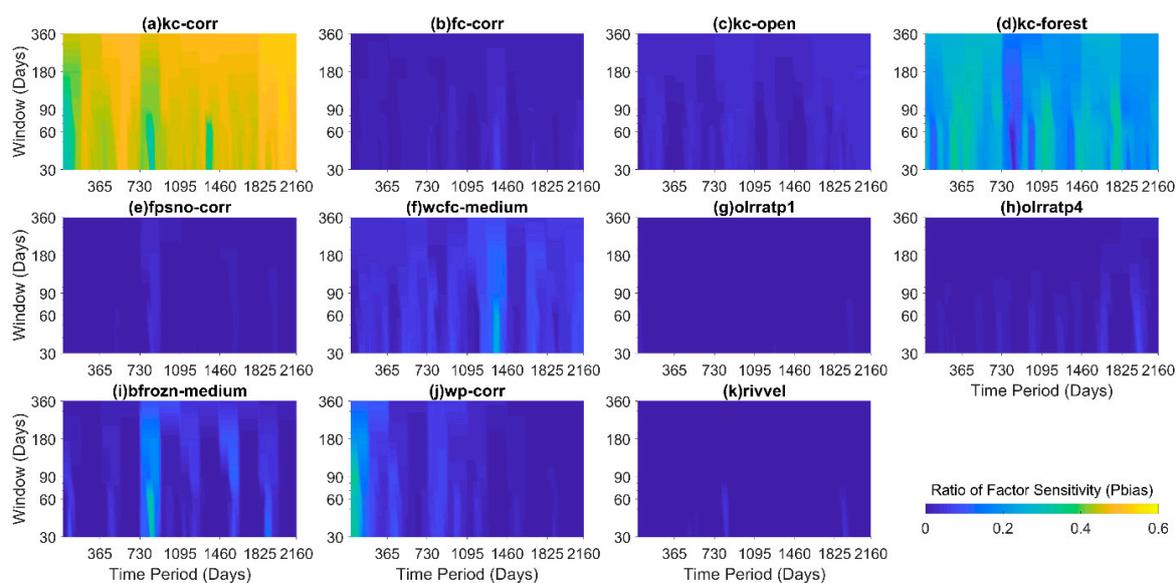
### 3.3. Time Variant Sensitivity Analysis

Figures 8 and 9 show the graphical representation of 11 of the most influential parameters, using conventional metrics (*NSE* and *PBIAS*) as evaluation criteria. In these figures, the x-axis represents the time period in days (e.g., 01/01/2000 to 31/12/2005), and the y-axis represents the window period (days) in the logarithmic scale. The RFS varies from a minimum value of zero to a maximum of 0.6, depending upon the parameter, window size, time period and evaluation metric used. TVSA confirms that *kc\_corr* is the most influential parameter throughout all six years of analysis. The *NSE*-based sensitivity factor for *kc\_corr* (Figure 8a) is lower relative to those obtained using *PBIAS* (Figure 9a) and *NRMSE* (Figure S2a), indicating that the crop coefficient parameter associated with ET process is more sensitive to changes in the overall flow volume compared to changes in peak flow.



**Figure 8.** Ratio of factor sensitivity for moving windows of 30 days, 60 days, 90 days, 180 days and 360 days using *NSE* as the evaluation criteria from 01/01/2000 to 29/11/2005.

These findings agree with the results from the previous two methods, however additional information regarding the relative sensitivity of the parameters is provided by the TVSA. For instance, as opposed to its normal seasonal sensitivity behavior demonstrated in monthly SA, the TVSA identifies the *kc* parameters to be the most influential during the winter season in some dry years (2000 and 2001/2002) as well as (Figure 8). One possible explanation for this seemingly contradictory finding is the relative insensitivity of soil moisture parameters prior to freeze up and over winter during drier years. During drier years, the model is relatively insensitive to *wfc* and *wp* parameters that are related to the water holding capacity of soils at different depths, which is a function of soil properties. During smaller precipitation events and longer dry periods, only the upper-layer (shallow) soil moisture contents are affected [64]. As precipitation ceases (or is minimal) for longer durations, the soil moisture of the other (lower) layers becomes uninfluential for the water balance, and hence the *kc* parameters become relatively more influential during the driest period of the year compared with other parameters.



**Figure 9.** Ratio of factor sensitivity for moving windows of 30 days, 60 days, 90 days, 180 days and 360 days using *PBIAS* as the evaluation criteria from 01/01/2000 to 29/11/2005.

Likewise, though *rivvel* identifies as an influential parameter by monthly SA, TVSA shows that the sensitivity of the model to this parameter varies from one year to another depending on the error metric used for evaluation. For instance, as compared to monthly SA using the *NSE* error metric, which identifies the *rivvel* parameter to be influential during only spring and peak flow seasons, TVSA identifies it to be influential only in certain years (2001, 2002 and 2005). The results are consistent with the highest peak flows in the measured streamflow record for the Nelson and Churchill River outlets. For instance, the measured streamflow at the Nelson outlet is observed to be maximum during the years of 2001 and 2005 with the values of 5631 m<sup>3</sup>/s and 7346 m<sup>3</sup>/s respectively. Likewise, maximum peak streamflow is observed at the Churchill outlet during the year 2005 (2990 m<sup>3</sup>/s) (Figure S3). Moreover, the model's sensitivity to the routing parameter (*rivvel*) is identifiable only during the smaller moving window period (less than 90 days) (Figure 8k). The use of larger window sizes (greater than 90 days) suppresses the sensitivity of this parameter, as the high flow condition tends not to persist for long durations. Freshet events contributing to increasing discharge, and subsequent water velocity, typically occur within a maximum window period of four months (120 days approximately) [65].

These results demonstrate that TVSA is effective in identifying process- and event-based sensitivity of the parameters, which are inherent limitations of the previous SA approaches. In addition, TVSA is able to detect the sensitivity of some parameters not considered influential when using a long-term or monthly SA approach. For instance, by evaluating the aggregated monthly sensitivity based on 3day-*Q*<sub>95</sub> in Figure 7b, it suggests that this metric is insensitive to routing parameter (*rivvel*) and lake parameter (*olrratp1*). Such parameters particularly bear high significance in terms of event-based calibration such as is done for flood forecasting. The runoff process in cold regions such as NCRB is governed by snowmelt during spring season, as the snowmelt alone accounts for more than 80% of annual local runoff in this region [66]. Hydrological models used for flood forecasting requires reliable simulation of high flows and runoff generated from snowmelt [67]. Therefore, it is essential to consider event based parameters contributing to high flows and snowmelt such as *rivvel* and *fpsno* parameters in model calibration.

### 3.4. Impact of the Choice of Evaluation Metrics

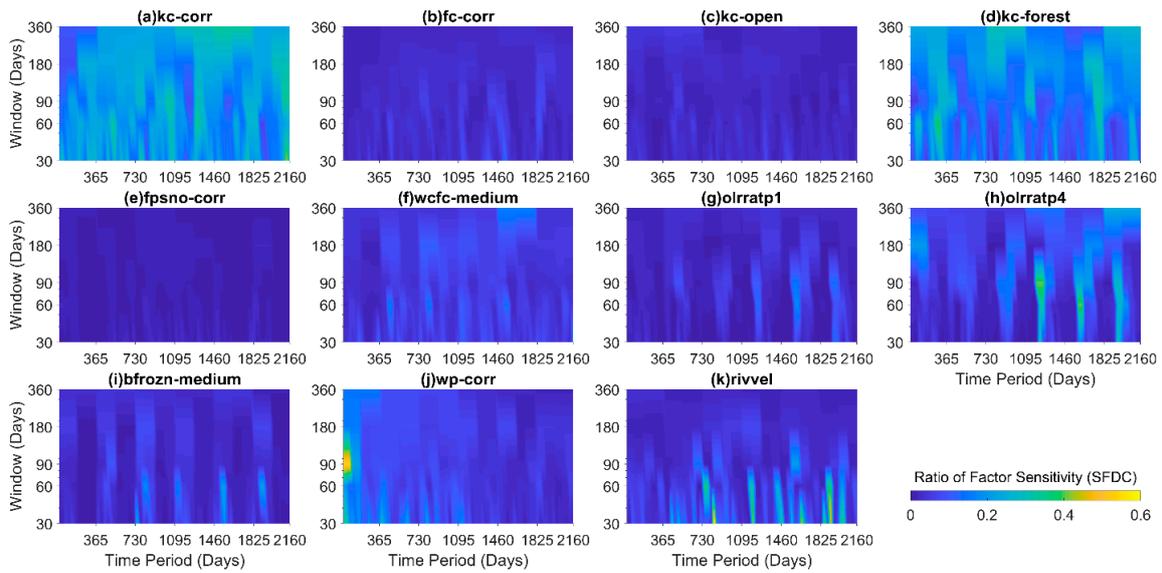
The results from long-term SA show a reduction in the parameter space ranging from 25% to 32% across all evaluation metrics (Table 3), with *NSE* having the largest reduction (32%), and *SFDC* having the lowest (25%). To analyze the percentage reduction in parameter range for monthly SA and TVSA,

we have taken the maximum RFS values of each parameter from the time scale of analysis (Table 4). Our comparison shows reduction in the parameter space for monthly SA varying from 14% to 26%, with *SFDC* also showing associated with the smallest reduction. The conventional error metrics based SA have larger percentage reductions varying from 20% to 26%. Table S1 shows the six most influential parameters at the monthly analysis timescale based on different evaluation metrics. The overall crop coefficient factor (*kc*) associated with ET remains dominant across all evaluation metrics, regardless of the SA method used. Conversely, the model is most sensitive to parameters associated with lake discharge (*olrratp1*) and routing (*rivvel*) during the high flow months of May and June when using *NSE*, which emphasizes peak flows (Figure 6a).

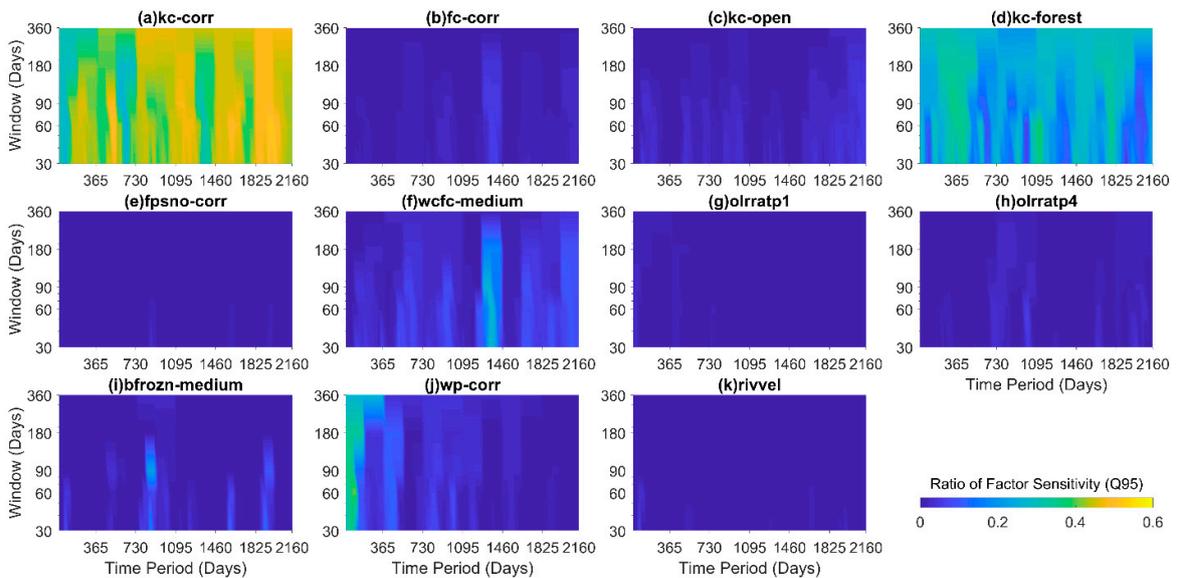
In contrast to long-term SA and monthly SA, the reduction in parameter space for TVSA varies from one window period to another for all evaluation criteria. Our results show a decrease in the parameter space ranging from 14% to 32%, with *SFDC* associated with the lowest reduction. This demonstrates the capability of *SFDC* to detect additional sensitivity among parameters, which is overlooked by the conventional error metrics and the larger reductions in parameter space (commonly 23% to 26% here). In contrast to conventional peak flow or volume-based metrics, the sensitivity of all 11 most influential parameters are identifiable throughout the six-year analysis period when using *SFDC* as evaluation criteria and the TVSA approach (Figure 10). The sensitivities of the *kc* parameters based on *SFDC* are less dominant compared to the results obtained from the other residual based metrics; but the sensitivities of parameters related to frozen soil, lakes and routing are more recognizable throughout different seasons and time period.

The RFS based on *NSE* for *kc\_corr* varies from 0.04 to 0.44 with a median value of 0.34 (Figure 8a). The frozen soil parameter (*bfrozn\_medium*) (Figure 8i) and the snow sublimation parameter (*fpsno\_corr*) (Figure 8e) are detectable only during the snowmelt period of 2002 and 2005 with median RFS values of 0.002 and 0.00018 respectively. When the *SFDC* evaluation metric is used, however, the *kc\_corr* shows a variation of RFS ranging from 0.06 to 0.44 with a median value of 0.21 (Figure 10a). This is significantly lower than the *NSE* based RFS value. In contrast, the *fpsno\_corr* (Figure 10e) and *bfrozn\_medium* (Figure 10i) parameters are detectable during every snowmelt period with comparatively high RFS median values of 0.008 and 0.02 respectively.

Regardless of the sensitivity method implemented, *SFDC* has low reliability compared to the conventional metrics. Therefore, for long-term SA we suggest that conventional metrics with high reliability such as *NSE* and *PBIAS* should be used since *SFDC* does not contribute significantly to the detection of additionally influential parameters. However, for the monthly SA and TVSA, *SFDC* is recommended despite its comparatively low reliability because it can detect the sensitivity of additional parameters that are overlooked while using the conventional error metrics. The application of *SFDC* as an evaluation metric for SA reduces the uncertainty range of streamflow, as the slope of the mid-segment of the FDC between 30% and 70% is milder compared to the high flow and low flow segments (Figure 2). This results in the detection of the sensitivity of additional parameters based on the changes in the rising and falling limb responses, which are not recognized by other evaluation metrics. Furthermore, since different parameters dominate at different times of the year, hydrographs in this region are non-uniform. This is precisely why we need to incorporate flow signatures such as *SFDC* into time varying SA. In comparison to *SFDC*, the application of flow signatures based on low (i.e., 3day- $Q_{95}$ ) and high flows (i.e., 3day- $Q_5$ ) are not as effective in representing the variation in sensitivity of some of the most influential parameters such as *fpsno\_corr* (Figure 11e, Figure S4e), *rivvel* (Figure 11k, Figure S4k) and *olrratp1* (Figure 11g, Figure S4g). This could be because *SFDC* is calculated as the slope of the mid-segment of the flow duration curve, whereas 3day- $Q_{95}$  and 3day- $Q_5$  are taken as the average of only three measured streamflow data sets corresponding to their respective percentage of exceedance from the flow duration curve. Not surprisingly, the 3day- $Q_5$  yields a variation in parameter sensitivity similar to that obtained using *NSE*, as both metrics emphasize peak flow [68].



**Figure 10.** Ratio of factor sensitivity for moving windows of 30 days, 60 days, 90 days, 180 days and 360 days using slope of the flow duration curve (*SFDC*) as the evaluation criteria from 01/01/2000 to 29/11/2005.

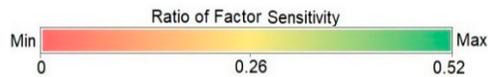


**Figure 11.** Ratio of factor sensitivity for moving windows of 30 days, 60 days, 90 days, 180 days and 360 days using 3-day averaged  $Q_{95}$  (low flow) as the evaluation criteria from 01/01/2000 to 29/11/2005.



**Table 4.** Comparison of the ranking of model parameters based on maximum ratio of factor sensitivity (RFS), from most influential (green) to least influential parameter (red) determined from time variant sensitivity analysis (TVSA) on 30-day window period (IVARS<sub>50</sub>), and monthly SA using different evaluation metrics.

Parameters	TVSA for a 30-Day Window Period						Monthly SA					
	NSE	PBIAS	NRMSE	Q <sub>5</sub>	Q <sub>95</sub>	SFDC	NSE	PBIAS	NRMSE	Q <sub>5</sub>	Q <sub>95</sub>	SFDC
kc_corr	31	34	33	34	34	33	33	34	34	34	34	34
wcfc-medium	28	30	27	30	30	25	29	29	29	30	27	29
kc_forest	34	33	34	32	32	29	34	33	33	32	32	32
kc_crops	33	28	28	28	28	27	31	31	30	31	30	28
kc_open	32	24	25	23	24	20	28	28	27	27	25	25
olrratp(4)	30	27	30	26	25	31	30	26	31	29	24	27
kc_lake	26	29	29	27	29	26	26	30	28	24	31	33
fc_corr	19	26	23	25	26	22	23	24	24	25	23	24
bfrozn_medium	29	32	32	33	27	28	32	32	32	33	28	30
fpsno_corr	24	25	24	24	20	16	22	22	21	21	21	23
rrc_corr	22	21	20	21	23	23	25	23	23	23	26	21
wp_cor	27	31	31	29	33	32	21	27	22	26	33	31
olrratp(1)	25	20	26	22	21	24	27	21	26	22	20	20
wcfc_coarse	17	19	21	18	18	21	20	20	20	20	19	22
ilrratk(1)	21	22	19	19	31	19	19	19	19	17	29	18
rivvel	20	23	22	31	22	34	24	25	25	28	22	26
deprl_corr	15	17	15	17	17	18	18	18	18	19	18	16
damp	14	16	16	15	19	14	16	14	16	16	14	17
ilrratk(3)	13	9	11	9	10	10	14	10	13	11	9	10
wcfc_organic	11	13	13	12	12	13	11	12	11	12	11	13
wcfc_fine	12	10	10	11	11	11	12	9	10	13	17	11
wcfc_shallow	23	18	14	20	15	30	17	17	15	18	13	19
kc_wetland	16	14	17	13	16	17	13	15	14	14	16	14
ilrratp(3)	8	8	8	8	6	7	10	8	8	8	8	6
olrratk(3)	18	15	18	16	14	12	15	16	17	15	12	15
bfrozn_coarse	7	11	9	10	9	9	8	11	9	10	7	8
bfrozn_organic	9	12	12	14	13	15	9	13	12	9	10	12
bfrozn_fine	6	7	7	7	8	8	6	7	7	7	15	9
bfrozn_shallow	10	6	6	6	7	6	7	6	6	6	6	7
bcosby_shallow	4	5	5	5	3	5	5	5	5	4	5	4
bcosby_organic	5	4	4	4	5	4	4	3	4	5	4	2
bcosby_medium	2	3	3	3	2	3	3	2	3	2	3	5
bcosby_coarse	3	1	1	1	4	1	1	1	1	1	1	1
bcosby_fine	1	2	2	2	1	2	2	4	2	3	2	3



#### 4. Conclusions

We have applied three different SA methods (long-term SA, monthly SA and TVSA) based on differing levels of temporal aggregation and different error metrics to analyze the sensitivity of the model parameters in HYPE for cold region catchment (NCRB). When applied to a set of 34 model parameters, long-term SA results suggest that up to ~32% of model parameters are uninfluential, and therefore can be eliminated from model calibration. However, in addition to exclusion of some (seasonally) influential parameters, it provides limited information regarding the relative sensitivity of individual model parameters due to the information loss that occurs during data aggregation over a longer time period. In contrast, monthly SA can reveal shorter-term, seasonal variations in the sensitivity of the model parameters associated with seasonally variant hydrological processes. Parameters related to cold region processes such as frozen soil and snow sublimation are found to be highly influential during snowmelt seasons as revealed by the monthly SA. The snow and ice accumulated during the winter season melts during the spring season, which contributes to high volume runoff and flooding in cold region catchments [69,70]. Intuitively, snowmelt runoff in the cold regions of the northern hemisphere is very sensitive to any alternation in parameters related to temperature [71]. The application of flood forecasting plays an integral role in foreseeing such events to minimize the extent of damage caused by heavy runoff [72]. Therefore, such parameters should not be discarded from model calibration, despite their low influence in long-term simulation, because of the high flows and volumes associated with spring freshet in nival-dominated regimes. Robust calibration of parameters associated with cold regions is essential to reliably simulate the snowmelt runoff peak [62,73,74]. Furthermore, the monthly SA approach to model calibration actually makes the model parameter choice more physically relevant too. This is because the texture of soils is accounted for explicitly by breaking up the seasons and making the calibration better informed. The calibrated parameters then reflect the physical environment (and changes in it) more realistically.

Although a monthly SA can detect the sensitivity of additional parameters with a comparative parameter space reduction (14% to 28%), the level of data aggregation to a monthly time step is still an issue as the monthly sensitivity of the parameters are considered the same throughout the time period of analysis. To address this issue, we have carried out a TVSA at various window periods to account for the event-based sensitivity of the parameters. TVSA is able to detect strong sensitivity signals for seasonal parameters, which would have been screened out when sensitivity information is aggregated over a longer time period. Despite the larger computational demand, implementation of TVSA can be useful in model calibration by identifying event-based hydrological processes. This is particularly important for event-based calibration useful for flood forecasting in cold regions dominated by snowmelt runoff [67,75]. Furthermore, the choice of error metric has significant influence on the rankings and selection of parameter sensitivity. The application of *SFDC* is found to be most effective for TVSA and monthly SA, as the retention of influential parameters are higher compared to the long-term SA, despite having comparatively low reliability. We also found that *SFDC* consistently represented the parameter sensitivity throughout all time period, and across the dry, wet, and normal climate conditions.

In conclusion, TVSA is essential for complex, multi-dimensional models applied in seasonal and cold region environment to assist with the identification of highly influential parameters and parameter interactions prior to model calibration. Since the sensitivity of the parameters is strongly influenced by climate forcing such as air temperature and precipitation, we recommend the mix of wet and dry periods [55] while implementing a TVSA of hydrological model parameters. Regardless of the purpose of the model use, event based calibration is recommended for improved simulation of cold region hydrological processes. Finally, we also recommend the use of *SFDC* as an objective function for event based model calibration due to its capability to capture variation in the sensitivity of the parameters throughout the time period of analysis.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4441/12/4/961/s1>, Figure S1: Delineation of Sub-basins in HYPE for Nelson Churchill River Basin, Figure S2: Ratio of factor sensitivity for moving windows of 30 days, 60 days, 90 days, 180 days and 360 days using NRMSE as the evaluation criteria from 2000-01-01 to 2005-11-29, Figure S3: Daily Average Annual Hydrograph (1981-2010) for simulated streamflow (best parameter set from sampling for Sensitivity Analysis) and observed streamflow for major river system within NCRB, Figure S4: Ratio of factor sensitivity for moving windows of 30 days, 60 days, 90 days, 180 days and 360 days using 3-day averaged  $Q_5$  (high flow) as the evaluation criteria from 2000-01-01 to 2005-11-29, Table S1: Ranking of six most influential parameters (monthly) using different evaluation metrics (a) NSE (b) PBIAS (c)  $Q_{95}$ , and (d) SFDC.

**Author Contributions:** Conceptualization, A.B. and T.S.; methodology, A.B., T.S. and M.A.; validation, H.A., T.A. and S.A.; formal analysis, A.B. and T.S.; writing—original draft preparation, A.B.; writing—review and editing, H.A, T.S. and M.A.; supervision, T.S. and M.A.; project administration, T.S.; funding acquisition, T.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Manitoba Hydro, Natural Sciences and Engineering Research Council of Canada, and Global Water Futures.

**Acknowledgments:** We are thankful to Saman Razavi for providing the VARS toolbox. We are also grateful to SMHI for providing the HYPE model and Hydro-GFD dataset, ECMWF for ERA-Interim dataset, and NCEP for providing the NARR dataset.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Benke, K.K.; Lowell, K.E.; Hamilton, A.J. Parameter uncertainty, sensitivity analysis and prediction error in a water-balance hydrological model. *Math. Comput. Model.* **2008**, *47*, 1134–1149. [[CrossRef](#)]
2. Shen, Z.Y.; Chen, L.; Chen, T. Analysis of parameter uncertainty in hydrological and sediment modeling using GLUE method: A case study of SWAT model applied to Three Gorges Reservoir Region, China. *Hydrol. Earth Syst. Sci.* **2012**, *16*, 121–132. [[CrossRef](#)]
3. Feyen, L.; Kalas, M.; Vrugt, J.A. Semi-distributed parameter optimization and uncertainty assessment for large-scale streamflow simulation using global optimization/Optimisation de paramètres semi-distribués et évaluation de l’incertitude pour la simulation de débits à grande échelle pa. *Hydrol. Sci. J.* **2008**, *53*, 293–308. [[CrossRef](#)]
4. van Zelm, R.; Huijbregts, M.A.J. Quantifying the Trade-off between Parameter and Model Structure Uncertainty in Life Cycle Impact Assessment. *Environ. Sci. Technol.* **2013**, *47*, 9274–9280. [[CrossRef](#)] [[PubMed](#)]
5. Pianosi, F.; Wagener, T. Understanding the time-varying importance of different uncertainty sources in hydrological modelling using global sensitivity analysis. *Hydrol. Process.* **2016**, *30*, 3991–4003. [[CrossRef](#)]
6. Höllering, S.; Wienhöfer, J.; Ihringer, J.; Samaniego, L.; Zehe, E. Regional analysis of parameter sensitivity for simulation of streamflow and hydrological fingerprints. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 203. [[CrossRef](#)]
7. Razavi, S.; Gupta, H.V. What do we mean by sensitivity analysis? The need for comprehensive characterization of “global” sensitivity in Earth and Environmental systems models. *Water Resour. Res.* **2015**, *51*, 3070–3092. [[CrossRef](#)]
8. Tian, W. A review of sensitivity analysis methods in building energy analysis. *Renew. Sustain. Energy Rev.* **2013**, *20*, 411–419. [[CrossRef](#)]
9. Sobol, I.M. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simul.* **2001**, *55*, 271–280. [[CrossRef](#)]
10. Morris, M.D. Factorial sampling plans for preliminary computational experiments. *Technometrics* **1991**, *32*, 161–174. [[CrossRef](#)]
11. Rosolem, R.; Gupta, H.V.; Shuttleworth, W.J.; Zeng, X.; De Gonçalves, L.G.G. A fully multiple-criteria implementation of the Sobol’ method for parameter sensitivity analysis. *J. Geophys. Res. Atmos.* **2012**, *117*, D07103. [[CrossRef](#)]
12. Baroni, G.; Tarantola, S. A General Probabilistic Framework for uncertainty and global sensitivity analysis of deterministic models: A hydrological case study. *Environ. Model. Softw.* **2014**, *51*, 26–34. [[CrossRef](#)]
13. Razavi, S.; Gupta, H. A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1. Theory. *Water Resour. Res.* **2016**, *52*, 423–439. [[CrossRef](#)]

14. Razavi, S.; Tolson, B.A.; Burn, D.H. Review of surrogate modeling in water resources. *Water Resour. Res.* **2012**, *48*, W07401. [[CrossRef](#)]
15. Zhan, C.S.; Song, X.M.; Xia, J.; Tong, C. An efficient integrated approach for global sensitivity analysis of hydrological model parameters. *Environ. Model. Softw.* **2013**, *41*, 39–52. [[CrossRef](#)]
16. Song, X.; Zhang, J.; Zhan, C.; Xuan, Y.; Ye, M.; Xu, C. Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications. *J. Hydrol.* **2015**, *523*, 739–757. [[CrossRef](#)]
17. Asher, M.J.; Croke, B.F.W.; Jakeman, A.J.; Peeters, L.J.M. A review of surrogate models and their application to groundwater modeling. *Water Resour. Res.* **2015**, *51*, 5957–5973. [[CrossRef](#)]
18. Lilhare, R.; Pokorny, S.; Déry, S.J.; Stadnyk, T.A.; Koenig, K.A. Sensitivity Analysis and Uncertainty Assessment in Water Budget Simulated by the Variable Infiltration Capacity Model across the Lower Nelson River Basin, Manitoba, Canada. *Hydrol. Process.* **2020**. [[CrossRef](#)]
19. Liang, X.; Lettenmaier, D.P.; Wood, E.F.; Burges, S.J. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *J. Geophys. Res.* **1994**, *99*, 14115–14128. [[CrossRef](#)]
20. Rasouli, K.; Pomeroy, J.W.; Janowicz, J.R.; Carey, S.K.; Williams, T.J. Hydrological sensitivity of a northern mountain basin to climate change. *Hydrol. Process.* **2014**, *28*, 4191–4208. [[CrossRef](#)]
21. Fang, X.; Pomeroy, J.W. Snowmelt runoff sensitivity analysis to drought on the Canadian prairies. *Hydrol. Process.* **2007**, *19*, 2594–2609. [[CrossRef](#)]
22. Loucks, D.P.; van Beek, E.; Stedinger, J.R.; Dijkman, J.P.M.; Villars, M.T. Model Sensitivity and Uncertainty Analysis. *Water Resour. Syte. Plan. Manage.* **2005**, 255–290.
23. Muleta, M.K.; Nicklow, J.W. Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. *J. Hydrol.* **2005**, *306*, 127–145. [[CrossRef](#)]
24. Kannan, N.; White, S.M.; Worrall, F.; Whelan, M.J. Sensitivity analysis and identification of the best evapotranspiration and runoff options for hydrological modelling in SWAT-2000. *J. Hydrol.* **2007**, *332*, 456–466. [[CrossRef](#)]
25. Xie, H.; Shen, Z.; Chen, L.; Qiu, J.; Dong, J. Time-varying sensitivity analysis of hydrologic and sediment parameters at multiple timescales: Implications for conservation practices. *Sci. Total Environ.* **2017**, *598*, 353–364. [[CrossRef](#)]
26. Razavi, S.; Gupta, H.V. A multi-method Generalized Global Sensitivity Matrix approach to accounting for the dynamical nature of earth and environmental systems models. *Environ. Model. Softw.* **2019**, *114*, 1–11. [[CrossRef](#)]
27. Viglione, A.; Parajka, J.; Rogger, M.; Salinas, J.L.; Laaha, G.; Sivapalan, M.; Blöschl, G. Comparative assessment of predictions in ungauged basins—Part 3: Runoff signatures in Austria. *Hydrol. Earth Syst. Sci.* **2013**, *17*, 2263. [[CrossRef](#)]
28. Shafii, M.; Tolson, B.A. Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives. *Water Resour. Res.* **2015**, *51*, 3796–3814. [[CrossRef](#)]
29. Donnelly, C.; Andersson, J.C.M.; Arheimer, B. Using flow signatures and catchment similarities to evaluate the E-HYPE multi-basin model across Europe. *Hydrol. Sci. J.* **2016**, *61*, 255–273. [[CrossRef](#)]
30. Pan, S.; Fu, G.; Chiang, Y.M.; Ran, Q.; Xu, Y.P. A two-step sensitivity analysis for hydrological signatures in Jinhua River Basin, East China. *Hydrol. Sci. J.* **2017**, *62*, 2511–2530. [[CrossRef](#)]
31. Lindström, G.; Pers, C.; Rosberg, J.; Strömqvist, J.; Arheimer, B. Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales. *Hydrol. Res.* **2010**, *41*, 295–319. [[CrossRef](#)]
32. Zubrycki, K.; Roy, D.; Osman, H.; Lewtas, K.; Gunn, G.; Grosshans, R. *Large Area Planning in the Nelson-Churchill River Basin (NCRB): Laying a Foundation in Northern Manitoba*; International Institute for Sustainable Development: Winnipeg, MB, Canada, 2016.
33. Manitoba Hydro. Climate Change Report Fiscal Year 2014–2015. Available online: [https://www.hydro.mb.ca/environment/pdf/climate\\_change\\_report\\_2014\\_15.pdf](https://www.hydro.mb.ca/environment/pdf/climate_change_report_2014_15.pdf) (accessed on 15 June 2019).
34. Gray, D.M.; Toth, B.; Zhao, L.; Pomeroy, J.W.; Granger, R.J. Estimating areal snowmelt infiltration into frozen soils. *Hydrol. Process.* **2001**, *15*, 3095–3111. [[CrossRef](#)]
35. Zhang, T.; Barry, R.G.; Knowles, K.; Heginbottom, J.A.; Brown, J. Statistics and characteristics of permafrost and ground ice distribution in the Northern Hemisphere. *Polar Geogr.* **1999**, *23*, 132–154. [[CrossRef](#)]

36. Smith, M.W.; Riseborough, D.W. Climate and the limits of permafrost: A zonal analysis. *Permafr. Periglac. Process.* **2002**, *13*, 1–15. [[CrossRef](#)]
37. Kurz, D.; Alfaro, M.; Graham, J. Thermal conductivities of frozen and unfrozen soils at three project sites in northern Manitoba. *Cold Reg. Sci. Technol.* **2017**, *140*, 30–38. [[CrossRef](#)]
38. Carey, S.K.; Pomeroy, J.W. Progress in Canadian snow and frozen ground hydrology, 2003–2007. *Can. Water Resour. J.* **2009**, *34*, 127–138. [[CrossRef](#)]
39. Tefs, A.A.G.; MacDonald, M.K.; Stadnyk, T.A.; Koenig, K.A.; Hamilton, M.; Slota, P.; Crawford, J. Simulating effects of Nelson-Churchill River regulation controls on reservoir performance in HYPE. Unpublished work, 2019.
40. MacDonald, M.K.; Stadnyk, T.A.; Dery, S.; Gustafsson, D.; Isberg, K. Improved high-latitude water storage for hydrological modelling of the Hudson Bay Drainage Basin. Unpublished work. 2019.
41. Déry, S.J.; Stadnyk, T.A.; MacDonald, M.K.; Gauli-Sharma, B. Recent trends and variability in river discharge across northern Canada. *Hydrol. Earth Syst. Sci.* **2016**, *20*, 4801–4818. [[CrossRef](#)]
42. Sabarly, F.; Essou, G.; Lucas-Picher, P.; Poulin, A.; Brissette, F. Use of Four Reanalysis Datasets to Assess the Terrestrial Branch of the Water Cycle over Quebec, Canada. *J. Hydrometeorol.* **2016**, *17*, 1447–1466. [[CrossRef](#)]
43. Nash, J.E.; Sutcliffe, J.V. River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.* **1970**, *10*, 282–290. [[CrossRef](#)]
44. Gupta, H.V.; Sorooshian, S.; Yapo, P.O. Status of Automatic Calibration for Hydrologic Models: Comparison with Multilevel Expert Calibration. *J. Hydrol. Eng.* **1999**, *4*, 135–143. [[CrossRef](#)]
45. U.S. Environmental Protection Agency. *EPA's Report on the Environment (ROE) (2008 Final Report)*; U.S. Environmental Protection Agency: Washington, DC, USA, 2008.
46. Demirel, M.C.; Booij, M.J.; Hoekstra, A.Y. Effect of different uncertainty sources on the skill of 10 day ensemble low flow forecasts for two hydrological models. *Water Resour. Res.* **2013**, *49*, 4035–4053. [[CrossRef](#)]
47. Jiang, S. Hydrological Water Quality Modelling of Nested Meso Scale Catchments. Doctoral Thesis, Technical University of Braunschweig, Braunschweig, Germany, 2014.
48. Arheimer, B.; Pimentel, R.; Isberg, K.; Crochemore, L.; Andersson, J.C.M.; Hasan, A.; Pineda, L. Global catchment modelling using World-Wide HYPE (WWH), open data and stepwise parameter estimation. *Hydrol. Earth Syst. Sci. Discuss.* **2019**, *2*, 535–559. [[CrossRef](#)]
49. Gelfan, A.; Gustafsson, D.; Motovilov, Y.; Arheimer, B.; Kalugin, A.; Krylenko, I.; Lavrenov, A. Climate change impact on the water regime of two great Arctic rivers: Modeling and uncertainty issues. *Clim. Chang.* **2017**, *141*, 499–515. [[CrossRef](#)]
50. Razavi, S.; Gupta, H. A new framework for comprehensive, robust, and efficient global sensitivity analysis: 2. Application. *Water Resour. Res.* **2016**, *1*, 440–455. [[CrossRef](#)]
51. Castaings, W.; Dartus, D.; Le Dimet, F.X.; Saulnier, G.M. Sensitivity analysis and parameter estimation for distributed hydrological modeling: Potential of variational methods. *Hydrol. Earth Syst. Sci.* **2009**, *13*, 503–517. [[CrossRef](#)]
52. Gonzaga, M.; Oliveira, A.D.; Netto, D.A.; Joaquim, R.; Neves, D.J.; Nascimento, A.; Almeida, C. Sensitivity Analysis and Calibration of Hydrological Modeling of the Watershed Northeast Brazil. *J. Environ. Prot.* **2015**, *6*, 837–850.
53. Rosenberg, L.; Hammer, T.; Shaw, J. Software metrics and reliability. In Proceedings of the 9th International Symposium on Software Reliability Engineering, Paderborn, Germany, 4–7 November 1998.
54. Razavi, S.; Sheikholeslami, R.; Gupta, H.V.; Haghnegahdar, A. VARS-TOOL: A toolbox for comprehensive, efficient, and robust sensitivity and uncertainty analysis. *Environ. Model. Softw.* **2019**, *112*, 95–107. [[CrossRef](#)]
55. Zhao, Y.; Nan, Z.; Yu, W.; Zhang, L. Calibrating a hydrological model by stratifying frozen ground types and seasons in a cold alpine basin. *Water* **2019**, *11*, 985. [[CrossRef](#)]
56. Ghasemizade, M.; Baroni, G.; Abbaspour, K.; Schirmer, M. Combined analysis of time-varying sensitivity and identifiability indices to diagnose the response of a complex environmental model. *Environ. Model. Softw.* **2017**, *88*, 22–34. [[CrossRef](#)]
57. Massmann, C.; Wagener, T.; Holzmann, H. A new approach to visualizing time-varying sensitivity indices for environmental model diagnostics across evaluation time-scales. *Environ. Model. Softw.* **2014**, *51*, 190–194. [[CrossRef](#)]
58. Hong, Y. Actual evapotranspiration estimation for different land use and land cover in urban regions using Landsat 5 data. *J. Appl. Remote Sens.* **2010**, *4*, 041873. [[CrossRef](#)]

59. Verstraeten, W.W.; Muys, B.; Feyen, J.; Veroustraete, F.; Minnaert, M.; Meiresonne, L.; De Schrijver, A. Comparative analysis of the actual evapotranspiration of Flemish forest and cropland, using the soil water balance model WAVE. *Hydrol. Earth Syst. Sci. Discuss.* **2010**, *9*, 225–241. [CrossRef]
60. Klyatis, L.M.; Anderson, E. *Reliability Prediction and Testing Textbook*; Wiley: Hoboken, NJ, USA, 2018.
61. Campos, C.A.; Hernández, M.E.; Moreno-Casasola, P.; Cejudo Espinosa, E.; Robledo, R.A.; Infante Mata, D. Soil water retention and carbon pools in tropical forested wetlands and marshes of the Gulf of Mexico. *Hydrol. Sci. J.* **2011**, *56*, 1388–1406. [CrossRef]
62. Gray, D.M.; Landine, P.G.; Granger, R.J. Simulating infiltration into frozen Prairie soils in streamflow models. *Can. J. Earth Sci.* **1985**, *22*, 464–472. [CrossRef]
63. Watanabe, K.; Kito, T.; Dun, S.; Wu, J.Q.; Greer, R.C.; Flury, M. Water Infiltration into a Frozen Soil with Simultaneous Melting of the Frozen Layer. *Vadose Zone J.* **2013**, *12*, 1–7. [CrossRef]
64. Wang, S.; Fu, B.J.; Gao, G.Y.; Yao, X.L.; Zhou, J. Soil moisture and evapotranspiration of different land cover types in the Loess Plateau, China. *Hydrol. Earth Syst. Sci.* **2012**, *16*, 2883–2892. [CrossRef]
65. Jones, N.E.; Petreman, I.C.; Schmidt, B.J. *High Flows and Freshet Timing in Canada: Observed Trends*; Ontario Ministry of Natural Resources and Forestry: Peterborough, ON, Canada, 2015.
66. Fang, X.; Minke, A.; Pomeroy, J.W.; Brown, T.; Westbrook, C.; Guo, X.; Guangul, S. *A Review of Canadian Prairie Hydrology: Principles, Modelling and Response to Land Use and Drainage Change*; University of Saskatchewan: Saskatoon, SK, Canada, 2007.
67. Unduche, F.; Tolossa, H.; Senbeta, D.; Zhu, E. Evaluation of four hydrological models for operational flood forecasting in a Canadian Prairie watershed. *Hydrol. Sci. J.* **2018**, *63*, 1133–1149. [CrossRef]
68. Price, K.; Purucker, S.T.; Kraemer, S.R.; Babendreier, J.E. Tradeoffs among watershed model calibration targets for parameter estimation. *Water Resour. Res.* **2012**, *48*, W10542. [CrossRef]
69. Lawford, R.G.; Prowse, T.D.; Hogg, W.D.; Warkentin, A.A.; Pilon, P.J. Hydrometeorological aspects of flood hazards in Canada. *Atmos. Ocean* **1995**, *33*, 303–328. [CrossRef]
70. Environment Canada. Causes of Flooding. Available online: <https://www.canada.ca/en/environment-climate-change/services/water-overview/quantity/causes-of-flooding.html> (accessed on 2 February 2020).
71. Van Der Linden, S.; Virtanen, T.; Oberman, N.; Kuhry, P. Sensitivity analysis of discharge in the Arctic USA basin, East-European Russia. *Clim. Chang.* **2003**, *57*, 139–161. [CrossRef]
72. Zahmatkesh, Z.; Kumar Jha, S.; Coulibaly, P.; Stadnyk, T. An overview of river flood forecasting procedures in Canadian watersheds. *Can. Water Resour. J.* **2019**, *44*, 213–229. [CrossRef]
73. Schöber, J.; Achleitner, S.; Kirnbauer, R.; Schöberl, F.; Schönlaub, H. Hydrological modelling of glacierized catchments focussing on the validation of simulated snow patterns—Applications within the flood forecasting system of the Tyrolean river Inn. *Adv. Geosci.* **2010**, *27*, 99–109. [CrossRef]
74. Cordeiro, M.R.C.; Wilson, H.F.; Vanrobaeys, J.; Pomeroy, J.W.; Fang, X. Simulating cold-region hydrology in an intensively drained agricultural watershed in Manitoba, Canada, using the Cold Regions Hydrological Model. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 3483–3506. [CrossRef]
75. Bhuiyan, H.A.K.M.; McNairn, H.; Powers, J.; Merzouki, A. Application of HEC-HMS in a cold region watershed and use of RADARSAT-2 soil moisture in initializing the model. *Hydrology* **2017**, *4*, 9. [CrossRef]

