Responses to Reviewer 1 comments on Manuscript HESS-2016-511

Title: Evaluation of various daily precipitation products for large-scale hydro-climatic applications over Canada

Authors: Jefferson Wong et al

Manuscript No: hess-2016-511

The review comments are in regular bold typeface, while all responses are in italics and indented paragraphs.

Response to Reviewer 1

The authors evaluated various gridded precipitation datasets against long-term station data in order to assess accuracy of each datasets. The presence of multiple gridded precipitation datasets available to researchers these days, assessing the accuracy of these datasets (or more appropriately uncertainties in these available datasets) is a very important concern in hydrological research. From the point, this work can be a significant contribution to the hydrologic research for Canada. Despite its importance, I found a number of questions that need to be answered before this one is accepted for publication.

> *We are grateful to the reviewer for his/her review and comments and suggestions to improve our paper. We have now addressed all of the comments and presented our responses below.*

Specific comments:

1. **The authors compare gridded precipitation products against data at individual stations. A rain gauge data represents only a small area but the gridded data evaluated in this study, especially those based on model products, represent values at much larger area, essentially averages over individual grid boxes. How can we expect that a rain gauge data can represent an average value of hundreds of square kilometers? This must be thoroughly discussed to justify their methodology.**

   The reviewer's point is well-taken. We are aware of the challenges and issues with comparing point measurements and area-averaged estimates. However, in the absence of a sufficiently dense precipitation gauge network in Canada, our options for assessing the different gridded products would be limited. The only gridded product that is basically representing areal averages of precipitation (via interpolation) based on ground observations is ANUSPLIN. This product, however, may not be qualified to be considered as the ground truth, as it has its own limitations which has already raised in the original manuscript (see Section 3.2.1). Therefore, we also included ANUSPLIN in the pool of gridded products to be evaluated.

Notwithstanding the issues, we found that using the selected gauge measurements would remain the best way for the evaluation of the multiple gridded products. First, this is because the set of gauges used had been adjusted (e.g. for undercatch) and are the most accurate source of information on precipitation in Canada (although small with limited spatial coverage). Second, given that we compared all the gridded data products against this common set of point-based measurements, it may be safe to assume that the bias that the difference between point and areal data inserts into comparisons is pretty much consistent for all the products. In other words, it would not work in favor of one product and against one other product. Parts of this discussion is already in the original manuscript (Section 6, P26:L8-13). We will certainly better reflect on this in the revised manuscript, which is shown as follows (coloured in red):

> *In addition, results from the above analysis should be interpreted with care because the precipitation-gauge station data are point measurements whereas the gridded precipitation products are areal averages, of which the accuracy and precision of the estimates could be very different given the non-linear responses of precipitation (Ebert et al., 2007). When comparing point measurements and areal-average estimates, fundamental challenges occur because of the sampling errors arising from different sampling schemes and errors related to gauge instrumentation (Bowman, 2005). It is therefore difficult to have perfect spatial matching between point measurements (gauge stations) and areal-averaged estimates (gridded products) (Sapiano and Arkin, 2009;Hong et al., 2007). However, in the absence of a sufficiently dense precipitation gauge network in Canada, the options for assessing different gridded products are limited. The only gridded product that is basically representing areal averages of precipitation (via interpolation) based on ground observations is ANUSPLIN. As aforementioned (see Sect. 3.2.1), this product has its own limitations and may not be qualified to be considered as the ground truth. Therefore, ANUSPLIN is also included in the pool of gridded products to be evaluated. Notwithstanding the issues, using the selected gauge measurements would remain the best way for the evaluation of the multiple gridded products because the set of gauges used had been adjusted (e.g. for undercatch) and are the most accurate source of information on precipitation in Canada (although small with limited spatial coverage). Also, given that all the gridded products are compared against this common set of point-based measurements, it is assumed that the bias that the difference between point and areal data inserts into comparisons is pretty much consistent for all the products. In other words, it would not work in favour of one product and against on other product. ~~However,~~ Therefore, the authors believe that given the current data situation, the preceding was the best methodology for ~~evaluating~~ comparing the performance of different daily gridded precipitation products.*

2. **To prepare for evaluations, the authors first interpolated all gridded data into a common grid of 0.5deg resolutions, then they re-interpolated from the grid to the location of individual rain gauges. This data processing includes two spatial interpolations. Because every interpolation step introduces its own errors or uncertainties, the number of interpolation steps must be as small as possible. I wonder why they did not directly interpolate each data set to the rain gauge locations without going through the intermediate grid? This can simply data processing and can reduce interpolation-related uncertainties.**

This point is also well-taken. But there are two things to consider. First of all, upscaling to a coarser grid size (e.g., from 10km to 50km) is mainly by averaging, and therefore, it would not introduce any significant errors into the upscaled data (unlike interpolation). Second, as the reviewer suggested, we could easily compare each gridded product against the point observations at their original resolutions. However, the main focus of this study is to inter-compare various gridded precipitation products using precipitation-gauge station data as a reference/benchmark but not to assess the individual accuracy of each product against the reference dataset. In other words, evaluating each product at its original spatial resolution can only provide the errors associated with that product at its original resolution but cannot directly compare with other products of different spatial resolutions. Therefore, we opted to upscale them all to a common (coarser) grid size first. This way the inter-comparison would be more consistent, as the different products when brought to a common scale are expected to show more similar statistical properties. For example, by coarsening, you expect to reduce the temporal variability of data (manifested in variance) as well, while you may not want these differences due to having different spatial resolutions obscure your inter-comparisons.

Moreover, the original spatial resolution of the climate forcing data does not always match with the spatial resolution of a large-scale hydrological model. Thus, under this circumstance, upscaling a fine resolution product is a necessary process for large-scale hydrological applications. The results of this study could therefore better reveal the errors incorporated in the rescaled precipitation products as the errors include the interpolation-related uncertainties. Also note that our methodology is consistent with the similar studies in the literature (e.g. Janowiak et al., 1998;Rauscher et al., 2010;Kimoto et al., 2005).

Accordingly, we will change the title from "Evaluation" to "Inter-comparison" to better reflect our aim of the study. Also, we will insert the above justification in the revised manuscript following the last paragraph of Section 4 in the original manuscript [P17:L4], which is shown as follows:

> *It is acknowledged that re-gridding products onto a common spatial resolution and then interpolating at the coordinates of the precipitation-gauge stations might*

introduce more errors or uncertainties and the number of interpolation steps should be minimized. However, the main focus of this study is to inter-compare various gridded precipitation products using precipitation-gauge station data as a reference/benchmark but not to assess the individual accuracy of each product against the reference dataset. Therefore, upscaling to a common resolution provides a direct and more consistent inter-comparison, as the different products when brought to a common scale are expected to show more similar statistical properties. For instance, by coarsening, the temporal variability of data (manifested in variance) are expected to be reduced, while these differences due to having different spatial resolutions are not desirable to obscure the inter-comparison. Moreover, upscaling to a coarser grid size (e.g., from 10km to 50km) is mainly by averaging, and therefore, it would not introduce any significant errors into the upscaled data (unlike interpolation). In addition, the original spatial resolution of the climate forcing data does not always match with the spatial resolution of a large-scale hydrological model. Thus, under this circumstance, upscaling a fine resolution product is a necessary process for large-scale hydrological applications. Therefore, we believe that re-gridding each product onto a common spatial and temporal resolution provide a fair ground for inter-comparison and better reveal the errors incorporated in the rescaled precipitation products as the errors include the interpolation-related uncertainties.

Lastly, motivated by this review comment, we conducted an inter-comparison test at the original scale (0.0833°) of ANUSPLIN against the upscale resolution (0.5°) in two ecozones (Ecozone 3 Southern Arctic and Ecozone 6 Boreal Shield where the numbers of precipitation-gauge stations are the least and largest, respectively), as shown in the following tables. The results show that the original and upscale resolutions produce performance measures of similar magnitude and the differences are not significantly large. Therefore, we believe that the interpolation-related uncertainties will be relatively smaller than the uncertainties arisen from other sources such as model structure, equifinality of parameters, and process representations.

Region (Ecozone)		*Resolution*							
		original (0.0833°)				*Upscale (0.5°)*			
		1979-2012		*2002-2012*		*1979-2012*		*2002-2012*	
		PBias	*RMSE*	*PBias*	*RMSE*	*PBias*	*RMSE*	*PBias*	*RMSE*
3	Southern Arctic	*-24.16*	*1.83*	*-36.88*	*2.37*	*-27.89*	*1.84*	*-40.70*	*2.38*
6	Boreal Shield	*-12.46*	*3.34*	*-14.15*	*3.39*	*-13.87*	*3.35*	*-15.67*	*3.40*

We believe including these results in the revised manuscript might be divergent from the core objective of this study. However, we would be open to suggestions by the reviewer.

3. **Model products based on RCP scenarios includes the effects of hypothetical emissions pathways implemented in these simulations. How can these model data be compared against the reference data in the same wat as other assimilated and/or station-based gridded data? The authors evaluated these data sets for two periods, 1979-2012 and 2002-2012. The CMIP5 experiment that seem relevant to the model data used in this study was designed in such a way that the present-day period simulation based on the realistic GHG concentration for the period from mid-19th century to 2005. Future projections based on specific RCP scenarios starts from 2005 up to 2300 with the initial condition taken at the end of the present-day simulation period. Thus all model data after 2005 are affected by hypothetical emissions pathways. It's pretty unusual discussing "accuracy" of the data generated to project future on the basis of hypothetical GHG concentrations. If the authors are interested in evaluating the model-generated data, the comparison must end in 2005, the end of the present-day period for which the observed external forcing and GHG concentration are implemented. For such periods, there is no need to distinguish runs according to RCP scenarios because the hypothetical emissions pathways are not implemented. My suggestion is to drop model data from the evaluation, or include the model data and limited the evaluation period to 2005 instead of 2012.**

We appreciate the value of the reviewer's comment on the climate model products and we agree that evaluating the model-generated products with different RCP scenarios is not appropriate. We decide to include the climate model products for evaluation and limit the evaluation period to 2005 instead of 2012. The reason of not dropping all the climate model products from the evaluation is we think it is still worthwhile to compare different climate model products to see which downscaling methods/which GCMs/RCMs provide better historical estimates so that potential users could use the results as a reference for their future climate change studies. Accordingly, the description of the evaluation period in the original manuscript [P16:L14-17] will be revised to reflect the change, with deleted materials being crossed out by drawing a line through them (and revised sentences being coloured in red):

> *Two common time spans were selected since CaPA covered a shorter time frame when compared to the rest of the products: (1) long-term comparison from January 1979 to December 2012 with the exclusion of CaPA (from January 1979 to December 2005 for PCIC and NA-CORDEX); and (2) short-term comparison from January 2002 to December 2012 when CaPA are available.*

Also, the percentage of reliability will be re-calculated for the climate model products in the revised manuscript and the results will be revised in the Results Section (Section 5.1), which is shown as follows:

Regarding the PCIC ensembles, the different GCMs provided a range of reliabilities for the individual seasons. ~~GFDL-ESM2G performed the best in spring (58.6 %) while CanESM2 in autumn (43.8 %). MPI-ESM-LR generally gave more reliable estimates in summer and winter (64.5 % and 38.3 %).~~MPI-ESM-LR performed the best in summer (70.2 %) while CanESM2 in autumn (45.5 %). GFDL-ESM2G generally gave more reliable estimates in spring and winter (57.4 % and 41.7 %). ~~The performance of HadGEM2-ES RCP 8.5 with BCCAQ statistical downscaling method was significantly poorer than the rest of the GCM ensembles, especially in summer (13.1 %).~~ Overall, the performance of MPI-ESM-LR (~~49.1 %~~52.0 %) was the best among the GCMs, followed by GFDL-ESM2G (~~47.0 %~~50.1 %), CanESM2 (~~42.2 %~~47.8 %), and HadGEM2 (~~36.7 %~~36.2 %). In terms of statistical downscaling methods, the BCCAQ method was on average slightly better than BCSD (~~47.5% versus 45.4 %~~49.5 % versus 44.0 %) with the former having a greater similarity in spring and summer as opposed to autumn and winter. These small differences therefore suggest that both methods are similar. With respect to the NA-CORDEX ensembles, the CRCM5 RCM gave the most reliable estimates in summer and autumn regardless of the GCM used. CanRCM4 had the best reliability in spring (~~46.9 %~~49.4 %) whereas RegCM4 had the poorest reliability in spring and summer (~~22.1 % and 36.6 %~~24.4 % and 34.0 %). ~~In addition, the CanESM2-driven CanRCM4 with RCP 4.5 and RCP 8.5 were equally reliable in four seasons.~~ Overall, the reliability of MPI-ESM-LR (~~44.8 %~~44.7 %) was better than that of CanESM2 (~~40.6 %~~42.5 %) regardless of the RCMs used whereas the reliability of CRCM5 (~~43.3 %~~43.6 %) was the best among the RCMs, followed by CanRCM4 (~~39.5 %~~41.2 %), and RegCM4 (~~33.3 %~~32.5 %). It should also be noted that in all cases, the station-based and reanalysis-based products outperformed the climate model-simulated products.

With regard to the short-term comparison (Fig. 2 right panel), ANUSPLIN had the best performance in summer with 94.1 % of reliability among the 137 precipitation-gauge stations while CaPA was the best in winter with 68.6 % of reliability. Again, WFDEI [GPCC] in general provided the most consistent and reliable estimates with over 65 % of reliability in four seasons. ~~Similar performances were seen among the PCIC ensembles and the NA-CORDEX ensembles in the period of 2002 to 2012 as compared with the long-term performance.~~ It is interesting to note that for the most part, there is a higher percentage of reliability in short-term period compared to long-term period. Reasons for this are not clear but can be partly attributed to the fact that the power of K-S test (i.e. the probability of rejecting the null hypothesis when the alternative is true) decreases with the number of samples.

Figures ~~3 and 4~~3, 4 and 5 display the seasonal distributions of p-value using the K-S test in the 15 ecozones for long-term and short-term comparison, respectively.

Due to the uneven distribution of precipitation-gauge stations across Canada, the numbers of stations in each ecozone are different (Table 4), with no stations in Region 1 (Arctic Cordillera), and Regions 2 to 5, 10, 12, and 15 have less than 10 stations. The percentage of missing values in precipitation-gauge station in Region 11 exceeded 10 % in the period of 2002 to 2012 and thus the station was dropped out for analysis, resulting in no stations in Region 11 for short-term comparison. As a result, two representations were used to show the distributions of p-values. Regions having more than or equal to 10 stations (6 to 9 and 13, 14) were shown in box-whisker plots with bottom, band (thick black line), and top of the box indicating the 25th, 50th (median), and 75th percentiles, respectively. Regions having less than 10 stations were given by hollow circles with each representing one p-value at one precipitation-gauge station. Different colours in the figures correspond to the various precipitation products. The more numbers of high p-values (> 0.05) are in one ecozone (either represented by a cluster of hollow circles or a thick black line in box-whisker plots towards 1 in y-axis in ~~3 and 4~~3, 4 and 5), the more confidence (more consistent) one has that the gridded precipitation datasets provide reliable estimates in that ecozone.

From 1979 to 2012 (Fig. 3), in regions where more precipitation-gauge stations were available (6 to 10, 13, and 14), the consistency of each type of precipitation products is explored by assessing the median of the p-values. Overall, all the precipitation products showed very low reliability and consistency in winter among these ecozones and in every season in Regions 13 and 14 (Pacific Maritime and Montane Cordillera) as the medians were close to zero, despite a couple of locations having higher chance of same CDFs as in the precipitation-gauge station data. The WFDEI [GPCC] dataset provided the highest consistency in the remaining three seasons except for Region 7 (Atlantic Maritime) where ANUSPLIN showed higher medians (0.51 and 0.46) than WFDEI [GPCC] (0.42 and 0.42) in spring and autumn respectively. Noticeably NARR provided the lowest median among the reanalysis-based datasets in all four seasons in Regions 6 to 8 but gave fairly consistent estimates in Regions 9 and 10, especially in summer in Region 9 (Boreal Plain) where it came second after WFDEI [GPCC]. The medians of Princeton were similar with that of ANUSPLIN on average in these regions except for summer in which ANUSPLIN offered higher medians than Princeton. WFDEI [CRU] generally showed consistent estimates among these ecozones with medians well above 0.05 except for Region 7 (Atlantic Maritime) in spring and autumn. ~~The~~ From 1979 to 2005 (Fig. 5), the PCIC ensembles and the NA-CORDEX ensembles showed different degrees of consistency among their GCM members with generally higher p-values using BCCAQ method than BCSD method in spring and summer regardless of GCMs in the PCIC datasets, whereas CanESM2 was generally having higher consistency and reliable estimates than

MPI-ESM-LR in spring and summer but opposite case in autumn in the NA-CORDEX ensembles.

In ecozones above 60° N (Regions 2 to 5, 11, and 12), almost all the precipitation products had lower chance of having same CDFs as the precipitation-gauge stations, especially in spring, autumn, and winter in Region 3 (Southern Arctic) and spring and summer in Region 11 (Taiga Cordillera). The WFDEI [GPCC] and WFDEI [CRU] generally tended to provide higher p-values in these regions in spring and summer, followed by the NARR dataset. The NA-CORDEX ensembles provided slightly higher chance of having same CDFs as the precipitation-gauge stations than the PCIC ensembles in Regions 2 to 5 in spring and autumn whereas the opposite case was shown in Region 12 (Boreal Cordillera) in spring.

For the shorter time period of 2002 to 2012 (Fig. 4), CaPA showed the highest consistency in winter in Regions 6, 8, 9, and 13 whereas ANUSPLIN was the highest in summer in Regions 8, 13, and 14, echoing the results found in Fig. 2. However, the reliability and consistency of CaPA in summer was not particularly high, especially in Regions 8 and 13 where the medians were approaching zero. In addition, in ecozones above 60° N, the performances of CaPA were generally similar to that of the WFDEI [GPCC] with higher chance of providing reliable estimates in autumn. Similar performances were seen among the other precipitation products in the period of 2002 to 2012 as compared with the long-term performance, despite some regional and seasonal differences.

The Discussion Section (Section 6) related to the climate model products [P25:L11-24] will also be revised to reflect the change, which is shown as follows:

By matching the statistical property of the adjusted gauge measurements at monthly time scale, one could establish the confidence in using the climate model-simulated products for long-term hydro-climatic studies. Comparing the overall reliability of the PCIC and NA-CORDEX datasets, it was found that for the individual seasons the PCIC ensembles (~~from spring to winter: 52.2 %, 56.0 %, 41.9 %, and 32.4 %~~ spring, summer, and winter: 54.0 %, 64.7 %, and 35.7 %) outperformed the NA-CORDEX ensembles (~~34.5 %, 41.4 %, 38.3 %, and 31.7 %~~39.1 %, 45.0 %, and 31.3 %) ~~under RCP 8.5 scenario. This result was the same under RCP 4.5 scenario~~ except in autumn when the NA-CORDEX ensembles (~~46.2 %~~45.5 %) provided slightly higher reliability than the PCIC ensembles (~~42.5 %~~45.2 %). The better reliability of the PCIC datasets could be due to the use of ANUSPLIN to train the GCMs and thus, the statistical properties of the downscaled outputs are guided by those of the ANUSPLIN. Similarly, for ecozones where more than 10 precipitation-gauge stations could be found (Regions 6 to 9, 13 and 14), the PCIC ensembles (reliability ranging from ~~36.4 % to 68.1 %~~35.7 % to 64.4 %) also outperformed the

NA-CORDEX ensembles (from ~~16.8 % to 49.9 %~~17.2 % to 61.6 %). This would suggest that the PCIC ensembles may be the preferred choice for long-term climate change impact assessment over Canada, although further research is required.

Please note that the re-calculation does not affect the overall conclusion we made in the original manuscript.

4. **The authors provide lengthy descriptions on the details of the data sets used in this study. Much of these discussions are unnecessary because there were developed by other research groups and relevant publications on the details of these data sets are already available. Sections 3.1 and 3.2 can be reduced by referencing suitable publications.**

The details in Sections 3.1 and 3.2 will be greatly reduced in the revised manuscript. In short, the spatial and temporal resolutions of each product, their compositions, and examples of their applications will be remained and other details will be deleted. The following shows the revised Sections 3.1 and 3.2, with deleted materials being crossed out by drawing a line through them (and revised sentences being coloured in red):

3.1 Precipitation-gauge station data

Climate data collection is coordinated by the Federal government of Canada. Agriculture and Agri-Food Canada maintains a few stations nationally especially in Alberta province. Also, most hydro-power companies collect their own data. However, their data are not made available to the public but are sent to Environment and Climate Change Canada for archiving prior to release. In other words, the National Climate Data Archive of Environment Canada provide the basis for all the available climate data. Based on the National Climate Data Archive of Environment Canada, there are a total of 1499 precipitation-gauge stations (as in 2012) across Canada. However, due to the addition and subtraction of climate stations over the past few decades, the number of stations with available precipitation data for specified time intervals varies greatly. For instance, the numbers of precipitation-gauge stations that were active in any given years over the period of 1961 to 2003 ranged from 2000 to 3000 (see Hutchinson et al. (2009) Figs 1 and 2 for details). The issue with these data is they are subject to various errors, among which the errors due undercatch are quite significant in Canada (Mekis and Hogg, 1999). In order to account for various measurement issues, Mekis and Vincent (2011) provided adjusted daily rainfall and snowfall data for 464 stations over Canada that were based on the Adjusted Precipitation for Canada dataset (Mekis and Hogg, 1999). The data extend back to 1895 for a few long-term stations and run through 2014. For these data, daily rainfall gauge and snowfall ruler data were extracted from the National Climate Data Archive of Environment Canada and adjustments of rain and snow were done *separately~~. Regarding each rain gauge type, corrections for wind undercatch, evaporation and wetting losses were performed based on field experiments at various locations~~ (Devine and Mekis, 2008)~~. For snowfall, a density correction based on coincident ruler and Nipher gauge observations was applied to all snow measurements~~ (Mekis and Brown, 2010). ~~Adjustments were also implemented to account for trace precipitations and accumulated amounts from multiple days were distributed over the affected days to minimize the impact on extreme values and preserve the monthly totals.~~ Observations from nearby stations were sometimes combined to create longer time series ~~and adjustments were done either based on overlapping observations or standardized ratios between test sites and their neighbours~~ (Vincent and Mekis, 2009). As a result of adjustments, total rainfall amounts were concluded to be 5 to 10 % higher in southern Canada and more than 20 % in the Canadian Arctic than the original observations. The effect of the adjustments on snowfall were larger and more variable throughout the country. Despite the lack of a measure of associated uncertainty, this adjusted precipitation-gauge station dataset has been recognized and widely used for different analyses (e.g. Nalley et al., 2012;Shook and Pomeroy, 2012;Wan et al., 2013). Therefore, this dataset was used in this study as the reference to represent the best available precipitation measurement and as the benchmark for all gridded precipitation product comparisons.*

3.2 Gridded precipitation products

Seven precipitation datasets were assessed. Table 1 provides a concise summary of these datasets, including their full names, and original spatial and temporal resolutions for the versions used. These particular datasets were chosen based on the following criteria: (1) a complete coverage of Canada; (2) minimum of daily temporal and 0.5° (~50 km) spatial resolutions; (3) sufficient lengths of data (>30 years) for long-term study and cover recent years up to 2012; and (4) representation of a range of sources/methodologies (e.g. station based, remote sensing, model, blended products). Note that other commonly used datasets including the monthly Canadian Gridded temperature and precipitation (CANGRD) dataset (Zhang et al., 2000) and the coarser resolution Japan Meteorological Agency 55-year Reanalysis (JRA-55) (Onogi et al., 2007;Kobayashi et al., 2015) and the Modern-Era Retrospective Analysis for Research and Applications (MERRA) (Rienecker et al., 2011) products were excluded as they do not meet criteria # 2 above.

3.2.1 Station-based product – ANUSPLIN

With the application of the Australian National University Spline (ANUSPLIN) model (Hutchinson, 1995;Hutchinson, 2004), Hutchinson et al. (2009) developed a climate dataset of daily precipitation and daily minimum and maximum air temperature over Canada at a spatial resolution of 300 arc-second of latitude and longitude (0.0833° or ~10 km) for the period of 1961 to 2003, using observed stations (from 2000 to 3000 in any given years over the period) recorded in the National Canadian Climate Data Archives of Environment Canada. However, to retain a better spatial coverage, no adjustments were done on the archive station data before the generation of the product. The dataset was generated to model the complex spatial patterns by using tri-variate thin-plate smoothing splines method that incorporated spatially continuous functions of latitude, longitude, and elevation. Hopkinson et al. (2011) subsequently extended this original dataset to include the period of 1950 to 2011. This ANUSPLIN product for Canada (hereafter the ANUSPLIN) ~~has first been quality controlled with various flags indicating trace values, accumulated values over multiple days, and missing and estimated values. The accuracy of the product was then assessed by withholding from the analyses 50 stations broadly representing the southern half of Canada and by examining the error statistics for the withheld stations. The ANUSPLIN dataset~~ has further been updated to 2013 and has recently been used as the basis of 'observed' data for evaluating different climate datasets (e.g. Eum et al., 2012) and for assessing the effects of different climate products in hydrological applications (e.g. Eum et al., 2014;Bonsal et al., 2013;Shrestha et al., 2012a).

3.2.2 Station-based model-derived product – CaPA

Initiated in November 2003 through collaborations within the Meteorological Service of Canada, the Canadian Precipitation Analysis (CaPA) was developed to produce a dataset of 6-hourly precipitation accumulation over North America in real-time at a spatial resolution of 15 km from 2002 onwards (Mahfouf et al., 2007). The dataset was generated based on an optimum interpolation technique (Daley, 1993), which required a background field and a specification of error statistics between the observations and the background field (e.g. Bhargava and Danard, 1994;Garand and Grassotti, 1995). For Canada, the short-term precipitation forecasts from the Canadian Meteorological Centre (CMC)'s regional model, the Global Environmental Multiscale (GEM) (Cote et al., 1998a;1998b), were used as the background field with the rain-gauge measurements from the observational network as the observations. ~~The analysis was created by simple kriging to interpolate the differences between the transformed data of GEM and stations, which was then re-transformed and applied back to GEM. The quality of rain-gauge stations was controlled by cross-checking with the neighbouring stations and by comparing with the radar-derived precipitation. The accuracy of the product was assessed by generating an analysis error that represented the amount of additional information gained from the multiple observations with regard to the background field.~~ CaPA has become operational at the CMC in April 2011, with updates to the statistical interpolation method (Lespinas et al., 2015), increase of spatial resolution to 10 km and the assimilation of Quantitative Precipitation Estimates from the Canadian Weather Radar Network as an additional source of observations (Fortin et al., 2015b). With its continuous improvement and different configurations, CaPA has been employed in Canada for various environmental *prediction applications (e.g. Eum et al., 2014;Fortin et al., 2015a;Pietroniro et al., 2007;Carrera et al., 2015). However, the study period of these applications only extended back to 2002.*

3.2.3 Reanalysis-based multiple-source products – Princeton, WFDEI, and NARR

Princeton

The Terrestrial Hydrology Research Group at the Princeton University initially developed a dataset of 3-hourly near-surface meteorology with global coverage at a 1.0° spatial resolution (~120 km) from 1948 to 2000 for driving land surface models and other terrestrial systems (Sheffield et al., 2006). The global dataset at the Princeton University (called hereafter the "Princeton") was constructed based on the National Centers for Environmental Prediction-National Center for Atmospheric Research (NCEP-NCAR) reanalysis (2.0° and 6-hourly) (Kalnay et al., 1996;Kistler et al., 2001), combining with a suite of global observation-based data including the Climatic Research Unit (CRU) monthly climate variables (New et al., 2000, 1999), the Global Precipitation Climatology Project (GPCP) daily precipitation (Huffman et al., 2001), the Tropical Rainfall Measuring Mission (TRMM) 3-hourly precipitation (Huffman et al., 2002), and the NASA Langley Research Center monthly surface radiation budget (Gupta et al., 1999). ~~Regarding precipitation, the dataset has undergone several stages in terms of spatial downscaling with the use of GPCP data, temporal downscaling based on sampling from TRMM data, and the sophistication of the correction methods (a correction to the wet-day statistics (Sheffield et al., 2004), and monthly bias corrections to match those of the CRU data (Adam and Lettenmaier, 2003)). The Princeton dataset has been evaluated against the Second Global Soil Wetness Project (GSWP-2) product (Zhao and Dirmeyer, 2003).~~ With the inclusion of new temperature and precipitation data (e.g. Willmott et al., 2001), Princeton has been updated and is currently available at 1.0° (plus 0.5° and 0.25°), 3-hourly (plus daily and monthly) resolution globally for 1948 to 2008. Experimental updates including a 1901-2012 version at 1.0° (plus 0.5°), 3-hourly (plus daily and monthly) resolution are also available. Studies employing Princeton to study different hydrological aspects have been carried out over different parts of Canada (e.g. Kang et al., 2014;Su et al., 2013;Wang et al., 2013;Wang et al., 2014).

WFDEI

To simulate the terrestrial water cycle using different land surface models and general hydrological models, the European Union Water and Global Change (WATCH) Forcing Data (WFD) were created to provide datasets of sub-daily (3-hourly or 6-hourly) and daily meteorological data with global coverage at a 0.5° spatial resolution (~50 km) from 1901 to 2001 (Weedon et al., 2011). Similar to the composition of the Princeton dataset, the WFD were derived from the 40-year European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40) (1.0° and 3-hourly) (Uppala et al., 2005) and *combined with the CRU monthly variables and the Global Precipitation Climatology Centre (GPCC) monthly data (Rudolf and Schneider, 2005;Schneider et al., 2008;Fuchs, 2009). ~~The generation of the WFD for 1958 to 2001, which was based on the ERA-40, followed the procedures developed by Ngo-Duc et al. (2005) and Sheffield et al. (2006) whereas the dataset for 1901 to 1957 was generated by using the reordered ERA-40 a year at a time. With respect to precipitation, the creation of the data (Weedon et al., 2010) involved spatially downscaling using the CRU data, sequential elevation correction, wet-day correction, monthly precipitation bias correction to match the GPCC data, and adjustment for gauge undercatch (Adam and Lettenmaier, 2003), however no corrections were made for orography effect (Adam et al., 2006). The same monthly bias corrections were also done using the CRU precipitation totals, resulting in two sets of precipitation data. The WFD were assessed by the FLUXNET data for selected years at seven sites (Araujo et al., 2002;Persson et al., 2000;Suni et al., 2003;Meyers and Hollinger, 2004;Grunwald and Bernhofer, 2007;Urbanski et al., 2007;Gockede et al., 2008).~~ The WATCH Forcing Data methodology applied to ERA-Interim (WFDEI) dataset has further been generated covering the period of 1979 to 2012 (Weedon et al., 2014). The WFDEI used the same methodology as the WFD, but based on the ERA-Interim (Dee et al., 2011) with higher spatial resolution (0.7°)~~, better data assimilation technique, updated monthly observation-based data, more extensive incorporation of observations, and correction of the most extreme cases of inappropriate precipitation phase~~. As for the WFD, the WFDEI had two sets of rainfall and snowfall data generated by using either CRU or GPCC precipitation totals (hereafter the WFDEI [CRU] and WFDEI [GPCC] respectively). To date, specific studies using the WFDEI related to Canada has been limited to the studies of permafrost in the Arctic regions (e.g. Chadburn et al., 2015;Park et al., 2015;Park et al., 2016) but the WFDEI could be a potential source in other environmental applications in Canada.*

NARR

Concerning the spatial and temporal water availability in the atmosphere, the North American Regional Reanalysis (NARR) was developed to provide datasets of 3-hourly meteorological data for the North America domain at a spatial resolution of 32 km (~0.3°) covering the period of 1979 to 2003 as the retrospective system and is being continued in near real-time (currently up to 2015) as the Regional Climate Data Assimilation System (R-CDAS) (Mesinger et al., 2006). The components in generating NARR included the NCEP-DOE reanalysis (Kanamitsu et al., 2002), the NCEP regional Eta Model (Mesinger et al., 1988;Black, 1988) and ~~its Data Assimilation System, a recent version of~~ the Noah land-surface model (Mitchell et al., 2004;Ek et al., 2003), and the use of numerous additional data sources (see Mesinger et al., 2006 Table 2). ~~The use of NCEP-DOE reanalysis was a major improvement upon the earlier NCEP-NCAR reanalysis in both resolution and accuracy to provide lateral boundary conditions. Regarding~~

~~precipitation assimilation scheme, the NARR adjusted the accumulated convective and grid-scale precipitation, assimilated the precipitation observations as latent heating profiles based on the differences between the modelled and observed precipitation (Lin et al., 1999), and disaggregated into hourly resolution using different sources over lands and oceans. For the period from 1979 to 2003 when NARR was run as the retrospective system, precipitation analyses over the continental United States (CONUS), Mexico, and Canada were derived solely from a gridded analysis of 24-hour rain-gauge measurements. For the period from 2004 onwards, NARR was generated in near-real time by the R-CDAS, which was identical to the retrospective NARR except for changes in input sources and their processing because of the real-time production constraints. One of the major differences was the use of radar-dominated precipitation analyses derived from the National Land Data Assimilation System (NLDAS) (Mitchell et al., 2004) over CONUS to disaggregate the 24-hour rain-gauge analysis to hourly precipitation whereas no assimilation was done over Canada due to the paucity of rain-gauge observations.~~ On the basis of hydrological modelling in Canada, Choi et al. (2009) found that NARR provided reliable climate inputs for northern Manitoba while Woo and Thorne (2006) concluded that NARR had a cold bias resulting in later snowmelt peaks in subarctic Canada. In addition, Eum et al. (2012) identified a structural break point in the NARR dataset over the Athabasca River basin.

3.2.4 GCM statistically downscaled products – PCIC

The Pacific Climate Impacts Consortium (PCIC), which is a regional climate service centre at the University of Victoria, British Columbia, has offered datasets of statistically downscaled daily precipitation and daily minimum and maximum air temperature under three different Representative Concentration Pathways (RCPs) scenarios (RCP 2.6, RCP 4.5, and RCP 8.5) (Meinshausen et al., 2011) over Canada at a spatial resolution of 300 arc-second (0.833° or ~10 km) for the historical and projected period of 1950 to 2100 (Pacific Climate Impacts Consortium; University of Victoria, Jan 2014). These downscaled datasets were a composite of 12 GCM projections from the Coupled Model Inter-comparison Project Phase 5 (CMIP5) (Taylor et al., 2012) and the ANUSPLIN dataset. The historical 1950 to 2005 period of the ANUSPLIN was used to drive the GCMs and the statistical properties and spatial patterns of the downscaled outputs tended to resemble those of the ANUSPLIN. ~~However, the timing of natural climate variability (e.g. El Niño-Southern Oscillation) in the observational record were not considered since GCMs were solved as a 'boundary value problem'.~~ Two different downscaling methods were used to downscale to a finer resolution. The first one was Bias Correction Spatial Disaggregation (BCSD) (Wood et al., 2004) following Maurer and Hidalgo (2008) and the second was Bias Correction Constructed Analogues (BCCA) with Quantile mapping reordering (BCCAQ) which was a post-processed version of BCCA (Maurer et al., 2010) . ~~In general, the most important distinction between the two methods was BCCAQ~~

~~obtained spatial information from a linear combination of historical analogues for daily values and retained the daily sequencing of weather events from the coarse resolution, while BCSD only used monthly averages to reconstruct daily patterns by randomly resampling a historic month and scaling its daily values to match the monthly projected values.~~ The ensemble of the PCIC dataset has currently been used in studying the hydrological impacts of climate change on river basins mainly in British Columbia (e.g. Shrestha et al., 2011;Shrestha et al., 2012b;Schnorbus et al., 2014) and Alberta (e.g. Kienzle et al., 2012;Forbes et al., 2011) in Canada. In this study, only four GCMs with two respective statistically downscaling methods under RCP 4.5 and 8.5 were chosen for comparison (see Table 2 for details). The choice of selecting the four GCMs under RCP 4.5 and 8.5 only in the PCIC dataset was to match those GCMs available in the NA-CORDEX dataset (see next section for details).

3.2.5 GCM-driven RCM dynamically downscaled products – NA-CORDEX

Sponsored by the World Climate Research Programme (WCRP), the COordinated Regional climate Downscaling EXperiment (CORDEX) over North America domain (NA-CORDEX) was launched to provide dynamically downscaled datasets of 3-hourly or daily meteorological data over most of North America (below 80° N) at two spatial resolutions of 0.22° and 0.44° (or 25 and 50 km) under two different RCPs (RCP 4.5, and RCP 8.5) for the historical and projected period of 1950 to 2100 (Giorgi et al., 2009). ~~Within the NA-CORDEX framework, a matrix of six GCMs from the CMIP5 driving six different RCMs was selected to compare the performance of RCMs and characterize the uncertainties underlying regional climate change projections and thus provided climate scenarios for further impact and adaption studies.~~ On top of the knowledge and experience gained from the North American Regional Climate Change Assessment Program (NARCCAP) (Mearns et al., 2012), a matrix of six GCMs from the CMIP5 driving six different RCMs was selected to compare the performance of RCMs and characterize the uncertainties underlying regional climate change projections and thus provided climate scenarios for further impact and adaption studies. ~~the selection of GCM-RCM matrix of simulations, with higher spatial resolution and greater sampling of uncertainty, was based on model climate sensitivity and quality of boundary conditions. In addition, to determine the large variations in future climate due to internal variability of the GCMs on downscaled outputs, samples among multiple realizations of GCM simulations were used to drive the RCMs. The performance of participating RCMs in reproducing historical and projected climate was then assessed by comparing the ERA-Interim-driven RCM simulations.~~ Current studies using NA-CORDEX datasets were mainly focused on evaluating the model performance of different GCM-driven RCM simulations over North America (e.g. Lucas-Picher et al., 2013;Martynov et al., 2013;Separovic et al., 2013) but the NA-CORDEX dataset could also be a potential source in hydro-climatic studies in Canada. In this study, only two GCMs with three

RCMs were chosen for comparison due to the availability of the NA-CORDEX dataset (see Table 3 for details).

5. **All figures are too busy to read. Need to make them bigger.**

   We believe that Figures 1, 5 to 8 are clear enough to show the messages and therefore we will only enlarge the figures as much as possible in the revised manuscript. In response to comment 3, we decide to include the climate model products for evaluation and limit the evaluation period to 2005 instead of 2012. Accordingly, Figures 2, 3, and 4 in the original manuscript will be reproduced to reflect the change. In short, the evaluation for the climate model products from the period of 1979 to 2005 will be shown separately from that of station-based and reanalysis-based products. Thus, Figures 3 and 4 will only show the distributions of p-value of the K-S test for the station-based and reanalysis-based products and a new Figure 5 will be created to show the distributions of p-value of the K-S test for climate model products in the revised manuscript. The numbering of Figures 5 to 8 will also be changed accordingly. Note that all the figures in the supplementary materials will also be subject to the same changes as aforementioned but will not be shown here. The revised figures are shown as follows:

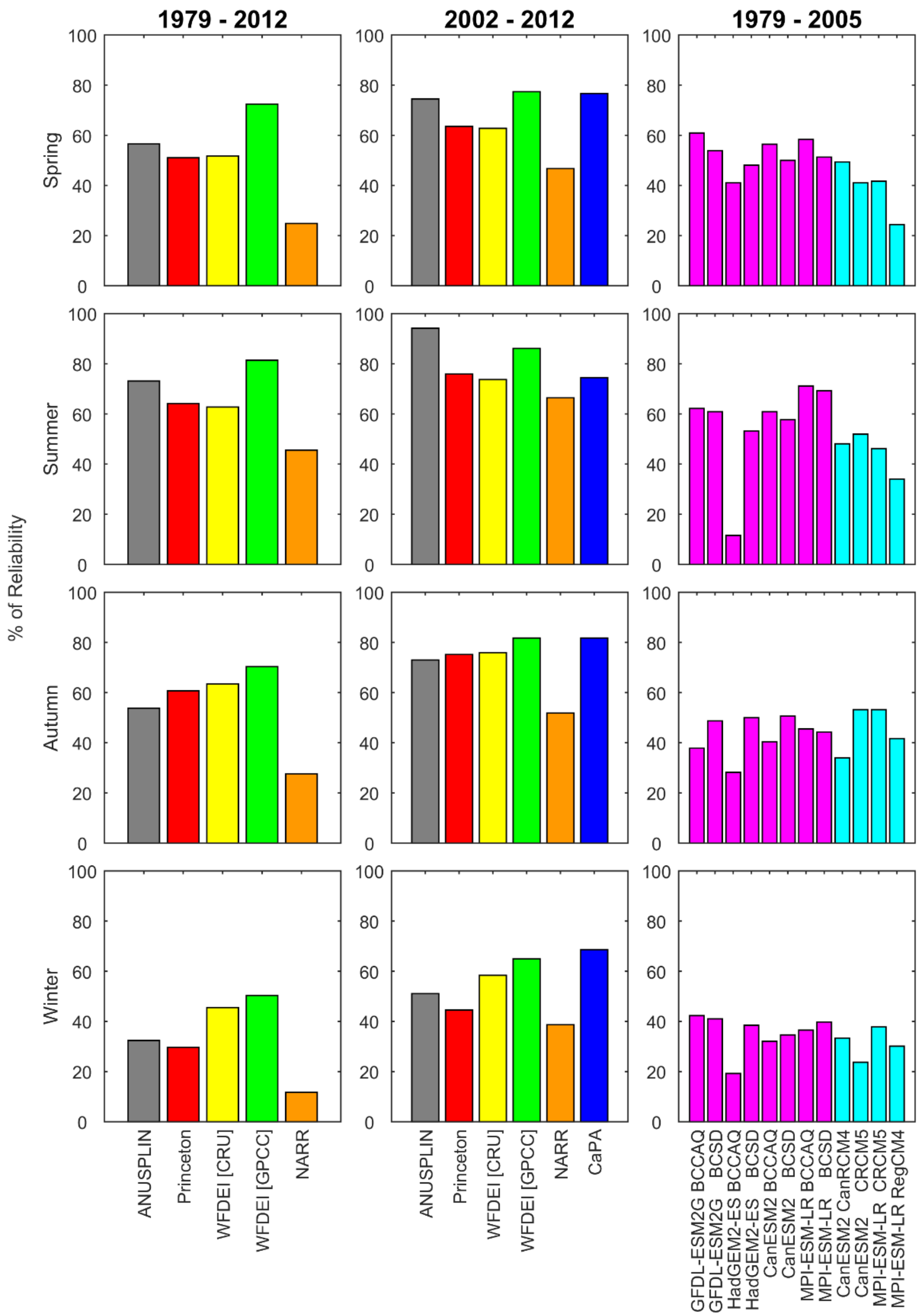

Figure 1. The percentage of reliability, calculated by the Eq. (1), of each precipitation dataset in four seasons for the period of 1979 to 2012 (left panel) and 2002 to 2012 (right panel) across Canada. The higher the percentage, the more reliable the precipitation dataset. Different colours represent different precipitation products, with magenta representing the whole PCIC datasets and cyan representing the whole NA-CORDEX datasets. The full names of the precipitation products are provided in Tables 1, 2, and 3.

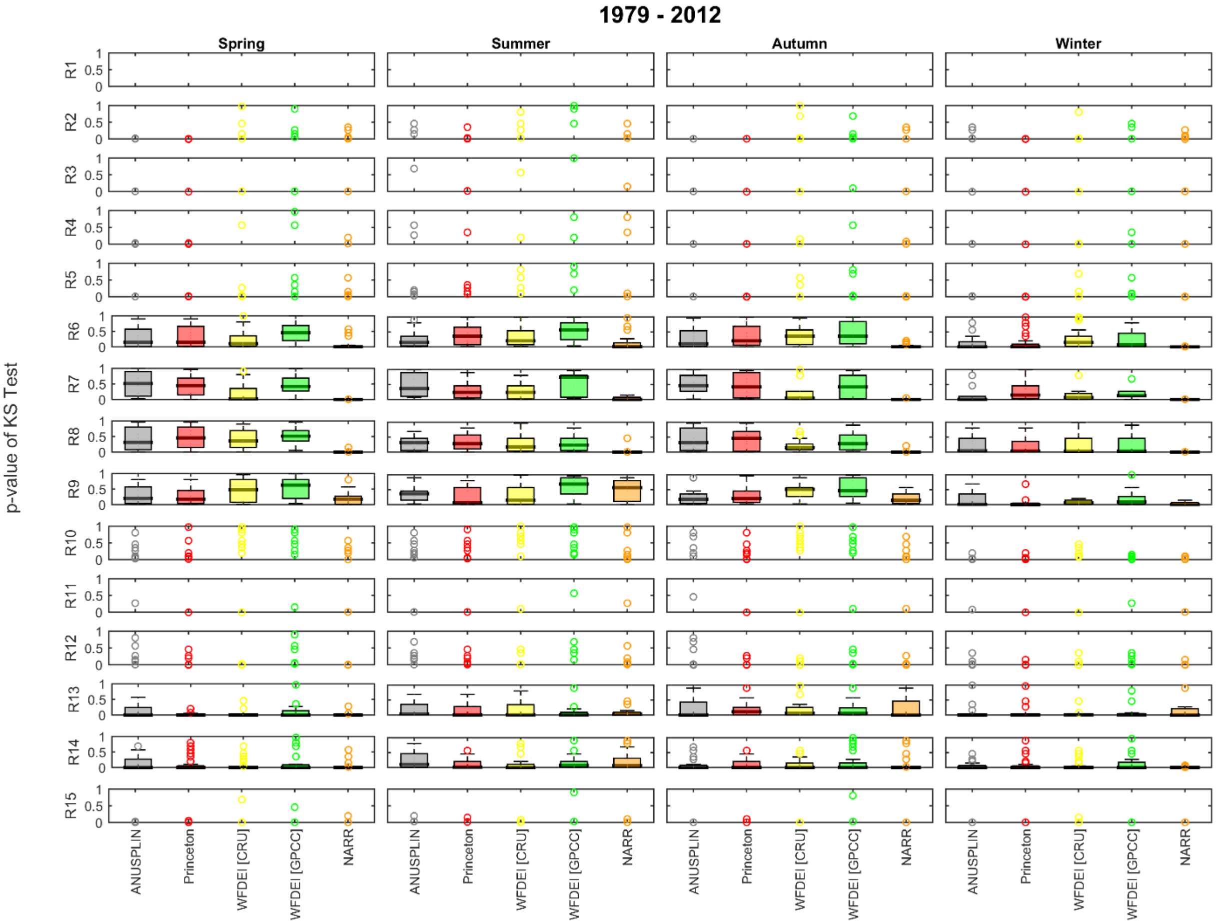

Figure 2. Distributions of p-value of the K-S test in the 15 ecozones in four seasons for the period of 1979 to 2012 (long-term comparison without CaPA). Note that the numbers of precipitation-gauge stations in each ecozone are different (see Table 4). Each hollow circle represents one p-value of the K-S test conducted at one precipitation-gauge station, with no stations in Region 1 (R1). The p-values of Regions 6 to 9, and 13 to 14 (R6-R9, and R13-R14), which have more than or equal to 10 stations, were shown in box-whisker plots with bottom, band (black thick line) and top of the box indicating the 25th, 50th (median), and 75th percentiles, respectively.

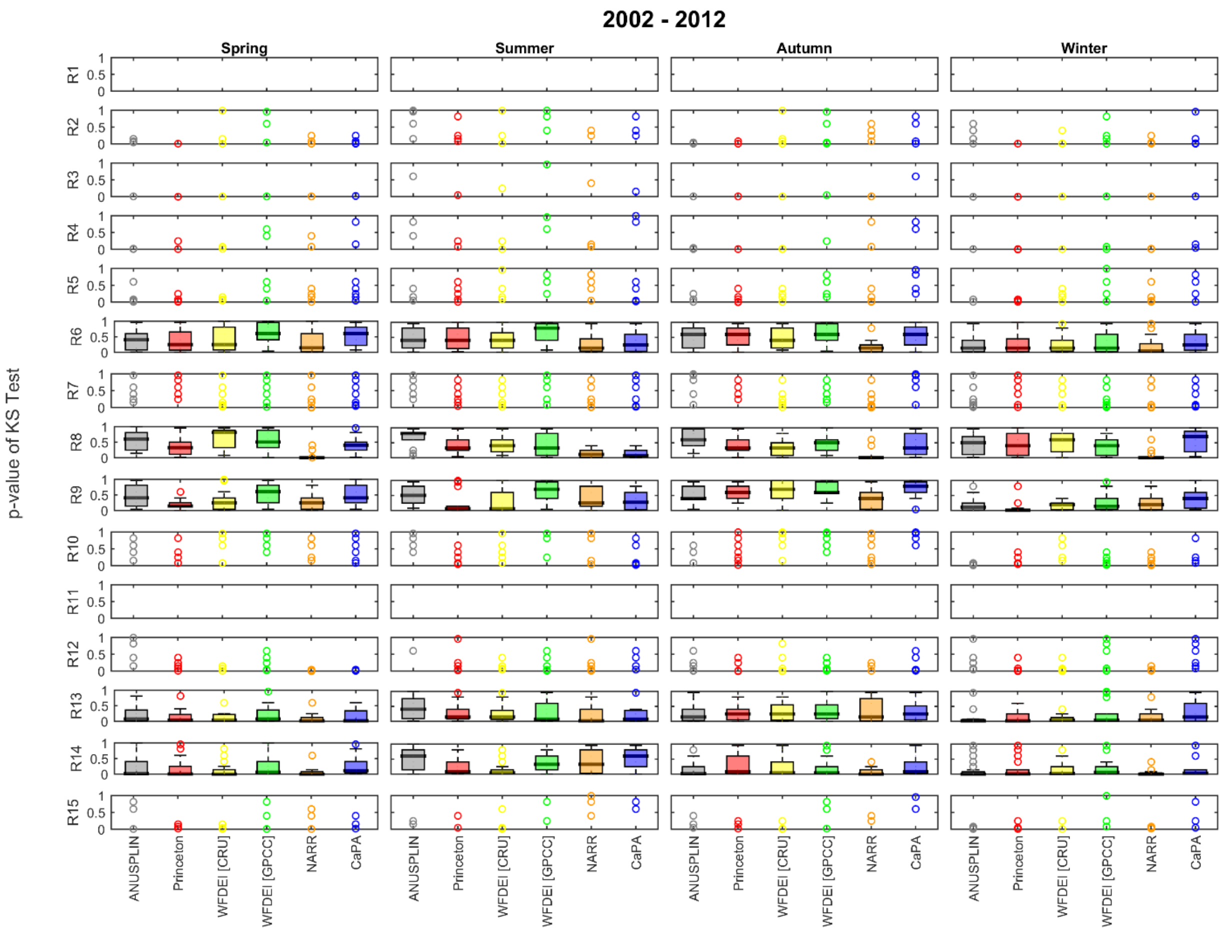

Figure 3. Distributions of p-value of the K-S test in the 15 ecozones in four seasons for the period of 2002 to 2012 (short-term comparison with the inclusion of CaPA). Note that the numbers of precipitation-gauge stations in each ecozone are different (see Table 4). Each hollow circle represents one p-value of the K-S test conducted at one precipitation-gauge station. The percentage of missing values in precipitation-gauge station in Region 11 (R11) exceeded 10% and thus no K-S test was conducted. The p-values of Regions 6, 8 to 9, and 13 to 14 (R6, R8-R9, and R13-R14), which have more than or equal to 10 stations, were shown in box-whisker plots with bottom, band (black thick line) and top of the box indicating the 25th, 50th (median), and 75th percentiles, respectively.

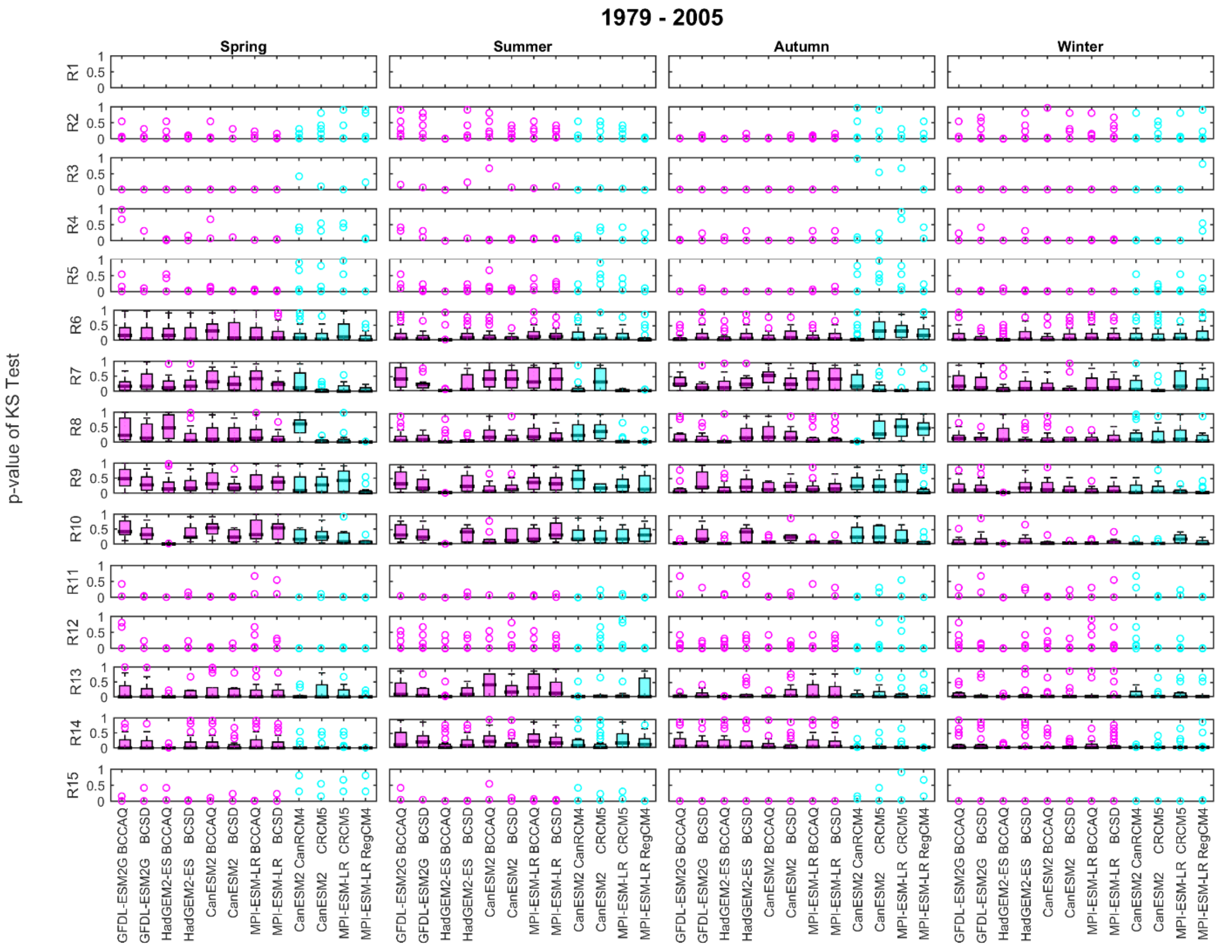

Figure 4. Distributions of p-value of the K-S test in the 15 ecozones in four seasons for the period of 1979 to 2005 (long-term comparison of PCIC and NA-CORDEX). Note that the numbers of precipitation-gauge stations in each ecozone are different (see Table 4). Each hollow circle represents one p-value of the K-S test conducted at one precipitation-gauge station, with no stations in Region 1 (R1). The p-values of Regions 6 to 9, and 13 to 14 (R6-R9, and R13-R14), which have more than or equal to 10 stations, were shown in box-whisker plots with bottom, band (black thick line) and top of the box indicating the 25th, 50th (median), and 75th percentiles, respectively.

References:

Adam, J. C., and Lettenmaier, D. P.: Adjustment of global gridded precipitation for systematic bias, J Geophys Res-Atmos, 108, Artn 4257 10.1029/2002jd002499, 2003.

Adam, J. C., Clark, E. A., Lettenmaier, D. P., and Wood, E. F.: Correction of global precipitation products for orographic effects, J Climate, 19, 15-38, Doi 10.1175/Jcli3604.1, 2006.

Araujo, A. C., Nobre, A. D., Kruijt, B., Elbers, J. A., Dallarosa, R., Stefani, P., von Randow, C., Manzi, A. O., Culf, A. D., Gash, J. H. C., Valentini, R., and Kabat, P.: Comparative measurements of carbon dioxide fluxes from two nearby towers in a central Amazonian rainforest: The Manaus LBA site, J Geophys Res-Atmos, 107, 2002.

Bhargava, M., and Danard, M.: Application of Optimum Interpolation to the Analysis of Precipitation in Complex Terrain, J Appl Meteorol, 33, 508-518, Doi 10.1175/1520-0450(1994)033<0508:Aooitt>2.0.Co;2, 1994.

Black, T. L.: The step-mountain, eta coordinate regional model: A documentation, National Meteorological Center, Development Division, 1988.

Bonsal, B. R., Aider, R., Gachon, P., and Lapp, S.: An assessment of Canadian prairie drought: past, present, and future, Clim Dynam, 41, 501-516, 10.1007/s00382-012-1422-0, 2013.

Bowman, K. P.: Comparison of TRMM precipitation retrievals with rain gauge data from ocean buoys, J Climate, 18, 178-190, Doi 10.1175/Jcli3259.1, 2005.

Carrera, M. L., Belair, S., and Bilodeau, B.: The Canadian Land Data Assimilation System (CaLDAS): Description and Synthetic Evaluation Study, J Hydrometeorol, 16, 1293-1314, 10.1175/Jhm-D-14-0089.1, 2015.

Chadburn, S. E., Burke, E. J., Essery, R. L. H., Boike, J., Langer, M., Heikenfeld, M., Cox, P. M., and Friedlingstein, P.: Impact of model developments on present and future simulations of permafrost in a global land-surface model, Cryosphere, 9, 1505-1521, 10.5194/tc-9-1505-2015, 2015.

Choi, W., Kim, S. J., Rasmussen, P. F., and Moore, A. R.: Use of the North American Regional Reanalysis for Hydrological Modelling in Manitoba, Can Water Resour J, 34, 17-36, 2009.

Cote, J., Desmarais, J. G., Gravel, S., Methot, A., Patoine, A., Roch, M., and Staniforth, A.: The operational CMC-MRB Global Environmental Multiscale (GEM) model. Part II: Results, Monthly Weather Review, 126, 1397-1418, Doi 10.1175/1520-0493(1998)126<1397:Tocmge>2.0.Co;2, 1998a.

Cote, J., Gravel, S., Methot, A., Patoine, A., Roch, M., and Staniforth, A.: The operational CMC-MRB Global Environmental Multiscale (GEM) model. Part I: Design considerations and formulation, Monthly Weather Review, 126, 1373-1395, Doi 10.1175/1520-0493(1998)126<1373:Tocmge>2.0.Co;2, 1998b.

Daley, R.: Atmospheric data analysis, 2, Cambridge university press, 1993.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Holm, E. V., Isaksen, L., Kallberg, P., Kohler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J. N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q J Roy Meteor Soc, 137, 553-597, 10.1002/qj.828, 2011.

Devine, K. A., and Mekis, E.: Field accuracy of Canadian rain measurements, Atmos Ocean, 46, 213-227, 10.3137/ao.460202, 2008.

Ebert, E. E., Janowiak, J. E., and Kidd, C.: Comparison of near-real-time precipitation estimates from satellite observations and numerical models, B Am Meteorol Soc, 88, 47-+, 10.1175/Bams-88-1-47, 2007.

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., and Tarpley, J. D.: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model, J Geophys Res-Atmos, 108, Artn 8851 10.1029/2002jd003296, 2003.

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

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

Forbes, K. A., Kienzle, S. W., Coburn, C. A., Byrne, J. M., and Rasmussen, J.: Simulating the hydrological response to predicted climate change on a watershed in southern Alberta, Canada, Climatic Change, 105, 555-576, 10.1007/s10584-010-9890-x, 2011.

Fortin, V., Jean, M., Brown, R., and Payette, S.: Predicting Snow Depth in a Forest-Tundra Landscape using a Conceptual Model Allowing for Snow Redistribution and Constrained by Observations from a Digital Camera, Atmos Ocean, 53, 200-211, 10.1080/07055900.2015.1022708, 2015a.

Fortin, V., Roy, G., Donaldson, N., and Mahidjiba, A.: Assimilation of radar quantitative precipitation estimations in the Canadian Precipitation Analysis (CaPA), J Hydrol, 531, 296-307, 2015b.
Fuchs, T.: GPCC annual report for year 2008: Development of the GPCC data base and analysis products, DWD Rep, 2009.
Garand, L., and Grassotti, C.: Toward an Objective Analysis of Rainfall Rate Combining Observations and Short-Term Forecast Model Estimates, J Appl Meteorol, 34, 1962-1977, Doi 10.1175/1520-0450(1995)034<1962:Taoaor>2.0.Co;2, 1995.
Giorgi, F., Jones, C., and Asrar, G. R.: Addressing climate information needs at the regional level: the CORDEX framework, World Meteorological Organization (WMO) Bulletin, 58, 175, 2009.
Gockede, M., Foken, T., Aubinet, M., Aurela, M., Banza, J., Bernhofer, C., Bonnefond, J. M., Brunet, Y., Carrara, A., Clement, R., Dellwik, E., Elbers, J., Eugster, W., Fuhrer, J., Granier, A., Grunwald, T., Heinesch, B., Janssens, I. A., Knohl, A., Koeble, R., Laurila, T., Longdoz, B., Manca, G., Marek, M., Markkanen, T., Mateus, J., Matteucci, G., Mauder, M., Migliavacca, M., Minerbi, S., Moncrieff, J., Montagnani, L., Moors, E., Ourcival, J. M., Papale, D., Pereira, J., Pilegaard, K., Pita, G., Rambal, S., Rebmann, C., Rodrigues, A., Rotenberg, E., Sanz, M. J., Sedlak, P., Seufert, G., Siebicke, L., Soussana, J. F., Valentini, R., Vesala, T., Verbeeck, H., and Yakir, D.: Quality control of CarboEurope flux data - Part 1: Coupling footprint analyses with flux data quality assessment to evaluate sites in forest ecosystems, Biogeosciences, 5, 433-450, 2008.
Grunwald, T., and Bernhofer, C.: A decade of carbon, water and energy flux measurements of an old spruce forest at the Anchor Station Tharandt, Tellus B, 59, 387-396, 10.1111/j.1600-0889.2007.00259.x, 2007.
Gupta, S. K., Ritchey, N. A., Wilber, A. C., Whitlock, C. H., Gibson, G. G., and Stackhouse, P. W.: A climatology of surface radiation budget derived from satellite data, J Climate, 12, 2691-2710, Doi 10.1175/1520-0442(1999)012<2691:Acosrb>2.0.Co;2, 1999.
Hong, Y., Gochis, D., Cheng, J. T., Hsu, K. L., and Sorooshian, S.: Evaluation of PERSIANN-CCS rainfall measurement using the NAME Event Rain Gauge Network, J Hydrometeorol, 8, 469-482, 10.1175/Jhm574.1, 2007.
Hopkinson, R. F., McKenney, D. W., Milewska, E. J., Hutchinson, M. F., Papadopol, P., and Vincent, L. A.: Impact of Aligning Climatological Day on Gridding Daily Maximum-Minimum Temperature and Precipitation over Canada, J Appl Meteorol Clim, 50, 1654-1665, 10.1175/2011jamc2684.1, 2011.
Huffman, G. J., Adler, R. F., Morrissey, M. M., Bolvin, D. T., Curtis, S., Joyce, R., McGavock, B., and Susskind, J.: Global precipitation at one-degree daily resolution from multisatellite observations, J Hydrometeorol, 2, 36-50, Doi 10.1175/1525-7541(2001)002<0036:Gpaodd>2.0.Co;2, 2001.
Huffman, G. J., Adler, R. F., Stocker, E., Bolvin, D. T., and Nelkin, E. J.: Analysis of TRMM 3-hourly multi-satellite precipitation estimates computed in both real and post-real time, 2002.
Hutchinson, M.: ANUSPLIN Version4. 3 User Guide. Canberra: The Australia National University, Center for Resource and Environment Studies, 2004.
Hutchinson, M. F.: Interpolating Mean Rainfall Using Thin-Plate Smoothing Splines, Int J Geogr Inf Syst, 9, 385-403, Doi 10.1080/02693799508902045, 1995.
Hutchinson, M. F., Mckenney, D. W., Lawrence, K., Pedlar, J. H., Hopkinson, R. F., Milewska, E., and Papadopol, P.: Development and Testing of Canada-Wide Interpolated Spatial Models of Daily Minimum-Maximum Temperature and Precipitation for 1961-2003, J Appl Meteorol Clim, 48, 725-741, 10.1175/2008jamc1979.1, 2009.
Janowiak, J. E., Gruber, A., Kondragunta, C. R., Livezey, R. E., and Huffman, G. J.: A comparison of the NCEP-NCAR reanalysis precipitation and the GPCP rain gauge-satellite combined dataset with observational error considerations, J Climate, 11, 2960-2979, Doi 10.1175/1520-0442(1998)011<2960:Acotnn>2.0.Co;2, 1998.
Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, B Am Meteorol Soc, 77, 437-471, Doi 10.1175/1520-0477(1996)077<0437:Tnyrp>2.0.Co;2, 1996.
Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S. K., Hnilo, J. J., Fiorino, M., and Potter, G. L.: NCEP-DOE AMIP-II reanalysis (R-2), B Am Meteorol Soc, 83, 1631-1643, 10.1175/Bams-83-11-1631, 2002.
Kang, D. H., Shi, X. G., Gao, H. L., and Dery, S. J.: On the Changing Contribution of Snow to the Hydrology of the Fraser River Basin, Canada, J Hydrometeorol, 15, 1344-1365, 10.1175/Jhm-D-13-0120.1, 2014.
Kienzle, S. W., Nemeth, M. W., Byrne, J. M., and MacDonald, R. J.: Simulating the hydrological impacts of climate change in the upper North Saskatchewan River basin, Alberta, Canada, J Hydrol, 412, 76-89, 10.1016/j.jhydrol.2011.01.058, 2012.
Kimoto, M., Yasutomi, N., Yokoyama, C., and Emori, S.: Projected Changes in Precipitation Characteristics around Japan under the Global Warming, Sola, 1, 85-88, 10.2151/sola.2005.023, 2005.

Kistler, R., Kalnay, E., Collins, W., Saha, S., White, G., Woollen, J., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., van den Dool, H., Jenne, R., and Fiorino, M.: The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation, B Am Meteorol Soc, 82, 247-267, Doi 10.1175/1520-0477(2001)082<0247:Tnnyrm>2.3.Co;2, 2001.
Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., and Takahashi, K.: The JRA-55 Reanalysis: General Specifications and Basic Characteristics, J Meteorol Soc Jpn, 93, 5-48, 10.2151/jmsj.2015-001, 2015.
Lespinas, F., Fortin, V., Roy, G., Rasmussen, P., and Stadnyk, T.: Performance Evaluation of the Canadian Precipitation Analysis (CaPA), J Hydrometeorol, 16, 2045-2064, 10.1175/Jhm-D-14-0191.1, 2015.
Lin, Y., Mitchell, K., Rogers, E., Baldwin, M., and DiMego, G.: Test assimilations of the real-time, multi-sensor hourly precipitation analysis into the NCEP Eta model, Preprints, 8th Conf. on Mesoscale Meteorology, Boulder, CO, Amer. Meteor. Soc, 1999, 341-344,
Lucas-Picher, P., Somot, S., Deque, M., Decharme, B., and Alias, A.: Evaluation of the regional climate model ALADIN to simulate the climate over North America in the CORDEX framework, Clim Dynam, 41, 1117-1137, 10.1007/s00382-012-1613-8, 2013.
Mahfouf, J. F., Brasnett, B., and Gagnon, S.: A Canadian precipitation analysis (CaPA) project: Description and preliminary results, Atmos Ocean, 45, 1-17, 2007.
Martynov, A., Laprise, R., Sushama, L., Winger, K., Separovic, L., and Dugas, B.: Reanalysis-driven climate simulation over CORDEX North America domain using the Canadian Regional Climate Model, version 5: model performance evaluation, Clim Dynam, 41, 2973-3005, 10.1007/s00382-013-1778-9, 2013.
Maurer, E. P., and Hidalgo, H. G.: Utility of daily vs. monthly large-scale climate data: an intercomparison of two statistical downscaling methods, Hydrol Earth Syst Sc, 12, 551-563, 2008.
Maurer, E. P., Hidalgo, H. G., Das, T., Dettinger, M. D., and Cayan, D. R.: The utility of daily large-scale climate data in the assessment of climate change impacts on daily streamflow in California, Hydrol Earth Syst Sc, 14, 1125-1138, 10.5194/hess-14-1125-2010, 2010.
Mearns, L. O., Arritt, R., Biner, S., Bukovsky, M. S., McGinnis, S., Sain, S., Caya, D., Correia, J., Flory, D., Gutowski, W., Takle, E. S., Jones, R., Leung, R., Moufouma-Okia, W., McDaniel, L., Nunes, A. M. B., Qian, Y., Roads, J., Sloan, L., and Snyder, M.: THE NORTH AMERICAN REGIONAL CLIMATE CHANGE ASSESSMENT PROGRAM Overview of Phase I Results, B Am Meteorol Soc, 93, 1337-1362, 2012.
Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M. L. T., Lamarque, J. F., Matsumoto, K., Montzka, S. A., Raper, S. C. B., Riahi, K., Thomson, A., Velders, G. J. M., and van Vuuren, D. P. P.: The RCP greenhouse gas concentrations and their extensions from 1765 to 2300, Climatic Change, 109, 213-241, 10.1007/s10584-011-0156-z, 2011.
Mekis, E., and Hogg, W. D.: Rehabilitation and analysis of Canadian daily precipitation time series, Atmos Ocean, 37, 53-85, 1999.
Mekis, E., and Brown, R.: Derivation of an Adjustment Factor Map for the Estimation of the Water Equivalent of Snowfall from Ruler Measurements in Canada, Atmos Ocean, 48, 284-293, 10.3137/Ao1104.2010, 2010.
Mekis, E., and Vincent, L. A.: An Overview of the Second Generation Adjusted Daily Precipitation Dataset for Trend Analysis in Canada, Atmos Ocean, 49, 163-177, Pii 938569134 10.1080/07055900.2011.583910, 2011.
Mesinger, F., Janjic, Z. I., Nickovic, S., Gavrilov, D., and Deaven, D. G.: The Step-Mountain Coordinate - Model Description and Performance for Cases of Alpine Lee Cyclogenesis and for a Case of an Appalachian Redevelopment, Monthly Weather Review, 116, 1493-1518, Doi 10.1175/1520-0493(1988)116<1493:Tsmcmd>2.0.Co;2, 1988.
Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., Jovic, D., Woollen, J., Rogers, E., Berbery, E. H., Ek, M. B., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y., Manikin, G., Parrish, D., and Shi, W.: North American regional reanalysis, B Am Meteorol Soc, 87, 343-360, 10.1175/Bams-87-3-343, 2006.
Meyers, T. P., and Hollinger, S. E.: An assessment of storage terms in the surface energy balance of maize and soybean, Agr Forest Meteorol, 125, 105-115, 10.1016/j.agrformet.2004.03.001, 2004.
Mitchell, K. E., Lohmann, D., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Cosgrove, B. A., Sheffield, J., Duan, Q. Y., Luo, L. F., Higgins, R. W., Pinker, R. T., Tarpley, J. D., Lettenmaier, D. P., Marshall, C. H., Entin, J. K., Pan, M., Shi, W., Koren, V., Meng, J., Ramsay, B. H., and Bailey, A. A.: The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system, J Geophys Res-Atmos, 109, Artn D07s90 10.1029/2003jd003823, 2004.
Nalley, D., Adamowski, J., and Khalil, B.: Using discrete wavelet transforms to analyze trends in streamflow and precipitation in Quebec and Ontario (1954-2008), J Hydrol, 475, 204-228, 10.1016/j.jhydrol.2012.09.049, 2012.

New, M., Hulme, M., and Jones, P.: Representing twentieth-century space-time climate variability. Part I: Development of a 1961-90 mean monthly terrestrial climatology, J Climate, 12, 829-856, Doi 10.1175/1520-0442(1999)012<0829:Rtcstc>2.0.Co;2, 1999.
New, M., Hulme, M., and Jones, P.: Representing twentieth-century space-time climate variability. Part II: Development of 1901-96 monthly grids of terrestrial surface climate, J Climate, 13, 2217-2238, Doi 10.1175/1520-0442(2000)013<2217:Rtcstc>2.0.Co;2, 2000.
Ngo-Duc, T., Polcher, J., and Laval, K.: A 53-year forcing data set for land surface models, J Geophys Res-Atmos, 110, Artn D06116 10.1029/2004jd005434, 2005.
Onogi, K., Tslttsui, J., Koide, H., Sakamoto, M., Kobayashi, S., Hatsushika, H., Matsumoto, T., Yamazaki, N., Kaalhori, H., Takahashi, K., Kadokura, S., Wada, K., Kato, K., Oyama, R., Ose, T., Mannoji, N., and Taira, R.: The JRA-25 reanalysis, J Meteorol Soc Jpn, 85, 369-432, DOI 10.2151/jmsj.85.369, 2007.
Pacific Climate Impacts Consortium; University of Victoria: Statistically Downscaled Climate Scenarios, in, 20th April 2016 ed., Downloaded from https://www.pacificclimate.org/data/statistically-downscaled-climate-scenarios on 20th April 2016, Jan 2014.
Park, H., Fedorov, A. N., Zheleznyak, M. N., Konstantinov, P. Y., and Walsh, J. E.: Effect of snow cover on pan-Arctic permafrost thermal regimes, Clim Dynam, 44, 2873-2895, 10.1007/s00382-014-2356-5, 2015.
Park, H., Yoshikawa, Y., Oshima, K., Kim, Y., Thanh, N. D., Kimball, J. S., and Yang, D. Q.: Quantification of Warming Climate-Induced Changes in Terrestrial Arctic River Ice Thickness and Phenology, J Climate, 29, 1733-1754, 10.1175/Jcli-D-15-0569.1, 2016.
Persson, T., Van Oene, H., Harrison, A., Karlsson, P., Bauer, G., Cerny, J., Coûteaux, M.-M., Dambrine, E., Högberg, P., and Kjøller, A.: Experimental sites in the NIPHYS/CANIF project, Springer, 2000.
Pietroniro, A., Fortin, V., Kouwen, N., Neal, C., Turcotte, R., Davison, B., Verseghy, D., Soulis, E. D., Caldwell, R., Evora, N., and Pellerin, P.: Development of the MESH modelling system for hydrological ensemble forecasting of the Laurentian Great Lakes at the regional scale, Hydrol Earth Syst Sc, 11, 1279-1294, 2007.
Rauscher, S. A., Coppola, E., Piani, C., and Giorgi, F.: Resolution effects on regional climate model simulations of seasonal precipitation over Europe, Clim Dynam, 35, 685-711, 10.1007/s00382-009-0607-7, 2010.
Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G. K., Bloom, S., Chen, J. Y., Collins, D., Conaty, A., Da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J.: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications, J Climate, 24, 3624-3648, 10.1175/Jcli-D-11-00015.1, 2011.
Rudolf, B., and Schneider, U.: Calculation of gridded precipitation data for the global land-surface using in-situ gauge observations, Proc. Second Workshop of the Int. Precipitation Working Group, 2005, 231-247,
Sapiano, M. R. P., and Arkin, P. A.: An Intercomparison and Validation of High-Resolution Satellite Precipitation Estimates with 3-Hourly Gauge Data, J Hydrometeorol, 10, 149-166, 10.1175/2008jhm1052.1, 2009.
Schneider, U., Fuchs, T., Meyer-Christoffer, A., and Rudolf, B.: Global precipitation analysis products of the GPCC, Global Precipitation Climatology Centre (GPCC), DWD, Internet Publikation, 112, 2008.
Schnorbus, M., Werner, A., and Bennett, K.: Impacts of climate change in three hydrologic regimes in British Columbia, Canada, Hydrol Process, 28, 1170-1189, 10.1002/hyp.9661, 2014.
Separovic, L., Alexandru, A., Laprise, R., Martynov, A., Sushama, L., Winger, K., Tete, K., and Valin, M.: Present climate and climate change over North America as simulated by the fifth-generation Canadian regional climate model, Clim Dynam, 41, 3167-3201, 10.1007/s00382-013-1737-5, 2013.
Sheffield, J., Ziegler, A. D., Wood, E. F., and Chen, Y. B.: Correction of the high-latitude rain day anomaly in the NCEP-NCAR reanalysis for land surface hydrological modeling, J Climate, 17, 3814-3828, Doi 10.1175/1520-0442(2004)017<3814:Cothrd>2.0.Co;2, 2004.
Sheffield, J., Goteti, G., and Wood, E. F.: Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling, J Climate, 19, 3088-3111, Doi 10.1175/Jcli3790.1, 2006.
Shook, K., and Pomeroy, J.: Changes in the hydrological character of rainfall on the Canadian prairies, Hydrol Process, 26, 1752-1766, 10.1002/hyp.9383, 2012.
Shrestha, R., Berland, A., Schnorbus, M., and Werner, A.: Climate change impacts on hydro-climatic regimes in the Peace and Columbia watersheds, British Columbia, Canada, Pacific climate impacts consortium, University of Victoria, 37, 2011.
Shrestha, R. R., Dibike, Y. B., and Prowse, T. D.: Modelling of climate-induced hydrologic changes in the Lake Winnipeg watershed, J Great Lakes Res, 38, 83-94, 10.1016/j.jglr.2011.02.004, 2012a.

Shrestha, R. R., Schnorbus, M. A., Werner, A. T., and Berland, A. J.: Modelling spatial and temporal variability of hydrologic impacts of climate change in the Fraser River basin, British Columbia, Canada, Hydrol Process, 26, 1841-1861, 10.1002/hyp.9283, 2012b.
Su, H., Dickinson, R. E., Findell, K. L., and Lintner, B. R.: How Are Spring Snow Conditions in Central Canada Related to Early Warm-Season Precipitation?, J Hydrometeorol, 14, 787-807, 10.1175/Jhm-D-12-029.1, 2013.
Suni, T., Rinne, J., Reissell, A., Altimir, N., Keronen, P., Rannik, U., Dal Maso, M., Kulmala, M., and Vesala, T.: Long-term measurements of surface fluxes above a Scots pine forest in Hyytiala, southern Finland, 1996-2001, Boreal Environ Res, 8, 287-301, 2003.
Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An Overview of Cmip5 and the Experiment Design, B Am Meteorol Soc, 93, 485-498, 10.1175/Bams-D-11-00094.1, 2012.
Uppala, S. M., Kållberg, P., Simmons, A., Andrae, U., Bechtold, V. d., Fiorino, M., Gibson, J., Haseler, J., Hernandez, A., and Kelly, G.: The ERA-40 re-analysis, Q J Roy Meteor Soc, 131, 2961-3012, 2005.
Urbanski, S., Barford, C., Wofsy, S., Kucharik, C., Pyle, E., Budney, J., McKain, K., Fitzjarrald, D., Czikowsky, M., and Munger, J. W.: Factors controlling CO2 exchange on timescales from hourly to decadal at Harvard Forest, J Geophys Res-Biogeo, 112, Artn G02020 10.1029/2006jg000293, 2007.
Vincent, L. A., and Mekis, E.: Discontinuities due to Joining Precipitation Station Observations in Canada, J Appl Meteorol Clim, 48, 156-166, 10.1175/2008jamc2031.1, 2009.
Wan, H., Zhang, X. B., Zwiers, F. W., and Shiogama, H.: Effect of data coverage on the estimation of mean and variability of precipitation at global and regional scales, J Geophys Res-Atmos, 118, 534-546, 10.1002/jgrd.50118, 2013.
Wang, S., Yang, Y., Luo, Y., and Rivera, A.: Spatial and seasonal variations in evapotranspiration over Canada's landmass, Hydrol Earth Syst Sc, 17, 3561-3575, 10.5194/hess-17-3561-2013, 2013.
Wang, S. S., Huang, J. L., Li, J. H., Rivera, A., McKenney, D. W., and Sheffield, J.: Assessment of water budget for sixteen large drainage basins in Canada, J Hydrol, 512, 1-15, 10.1016/j.jhydrol.2014.02.058, 2014.
Weedon, G., Gomes, S., Viterbo, P., Österle, H., Adam, J., Bellouin, N., Boucher, O., and Best, M.: The WATCH forcing data 1958–2001: A meteorological forcing dataset for land surface and hydrological models, Watch Ed Watch Tech Rep, 22, 41, 2010.
Weedon, G. P., Gomes, S., Viterbo, P., Shuttleworth, W. J., Blyth, E., Osterle, H., Adam, J. C., Bellouin, N., Boucher, O., and Best, M.: Creation of the WATCH Forcing Data and Its Use to Assess Global and Regional Reference Crop Evaporation over Land during the Twentieth Century, J Hydrometeorol, 12, 823-848, 10.1175/2011jhm1369.1, 2011.
Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J., and Viterbo, P.: The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, Water Resour Res, 50, 7505-7514, 10.1002/2014wr015638, 2014.
Willmott, C. J., Matsuura, K., and Legates, D.: Terrestrial air temperature and precipitation: monthly and annual time series (1950–1999), Center for climate research version, 1, 2001.
Woo, M. K., and Thorne, R.: Snowmelt contribution to discharge from a large mountainous catchment in subarctic Canada, Hydrol Process, 20, 2129-2139, 10.1002/hyp.6205, 2006.
Wood, A. W., Leung, L. R., Sridhar, V., and Lettenmaier, D. P.: Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs, Climatic Change, 62, 189-216, DOI 10.1023/B:CLIM.0000013685.99609.9e, 2004.
Zhang, X. B., Vincent, L. A., Hogg, W. D., and Niitsoo, A.: Temperature and precipitation trends in Canada during the 20th century, Atmos Ocean, 38, 395-429, 2000.
Zhao, M., and Dirmeyer, P. A.: Production and analysis of GSWP-2 near-surface meteorology data sets, Center for Ocean-Land-Atmosphere Studies Calverton, 2003.