Responses to Editor final 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

Dear Prof. Jan Seibert, thank you very much again for your comments and recommendations. We have addressed all of the comments and presented our responses below.

The review comments are in regular bold typeface, while all responses are in italics and indented paragraphs, with deleted materials being crossed out by drawing a line through them and revised sentences being coloured in red.

Response to Editor

Editor Decision: Publish subject to revisions (further review by Editor and Referees) (31 Dec 2016) by Prof. Jan Seibert

Comments to the Author:

The reviews list a number of important points and based on the responses in the discussion phase, I am confident that the authors will be able to address these. As also indicated by the reviews, the original submission suffered from language issues and from poor figures. The authors need to address these two issues also beyond the concrete suggestions throughout the manuscript. Here it is important that the senior authors (native speakers!) look carefully at the manuscript before resubmission!

> *In response to the Editor's comments, we have focused on proofreading the manuscript and made some additional modifications. We have gone through the entire manuscript and where applicable, have improved the language and flow (e.g., removed repetitive phrases, re-organized sub-sections). Specific areas where text has been modified include:*
>
> 1) *Sect. 3.1 [L227-270] now provides a better description of the precipitation-gauge station observations.*
> 2) *A new sub-section heading [Sect. 4.1; L483] has been added to include the description of the pre-processing of the gridded products and the precipitation-gauge station observations.*
> 3) *Two paragraphs have been re-arranged for better logical flow of the manuscript.*
>   a. *Paragraph describing the precipitation measurements and their limitations in the Canadian context has been moved from Study Area Section [L203-224] to Introduction Section [L152-173].*
>   b. *Paragraph describing the selection of the study period has been moved from [L519-526] to [L484-499].*

__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 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 has been adjusted (e.g. for undercatch) and are the most accurate source of information on precipitation in Canada (although 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 biases in differences between point and areal data 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, *L827-830). We have revised the manuscript to better reflect on this issue, 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 essentially represents 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 as the "ground truth". Therefore, ANUSPLIN is also included in the pool of gridded products to be evaluated. Notwithstanding the issues, the authors feel that using the selected gauge measurements is best for the evaluation of the multiple gridded products because the set of gauges used has been adjusted (e.g. for undercatch) and are the most accurate source of information on precipitation in Canada (although 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 biases in differences between point and areal data is consistent for all the products. ~~However, the authors believe that given the current data situation, the preceding was the best methodology for evaluating 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, this study does not intend to assess different products for reproducing observed individual precipitation events generated by a given weather system but to examine the combined precipitation distribution over a period of time. 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 have changed the title from "Evaluation" to "Inter-comparison" to better reflect our aim of the study. Also, we have summarized and inserted the above justification in the revised manuscript [L484-491], which is shown as follows:

> *Given that the main focus of this study was to inter-compare the various gridded precipitation products using precipitation-gauge station data as a reference/benchmark (and not to assess the individual accuracy of each product against this reference), it was decided to re-grid each product onto a common 0.5° x 0.5° resolution to match the lowest-resolution dataset. It was acknowledged that re-gridding can introduce uncertainties due to the extra interpolations, however, the authors believe that upscaling to a common resolution provided a direct and more consistent inter-comparison. Furthermore, this methodology was consistent with similar studies in the literature (e.g. Janowiak et al., 1998;Rauscher et al., 2010;Kimoto et al., 2005).*

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 has been changed in the revised manuscript [L492-496], which is shown as follows:

> *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 as the historical period of the datasets ends in 2005); and (2) short-term comparison from January 2002 to December 2012 when CaPA are available.*

Also, the percentage of reliability has been re-calculated for the climate model products in the revised manuscript and the results have been 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 gridded station-based and reanalysis-based products outperformed the climate model-simulated products.

With regard to the short-term comparison (Fig. 2 ~~right~~ middle panel), ANUSPLIN ~~had the best~~ showed better performance in summer with 94.1 % of reliability among the 137 precipitation-gauge stations while CaPA ~~was the best~~ indicated better skill 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 was excluded in the ~~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 corresponded to the various precipitation products. The ~~more~~ higher the numbers of high p-values (> 0.05) ~~are~~ in ~~one~~ each 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~~ of each 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~~ those 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 has also been revised to reflect the change [L797-812], which is shown as follows:

By matching the statistical ~~property~~ properties 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 have been greatly reduced in the revised manuscript. In short, the spatial and temporal resolutions of each product, their compositions, and examples of their applications are remained and other details have been 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~~ observations

In Canada, ~~Climate~~ climate data collection is coordinated by the Federal government, which is ~~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 by ~~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 and Climate Change Canada (NCDA). These data provide the basis for all the available quality controlled climate ~~data~~ observations. ~~Based on the National Climate Data Archive of Environment Canada, there~~ There are a total of 1499 precipitation-gauge stations (as ~~in~~ of 2012) across Canada. However, ~~due to the~~ given the frequent addition and subtraction of climate stations, these numbers have greatly varied through time with peak reporting in the 1970s followed by a general decline to the present ~~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~~ Furthermore, the existing precipitation observations are often subject to various errors, with gauge undercatch being of significant concern ~~among which the errors due undercatch are quite significant in Canada~~ (Mekis and Hogg, 1999). ~~In order to~~ To account for various measurement issues, Mekis and Hogg (1999) first produced the Adjusted and Homogenized Canadian Climate Data (AHCCD) including adjusted daily rainfall and snowfall values and Mekis and Vincent (2011) then updated the data for a subset of 464 stations over Canada. ~~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~~ on the order of 5 to 10 % higher in southern Canada and more than 20 % in the Canadian Arctic when compared to ~~than~~ the original observations. ~~The effect of the adjustments on~~ Adjustments to snowfall were even larger and ~~more variable~~ varied 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~~ These adjusted values are considered as better estimates of actual precipitation and therefore have been used in numerous 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~~ Given the lack of an adjusted daily gridded precipitation data for Canada, the AHCCD station precipitation is considered to be the best available data for Canada and thus is used 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 for assessment 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~~ including recent years up to 2012; and (4) ~~representation of~~ representing a range of sources/methodologies (e.g. station based, remote sensing, model, blended products). Table 1 summarizes these datasets, including their full names and original spatial and temporal resolutions for the versions used. 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~~ (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) used the Australian National University Spline (ANUSPLIN) model to ~~developed~~ develop a ~~climate~~ dataset of daily precipitation and daily minimum and maximum air temperature over Canada at a spatial resolution of 300 arc-seconds ~~of latitude and longitude~~ (0.0833° or ~10 km) for

the period of 1961 to 2003~~, using observed stations~~. All available NCDA stations (that ranged from 2000 to 3000 ~~in~~ for any given year~~s over the~~ during this period) were used as an input to the gridding procedure. ~~recorded in the National Canadian Climate Data Archives of Environment Canada. However, to~~ To retain ~~a better~~ maximum spatial coverage, the smaller number of stations in AHCCD were not incorporated (i.e. only unadjusted archive values were used). ~~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~~ Interpolation procedures included incorporation of tri-variate thin-plate smoothing splines method ~~that incorporated~~ using 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~~ The Canadian ANUSPLIN has now 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~~ multiple-source product – CaPA

~~Initiated in~~ 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~~ Data were generated ~~based on~~ using an optimum interpolation technique (Daley, 1993), which required ~~a background field and~~ a specification of error statistics between the observations and ~~the~~ a 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) model (Cote et al., 1998a;1998b)~~,~~ were used as the background field with the rain-gauge measurements from ~~the observational network~~ NCDA as the observations to generate an analysis error at every grid point. ~~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~~ in the statistical interpolation method (Lespinas et al., 2015)~~,~~ and increase of spatial resolution to 10 km. ~~and the~~ The assimilation of Quantitative Precipitation Estimates from the Canadian Weather Radar Network is also used 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~~ started in 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~~ This dataset (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~~ additional temperature and precipitation data (e.g. Willmott et al., 2001), Princeton has been updated and is currently available with two versions: 1) 1948 to 2008 at 1.0°, ~~(plus~~ 0.5°, and 0.25°~~),~~ at 3-hourly, ~~(plus~~ daily, and monthly~~) resolution globally for 1948 to 2008~~ time steps and 2)~~. Experimental updates including a~~ 1901-2012 experimental version at 1.0°, ~~(plus~~ 0.5°~~),~~ at 3-hourly, ~~(plus~~ daily, and monthly~~) resolution are also available~~ time steps (used in this study). Studies employing Princeton to ~~study~~ examine 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)*~~*. For instance, Kang et al. (2014) examined the changing contribution of snow to runoff generation in the Fraser River Basin while Su et al. (2013) investigated the relationships between spring snow and warm-season precipitation in central Canada. In addition, Wang et al. (2013) and Wang et al. (2014) used this dataset to characterize the spatial and seasonal variations of the surface water budget at Canada national scale.*

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~~ and 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~~ developed covering the period of 1979 to 2012 (Weedon et al., 2014). The WFDEI used the same methodology as the WFD, but was 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. Both sets of data were used in this study (hereafter ~~the~~ known as 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~~ With the aim of evaluating 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~~ For 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 beginning in January 2004 over the Athabasca River basin due to the assimilation of station observations over Canada being discontinued in 2003.

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, Canada, 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-seconds (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 for bias-correction and downscaling of the GCMs. ~~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 (Werner and Cannon, 2016). ~~The first one was~~ These included 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 (1950 – 2005) and ~~projected~~ future (2006 – 2100) 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~~ Drawing from the strengths of 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 and characterize the uncertainties of RCMs 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~~ and 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 have only enlarged 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 have been 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 have also been subject to the same changes as aforementioned but will not be shown here. The revised figures are shown as follows:

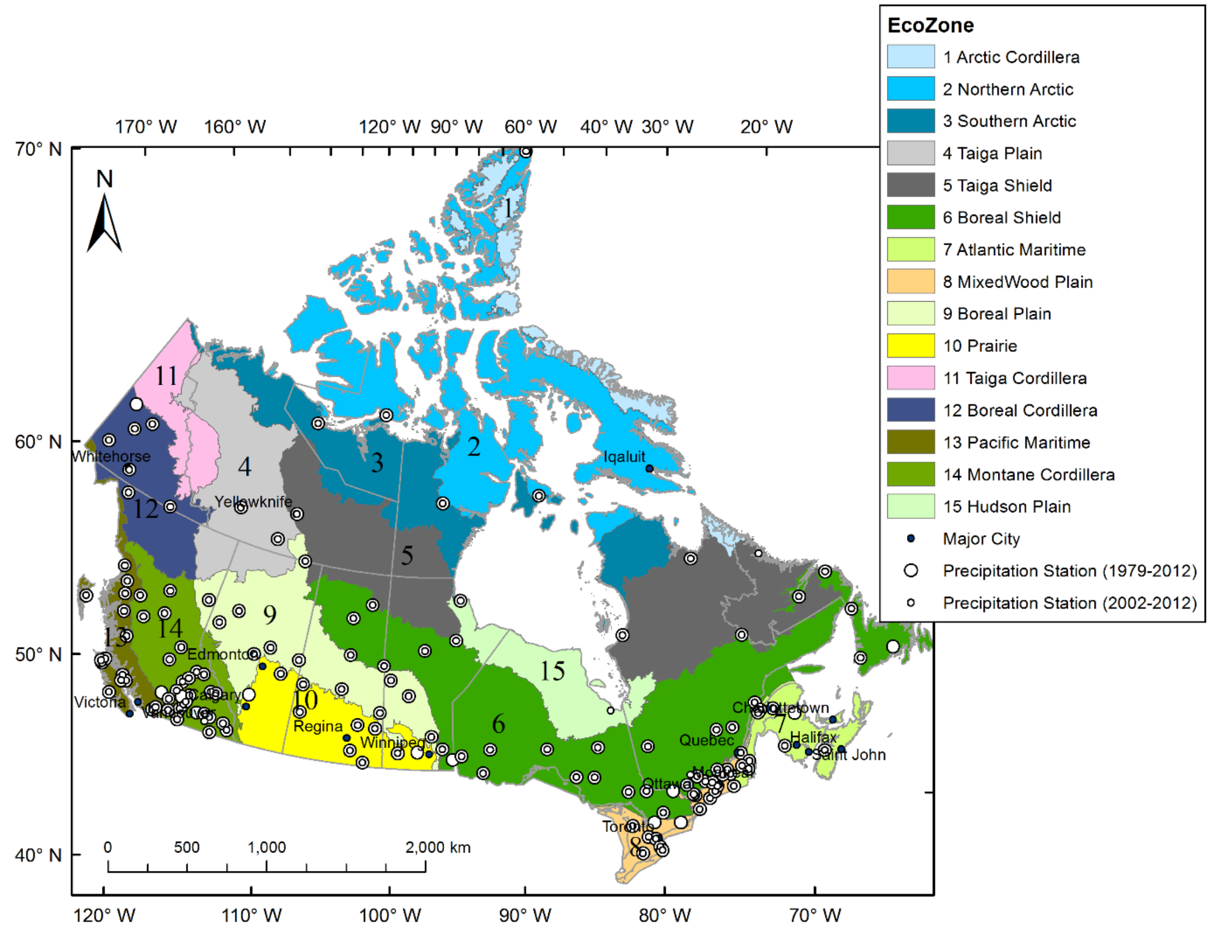


Figure 1. 15 terrestrial ecozones of Canada with numerical codes indicating Region from 1 Arctic Cordillera to 15 Hudson Plain. Big (a total of 145) and small (a total of 137) white dots are the extracted precipitation-gauge stations from the Canadian adjusted and homogenized precipitation datasets of Mekis and Vincent (2011) for the period of 1979 to 2012 and 2002 to 2012 respectively. Black dots are major cities in Canada.

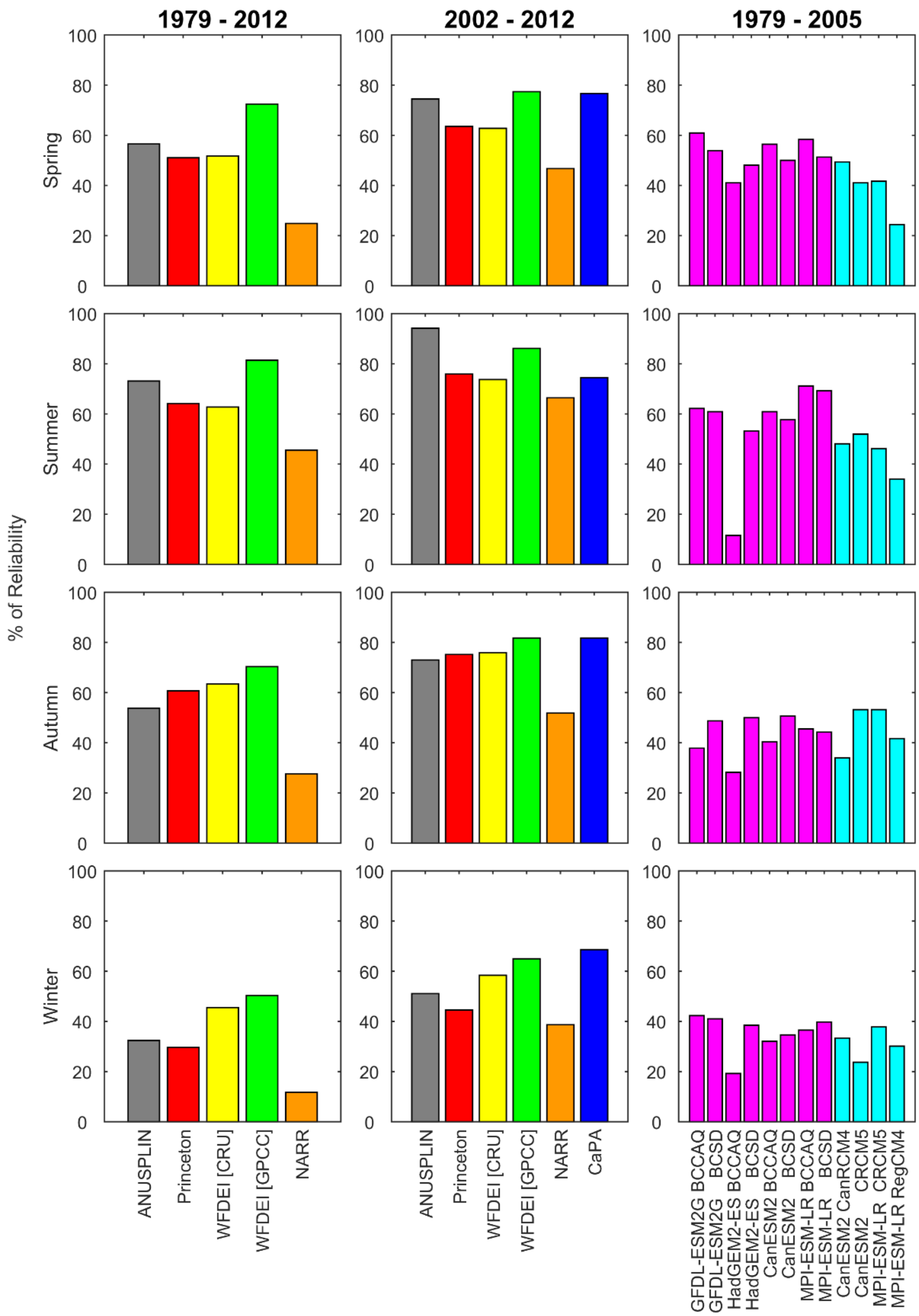


Figure 2. The percentage of reliability, calculated by the Eq. (1), of each precipitation dataset in four seasons for the period of 1979 to 2012 (left panel) and 2002 to 2012 (right panel) across Canada. The higher the percentage, the more reliable the precipitation dataset. Different colours represent 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.

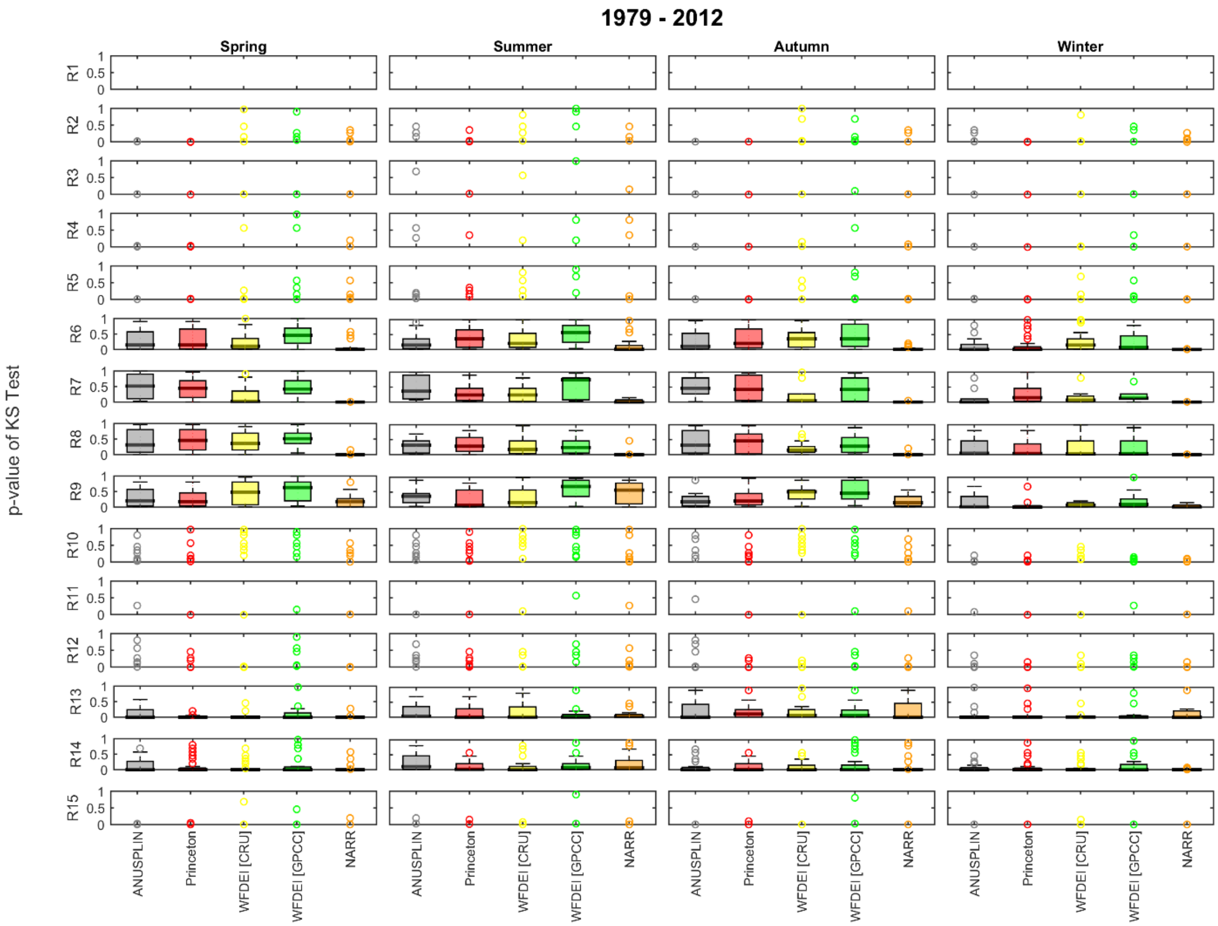


Figure 3. Distributions of p-value of the K-S test in the 15 ecozones in four seasons for the period of 1979 to 2012 (long-term comparison without CaPA). Note that the numbers of precipitation-gauge stations in each ecozone are different (see Table 4). Each hollow circle represents one p-value of the 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.

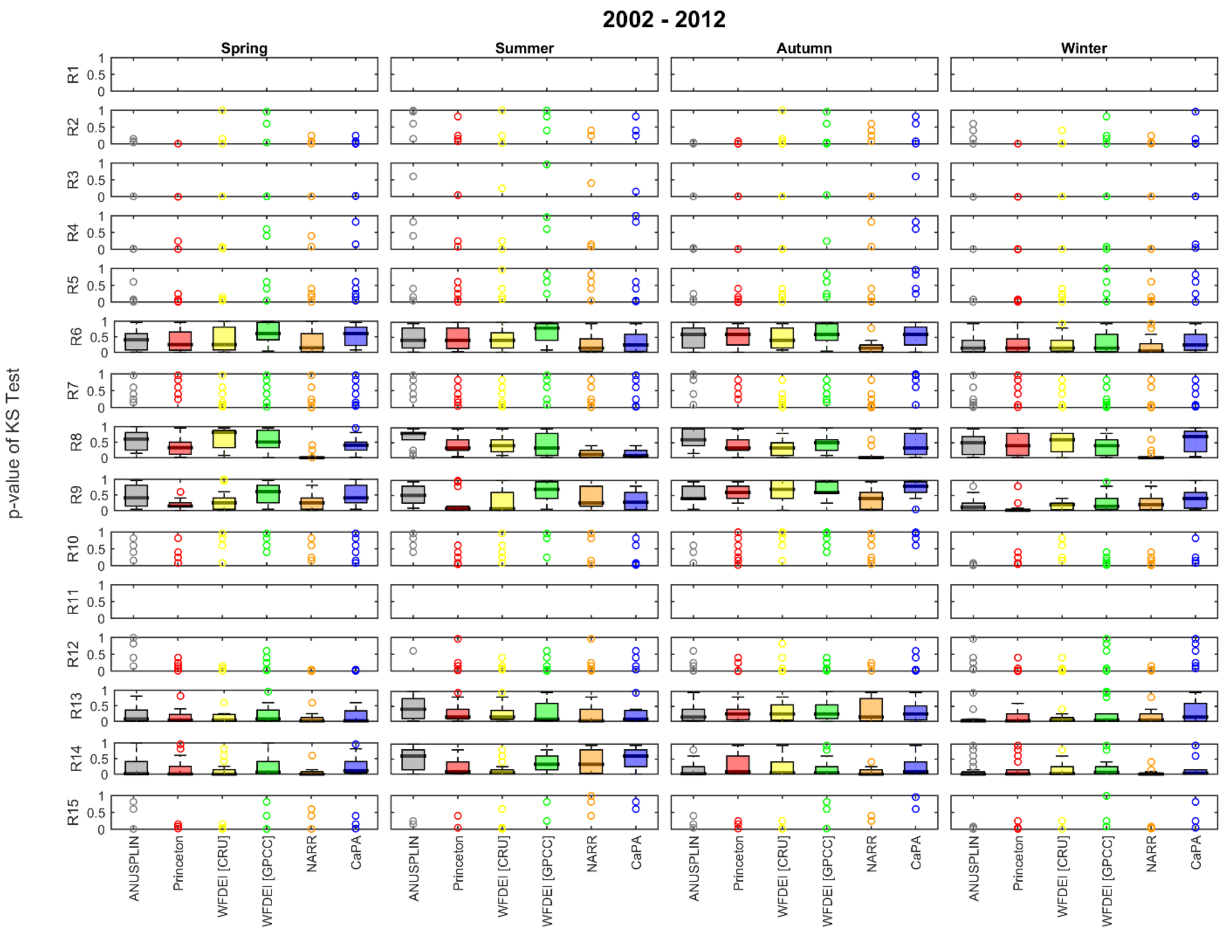


Figure 4. Distributions of p-value of the K-S test in the 15 ecozones in four seasons for the period of 2002 to 2012 (short-term comparison with the inclusion of CaPA). Note that the numbers of precipitation-gauge stations in each ecozone are different (see Table 4). Each hollow circle represents one 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.

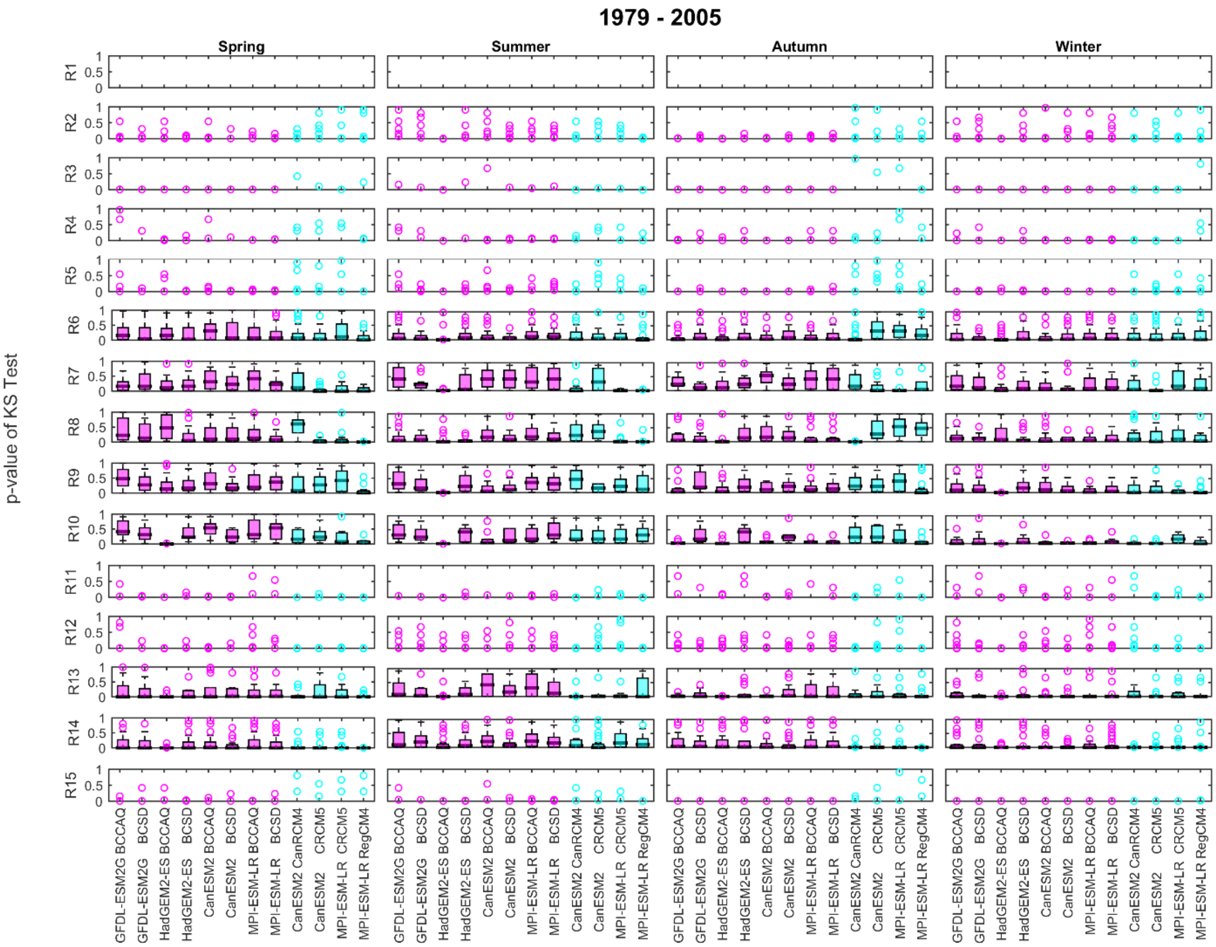


Figure 5. Distributions of p-value of the K-S test in the 15 ecozones in four seasons for the period of 1979 to 2005 (long-term comparison of PCIC and NA-CORDEX). Note that the numbers of precipitation-gauge stations in each ecozone are different (see Table 4). Each hollow circle represents one p-value 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.

<u>Response to Reviewer 2</u>

The study examines and compares 8 types of gridded precipitation sources (i.e. 22 precipitation products based on station, reanalysis, and GCM models) over 15 terrestrial ecozones in Canada. I think the results reported by this manuscript can be useful for hydrologists, meteorologists, and potential data users over Canada. In general, the paper is concise and well organized. The results are original and useful for both data developers and end-users, especially for large-scale hydrometeorological applications in Canada. The paper is thus worth to be published after the minor suggestions listed below.

We thank the reviewer for reviewing our manuscript and providing his/her valuable comments. We have now addressed all of the comments and presented our responses below.

Specific comments:

1. **The abstract seems too long and needs to be further condensed in the revision. Moreover, the spatiotemporal scales of evaluation (daily and 0.5 deg.) should be denoted in the abstract.**

The length of the abstract has been reduced and the spatiotemporal scales of evaluation have been included in the revised abstract, as shown in the following:

A number of global and regional gridded climate products based on multiple data sources ~~and models~~ are available that can potentially provide ~~better and~~ more reliable estimates of precipitation for climate and hydrological studies. However, research into the reliability of these products for various regions has been limited and in many cases non-existent. This study ~~identifies~~ inter-compares several gridded precipitation products and ~~over Canada and develops a systematic analysis framework to assess the characteristics of errors associated with the different datasets, using the best available adjusted precipitation-gauge data as a benchmark over the period 1979 to 2012. The framework~~ quantifies the spatial and temporal variability of the errors (relative to station observations) over 15 terrestrial ecozones in Canada for different seasons over the period 1979 to 2012 at 0.5° and daily spatiotemporal resolution ~~at the daily time scale~~. These datasets were assessed in their ability to represent the daily variability of precipitation amounts by four performance measures: percentage of bias, root-mean-square-error, correlation coefficient, and standard deviation ratio. Results showed that most of the ~~products~~ datasets were relatively skillful in central Canada. However, they tended to overestimate precipitation amounts in the west and underestimate in the north and east, with the underestimation being particularly dominant in northern Canada (above 60° N). ~~but tended to underestimate precipitation~~

~~amounts on the east coast and overestimate on the west~~. The global product by WATCH Forcing Data ERA-Interim (WFDEI) augmented by Global Precipitation Climatology Centre (GPCC) data (WFDEI [GPCC]) performed best with respect to different metrics. The Canadian Precipitation Analysis (CaPA) product ~~of Meteorological Service of Canada~~, performed comparably with WFDEI [GPCC], however it only provides data starting from 2002. All the ~~products~~ datasets performed best in summer, followed by autumn, spring, and winter in order of decreasing quality. ~~Due to the sparse observational network, northern Canada (above 60° N) was most difficult to assess with the majority of products tending to significantly underestimate total precipitation. Results~~ Findings from this study ~~can be used as a~~ provide guidance ~~for~~ to potential users regarding the performance of different precipitation products for a range of geographical regions and time periods.

2. **P4 Line 10-14: In terms of retrieval errors in satellite precipitation, the impact of the snow cover on passive microwave sensors is rather serious over high mountainous regions or high latitude areas, e.g. the Tibetan Plateau (Yong et al., 2015). The authors should address this issue here. Additionally, the Global Precipitation Measurement (GPM; Hou et al. 2014) has been coming and the authors should mention the GPM mission in describing the satellite precipitation estimates. Hou, A. Y., and Coauthors, 2014: The global precipitation measurement mission. Bull. Amer. Meteor. Soc., 95, 701-722. Yong, B., and Coauthors, 2015: Global view of real-time TRMM multisatellite precipitation analysis: Implications for its successor global precipitation measurement mission. Bull. Amer. Meteor. Soc., 96, 283-296.**

The impact of snow cover on passive microwave sensors has been addressed [L92-94] and the GPM mission has been mentioned [L86-88] in the revised manuscript. Accordingly, the corresponding references have also been added. The following shows the revised discussion of satellite-based estimates, with additional sentences being coloured in red:

Development of satellite-based precipitation estimates has provided coverage over vast gauged/ungauged regions with continuous observations regardless of time of day, terrain, and weather condition of the ground (Gebregiorgis and Hossain, 2015). The recently launched Global Precipitation Measurement (GPM) Core Observatory has further opened up new opportunities for observing worldwide precipitation from space (Hou et al., 2014). However, satellite-based estimates also contain inaccuracies resulting primarily from temporal sampling errors due to infrequent satellite visits to a particular location, instrumental errors due to calibration and measurement noise, and algorithm errors related to approximations to the cloud physics used (Nijssen and Lettenmaier, 2004;Gebremichael et al., 2005). In particular, the passive microwave overpasses *were shown to be unreliable over regions with snow cover and complex terrain such as the Tibetan Plateau (Yong et al., 2015).*

3. **P17 Line 10-14: Using the approach of Kolmogorov-Smirnov test to evaluate different precipitation products is an interesting way for readers. But here the equation (1) is not clear. I suggest that the authors may carefully re-modified the calculating equation and illustrate the meanings of parameters. If possible, an appendix that introduces the Kolmogorov-Smirnov test might be added at the end of the text. At least, the Eq. (1) should be revised again.**

   We have addressed this comment by providing better explanation of the calculation and revising the wordings in the equation for better clarity in the revised manuscript [L542-557]:

   > *A two-sample, non-parametric Kolmogorov-Smirnov (K-S) test ~~compared~~ was used to compare the cumulative distribution functions (CDFs) ~~for each type~~ of gridded precipitation product with the AHCCD.~~at 5 % significance level (α = 0.05) to support the~~ The null hypothesis (H_0) was that the two datasets came from same population. ~~Monthly total precipitation data were used and aggregated for each season because the existence~~ For each season, monthly total precipitation data were used to avoid commonly known issues of numerous zero values in the daily precipitation data that might ~~reduce the statistical identification of significant differences to support the null hypothesis~~ affect significance. The K-S test was repeated independently for all precipitation-gauge stations at 5 % significance level (α = 0.05)~~ and a measure of reliability (in percent) was calculated to show how reliable each type of precipitation products was among all the precipitation-gauge stations, as shown by Eq. (1)~~. A measure of reliability (in percent) was calculated based on counting the numbers of stations that do not reject the null hypothesis (any p-values greater than 0.05) over the total numbers of stations (145 and 137 stations in long-term and short-term comparison respectively), as shown in Eq. (1).*

$$\cancel{\%\ of\ reliability = \frac{no\ of\ station\ that\ support\ H_0}{total\ no\ of\ precipitation\ gauge\ station} \cdot 100 \qquad (1)}$$

$$\%\ of\ reliability = \frac{number\ of\ stations\ that\ support\ H_0}{total\ number\ of\ precipitation\ gauge\ stations} \cdot 100 \qquad (1)$$

   We appreciate the suggestion on having an appendix to introduce the Kolmogorov-Smirnov (K-S) test but we decide not to do so due to the following reasons: (1) K-S test is one of the commonly-used statistical tests and its basic theory, assumptions, and calculation is easily found in any statistical handbooks; (2) we only applied the standard two-sample non-parametric K-S test in our study without any modifications in *its assumptions or calculation; and (3) given the length of our manuscript, we prefer saving the space for better explanation or clarification in other parts of the manuscript (if necessary).*

4. **P27 Line 12-14: In the conclusion, please clarify and explain the reasons of the poorest performance of station-based and reanalysis-based products in Atlantic and Pacific regions.**

   We think that this statement in the Conclusion [L878-880] will cause some confusions and we decide to drop it from the conclusion and address the reasons of the poor performance in the Results Section (Section 5.2) [L714-717] in the revised manuscript, which is shown as follows:

   > *The resulting values of the RMSE metric in Regions 7 (Atlantic Maritime) and 13 (Pacific Maritime) tended to be larger than that of other areas. However, the other metrics such as correlation coefficient and PBias showed better performance in these regions. This suggests that higher RMSE values can be mainly attributed to the fact that precipitation amounts are higher in the maritime regions.*

5. **Some figures are not very clear and they should be modified or redrawn. For example, there is no whole Canada map (or North American map), no north arrow, no measuring scale in Fig. 1. Figure 2 is OK, but the plots in Fig. 3 and Fig. 4 are too small and not clear for reading. I really hope that these plots could be better displayed in the revised manuscript.**

   We agree that some of the figures are not very clear as it is also commented by Reviewer 1. Please refer to Reviewer 1's comment 5 for the revised figures. In short, we have enlarged the figures as much as possible and provided the missing map information in Figure 1 in the revised manuscript. In response to comment 3 of Reviewer 1, we decide to limit the evaluation period to 2005 instead of 2012 for the climate model products. 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.

<u>Response to Reviews by MSc student</u>

Paper Summary

This paper sought to evaluate the performance and reliability of daily gridded precipitation products for Canada – based on seasonality and eco/hydro-zones. The aim of defining specific climatic/hydrological regions and factoring in seasonality was to relay more usability and relatability with the results. The authors identified a need for such study as few had been done previously which looked at precipitation products for Canada – although they do make reference to a study being conducted previously for "North America".

7 datasets were assessed which fell under 1 of 5 types of precipitation products: station-based, station-based model-derived, Reanalysis-based multiple-source, GCM statistically downscaled and GCM-driven RCM dynamically downscaled. These products were compared against direct precipitation-gauge data from an adjusted and homogenized dataset covering Canada, with the authors acknowledging the scarcity of gauges and lack of quantification of the uncertainty associated with this benchmark dataset.

A Kolmogorov-Smirnov test was done to compare the probability distributions of the products and 4 performance measures were carried out: Percentage of Bias, Root-mean-square-error, Correlation coefficient and Standard Deviation. Ultimately, the results indicated a strong conclusion was not possible that would name one product superior to all others. Rather, 9 concluding points were presented which cover various regions, seasons and performance measures.

> *We thank the reviewer for reviewing our manuscript and providing a very nice summary of our work. We have now addressed all of the comments and presented our responses below, with deleted materials being crossed out by drawing a line through them and revised sentences being coloured in red.*

Main points:

1. **Overall this study does fall under the scope of HESS and has a meaningful aim in assessing the reliability of precipitation products as these same datasets are the ones which feed into hydrological models. This type of work appears to no have been carried out on such a large scale previously, but perhaps setting out to analyze and summarize 7 datasets, over 15 regions and for all seasons is too grand for a single paper. It is apparent that widespread results exist, as evidenced by the conclusions that the performance of the products depended on both season and eco-zone. An alternative approach to add greater clarity to a project of this size could be to re-structure the format of the paper to present the results based on the zones assessed, perhaps in a**

tabular format. This would also help users of this study to efficiently compare, contrast and determine the best dataset for their needs (which was an objective of this study). Although the results, discussion and conclusion sections are presented in a convoluted manner, the outcome is still thorough and definitive conclusions are presented. As well, the performance measure methodology is clearly presented and would be easy to reproduce.

We appreciate the value of the reviewer's suggestion on the format of the presentation of results. We agree that presenting the results based on ecozones in a tabular format would be very efficient to compare and contrast only when several datasets (e.g. three to four) over a few regions (e.g. up to five) are involved in the analysis (i.e. up to 20 numbers in a table). However, when more datasets and more regions are involved, such as in our case (six datasets over 15 ecozones), efficiency might be significantly reduced when going through a tabular table with 90 numbers. We have already thought about different ways to present and summarize our results (e.g. tabular table, Taylor diagram, line graph, box and whisker plot) and identified portrait diagram (Figures 6 and 7 in the revised manuscript), which is widely used in climate models comparison studies (e.g. Pincus et al., 2008;Sillmann et al., 2013), is the most suitable way to show the results which can highly condense information in one diagram.

2. **The precipitation data section is incredibly unclear. It would first be beneficial to break the section into further components, for example data sources, limitations and treatment. Secondly, the authors have presented a lengthy description on how data was gathered, complied and corrected, although all of this work was carried out in previous research.**

   We agree that we have a lengthy data description section as it is also commented by Reviewer 1. The details in Sections 3.1 and 3.2 have been greatly reduced in the revised manuscript. In short, the spatial and temporal resolutions of each product, their compositions, and examples of their applications have been remained and other details have been deleted. Please refer to Reviewer 1's comment 4 for details.

3. **What is lacking is a better description toward the end of the section to outline why exactly this reference dataset was selected despite it clearly having major deficiencies. Three studies are referenced with regards to this dataset being widely used yet no further information is presented. This reference dataset is an integral piece of the analysis, all of the datasets are being compared to it, therefore it is not enough to only state that it "has been recognized". It would make more sense to outline in detail why it is being used rather than how it came to be as that work has already been done.**

   We have further explained and justified the reasons of using Mekis and Vincent (2011) as our reference in the revised manuscript [L264-270], which is shown as follows:

~~Despite the lack of a measure of associated uncertainty, this adjusted precipitation-gauge station dataset has been recognized and widely used for different~~ These adjusted values are considered as better estimates of actual precipitation and therefore have been used in numerous 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~~ Given the lack of an adjusted daily gridded precipitation product for Canada, the AHCCD station precipitation is considered to be the best available data for Canada and thus is used as the benchmark for all gridded precipitation product comparisons.

4. **This study was done for a large scale and included a number of variables. Textually the results are quite difficult to follow and there is an abundance of figures provided to illustrate these results, but they too are quite dense. A solution would be to either separate, enlarge or regroup the figures to add clarity and meaning to the results, and by doing so much of the text can be condensed to include key references to the figures without spelling out each result.**

   We agree that some of the figures are too dense as it is also commented by both Reviewers 1 and 2. However, we believe that Figures 1, 5 to 8 are clear enough to show the messages and therefore we have only enlarged the figures as much as possible in the revised manuscript. Please refer to Reviewer 1's comment 5 for the revised figures. In response to comment 3 of Reviewer 1, we decide to limit the evaluation period to 2005 instead of 2012 for the climate model products. 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.

Minor Points.

5. **Title: the word various does not add any meaning. It can be removed or the count of precipitation products can be used in its place.**

   We agree that the word various does not add much meaning in the tile and we decide to remove the word in the revised manuscript. Also, in response to comment 2 of Reviewer 1, we have changed the title from "Evaluation" to "Inter-comparison" to better reflect our aim of the study. The title in the revised manuscript has become:

Inter-comparison of daily precipitation products for large-scale hydro-climatic applications over Canada

6. **Abstract: should list the precipitation products under review, as well, mentions a "systematic analysis framework" but the paper does not read as though any framework has been developed.**

   We fully understand that it is essential to list the precipitation products under review in the abstract. However, given the numbers of precipitation products we analyzed and the length of the full names of the products, listing the products in the abstract takes so much room which then limit the messages we can deliver from our study. Therefore, we prefer saving the space for telling the main findings of our study which are more important to the readers and decide not to add the list of the products in the revised abstract. We have deleted "systematic analysis framework" and reduced the length of the abstract when responding to comment 1 of Reviewer 2. Please refer to Reviewer 2's comment 1 for the revised abstract.

7. **Structure and Content: needs reworking.**

- **P15:L28: references Section 2.1 which does not exist. Should reference section 3.1 instead.**
  Thank you for spotting out this mistake. We have deleted the sentence and re-written the whole paragraph for better clarification in the revised manuscript [L500-516], which is shown as follows:

  > *To identify the most consistent gridded dataset corresponding to different seasons and regions ~~across Canada~~, comparisons of each gridded product with direct precipitation-gauge station data from the aforementioned AHCCD ~~the Canadian adjusted and homogenized precipitation datasets of Mekis and Vincent (2011) (see Sect. 2.1~~3.1~~)~~ were carried out.*

- **Study area includes a discussion of data collection.**
  We are unsure what the reviewer means by "a discussion of data collection" and we believe that we have discussed the overview of data availability in Canadian situation [L152-173], which has been moved from Study Area Section [L203-224] to Introduction Section [L152-173]*. We have also provided the data descriptions in Section 3 in the original manuscript. Also, we believe that it is better to separately describe the study area and data collection given the amount of datasets being analyzed which otherwise it will be too long for one section.*
- **Introduction should be presented on its own. "Precipitation measurements and their limitations" and "Objectives and Scope" should not be in the introduction.**
  Thank you for your suggestion. We think that having the subheadings in the introduction helps the readers to better understand and to faster grasp the ideas of the paragraphs. Therefore, we decide to keep the subheadings in the revised manuscript. Also, we have changed the subheading "Objectives and Scope" to "Scope and Objectives".

- **Most of section 3.2 can be removed and inserted as a summary table as it completely references the outcome of prior studies.**
  The details in Section 3.1 and 3.2 have been greatly reduced in the revised manuscript and the changes are shown in the response to Reviewer 1's comment 4. We do have a summary table (Table 1) in the original manuscript to provide an overview of the datasets being compared.

8. **Language: an edit should be conducted to check for grammar and sentence structure. Examples:**

The results point on P28:15 contains 3 sets of parentheses in a single sentence.

We have deleted one set of parentheses in the revised manuscript [L882], which is shown as follows:

> *In northern Canada (above 60° N), the different products tended to moderately (ranging from -0.6 % to -40.3 %) ~~(~~and in cases significantly (up to -60.3 % in Taiga Cordillera)~~)~~ underestimate total precipitation, while reproducing the timing of daily precipitation rather well. It should be noted that this assessment was based on only a limited number of precipitation-gauges in the north.*

The sentence on P7:L20 ends with "along the southern Canada".

We have changed the sentence in the revised manuscript [L162-164], which is shown as follows:

> *The Meteorological Service of Canada has implemented a network of 31 radars (radar coverage at full range of 256 km) along ~~the~~ southern Canada (see Fortin et al. (2015b) Fig. 1 for spatial distribution).*

P8:L4 refers to the province of Alberta as Alberta province.

We have deleted "Alberta province" and re-written the whole paragraph for better description of the data in the revised manuscript [L227-270], which is shown as follows:

> *In Canada, ~~Climate~~ climate data collection is coordinated by the Federal government, which is ~~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 by ~~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 and Climate Change Canada (NCDA). These data provide the basis for all the available quality controlled climate ~~data~~ observations.*

9. **References: ample amount of references but this is appropriate given the amount of datasets being analysed. Though several references appear dated, for example the Radar Reflectivity and Surface Rainfall paper likely had several further advances on the topic since 1987.**

We agree that the Austin (1987) reference is a bit outdated and there are further advances in addressing the errors in rain-rate reflectivity by the radar. We have updated and replaced Austin (1987) reference in the revised manuscript [L83-84] by Villarini and Krajewski (2010), which is shown as follows:

> Villarini, G., and Krajewski, W. F.: Review of the Different Sources of Uncertainty in Single Polarization Radar-Based Estimates of Rainfall, Surv Geophys, 31, 107-129, 10.1007/s10712-009-9079-x, 2010.

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.
Gebregiorgis, A. S., and Hossain, F.: How well can we estimate error variance of satellite precipitation data around the world?, Atmos Res, 154, 39-59, 10.1016/j.atmosres.2014.11.005, 2015.
Gebremichael, M., Krajewski, W. F., Morrissey, M. L., Huffman, G. J., and Adler, R. F.: A detailed evaluation of GPCP 1 degrees daily rainfall estimates over the Mississippi river basin, J Appl Meteorol, 44, 665-681, Doi 10.1175/Jam2233.1, 2005.
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.
Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., Oki, R., Nakamura, K., and Iguchi, T.: The Global Precipitation Measurement Mission, B Am Meteorol Soc, 95, 701-+, 10.1175/Bams-D-13-00164.1, 2014.
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.
Nijssen, B., and Lettenmaier, D. P.: Effect of precipitation sampling error on simulated hydrological fluxes and states: Anticipating the Global Precipitation Measurement satellites, J Geophys Res-Atmos, 109, Artn D02103 10.1029/2003jd003497, 2004.
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.
Pincus, R., Batstone, C. P., Hofmann, R. J. P., Taylor, K. E., and Glecker, P. J.: Evaluating the present-day simulation of clouds, precipitation, and radiation in climate models, J Geophys Res-Atmos, 113, Artn D14209

10.1029/2007jd009334, 2008.
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.
Sillmann, J., Kharin, V. V., Zhang, X., Zwiers, F. W., and Bronaugh, D.: Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate, J Geophys Res-Atmos, 118, 1716-1733, 10.1002/jgrd.50203, 2013.
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.
Villarini, G., and Krajewski, W. F.: Review of the Different Sources of Uncertainty in Single Polarization Radar-Based Estimates of Rainfall, Surv Geophys, 31, 107-129, 10.1007/s10712-009-9079-x, 2010.
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.
Werner, A. T., and Cannon, A. J.: Hydrologic extremes - an intercomparison of multiple gridded statistical downscaling methods, Hydrol Earth Syst Sc, 20, 1483-1508, 10.5194/hess-20-1483-2016, 2016.
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.
Yong, B., Liu, D., Gourley, J. J., Tian, Y. D., Huffman, G. J., Ren, L. L., and Hong, Y.: GLOBAL VIEW OF REAL-TIME TRMM MULTISATELLITE PRECIPITATION ANALYSIS Implications for Its Successor Global Precipitation Measurement Mission, B Am Meteorol Soc, 96, 283-296, 10.1175/Bams-D-14-00017.1, 2015.
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.

~~Evaluation~~ Inter-comparison of ~~various~~ daily precipitation products for large-scale hydro-climatic applications over Canada

Jefferson S. Wong[1*], Saman Razavi[1], Barrie R. Bonsal[2], Howard S. Wheater[1], & Zilefac E. Asong[1]

[1] *Global Institute for Water Security and School of Environment and Sustainability, University of Saskatchewan, 11 Innovation Blvd, Saskatoon, SK, Canada S7N 3H5*

[2] *Environment and Climate Change Canada, 11 Innovation Blvd, Saskatoon, SK, Canada S7N 3H5*

[*] Corresponding author:

Email: jefferson.wong@usask.ca

Phone: +1 306 966 7816

Abstract

A number of global and regional gridded climate products based on multiple data sources ~~and models~~ are available that can potentially provide ~~better and more~~ reliable estimates of precipitation for climate and hydrological studies. However, research into the reliability of these products for various regions has been limited and in many cases non-existent. This study ~~identifies~~ inter-compares several gridded precipitation products and ~~over Canada and develops a systematic analysis framework to assess the characteristics of errors associated with the different datasets, using the best available adjusted precipitation-gauge data as a benchmark over the period 1979 to 2012. The framework~~ quantifies the spatial and temporal variability of the errors (relative to station observations) over 15 terrestrial ecozones in Canada for different seasons over the period 1979 to 2012 at 0.5° and daily spatiotemporal resolution. These datasets were assessed in their ability to represent the daily variability of precipitation amounts by four performance measures: percentage of bias, root-mean-square-error, correlation coefficient, and standard deviation ratio. ~~at the daily time scale.~~ Results showed that most of the ~~products~~ datasets were relatively skillful in central Canada. However, they tended to overestimate precipitation amounts in the west and underestimate in the north and east, with the underestimation being particularly dominant in northern Canada (above 60° N). ~~but tended to underestimate precipitation amounts on the east coast and overestimate on the west~~. The global product by WATCH Forcing Data ERA-Interim (WFDEI) augmented by Global Precipitation Climatology Centre (GPCC) data (WFDEI [GPCC]) performed best with respect to different metrics. The Canadian Precipitation Analysis (CaPA) product ~~of Meteorological Service of Canada,~~ performed comparably with WFDEI [GPCC], however it only provides data starting from 2002. All the ~~products~~ datasets performed best in summer, followed by autumn, spring, and winter in order of decreasing quality. ~~Due to the sparse observational network, northern Canada (above 60° N) was most difficult to assess with the majority of products tending to significantly underestimate total precipitation. Results~~ Findings from this study can ~~be used as a~~provide guidance ~~for~~ to potential users regarding the performance of different precipitation products for a range of geographical regions and time periods.

Keywords: precipitation; evaluation and comparison; datasets; ~~reanalysis~~ecozones; hydro-climatology; Canada

1. Introduction

The availability of accurate data, especially precipitation~~,~~ is essential for understanding the climate system and hydrological processes~~, as precipitation~~ since it is a vital element of the water and energy cycles and a key forcing variable in driving hydrological models. ~~Precipitation~~ Reliable precipitation measurements provide valuable information for meteorologists, climatologists, hydrologists, and other decision makers in many applications, including climate ~~change~~ and/or land-use change studies (e.g. Cuo et al., 2011;Huisman et al., 2009;Dore, 2005), agricultural and environmental studies (e.g. Zhang et al., 2012;Hively et al., 2006), natural hazards (e.g. Taubenbock et al., 2011;Kay et al., 2009;Blenkinsop and Fowler, 2007), and hydrological and water resource planning (e.g. Middelkoop et al., 2001;Hong et al., 2010). With respect to land-surface hydrology, the increasing sophistication of distributed hydrological modeling has urged the requirement of better and more reliable gridded precipitation estimates ~~with~~ at a minimum, daily temporal resolution. Before incorporating precipitation measurements, quantifying their uncertainty becomes an essential prerequisite for hydrological applications and is increasingly critical for potential users who are left without guidance and/or confidence in the myriad of products for their specific hydrological problems over different geographical regions. This ~~paper~~ study attempts to address this issue by comparing and examining the error characteristics of different types of gridded precipitation products and assessing how these precipitation products perform geographically and temporally over Canada.

Precipitation measurements and their limitations

With ~~the~~ technological and scientific advancements over the past three decades, tremendous progress has been made in the various methods of precipitation measurement, each one with its own strengths and limitations. Conventional measurements through the use of rain gauges continue to play an important role in precipitation observations, as they are the only source that provide the direct physical readings and provide relatively accurate measurements at specific points. However, such measurements are subject to various errors arising from wind effects (Nešpor et al., 2000;Ciach, 2003), evaporation (Strangeways, 2004;Mekis and Hogg, 1999), undercatch (Yang et al., 1998;Adam and Lettenmaier, 2003;Mekis and Hogg, 1999), and instrumental problems like basic mechanical and electrical failure. Moreover, since many applications such as distributed hydrological models and hydraulic models require areal precipitation estimates, rain-gauge measurements are often spatially interpolated. Interpolation, however, may not capture the true spatial variability of precipitation field due to sparsity of gauge networks, particularly in complex terrains like mountainous regions or remote high latitude locations. Radars, as alternative ground-based measurements, can estimate precipitation over a relatively large area (radius of 200 to 300 km), but are also prone to inaccuracies as a result of beam spreading, curvature of the earth, and terrain blocking (Dinku et al., 2002;Young et al., 1999), and errors in the rain rate-reflectivity relationship, range effects, and clutter (Jameson and Kostinski, 2002;~~Austin, 1987~~;Villarini and Krajewski, 2010). Development of satellite-based precipitation estimates has provided coverage over vast gauged/ungauged regions with continuous observations regardless of time of day, terrain, and weather condition of the ground (Gebregiorgis and Hossain, 2015). The recently launched Global Precipitation Measurement (GPM) Core Observatory has further opened up new opportunities for observing worldwide precipitation from space (Hou et al., 2014). However, satellite-based estimates also contain inaccuracies resulting primarily from temporal sampling errors due to infrequent satellite visits to a particular location, instrumental errors due to calibration and measurement noise, and algorithm errors related to approximations ~~to the~~ in cloud physics ~~used~~ (Nijssen and Lettenmaier, 2004;Gebremichael et al., 2005). In particular, the passive microwave overpasses were shown to be unreliable over regions with snow cover and complex terrain such as the Tibetan Plateau (Yong et al., 2015).

Recognizing the limitations inherent in the individual sources of precipitation observation, a number of attempts to combine information from multiple sources have been undertaken (Xie and Arkin, 1996;Maggioni et al., 2014;Shen et al., 2010). Numerous approaches ~~have been~~were developed to produce high-resolution ~~precipitation~~ estimates through combining infrared and microwave data (e.g. Huffman et al., 2007;Turk et al., 2010), merging multi-satellite products with gauge observation (e.g. Huffman et al., 1997;Huffman et al., 2010;Adler et al., 2003;Xie and Arkin, 1997;Wang and Lin, 2015), and implementing different precipitation retrieval techniques (e.g. Joyce et al., 2004;Hsu et al., 2010). Reanalysis data provide an alternative source of precipitation estimates that mitigate the sparse distribution of precipitation observations by assimilating all available data (rain-gauge stations, aircraft, satellite, etc.) into a background forecast physical model. However, they are only an estimate of the real state of the atmosphere which do not necessarily match the observations (Bukovsky and Karoly, 2007;West et al., 2007). Inaccuracies in reanalysis precipitation might also arise from the complex interactions between the model and observations that depend on the specific analysis-forecast systems and the choice of physical parameterizations, especially in regions of missing observations (Betts et al., 2006). Numerical ~~coupled~~ climate models including Atmosphere-Ocean General Circulation Models (AOGCMs) and Regional Climate Models (RCMs) offer another potential source of precipitation estimates, as well as future precipitation simulations. ~~GCMs~~ AOGCMs remain relatively coarse in resolution (approximately 100 to 250 km) and are not able to resolve important sub-grid scale features such as topography, land cover, and clouds (Grotch and Maccracken, 1991), resulting in the requirement of downscaling to provide fine resolution climate parameters for hydrological analyses. ~~Two families of downscaling approaches are commonly used including statistical and dynamical approaches and they have their own advantages and disadvantages (Wilby and Wigley, 1997).~~ In general, precipitation estimates from climate models often produce systematic bias due to imperfect model-conceptualization~~ of the models~~, discretization and spatial averaging within grid cells (Teutschbein and Seibert, 2010;Xu et al., 2005).

~~Objectives and Scope~~Scope and Objectives

Numerous previous evaluation efforts among the precipitation products have been limited into three groups of inter-comparison of (1) satellite-derived products (e.g. Adler et al., 2001;Xie and Arkin, 1995;Turk et al., 2008); (2) reanalysis data (e.g. Janowiak et al., 1998;Bosilovich et al., 2008;Betts et al., 2006;Bukovsky and Karoly, 2007); and (3) climate model simulations (e.g. Covey et al., 2003;Christensen et al., 2007;Mearns et al., 2006;2012). Despite the ~~tremendous~~ aforementioned efforts, few studies have conducted a detailed inter-comparison among different types of precipitation products. Gottschalck et al. (2005) ~~was one of the very few studies which~~ compared the seasonal total precipitation of several satellite-derived, rain-gauge-based, and model-simulated datasets over contiguous United States ~~(CONUS)~~ and showed the spatial root mean square error of seasonal total precipitation and mean correlation of daily precipitation between each product and the impacts of these errors on land surface modelling. Additionally, Ebert et al. (2007) examined 12 satellite-derived precipitation products and four numerical weather prediction models over the United States, Australia, and northwestern Europe and found that satellite-derived precipitation estimates performed best in summer and model-induced ones performed best in winter. However, a number of questions regarding the reliability of the precipitation products remained in doubt, including: to what extent do the users have the knowledge about the error information associated with all these different types of precipitation products; how do the error distribution of precipitation products vary by location and season; and which product(s) should the users choose for their regions of interest. Answering these questions is, therefore, a crucial first step in quantifying the spatial and temporal variability of the precipitation products so as to ~~improve~~ better understand their reliability as forcing inputs in hydrological modelling and other related studies.

Given the emergence of various products derived from different methods and sources (Tapiador et al., 2012), accuracy comparison studies of ~~-~~precipitation products have been reported over several regions; examples include the globe (e.g. Gebregiorgis and Hossain, 2015;Adler et al., 2001;Tian and Peters-Lidard, 2010), Europe (e.g. Frei et al., 2006;Chen et al., 2006;Kidd et al., 2012), Africa (e.g. Dinku et al., 2008;Asadullah et al., 2008), North America (e.g. Tian et al., 2009;West et al., 2007), South America (e.g. Vila et al., 2009), China (e.g. Shen et al., 2010;Wetterhall et al., 2006). However, less attention has been paid to high-latitude regions like Canada where a considerable proportion of precipitation is in the form of snow (Behrangi et al., 2016). In many regions of Canada, precipitation-gauge stations are sparsely distributed and the information required for hydrological modelling may not be available at the site of interest. This is especially true in northern areas (north of 60°N) and over mountainous regions where precipitation-gauge stations are usually 500 to 700 km apart or at low elevations (Wang and Lin, 2015). Meanwhile, the decline and closure of manual observing precipitation-gauge stations further reduced the spatial coverage and availability of long-term precipitation measurements (Metcalfe et al., 1997;Mekis and Hogg, 1999;Rapaic et al., 2015). Of additional concern, the observations for solid precipitation (snow, snow pellets, ice pellets, and ice crystals) and precipitation phase (liquid or solid) changes make accurate measurement of precipitation more difficult and challenging, and the measurement errors have been found to range from 20 to 50 % for automated systems (Rasmussen et al., 2012). The Meteorological Service of Canada has implemented a network of 31 radars (radar coverage at full range of 256 km) along southern Canada (see Fortin et al. (2015b) Fig. 1 for spatial distribution). This Canadian radar network has been employed as an additional source of observations in generating the gridded product CaPA (see Sect. 3.2.2 for details). Yet, the shortcomings of using the radar data are twofold: (1) many areas of the country (north of 60°N) are not covered by this network; and (2) the implementation of the network began in 1997 and thus did not have sufficient lengths of data for any long-term hydro-climatic studies. The availability, coverage, and quality of precipitation-gauge measurements are thus obstacles to effective hydrological modelling and water management in Canada. However, the availability of several global and regional gridded precipitation products which provide complete coverage of the whole country at applicable time and spatial scales may provide a viable alternative for regional- to national-scale precipitation analyses in Canada.

Given the aforementioned, this study aims to (1) ~~evaluate~~ inter-compare various daily gridded precipitation products against the best available precipitation-gauge ~~measurements~~observations; and (2) characterize the error distributions of different types of precipitation products over time and different geographical regions in Canada. ~~Evaluation~~ Such inter-comparison ~~of the products over specific climatic/hydrological regions~~ will in turn help assess the performance of the precipitation products ~~under different circumstances~~over specific climatic/hydrological regions.

The rest of this paper is organized as follows: a brief description of the study area and precipitation data ~~used~~ is provided in Sect. 2 and 3. The methodology for evaluating precipitation products against the precipitation-gauge station ~~data~~ observations is described in Sect. 4. Results and discussion are provided in Sect. 5 and 6, respectively, with a summary and conclusion following in Sect. 7.

2. Study Area

Canada, which covers a land area of 9.9 million km^2, extends ~~northward~~ from 42° N to 83° N latitude and spans between 141° W to 52° W longitude. With substantial variations over its landmass, the country can be divided into many regions according to aspects such as climate, topography, vegetation, soil, geology, and land use. The National Ecological Framework for Canada classified ecologically distinct areas with four hierarchical levels of generalization (15 ecozones, 53 ecoprovinces, 194 ecoregions, and 1021 ecodistricts from broadest to the smallest) (Ecological Stratification Working Group, 1996;Marshall et al., 1999). Similarly, the Standard Drainage Area Classification (SDAC) ~~in 2003~~ was developed to delineate hydrographic areas to cover all the land and interior freshwater lakes of the country with three levels of classification (11 major drainage areas, 164 sub-drainage areas, and 974 sub-sub-drainage areas) (Brooks et al., 2002;Pearse et al., 1985). The precipitation comparisons in this study incorporated both the ecological and hydrological delineations. This involved classifying the Canadian landmass into 15 ecozones for the main study (Fig. 1) and 14 major drainage areas (the Arctic Major Drainage Area was further divided into Arctic and Mackenzie, whereas the St. Lawrence Major Drainage Area was further split into St. Lawrence, Great Lakes, and Newfoundland). Results ~~presented in the body of the paper~~ are based on the ecozone classification; while those based on drainage areas are reported in the supplementary material~~s, for the sake of brevity~~.

~~In many regions of Canada, precipitation-gauge stations are sparsely distributed and the information required for hydrological modelling may not be available at the site of interest. This is especially true in northern regions (north of 60° N) and over mountainous regions where rain-gauge stations are usually 500 to 700 km apart or at low elevations (Wang and Lin, 2015). Meanwhile, the decline and closure of manual observing rain-gauge stations further reduced the spatial coverage and availability of long-term precipitation measurements (Metcalfe et al., 1997;Mekis and Hogg, 1999;Rapaic et al., 2015). Of additional concern, the observations for solid precipitation (snow, snow pellets, ice pellets, and ice crystals) and precipitation phase (liquid or solid) changes make accurate measurement of precipitation more difficult and challenging, and the measurement errors have been found to range from 20 to 50 % for automated systems (Rasmussen et al., 2012). The Meteorological Service of Canada has implemented a network of 31 radars (radar coverage at full range of 256 km) along the southern Canada (see Fortin et al. (2015b) Fig. 1 for spatial distribution). This Canadian radar network has been employed as an additional source of observations in generating the gridded product CaPA (see Sect. 3.2.2 for details). Yet, the shortcomings of using the radar data are twofold: (1) many areas of the country (north of 60° N) are not covered by this network; and (2) the implementation of the network began in 1997 and thus did not have sufficient lengths of data for any long-term hydro-climatic studies. The availability, coverage, and quality of precipitation-gauge measurements are thus obstacles to effective hydrological modelling and water management in Canada. However, the availability of several global and regional gridded precipitation products which provide complete coverage of the whole country at applicable time and spatial scales may provide a viable alternative for regional- to national-scale precipitation analyses in Canada.~~

3. Precipitation Data

3.1. Precipitation-gauge station ~~data~~observations

In Canada, ~~Climate~~ climate data collection is coordinated by the Federal government, which is ~~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 by ~~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 and Climate Change Canada (NCDA).These data provide the basis for all the available quality controlled climate ~~data~~observations. ~~Based on the National Climate Data Archive of Environment Canada, there~~There are a total of 1499 precipitation-gauge stations (as ~~in~~ of 2012) across Canada. However, ~~due to the~~given the frequent addition and subtraction of climate stations, these numbers have greatly varied through time with peak reporting in the 1970s followed by a general decline to the present ~~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~~Furthermore, the existing precipitation observations are often subject to various errors, with gauge undercatch being of significant concern ~~among which the errors due undercatch are quite significant in Canada~~ (Mekis and Hogg, 1999). ~~In order to~~ To account for various measurement issues, Mekis and Hogg (1999) first produced the Adjusted and Homogenized Canadian Climate Data (AHCCD) including adjusted daily rainfall and snowfall values and Mekis and Vincent (2011) then updated the data for a subset of 464 stations over Canada. ~~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~~on the order of 5 to 10 % higher in southern Canada and more than 20 % in the Canadian Arctic when compared to ~~than~~ the original observations. ~~The effect of the adjustments on~~ Adjustments to snowfall were even larger and ~~more variable~~varied 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~~ These adjusted values are considered as better estimates of actual precipitation and therefore have been used in numerous 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~~ Given the lack of an adjusted daily gridded precipitation product for Canada, the AHCCD station precipitation is considered to be the best available data for Canada and thus is used 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 for assessment 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 length~~s~~ of data (>30 years) for long-term study ~~and cover~~including recent years up to 2012; and (4) ~~representation of~~representing a range of sources/methodologies (e.g. station based, remote sensing, model, blended products). Table 1 summarizes these datasets, including their full names and original spatial and temporal resolutions for the versions used. 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~~(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) used the Australian National University Spline (ANUSPLIN) model to ~~developed~~ develop a ~~climate~~ dataset of daily precipitation, and daily minimum and maximum air temperature over Canada at a spatial resolution of 300 arc-seconds ~~of latitude and longitude~~ (0.0833° or ~10 km) for the period of 1961 to 2003~~,~~ ~~using observed stations~~. All available NCDA stations (that ranged from 2000 to 3000 ~~in~~ for any given year~~s~~ ~~over the~~during this period) were used an input to the gridding procedure. ~~recorded in the National Canadian Climate Data Archives of Environment Canada. However, to~~ To retain ~~a better~~maximum spatial coverage, the smaller number of stations in AHCCD were not incorporated (i.e. only unadjusted archive values were used). ~~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~~ Interpolation procedures included incorporation of tri-variate thin-plate smoothing splines method ~~that incorporated~~ using 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~~ The Canadian ANUSPLIN has now 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~~multiple-source product – CaPA

~~Initiated in~~ 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~~Data were generated ~~based on~~using an optimum interpolation technique (Daley, 1993), which required ~~a background field and~~ a specification of error statistics between the observations and ~~the~~ a 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) model (Cote et al., 1998a;1998b)~~,~~ were used as the background field with the rain-gauge measurements from ~~the observational network~~NCDA as the observations to generate an analysis error at every grid point. ~~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~~ in the statistical interpolation method (Lespinas et al., 2015)~~,~~ and increase of spatial resolution to 10 km. ~~and the~~The assimilation of Quantitative Precipitation Estimates from the Canadian Weather Radar Network is also used 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~~started in 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~~This dataset (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 (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~~ additional temperature and precipitation data (e.g. Willmott et al., 2001), Princeton has been updated and is currently available with two versions: 1) 1948 to 2008 at 1.0°, ~~(plus~~ 0.5°, and 0.25°~~),~~ at 3-hourly, ~~(plus~~ daily, and monthly~~) resolution globally~~ time steps and 2) ~~for 1948 to 2008~~. ~~Experimental updates including a~~ 1901-2012 experimental version at 1.0° ~~(plus~~and 0.5°~~),~~ at 3-hourly, ~~(plus~~ daily, and monthly~~) resolution are also available~~ time steps (used in this study). Studies employing Princeton to ~~study~~ examine 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)~~. For instance, Kang et al. (2014) examined the changing contribution of snow to runoff generation in the Fraser River Basin while Su et al. (2013) investigated the relationships between spring snow and warm-season precipitation in central Canada. In addition, Wang et al. (2013) and Wang et al. (2014) used this dataset to characterize the spatial and seasonal variations of the surface water budget at Canada national scale.

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~~and 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~~ developed covering the period of 1979 to 2012 (Weedon et al., 2014). The WFDEI used the same methodology as the WFD, but was 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. Both sets of data were used in this study (hereafter ~~the~~ known as 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~~ With the aim of evaluating 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~~ For 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 beginning in January 2004 over the Athabasca River basin due to the assimilation of station observations over Canada being discontinued in 2003.

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, Canada, 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-seconds (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 for bias-correction and downscaling of the GCMs. ~~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 (Werner and Cannon, 2016). ~~The first one was~~ These included 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 (1950 – 2005) and ~~projected~~ future (2006 – 2100) 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~~ Drawing from the strengths of 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 and characterize the uncertainties of RCMs 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~~ and three RCMs were chosen for comparison due to the availability of the NA-CORDEX dataset (see Table 3 for details).

4. Methodology

~~3.3.~~4.1. Pre-processing

Given that the main focus of this study was to inter-compare the various gridded precipitation products using precipitation-gauge station data as a reference/benchmark (and not to assess the individual accuracy of each product against this reference), it was decided to re-grid each product onto a common 0.5° x 0.5° resolution to match the lowest-resolution dataset. It was acknowledged that re-gridding can introduce uncertainties due to the extra interpolations, however, the authors believe that upscaling to a common resolution provided a direct and more consistent inter-comparison. Furthermore, this methodology was consistent with similar studies in the literature (e.g. Janowiak et al., 1998;Rauscher et al., 2010;Kimoto et al., 2005). All data were accumulated to daily time scale for comparison. Two common time spans were selected since CaPA covered a shorter timeframe 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 as the historical period of the datasets ends in 2005); and (2) short-term comparison from January 2002 to December 2012 when CaPA data are available. Daily values were summed over the four standard seasons (spring: March to May – MAM, summer: June to August – JJA, autumn: September to November – SON, and winter: December to February – DJF) to inter-compare the precipitation products at a seasonal scale.

To identify the most consistent gridded dataset corresponding to different seasons and regions ~~across Canada~~, comparisons of each gridded product with direct precipitation-gauge station data from the aforementioned AHCCD ~~the Canadian adjusted and homogenized precipitation datasets of Mekis and Vincent (2011) (see Sect. 2.1)~~ were carried out. ~~It is recognized that the same gauged stations are utilized in both gridded precipitation products (ANUSPLIN and CaPA), however, the generation of these gridded data used archive (unadjusted) values from these stations. Also, as aforementioned, the Canadian radar network has been used in generating CaPA and thus could not be used as an independent source for evaluation of the gridded products. Two screening processes were done to select the suitable precipitation-gauge stations. The first was to eliminate those stations that did not cover the period from 1979 to 2012. This resulted in~~ For the period of 1979 to 2012, only169 ~~out~~ of the original 464 stations across Canada ~~being retained~~were available. ~~The~~ This drastic drop ~~in stations~~ was due to 271 ~~of them~~ stations ending before or after early 2000s and 23 not having a complete year of 2012. ~~The second step was to eliminate~~Subsequently, any of the 169 stations where the percentage of missing values exceeded 10 % ~~in the time series of~~ during the study period were also eliminated. This resulted in ~~a total of~~ 145 and 137 stations across Canada for long-term and short-term comparison respectively (see Fig. 1 for locations). Note that most of the stations are located in southern Canada with only 15 stations above 60° N.

~~Due to the different spatial and temporal resolutions of the various precipitation products, the first step was to re-grid each onto a common 0.5° x 0.5° resolution to match the lowest-resolution dataset. Those having sub-daily time scale were also aggregated to daily accumulation for comparison. 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; and (2) short-term comparison from January 2002 to December 2012 when CaPA are available. The analysis was performed by summing up the daily values for four seasons (spring: March to May, summer: June to August, autumn: September to November, and winter: December to February) to evaluate how well the precipitation products work in capturing the seasonal differences in precipitation.~~

Gridded-based precipitation estimates at the coordinates of the precipitation-gauge stations were then extracted by employing an inverse-distance-square weighting method (Cressman, 1959), which has been used to interpolate climate data for simple and efficient applications (Eum et al., 2014;Shen et al., 2001). This method assumes that an interpolated point is solely influenced by the nearby gridded points based on the inverse of the distance between the interpolated point and the gridded points. The interpolations ~~are~~ were carried out on an individual ecodistrict basis and ~~are~~ were based on both the number of precipitation-gauge stations and number of 0.5° x 0.5° grid cells within the ecodistrict in question. For instance, when a single precipitation-gauge station ~~is~~ was located within an ecodistrict, the value of the interpolated point ~~is~~ was calculated by using all of the gridded points within that ecodistrict. When two or more precipitation-gauge stations ~~are~~ were within the same ecodistrict, their interpolated values ~~are~~ were calculated by using the same numbers of gridded points but with different weightings based on inverse distance. In the case when an ecodistrict ~~contains~~ contained one grid cell, no weighting ~~is~~ was used and the interpolated value ~~is~~ was equal to the nearest ~~gridded~~ grid point.

~~3.4.~~4.2. Comparison of probability distributions using Kolmogorov-Smirnov test

A two-sample, non-parametric Kolmogorov-Smirnov (K-S) test ~~compared~~ was used to compare the cumulative distribution functions (CDFs) ~~for~~of ~~each type of~~ gridded precipitation product with the AHCCD. ~~at 5 % significance level ($\alpha = 0.05$) to support the~~ The null hypothesis (H_0) was that the two datasets came from same population. ~~Monthly total precipitation data were used and aggregated for each season because the existence~~ For each season, monthly total precipitation data were used to avoid commonly known issues of numerous zero values in the daily precipitation data that might affect significance. ~~reduce the statistical identification of significant differences to support the null hypothesis.~~ The K-S test was repeated independently for all precipitation-gauge stations at 5 % significance level ($\alpha = 0.05$). ~~and a measure of reliability (in percent) was calculated to show how reliable each type of precipitation products was among all the precipitation-gauge stations, as shown by Eq. (1)~~. A measure of reliability (in percent) was calculated based on counting the number of stations that do not reject the null hypothesis (any p-values greater than 0.05) over the total numbers of stations (145 and 137 stations in long-term and short-term comparison respectively), as shown in Eq. (1).

$$\cancel{\%\ of\ reliability = \frac{no\ of\ station\ that\ support\ H_0}{total\ no\ of\ precipitation\ gauge\ station} \cdot 100 \qquad (1)}$$

$$\%\ \text{of reliability} = \frac{number\ of\ stations\ that\ support\ H_0}{total\ number\ of\ precipitation\ gauge\ stations} \cdot 100 \qquad (1)$$

~~3.5.~~4.3.~~Evaluation~~ Comparison of gridded precipitation data using performance measures

Since the generation of the climate model-based precipitation products (PCIC dataset and NA-CORDEX dataset) only preserved the statistical properties without considering the ~~timing~~ day-by-day sequenceing of precipitation events in the observational record, these two datasets were excluded from the following ~~evaluation~~comparison, which only focused on the station-based and reanalysis-based gridded products. In particular, these ~~two~~ products were assessed in their ability to represent the daily variability of precipitation amounts ~~and occurrence~~ in different ecozones by four performance measures: percentage of bias ($PBias$) (P_{Bias}), root-mean-square-error ($RMSE$) (E_{rms}), correlation coefficient (r), and standard deviation ratio (σ_G/σ_R), as shown by Eqs (2) to (5), respectively.

$$P_{Bias;s} = \frac{\sum_i^N (G_i - R_i)}{\sum_i^N (R_i)} \cdot 100 \tag{2}$$

$$E_{rms;s} = \sqrt{\frac{\sum_i^N (G_i - R_i)^2}{N}} \tag{3}$$

$$r_s = \frac{\sum_i^N (G_i - \bar{G})(R_i - \bar{R})}{\sqrt{\sum_i^N (G_i - \bar{G})^2}\sqrt{\sum_i^N (R_i - \bar{R})^2}} \tag{4}$$

$$(\sigma_G/\sigma_R)_s = \frac{\sqrt{\frac{\sum_i^N (G_i - \bar{G})^2}{N}}}{\sqrt{\frac{\sum_i^N (R_i - \bar{R})^2}{N}}} \tag{5}$$

where s is the season, G and R are the spatial average of the daily gridded precipitation product and the reference observation dataset (precipitation-gauge stations) respectively, $\bar{G}$ and $\bar{R}$ are the daily mean of gridded precipitation product and point station data over the time spans (1979-2012 and 2002-2012), respectively, i is the i-th day of the season, and N is the total numbers of day in the season. These four performance measures examined different aspects of the gridded precipitation products, with $PBias$ for accuracy of product estimation, $RMSE$ for magnitude of the errors, r for strength and direction of the linear relationship between gridded products and precipitation-gauge station data, and σ_G/σ_R for amplitude of the variations.

~~4.~~5. Results

~~4.1.~~5.1. ~~Cumulative distribution function of all products~~Reliability of precipitation products

The percentage of reliability of each precipitation dataset ~~in each of the four seasons~~ during every season for the periods of 1979 to 2012 and 2002 to 2012 across Canada is shown in Fig. 2. The higher the percentage, the more reliable the precipitation dataset~~s are for the precipitation gauges~~ in question. In general, for long-term comparison (Fig. 2 left panel), WFDEI [GPCC] provided the highest percentage of reliability for the individual seasons (from spring to winter: 72.5 %, 81.4 %, 70.3 %, and 50.3 %) while NARR had the lowest percentage (24.8 %, 45.5 %, 27.6 %, and 11.7 %). Therefore in spring, WFDEI [GPCC] is not significantly different for 72.5 % of the 145 precipitation-gauge stations while for NARR it is only 24.8 %. ANUSPLIN is second in spring and summer (56.6 % and 73.1 %) and WFDEI [CRU] in autumn and winter (63.4 % and 45.5 %).

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 (4~~0.6 %~~2.5 %) regardless of the RCMs used whereas the reliability of CRCM5 (4~~3.3 %~~3.6 %) 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 gridded station-based and reanalysis-based products outperformed the climate model-simulated products.

With regard to the short-term comparison (Fig. 2 ~~right~~ middle panel), ANUSPLIN ~~had the best~~showed better performance in summer with 94.1 % of reliability among the 137 precipitation-gauge stations while CaPA ~~was the best~~indicated better skill 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 all 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 was excluded in the ~~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 corresponded to the various precipitation products. The ~~more~~ higher the numbers of high *p*-values (> 0.05) ~~are~~ in ~~one~~ each ecozone (either represented by a cluster of hollow circles or a thick black line in box-whisker plots towards 1 in y-axis in Figs ~~3 and 4~~3, 4 and 5), the more confidence (more consistent) ~~one has that the~~of each 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~~ those 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~~.

~~4.2.~~5.2. Daily variability of precipitation (Station-~~based~~ and reanalysis-based products)

The accuracy ($PBias$), magnitude of the errors ($RMSE$), strength and direction of the relationship between gridded products and precipitation-gauge station data (r), and amplitude of the variations (σ_G/σ_R) are shown in Figs ~~5~~ 6 and ~~6~~ 7 for the period of 1979 to 2012 and 2002 to 2012, respectively. In general, the gridded precipitation products that agree well with the precipitation-gauge station data should have relatively high correlation and low RMSE, low bias and similar standard deviation (indicated as light grey or dark grey square in Figs 5 and 6).

With respect to long-term comparison, in terms of overall accuracy among the four seasons, ANUSPLIN performed ~~the best~~relatively better in Region 11 (Taiga Cordillera) with smallest positive $PBias$ (+0.5 %) while the rest of the gridded products had negative $PBias$ ranging from -1.4 % (NARR) to -67.6 % (Princeton). However, ANUSPLIN was associated with a generally negative $PBias$ for the rest of the ecozones ranging from -5.3 % (Region 13 Pacific Maritime) to -29.6 % (Region 3 Southern Arctic), except for Regions 12 (Boreal Cordillera) and 14 (Montane Cordillera). On the other hand, WFDEI [CRU] and WFDEI [GPCC] had similar performances across different regions except in spring when the former underestimated the precipitation amounts by 63.0 % but the latter overestimated by 5.3 % in Region 11 (Taiga Cordillera). Differences could also be found in Region 7 (Atlantic Maritime) where WFDEI [CRU] overestimated precipitation amounts in spring, autumn, and winter by 10.6 %, 7.1 %, and 7.5 % while the accuracy of WFDEI [GPCC] was within -3.5 % to 0.5 % and it was the opposite case in Region 12 (Boreal Cordillera) in autumn and winter. With the exception of Regions 13 and 14, Princeton generally provided the overall largest underestimation of precipitation amounts across different ecozones by -25.9 %, -24.8 %, and -34.6 % in spring, autumn, and winter respectively. NARR ~~came second~~ performed second worst in spring (-19.0 %), autumn (-20.3 %), and winter (-27.1 %) and first in summer (-18.1 %). In general, all gridded products tended to overestimate total precipitation in Regions 12 to 14~~ and ~~, while Region 14 (Montane Cordillera) had the overall highest positive $PBias$ ranging from 17.1 % (WFDEI [GPCC]) to 44.2 % (WFDEI [CRU]).

When examining the magnitude of errors, ANUSPLIN, showed generally ~~agreed best~~better correspondence with precipitation-gauge station data, providing the overall lowest $RMSE$ across ecozones in four seasons (2.50 mm/day, 3.24 mm/day, 2.79 mm/day, and 2.45 mm/day) with the only exception in spring in Region 15 (Hudson Plain). Moreover, ANUSPLIN had the overall highest r across ecozones in four seasons (0.75, 0.78, 0.80, and 0.74). On the contrary, Princeton had the worst performance in both magnitude of errors and correlation with observations ~~no matter across different ecozones or among different seasons~~irrespective of ecozone or season, with the grand $RMSE$ and r of 5.65 mm/day and 0.17 respectively. The performances of WFDEI [CRU], WFDEI [GPCC], and NARR were in between ANUSPLIN and Princeton and they shared similar $RMSE$ and r across different regions and seasons, with very high magnitude of errors in Regions 6 to 8, and 13 and fair correlation in Regions 6 to 14 and minor regional and seasonal differences. The resulting values of the $RMSE$ metric in Regions 7 (Atlantic Maritime) and 13 (Pacific Maritime) tended to be larger than that of other ecozones. However, the other metrics such as $PBias$ and r showed better performance in these regions. This suggests that higher $RMSE$ values can be mainly attributed to the fact that precipitation amounts are higher in the maritime regions.

Regarding the amplitude of variations, NARR had the lowest variability across different regions in all four seasons (0.70, 0.67, 0.68, and 0.60), followed by ANUSPLIN (0.84, 0.77, 0.76, and 0.75). WFDEI [GPCC] had the most similar standard deviations as that of precipitation-gauge station data in Regions 5 to 8, 13, and 14 in autumn and winter while WFDEI [CRU] had about the same standard deviations in Regions 6 to 8 in autumn only. Unlike ANUSPLIN and NARR which were consistently having too little variability across different ecozones, Princeton estimated the amplitude of variations with more diversified regional and seasonal patterns. Princeton estimated σ_G/σ_R the best in Regions 4 to 10 in summer and Regions 9, 10, and 12 in autumn. However, the dataset had variations that were much larger than precipitation-gauge station data in Regions 7 and 8 in four seasons except summer, Region 13 in four seasons except winter, Region 14 in all seasons but too little variability in Regions 3, 11, and 15 in all seasons.

Concerning the short-term comparison, the performance of CaPA generally resembled that of ANUSPLIN in terms of accuracy, with general underestimation of precipitation amounts in Regions 4 to 10 in four seasons and overestimation in Region 12 and 13 especially in spring. CaPA had similar overestimation in Region 14 (Montane Cordillera) in winter as the rest of the gridded products but performed the best in estimating the precipitation amounts in other seasons of the region. CaPA also performed the best in Regions 5 and 15 in autumn among the gridded precipitation products. However, while all the gridded products experienced negative $PBias$ in Region 3 (Southern Arctic) in summer, CaPA performed the opposite with a positive $PBias$ of 10.8 %. Similar to ANUSPLIN, CaPA ~~was able to minimize the magnitude of errors and had strong association with precipitation-gauge station data, providing~~ had the second lowest overall $RMSE$ (2.70 mm/day, 3.74 mm/day, 3.35 mm/day, and 3.05 mm/day) and r (0.72, 0.73, 0.75, and 0.70) across ecozones in ~~four~~ all seasons, respectively. Despite its better performances in terms of $RMSE$ and r, CaPA was generally not able to capture ~~the right amount of~~satisfactorily the amplitude of variations, with consistently lower values across different regions for seasons~~less than that of the precipitation-gauge station data across different regions in four seasons~~ (0.83, 0.82, 0.85, and 0.72). ~~CaPA, however, estimated~~ In terms of σ_G/σ_R, CaPA showed more skill compared to ~~better than~~ ANUSPLIN (0.72, 0.76, 0.74, and 0.64) and NARR (0.75, 0.75, 0.72, and 0.63).

Some regional and seasonal differences could be seen in the other gridded precipitation products. For instance, seasonally, WFDEI [CRU] performed well in Region 8 (Mixedwood Plain) ~~in four seasons in terms of having~~as judged by low $PBias$ (~~within~~ -1.7 % to 4.3 %) for the period of 1979 to 2012 but ~~started to have~~showed higher positive $PBias$ in autumn and winter (7.1 % and 5.3 %) for the period of 2002 to 2012. WFDEI [GPCC] also ~~started to have~~had higher positive $PBias$ in Region 2 (Northern Arctic) in summer (7.4 % as compared to 1.2 %) and ~~in~~ winter (33.3 % as compared to 9.9 %). In terms of magnitude of errors and correlation with observations, the five gridded products in the long-term comparison performed similarly in the period of 2002 to 2012, with ANUSPLIN having the lowest grand $RMSE$ and r of 2.88 mm/day and 0.78 and Princeton being the worst again with the highest grand $RMSE$ and r of 6.12 mm/day and 0.16 respectively. Equally, the performances of ANUSPLIN and NARR in capturing the amplitude of variations were again consistently having too little variability across different ecozones. Princeton also demonstrated similar regional and seasonal differences as in the long-term comparison with higher variability in Regions 6 to 8 in all seasons except summer. WFDEI [CRU] and WFDEI [GPCC] both performed well in Regions 6 to 8, 12, and 14 in autumn.

~~5.~~6. Discussion

The preceding has provided insight into the relative performance of various gridded precipitation products over Canada ~~when compared~~relative to ~~adjusted~~ gauge measurements over different seasons and ~~geographical regions~~ecozones. Results showed that there is no particular product that is superior for all performance measures although ~~there are various~~some datasets ~~that do perform~~are consistently ~~better.~~

~~Based~~better. Based on the performances ~~in the four measures~~, one could broadly characterize the station-~~based~~ and reanalysis-based precipitation products into four groups, (1) ANUSPLIN and CaPA~~, as having~~ with negative $PBias$, low $RMSE$, high r, and small σ_G/σ_R; (2) WFDEI [CRU] and WFDEI [GPCC], ~~as~~ with relatively small $PBias$, high $RMSE$, fair r, and similar standard deviation; (3) Princeton, ~~as having~~with negative $PBias$, high $RMSE$, low r, and a mixture of large and small σ_G/σ_R; and (4) NARR, ~~as having~~with negative $PBias$, high $RMSE$, fair r, and small σ_G/σ_R. Among the reanalysis-based gridded products, Princeton performed the worst in all seasons and regions in terms of minimizing error magnitudes (Figs ~~7~~ 8 and ~~8~~9). Princeton was especially poor in winter (Fig. ~~7~~8) and showed significant underestimation in regions above 60° N (Fig. ~~8~~9). This could be due to the use of the NCEP-NCAR reanalysis as the basis to generate the dataset, which have been shown to be less accurate than NCEP-DOE reanalysis (used in NARR) and ERA-40 reanalysis (used in WFD) (Sheffield et al., 2006). The better performance of NARR in capturing the timings and amounts of precipitation than Princeton was probably because NCEP-DOE reanalysis was a major improvement upon the earlier NCEP-NCAR reanalysis in both resolution and accuracy. However, the overall reliability of NARR was among the poorest mainly because of non-assimilation of gauge precipitation observations over Canada from 2004 onwards, as reported by Mesinger et al. (2006). ANUSPLIN and CaPA performed well in capturing the timings and minimizing the error magnitudes of the precipitation, despite their general underestimation across Canada ($PBias$ ranging from -7.7 % (Region 13) to -40.7 % (Region 3) and -2.0 % (Region 15) to -17.1 % (Region 8) in the period of 2002 to 2012) (Fig. ~~8~~9) and too little variability (grand σ_G/σ_R of 0.72 and 0.80 of the same period). This was not surprising given the generation of the products was based on the unadjusted precipitation-gauge stations where the total rainfall amounts were increased after adjustment (Mekis and Vincent, 2011). WFDEI [CRU] and WFDEI [GPCC], on the other hand, performed well in estimating the accuracy and amplitude of variations, but not the timings and error magnitudes of the precipitation. This could probably due to the positive bias offsetting the negative bias resulting in small mean bias, but was picked up by $RMSE$ that gives more weights to the larger errors. The larger errors could be come from a mismatch of occurrence of precipitation in the time series, as reflected by the fair correlation coefficients (grand r of 0.52 and 0.50 for WFDEI [CRU], 0.54 and 0.53 for WFDEI [GPCC], for time periods of 1979 to 2012 and 2002 to 2012 respectively).

By matching the statistical ~~property~~ properties 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.

The evaluations of this comparison study ~~are~~ were impacted by the spatial distribution of adjusted precipitation-gauge stations (Mekis and Vincent, 2011), which were assumed to be the best representation of reality owing to the efforts in improving the raw archive of the precipitation-gauge stations ~~by accounting for various measurement issues like wind undercatch, evaporation and wetting loss, and snowfall adjustment~~. However, ~~this dataset was not error free and~~ the major limitation of this dataset was the number~~s~~ of precipitation-gauge stations that could be used for comparison~~ in this study~~. As aforementioned, due to temporal coverage not encompassing the entire study period and not having a complete year of 2012, over half of the precipitation-gauge stations were ~~dropped out for~~discarded from the analysis. Although the locations of the remaining stations covered much of Canada, there are only one or a few stations located in some of the ecozones (e.g. Region 3 to 5, 11, and 15). Even in Region 10 (Prairie) there are only nine precipitation-gauge stations for analysis. While the reliability of different types of gridded products could be tested in these ecozones, the consistency of the performance of each gridded product could not be established due to small sample sizes.

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 occurs 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 essentially represents 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 many not be qualified as the “ground truth”. Therefore, ANUSPLIN is also included in the pool of gridded products to be evaluated. Notwithstanding the issues, the authors feel that using the selected gauge measurements is best for the evaluation of the multiple gridded products because the set of gauges used has been adjusted (e.g. for undercatch) and are the most accurate source of information on precipitation in Canada (although 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 biases in differences between point and areal data is consistent for all the products. ~~However, the authors believe that given the current data situation, the preceding was the best methodology for evaluating the performance of different daily gridded precipitation products.~~

~~6.~~7. Conclusion

A number of gridded climate products incorporating multiple sources of data have recently been developed with the aim of providing better and more reliable measurements for climate and hydrological studies. There is a pressing need for characterizing the quality and error characteristics of various precipitation products and assessing how they perform at different spatial and temporal scales. This is particularly important in light of the fact that these products are the main driver of hydrological models in many regions, including Canadian watersheds where precipitation-gauge network is typically limited and sparse. This study was conducted to inter-compare several gridded precipitation products of their probability distributions ~~understand~~ and quantify the spatial and temporal variability of the errors ~~associated with five different types of gridded precipitation products~~relative to station observations in Canada, so as to provide some insights for potential users in selecting the products for their particular interests and applications. Based on the above analysis, the following conclusions can be drawn:

- In general, all the products performed best in summer, followed by autumn, spring, and winter in order of decreasing quality. The lower reliability in winter is likely the result of difficulty in accurately capturing solid precipitation.
- Overall, WFDEI [GPCC] and CaPA performed best with respect to different performance measures. WFDEI [GPCC], however, may be a better choice for long-term analyses as it covers a longer historical period. ANUSPLIN and WEDEI [CRU] also performed comparably, with considerably lower quality than WFDEI [GPCC] and CaPA. Princeton and NARR demonstrated the lowest quality in terms of different performance measures.
- Station-based and reanalysis-based products tended to underestimate total precipitation across Canada except in southwestern regions (Pacific Maritime and Montane Cordillera) where the tendency was towards overestimation. This may be the due to the fact that the majority of precipitation-gauge stations are located at lower altitudes which might not accurately reflect areal precipitation due to topographic effect.
- In southern Canada, WFDEI [GPCC] and CaPA demonstrated their best performance in the western cold interior (Boreal Plain, Prairie, Montane Cordillera) in terms of timing and magnitude of daily precipitation.

- ~~In Atlantic and Pacific coastal regions (Atlantic Maritime and Pacific Maritime) station-based and reanalysis-based products demonstrated their poorest performance in reproducing the timing and magnitude of daily precipitation.~~
- In northern Canada (above 60° N), the different products tended to moderately (ranging from -0.6 % to -40.3 %) ~~(~~and in cases significantly (up to -60.3 % in Taiga Cordillera)~~)~~ underestimate total precipitation, while reproducing the timing of daily precipitation rather well. It should be noted that this assessment was based on only a limited number of precipitation-gauges in the north.
- Comparing the climate model-simulated products, PCIC ensembles generally performed better than NA-CORDEX ensembles in terms of reliability and consistency in four seasons across Canada.
- In terms of statistical downscaling methods, the BCCAQ method was slightly more reliable than the BCSD method across Canada on the annual basis.
- Regarding GCMs, MPI-ESM-LR provide the highest reliability, followed by GFDL-ESM2G, CanESM2, and HadGEM2. With respect to RCMs, CRCM5 performed the best regardless of the GCM used, followed by CanRCM4, and RegCM4.

The findings from this analysis provide additional information for potential users to draw inferences about the relative performance of different gridded products. Although no clear-cut product was shown to be superior, researchers/users can use this information for selecting or excluding various datasets depending on their purpose of study. It is realized that this analysis only focused on the daily time scale at a relatively coarse 0.5° x 0.5° resolution suitable for large-scale hydro-climatic studies. ~~In addition, further~~Further research is thus required toward the performance assessment of various products with respect to precipitation extremes, which often have the greatest hydro-climatic impacts. As new products become available, similar comparisons should be conducted to assess their reliability.

Acknowledgements

The financial support from the Canada Excellence Research Chair in Water Security is gratefully acknowledged. Thanks are due to Melissa Bukovsky and Katja Winger from the NA-CORDEX modelling group for providing access to RegCM4 and CRCM5 data used in this study. The authors are also grateful to the various organizations that made the datasets freely available to the scientific community.

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.

Adler, R. F., Kidd, C., Petty, G., Morissey, M., and Goodman, H. M.: Intercomparison of global precipitation products: The third Precipitation Intercomparison Project (PIP-3), B Am Meteorol Soc, 82, 1377-1396, Doi 10.1175/1520-0477(2001)082<1377:Iogppt>2.3.Co;2, 2001.

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P. P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P., and Nelkin, E.: The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present), J Hydrometeorol, 4, 1147-1167, Doi 10.1175/1525-7541(2003)004<1147:Tvgpcp>2.0.Co;2, 2003.

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.

Asadullah, A., McIntyre, N., and Kigobe, M.: Evaluation of five satellite products for estimation of rainfall over Uganda, Hydrolog Sci J, 53, 1137-1150, DOI 10.1623/hysj.53.6.1137, 2008.

Austin, P. M.: Relation between Measured Radar Reflectivity and Surface Rainfall, Monthly Weather Review, 115, 1053-1071, Doi 10.1175/1520-0493(1987)115<1053:Rbmrra>2.0.Co;2, 1987.

Behrangi, A., Christensen, M., Lebsock, M. R., Stephens, G., Huffman, G. J., Bolvin, D., Adler, R. F., Gardner, A., Lambrigtsen, B., and Fetzer, E.: Status of High latitude precipitation estimates from observations and reanalyses, Journal of Geophysical Research: Atmospheres, 2016.

Betts, A. K., Zhao, M., Dirmeyer, P. A., and Beljaars, A. C. M.: Comparison of ERA40 and NCEP/DOE near-surface data sets with other ISLSCP-II data sets, J Geophys Res-Atmos, 111, Artn D22s04 10.1029/2006jd007174, 2006.

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.

Blenkinsop, S., and Fowler, H. J.: Changes in European drought characteristics projected by the PRUDENCE regional climate models, Int J Climatol, 27, 1595-1610, 10.1002/joc.1538, 2007.

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.

Bosilovich, M. G., Chen, J. Y., Robertson, F. R., and Adler, R. F.: Evaluation of global precipitation in reanalyses, J Appl Meteorol Clim, 47, 2279-2299, 10.1175/2008jamc1921.1, 2008.

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.

Brooks, R., Harvey, K., Kirk, D., Soulard, F., Paul, P., and Murray, A.: Building a Canadian Digital Drainage Area Framework, 55 th Annual CWRA Conference, Winnipeg, Manitoba, Canada, 2002,

Bukovsky, M. S., and Karoly, D. J.: A brief evaluation of precipitation from the North American Regional Reanalysis, J Hydrometeorol, 8, 837-846, 10.1175/Jhm595.1, 2007.

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.

Chen, D. L., Achberger, C., Raisanen, J., and Hellstrom, C.: Using statistical downscaling to quantify the GCM-related uncertainty in regional climate change scenarios: A case study of Swedish precipitation, Adv Atmos Sci, 23, 54-60, DOI 10.1007/s00376-006-0006-5, 2006.

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.

Christensen, J. H., Carter, T. R., Rummukainen, M., and Amanatidis, G.: Evaluating the performance and utility of regional climate models: the PRUDENCE project, Climatic Change, 81, 1-6, 10.1007/s10584-006-9211-6, 2007.

Ciach, G. J.: Local random errors in tipping-bucket rain gauge measurements, J Atmos Ocean Tech, 20, 752-759, Doi 10.1175/1520-0426(2003)20<752:Lreitb>2.0.Co;2, 2003.

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.

Covey, C., AchutaRao, K. M., Cubasch, U., Jones, P., Lambert, S. J., Mann, M. E., Phillips, T. J., and Taylor, K. E.: An overview of results from the Coupled Model Intercomparison Project, Global Planet Change, 37, 103-133, 10.1016/S0921-8181(02)00193-5, 2003.

Cressman, G. P.: An operational objective analysis system, Monthly Weather Review, 87, 367-374, 1959.

Cuo, L., Beyene, T. K., Voisin, N., Su, F. G., Lettenmaier, D. P., Alberti, M., and Richey, J. E.: Effects of mid-twenty-first century climate and land cover change on the hydrology of the Puget Sound basin, Washington, Hydrol Process, 25, 1729-1753, 10.1002/hyp.7932, 2011.

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.

Dinku, T., Anagnostou, E. N., and Borga, M.: Improving radar-based estimation of rainfall over complex terrain, J Appl Meteorol, 41, 1163-1178, Doi 10.1175/1520-0450(2002)041<1163:Irbeor>2.0.Co;2, 2002.

Dinku, T., Connor, S. J., Ceccato, P., and Ropelewski, C. F.: Comparison of global gridded precipitation products over a mountainous region of Africa, Int J Climatol, 28, 1627-1638, 10.1002/joc.1669, 2008.

Dore, M. H. I.: Climate change and changes in global precipitation patterns: What do we know?, Environ Int, 31, 1167-1181, 10.1016/j.envint.2005.03.004, 2005.

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.

Ecological Stratification Working Group: A national ecological framework for Canada, Centre for Land and Biological Resources Research, 1996.

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.

Frei, C., Scholl, R., Fukutome, S., Schmidli, J., and Vidale, P. L.: Future change of precipitation extremes in Europe: Intercomparison of scenarios from regional climate models, J Geophys Res-Atmos, 111, Artn D06105 10.1029/2005jd005965, 2006.

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.

Gebregiorgis, A. S., and Hossain, F.: How well can we estimate error variance of satellite precipitation data around the world?, Atmos Res, 154, 39-59, 10.1016/j.atmosres.2014.11.005, 2015.

Gebremichael, M., Krajewski, W. F., Morrissey, M. L., Huffman, G. J., and Adler, R. F.: A detailed evaluation of GPCP 1 degrees daily rainfall estimates over the Mississippi river basin, J Appl Meteorol, 44, 665-681, Doi 10.1175/Jam2233.1, 2005.

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.

Gottschalck, J., Meng, J., Rodell, M., and Houser, P.: Analysis of multiple precipitation products and preliminary assessment of their impact on global land data assimilation system land surface states, J Hydrometeorol, 6, 573-598, Doi 10.1175/Jhm437.1, 2005.

Grotch, S. L., and Maccracken, M. C.: The Use of General-Circulation Models to Predict Regional Climatic-Change, J Climate, 4, 286-303, Doi 10.1175/1520-0442(1991)004<0286:Tuogcm>2.0.Co;2, 1991.

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.

Hively, W. D., Gerard-Marchant, P., and Steenhuis, T. S.: Distributed hydrological modeling of total dissolved phosphorus transport in an agricultural landscape, part II: dissolved phosphorus transport, Hydrol Earth Syst Sc, 10, 263-276, 2006.

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.

Hong, Y., Adler, R. F., Huffman, G. J., and Pierce, H.: Applications of TRMM-Based Multi-Satellite Precipitation Estimation for Global Runoff Prediction: Prototyping a Global Flood Modeling System, Satellite Rainfall Applications for Surface Hydrology, 245-265, 10.1007/978-90-481-2915-7_15, 2010.

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.

Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., Oki, R., Nakamura, K., and Iguchi, T.: The Global Precipitation Measurement Mission, B Am Meteorol Soc, 95, 701-+, 10.1175/Bams-D-13-00164.1, 2014.

Hsu, K. L., Behrangi, A., Imam, B., and Sorooshian, S.: Extreme Precipitation Estimation Using Satellite-Based PERSIANN-CCS Algorithm, Satellite Rainfall Applications for Surface Hydrology, 49-67, 10.1007/978-90-481-2915-7_4, 2010.

Huffman, G. J., Adler, R. F., Arkin, P., Chang, A., Ferraro, R., Gruber, A., Janowiak, J., McNab, A., Rudolf, B., and Schneider, U.: The Global Precipitation Climatology Project (GPCP) Combined Precipitation Dataset, B Am Meteorol Soc, 78, 5-20, Doi 10.1175/1520-0477(1997)078<0005:Tgpcpg>2.0.Co;2, 1997.

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.

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P., and Stocker, E. F.: The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, J Hydrometeorol, 8, 38-55, 2007.

Huffman, G. J., Adler, R. F., Bolvin, D. T., and Nelkin, E. J.: The TRMM Multi-Satellite Precipitation Analysis (TMPA), Satellite Rainfall Applications for Surface Hydrology, 3-22, 10.1007/978-90-481-2915-7_1, 2010.

Huisman, J. A., Breuer, L., Bormann, H., Bronstert, A., Croke, B. F. W., Frede, H. G., Graff, T., Hubrechts, L., Jakeman, A. J., Kite, G., Lanini, J., Leavesley, G., Lettenmaier, D. P., Lindstrom, G., Seibert, J., Sivapalan, M., Viney, N. R., and Willems, P.: Assessing the impact of land use change on hydrology by ensemble modeling (LUCHEM) III: Scenario analysis, Adv Water Resour, 32, 159-170, 10.1016/j.advwatres.2008.06.009, 2009.

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.

Jameson, A. R., and Kostinski, A. B.: Spurious power-law relations among rainfall and radar parameters, Q J Roy Meteor Soc, 128, 2045-2058, Doi 10.1256/003590002320603520, 2002.

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.

Joyce, R. J., Janowiak, J. E., Arkin, P. A., and Xie, P. P.: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution, J Hydrometeorol, 5, 487-503, Doi 10.1175/1525-7541(2004)005<0487:Camtpg>2.0.Co;2, 2004.

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.

Kay, A. L., Davies, H. N., Bell, V. A., and Jones, R. G.: Comparison of uncertainty sources for climate change impacts: flood frequency in England, Climatic Change, 92, 41-63, 10.1007/s10584-008-9471-4, 2009.

Kidd, C., Bauer, P., Turk, J., Huffman, G. J., Joyce, R., Hsu, K. L., and Braithwaite, D.: Intercomparison of High-Resolution Precipitation Products over Northwest Europe, J Hydrometeorol, 13, 67-83, 10.1175/Jhm-D-11-042.1, 2012.

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.

Maggioni, V., Sapiano, M. R. P., Adler, R. F., Tian, Y. D., and Huffman, G. J.: An Error Model for Uncertainty Quantification in High-Time-Resolution Precipitation Products, J Hydrometeorol, 15, 1274-1292, 10.1175/Jhm-D-13-0112.1, 2014.

Mahfouf, J. F., Brasnett, B., and Gagnon, S.: A Canadian precipitation analysis (CaPA) project: Description and preliminary results, Atmos Ocean, 45, 1-17, 2007.

Marshall, I., Schut, P., and Ballard, M.: A national ecological framework for Canada: Attribute data. Ottawa, Ontario: Environmental Quality Branch, Ecosystems Science Directorate, Environment Canada and Research Branch, Agriculture and Agri-Food Canada, 1999.

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., 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.: Overview of the North American regional climate change assessment program, NOAA RISA-NCAR Meeting, 2006,

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.

Metcalfe, J. R., Routledge, B., and Devine, K.: Rainfall measurement in Canada: Changing observational methods and archive adjustment procedures, J Climate, 10, 92-101, Doi 10.1175/1520-0442(1997)010<0092:Rmicco>2.0.Co;2, 1997.

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.

Middelkoop, H., Daamen, K., Gellens, D., Grabs, W., Kwadijk, J. C. J., Lang, H., Parmet, B. W. A. H., Schadler, B., Schulla, J., and Wilke, K.: Impact of climate change on hydrological regimes and water resources management in the rhine basin, Climatic Change, 49, 105-128, Doi 10.1023/A:1010784727448, 2001.

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.

Nešpor, V., Krajewski, W. F., and Kruger, A.: Wind-induced error of raindrop size distribution measurement using a two-dimensional video disdrometer, J Atmos Ocean Tech, 17, 1483-1492, 2000.

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.

Nijssen, B., and Lettenmaier, D. P.: Effect of precipitation sampling error on simulated hydrological fluxes and states: Anticipating the Global Precipitation Measurement satellites, J Geophys Res-Atmos, 109, Artn D02103 10.1029/2003jd003497, 2004.

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.

Pearse, P. H., Bertrand, F., and MacLaren, J. W.: Currents of change; Final Report: inquiry on Federal water policy, Inquiry on Federal Water Policy, 1985.

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.

Rapaic, M., Brown, R., Markovic, M., and Chaumont, D.: An Evaluation of Temperature and Precipitation Surface-Based and Reanalysis Datasets for the Canadian Arctic, 1950-2010, Atmos Ocean, 53, 283-303, 10.1080/07055900.2015.1045825, 2015.

Rasmussen, R., Baker, B., Kochendorfer, J., Meyers, T., Landolt, S., Fischer, A. P., Black, J., Theriault, J. M., Kucera, P., Gochis, D., Smith, C., Nitu, R., Hall, M., Ikeda, K., and Gutmann, E.: HOW WELL ARE WE MEASURING SNOW? The NOAA/FAA/NCAR Winter Precipitation Test Bed, B Am Meteorol Soc, 93, 811-829, 10.1175/Bams-D-11-00052.1, 2012.

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.

Shen, S. S. P., Dzikowski, P., Li, G. L., and Griffith, D.: Interpolation of 1961-97 daily temperature and precipitation data onto Alberta polygons of ecodistrict and soil landscapes of Canada, J Appl Meteorol, 40, 2162-2177, Doi 10.1175/1520-0450(2001)040<2162:Iodtap>2.0.Co;2, 2001.

Shen, Y., Xiong, A. Y., Wang, Y., and Xie, P. P.: Performance of high-resolution satellite precipitation products over China, J Geophys Res-Atmos, 115, Artn D02114 10.1029/2009jd012097, 2010.

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.

Strangeways, I.: Improving precipitation measurement, Int J Climatol, 24, 1443-1460, 10.1002/joc.1075, 2004.

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.

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

Taubenbock, H., Wurm, M., Netzband, M., Zwenzner, H., Roth, A., Rahman, A., and Dech, S.: Flood risks in urbanized areas - multi-sensoral approaches using remotely sensed data for risk assessment, Nat Hazard Earth Sys, 11, 431-444, 10.5194/nhess-11-431-2011, 2011.

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.

Teutschbein, C., and Seibert, J.: Regional climate models for hydrological impact studies at the catchment scale: a review of recent modeling strategies, Geography Compass, 4, 834-860, 2010.

Tian, Y. D., Peters-Lidard, C. D., Eylander, J. B., Joyce, R. J., Huffman, G. J., Adler, R. F., Hsu, K. L., Turk, F. J., Garcia, M., and Zeng, J.: Component analysis of errors in satellite-based precipitation estimates, J Geophys Res-Atmos, 114, Artn D24101 10.1029/2009jd011949, 2009.

Tian, Y. D., and Peters-Lidard, C. D.: A global map of uncertainties in satellite-based precipitation measurements, Geophys Res Lett, 37, Artn L24407 10.1029/2010gl046008, 2010.

Turk, F. J., Arkin, P., Ebert, E. E., and Sapiano, M. R. P.: Evaluating High-Resolution Precipitation Products, B Am Meteorol Soc, 89, 1911-1916, 10.1175/2008bams2652.1, 2008.

Turk, J. T., Mostovoy, G. V., and Anantharaj, V.: The NRL-Blend High Resolution Precipitation Product and its Application to Land Surface Hydrology, Satellite Rainfall Applications for Surface Hydrology, 85-104, 10.1007/978-90-481-2915-7_6, 2010.

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.

Vila, D. A., de Goncalves, L. G. G., Toll, D. L., and Rozante, J. R.: Statistical Evaluation of Combined Daily Gauge Observations and Rainfall Satellite Estimates over Continental South America, J Hydrometeorol, 10, 533-543, 10.1175/2008jhm1048.1, 2009.

Villarini, G., and Krajewski, W. F.: Review of the Different Sources of Uncertainty in Single Polarization Radar-Based Estimates of Rainfall, Surv Geophys, 31, 107-129, 10.1007/s10712-009-9079-x, 2010.

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.

Wang, X. L. L., and Lin, A.: An algorithm for integrating satellite precipitation estimates with in situ precipitation data on a pentad time scale, J Geophys Res-Atmos, 120, 3728-3744, 10.1002/2014jd022788, 2015.

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.

Werner, A. T., and Cannon, A. J.: Hydrologic extremes - an intercomparison of multiple gridded statistical downscaling methods, Hydrol Earth Syst Sc, 20, 1483-1508, 10.5194/hess-20-1483-2016, 2016.

West, G. L., Steenburgh, W. J., and Cheng, W. Y. Y.: Spurious grid-scale precipitation in the North American regional reanalysis, Monthly Weather Review, 135, 2168-2184, 10.1175/Mwr3375.1, 2007.

Wetterhall, F., Bardossy, A., Chen, D. L., Halldin, S., and Xu, C. Y.: Daily precipitation-downscaling techniques in three Chinese regions, Water Resour Res, 42, Artn W11423 10.1029/2005wr004573, 2006.

Wilby, R. L., and Wigley, T. M. L.: Downscaling general circulation model output: a review of methods and limitations, Prog Phys Geog, 21, 530-548, Doi 10.1177/030913339702100403, 1997.

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.

Xie, P. P., and Arkin, P. A.: An Intercomparison of Gauge Observations and Satellite Estimates of Monthly Precipitation, J Appl Meteorol, 34, 1143-1160, Doi 10.1175/1520-0450(1995)034<1143:Aiogoa>2.0.Co;2, 1995.

Xie, P. P., and Arkin, P. A.: Global monthly precipitation: An intercomparison of several datasets based on gauge observations, satellite estimates and model predictions, Eighth Conference on Satellite Meteorology and Oceanography, 225-229, 1996.

Xie, P. P., and Arkin, P. A.: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs, B Am Meteorol Soc, 78, 2539-2558, Doi 10.1175/1520-0477(1997)078<2539:Gpayma>2.0.Co;2, 1997.

Xu, C. Y., Widen, E., and Halldin, S.: Modelling hydrological consequences of climate change - Progress and challenges, Adv Atmos Sci, 22, 789-797, Doi 10.1007/Bf02918679, 2005.

Yang, D. Q., Goodison, B. E., Metcalfe, J. R., Golubev, V. S., Bates, R., Pangburn, T., and Hanson, C. L.: Accuracy of NWS 8" standard nonrecording precipitation gauge: Results and application of WMO intercomparison, J Atmos Ocean Tech, 15, 54-68, Doi 10.1175/1520-0426(1998)015<0054:Aonsnp>2.0.Co;2, 1998.

Yong, B., Liu, D., Gourley, J. J., Tian, Y. D., Huffman, G. J., Ren, L. L., and Hong, Y.: GLOBAL VIEW OF REAL-TIME TRMM MULTISATELLITE PRECIPITATION ANALYSIS Implications for Its Successor Global Precipitation Measurement Mission, B Am Meteorol Soc, 96, 283-296, 10.1175/Bams-D-14-00017.1, 2015.

Young, C. B., Nelson, B. R., Bradley, A. A., Smith, J. A., Peters-Lidard, C. D., Kruger, A., and Baeck, M. L.: An evaluation of NEXRAD precipitation estimates in complex terrain, J Geophys Res-Atmos, 104, 19691-19703, Doi 10.1029/1999jd900123, 1999.

Zhang, Q., Sun, P., Singh, V. P., and Chen, X. H.: Spatial-temporal precipitation changes (1956-2000) and their implications for agriculture in China, Global Planet Change, 82-83, 86-95, 10.1016/j.gloplacha.2011.12.001, 2012.

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.

List of Tables

Table 1 ~~A summary of different types of precipitation~~Precipitation products used in this ~~comparison~~ study.

Dataset	Full Name	Type	Spatial Resolution	Temporal Resolution	Duration	Coverage	Reference
ANUSPLIN	Australian National University Spline	Station-based Interpolated	300 arc-second (~0.0833°/ ~10 km)	24 hr	1950 – 2013	Canada	Hutchinson et al. (2009)
CaPA	Canadian Precipitation Analysis	Station-based Model-derived	10 km (~0.0833°)	6 hr	2002 – 2014	North America	Mahfouf et al. (2007)
Princeton	Global dataset at the Princeton University	Reanalysis-based multiple source	0.5° (~50 km)	3 hr	1901 – 2012	Global	Sheffield et al. (2006)
WFDEI [CRU]	Water and Global Change Forcing Data methodology applied to ERA-Interim [Climate Research Unit]	Reanalysis-based multiple source	0.5° (~50 km)	3 hr	1979 – 2012	Global	Weedon et al. (2014)
WFDEI [GPCC]	Water and Global Change Forcing Data methodology applied to ERA-Interim [Global Precipitation Climatology Centre]	Reanalysis-based multiple source	0.5° (~50 km)	3 hr	1979 – 2012	Global	Weedon et al. (2014)
NARR	North American Regional Reanalysis	Reanalysis-based multiple source	32 km (0.3°)	3 hr	1979 – 2015	North America	Mesinger et al. (2006)
PCIC	Pacific Climate Impacts Consortium	Station-driven GCM	300 arc-second (~0.0833°/ ~10 km)	24 hr	Historical: 1950 – 2005 Projected: 2006 – 2100	Canada	Pacific Climate Impacts Consortium; University of Victoria (Jan 2014)
NA-CORDEX	North America COordinated Regional climate Downscaling EXperiment	GCM-driven RCM	0.22° (25 km)	3 hr	Historical: 1950 – 2005 Projected: 2006 – 2100	North America	Giorgi et al. (2009)

Table 2 A ~~summary of the~~ GCMs chosen in the Pacific Climate Impacts Consortium (PCIC) dataset.

~~PCIC~~	~~Full Name~~	~~Country~~	~~Statistical Downscaling Method~~	~~Representative Concentration Pathway (RCP)~~
~~GFDL-ESM2G_BCCAQ_RCP85~~	~~Geophysical Fluid Dynamics Laboratory Earth System Model 2G~~	~~USA~~	~~Bias Correction Constructed Analogues with Quantile mapping reordering~~	~~8.5~~
~~GFDL-ESM2G_BCSD_RCP85~~				~~8.5~~
~~HadGEM2-ES_BCCAQ_RCP85~~	~~Hadley Global Environmental Model 2 – Earth System~~	~~UK~~	~~Bias Correction Constructed Analogues with Quantile mapping reordering~~	~~8.5~~
~~HadGEM2-ES_BCSD_RCP85~~				~~8.5~~
~~CanESM2_BCCAQ_RCP45~~	~~Second generation Canadian Earth System Model~~	~~Canada~~	~~Bias Correction Constructed Analogues with Quantile mapping reordering~~	~~4.5~~
~~CanESM2_BCCAQ_RCP85~~				~~8.5~~
~~CanESM2_BCSD_RCP45~~			~~Bias Correction Spatial Disaggregation~~	~~4.5~~
~~CanESM2_BCSD_RCP85~~				~~8.5~~
~~MPI-ESM-LR_ BCCAQ_RCP45~~	~~Max-Planck-Institute Earth System Model running on low resolution~~	~~Germany~~	~~Bias Correction Constructed Analogues with Quantile mapping reordering~~	~~4.5~~
~~MPI-ESM-LR_ BCCAQ_RCP85~~				~~8.5~~
~~MPI-ESM-LR_ BCSD_RCP45~~			~~Bias Correction Spatial Disaggregation~~	~~4.5~~
~~MPI-ESM-LR_ BCSD_RCP85~~				~~8.5~~

PCIC	Full Name	Country	Statistical Downscaling Method
GFDL-ESM2G_BCCAQ	Geophysical Fluid Dynamics Laboratory Earth System Model 2G	USA	Bias Correction Constructed Analogues with Quantile mapping reordering
GFDL-ESM2G_BCSD			Bias Correction Spatial Disaggregation
HadGEM2-ES_BCCAQ	Hadley Global Environmental Model 2 – Earth System	UK	Bias Correction Constructed Analogues with Quantile mapping reordering
HadGEM2-ES_BCSD			Bias Correction Spatial Disaggregation
CanESM2_BCCAQ	Second generation Canadian Earth System Model	Canada	Bias Correction Constructed Analogues with Quantile mapping reordering
CanESM2_BCSD			Bias Correction Spatial Disaggregation
MPI-ESM-LR_ BCCAQ	Max-Planck-Institute Earth System Model running on low resolution	Germany	Bias Correction Constructed Analogues with Quantile mapping reordering
MPI-ESM-LR_ BCSD			Bias Correction Spatial Disaggregation

Table 3 ~~A summary of the~~ GCMs-RCMs chosen in the North America COordinated Regional climate Downscaling EXperiment (NA-CORDEX) dataset.

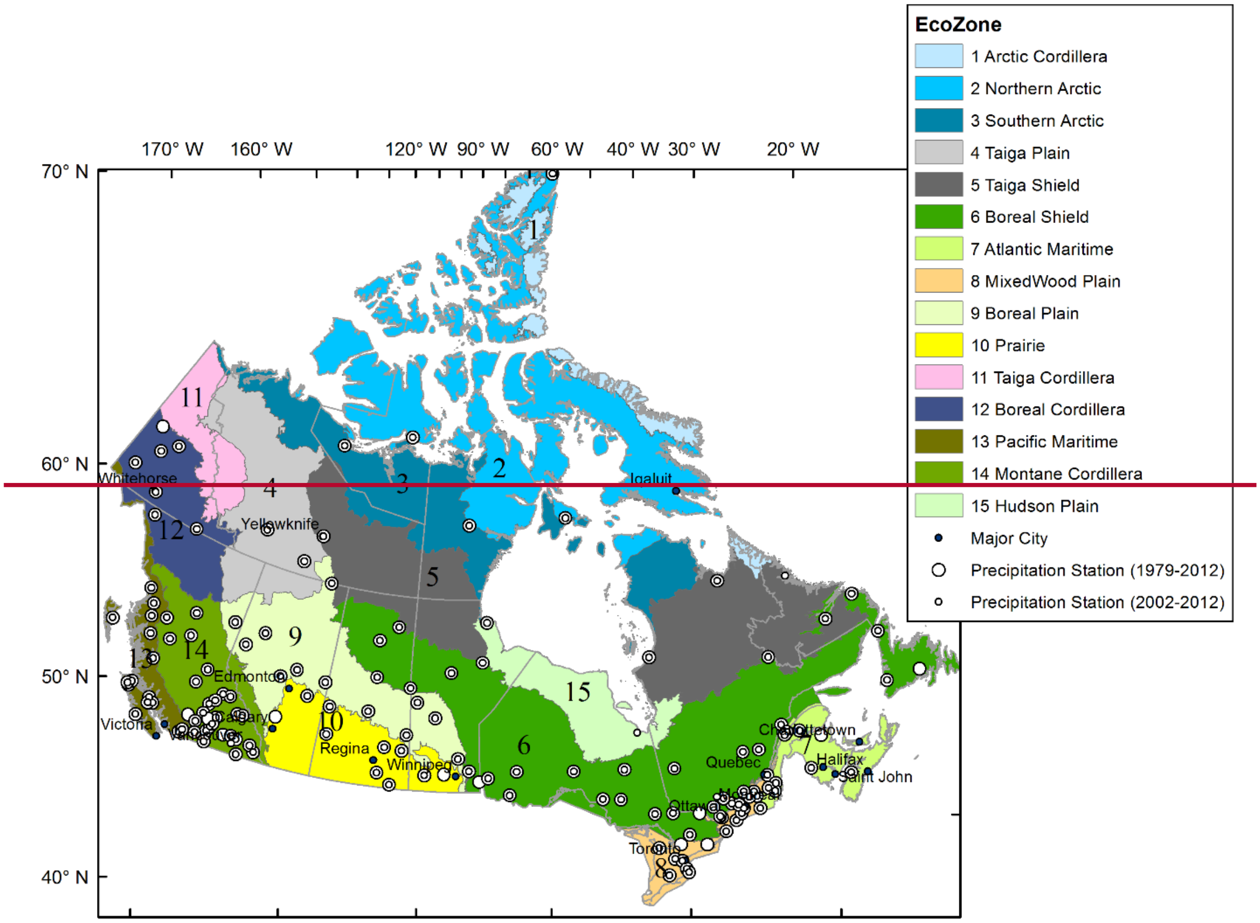
~~NA-CORDEX~~	~~Full Name~~		~~Representative Concentration Pathway (RCP)~~
	~~Global Circulation Model (GCM)~~	~~Regional Climate Model (RCM)~~	
~~CanESM2 – CanRCM4_RCP45~~	~~Second generation Canadian Earth System Model~~	~~Fourth generation Canadian Regional Climate Model~~	~~4.5~~
~~CanESM2 – CanRCM4_RCP85~~			~~8.5~~
~~CanESM2 – CRCM5_UQAM_RCP45~~		~~Fifth generation Canadian Regional Climate Model~~	~~4.5~~
~~MPI-ESM-LR – CRCM5_UQAM_RCP45~~	~~Max-Planck-Institute Earth System Model running on low resolution~~		~~4.5~~
~~MPI-ESM-LR – RegCM4_RCP85~~		~~Fourth generation Regional Climate Model~~	~~8.5~~

NA-CORDEX	Full Name	
	Global Circulation Model (GCM)	Regional Climate Model (RCM)
CanESM2 – CanRCM4	Second generation Canadian Earth System Model	Fourth generation Canadian Regional Climate Model
CanESM2 – CRCM5_UQAM		Fifth generation Canadian Regional Climate Model
MPI-ESM-LR – CRCM5_UQAM	Max-Planck-Institute Earth System Model running on low resolution	
MPI-ESM-LR – RegCM4		Fourth generation Regional Climate Model

Table 4 Number~~s~~ of precipitation-gauge stations within each Ecozone.

Region (Ecozone)		Number of Precipitation-gauge Stations	
		1979 – 2012	2002 – 2012
1	Arctic Cordillera	0	0
2	Northern Arctic	4	4
3	Southern Arctic	1	1
4	Taiga Plain	2	2
5	Taiga Shield	4	5
6	Boreal Shield	31	29
7	Atlantic Maritime	10	9
8	Mixedwood Plain	18	16
9	Boreal Plain	14	14
10	Prairie	9	7
11	Taiga Cordillera	1	0
12	Boreal Cordillera	6	6
13	Pacific Maritime	15	15
14	Montane Cordillera	28	26
15	Hudson Plain	2	3
Total		145	137

List of Figures



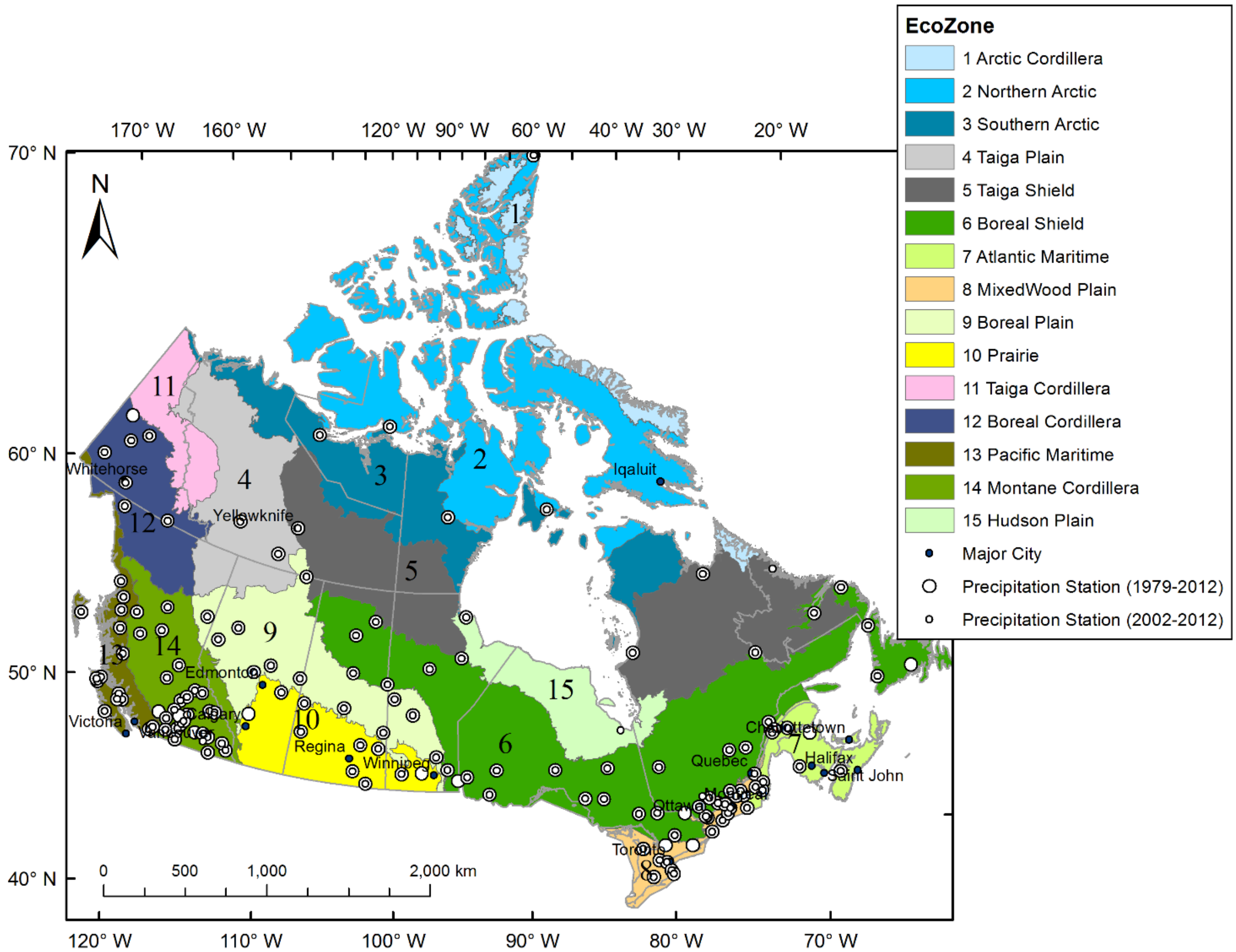
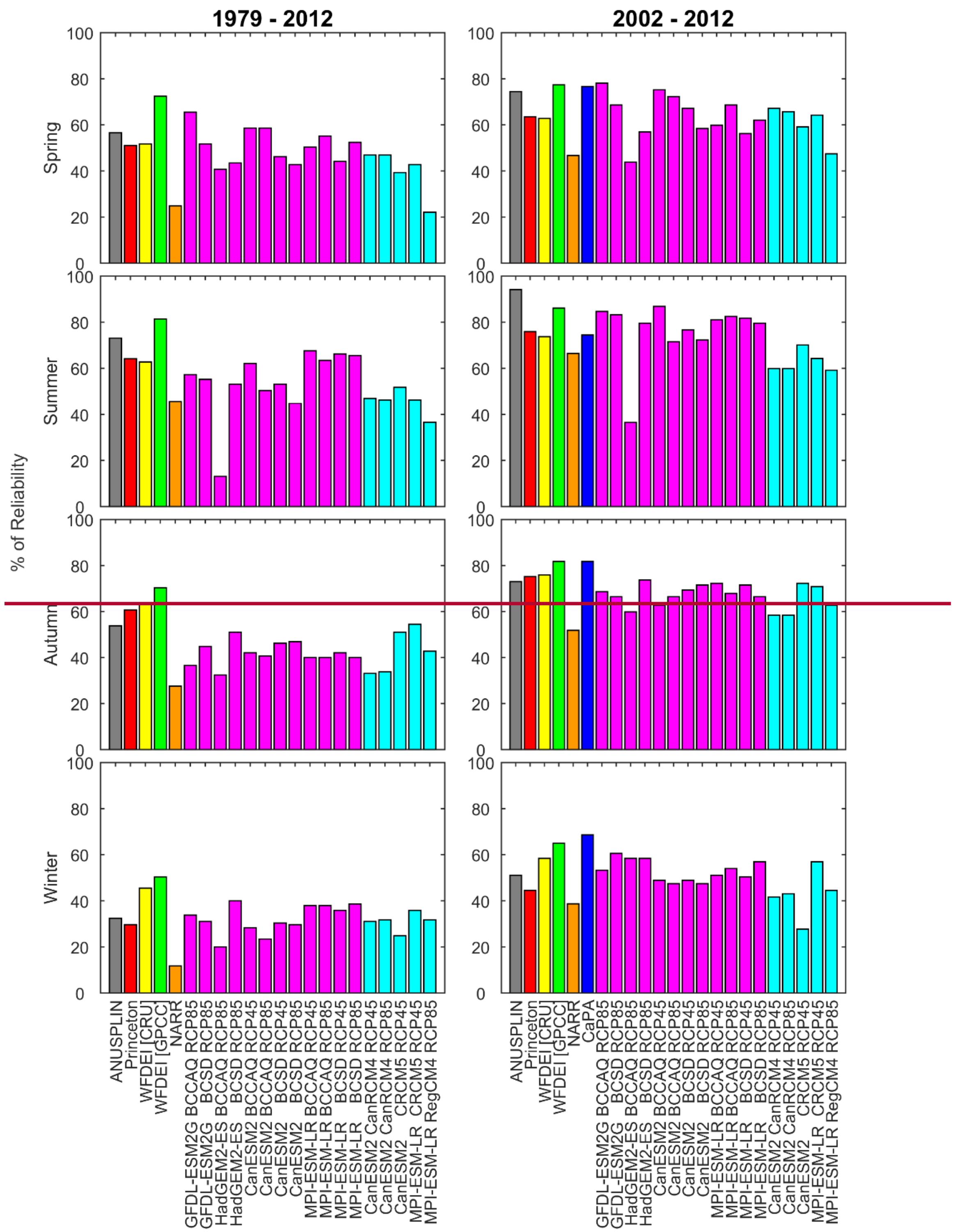


Figure 1. 15 terrestrial ecozones of Canada with numerical codes indicating Region from 1 Arctic Cordillera to 15 Hudson Plain. Big (a total of 145) and small (a total of 137) white dots are the extracted precipitation-gauge stations from the Canadian adjusted and homogenized precipitation datasets of Mekis and Vincent (2011) for the period of 1979 to 2012 and 2002 to 2012 respectively. Black dots are major cities in Canada.


1979 - 2012
2002 - 2012
Spring
Summer
Autumn
Winter
% of Reliability
ANUSPLIN
Princeton
WFDEI [CRU]
WFDEI [GPCC]
NARR
CaPA
GFDL-ESM2G BCCAQ RCP85
GFDL-ESM2G BCSD RCP85
HadGEM2-ES BCCAQ RCP85
HadGEM2-ES BCSD RCP85
CanESM2 BCCAQ RCP45
CanESM2 BCCAQ RCP85
CanESM2 BCSD RCP45
CanESM2 BCSD RCP85
MPI-ESM-LR BCCAQ RCP45
MPI-ESM-LR BCCAQ RCP85
MPI-ESM-LR BCSD RCP45
MPI-ESM-LR BCSD RCP85
CanESM2 CanRCM4 RCP45
CanESM2 CanRCM4 RCP85
CanESM2 CRCM5 RCP45
MPI-ESM-LR CRCM5 RCP45
MPI-ESM-LR RegCM4 RCP85

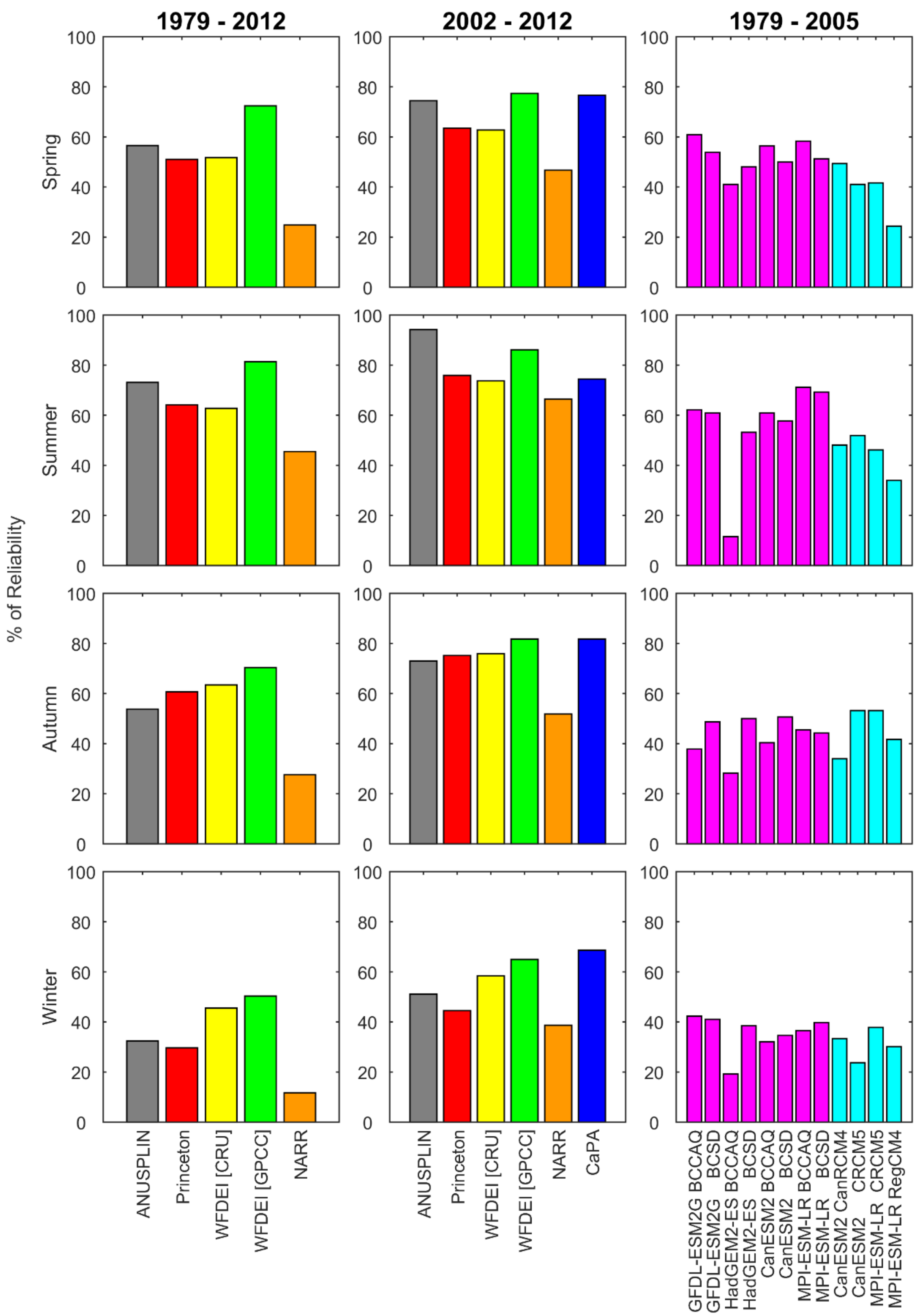
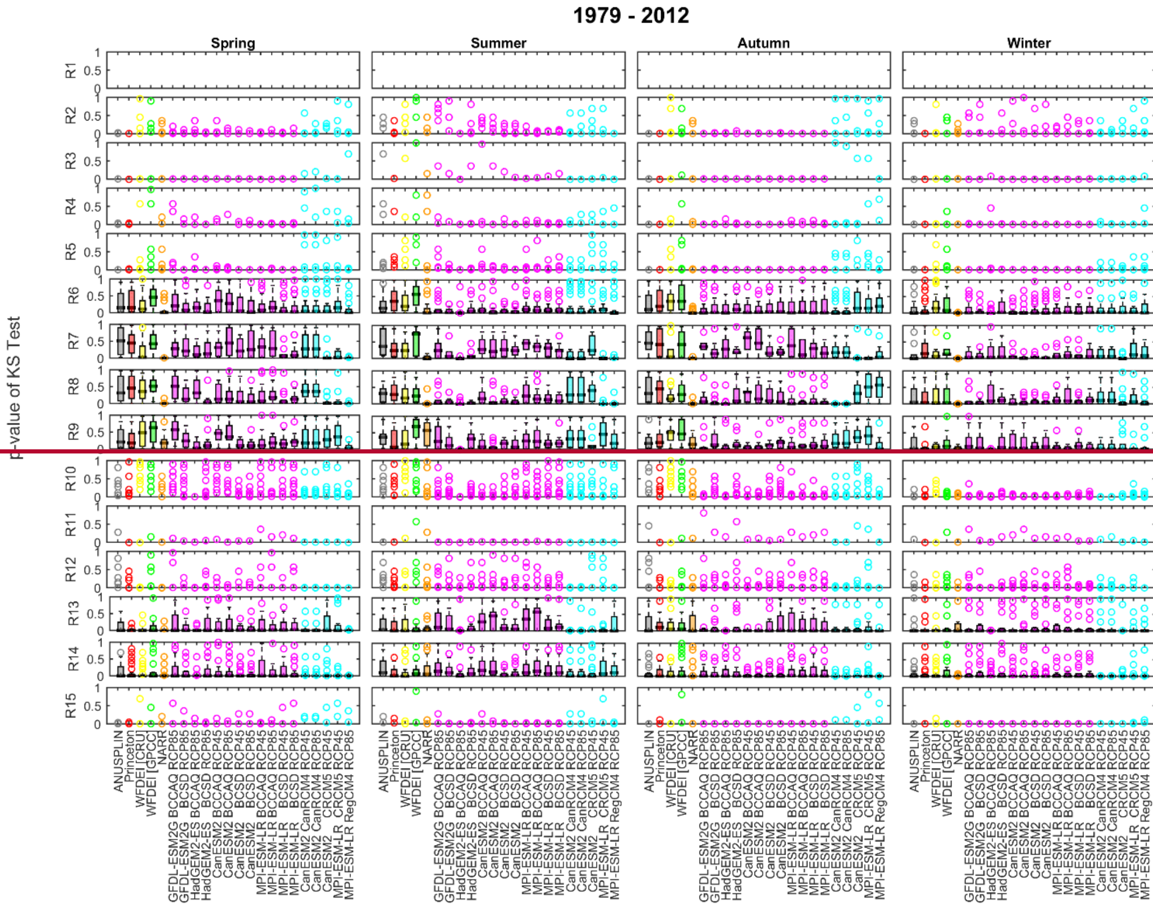


Figure 2. The percentage of reliability, calculated by the Eq. (1), of each precipitation dataset in four seasons for the period of 1979 to 2012 (left panel) ~~and~~, 2002 to 2012 (~~right~~ middle panel), and 1979 to 2005 (right panel) across Canada. The higher the percentage, the more reliable the 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.



1979 - 2012
Spring
Summer
Autumn
Winter
p-value of KS Test
R1
R2
R3
R4
R5
R6
R7
R8
R9
R10
R11
R12
R13
R14
R15
ANUSPLIN
Princeton
WFDEI [CRU]
WFDEI [GPCC]
NARR
GFDL-ESM2G BCCAQ RCP85
GFDL-ESM2G BCSD RCP85
HadGEM2-ES BCCAQ RCP85
HadGEM2-ES BCSD RCP85
CanESM2 BCCAQ RCP45
CanESM2 BCCAQ RCP85
CanESM2 BCSD RCP45
CanESM2 BCSD RCP85
MPI-ESM-LR BCCAQ RCP45
MPI-ESM-LR BCCAQ RCP85
MPI-ESM-LR BCSD RCP45
MPI-ESM-LR BCSD RCP85
CanESM2 CanRCM4 RCP45
CanESM2 CanRCM4 RCP85
CanESM2 CRCM5 RCP45
MPI-ESM-LR CRCM5 RCP45
MPI-ESM-LR RegCM4 RCP85

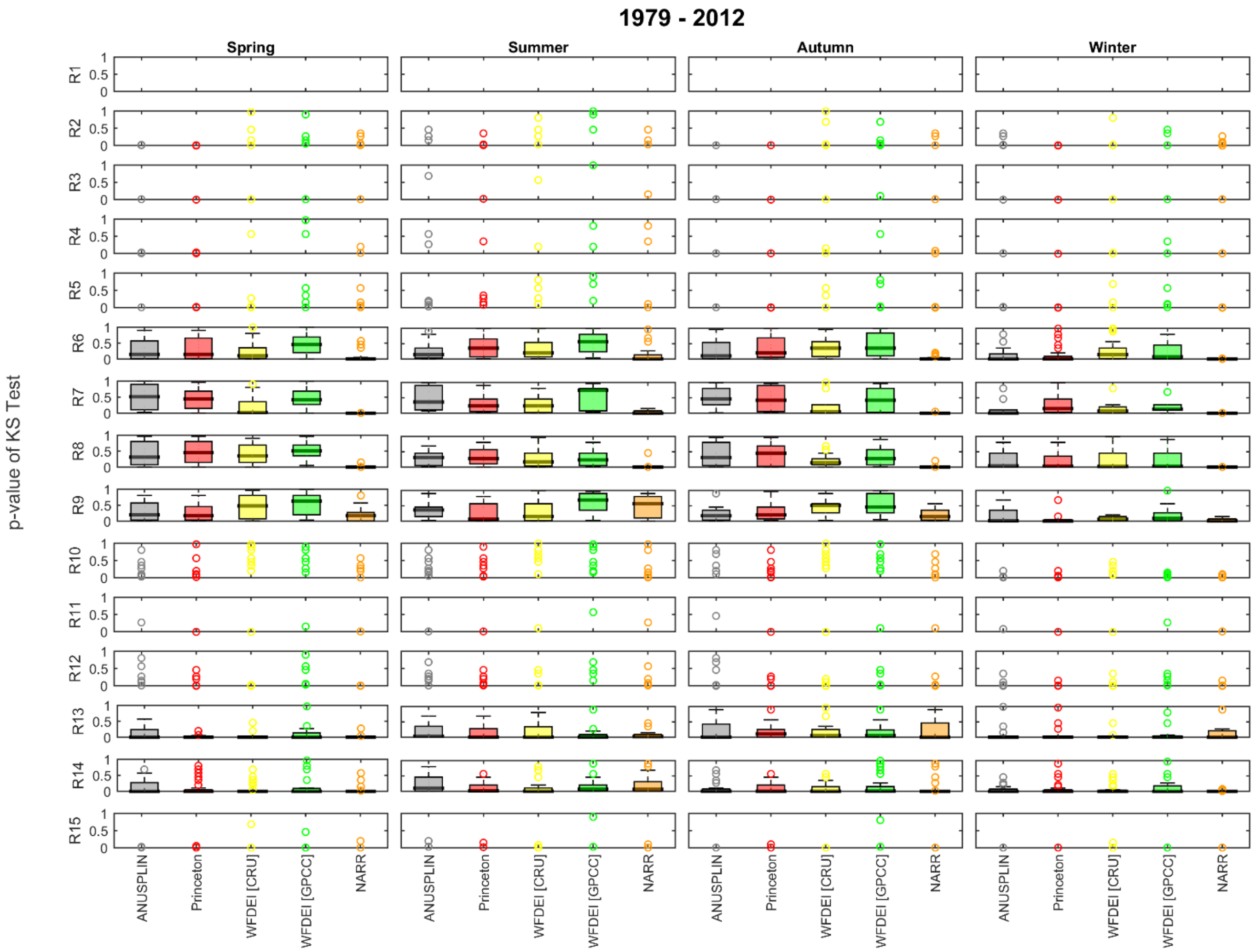
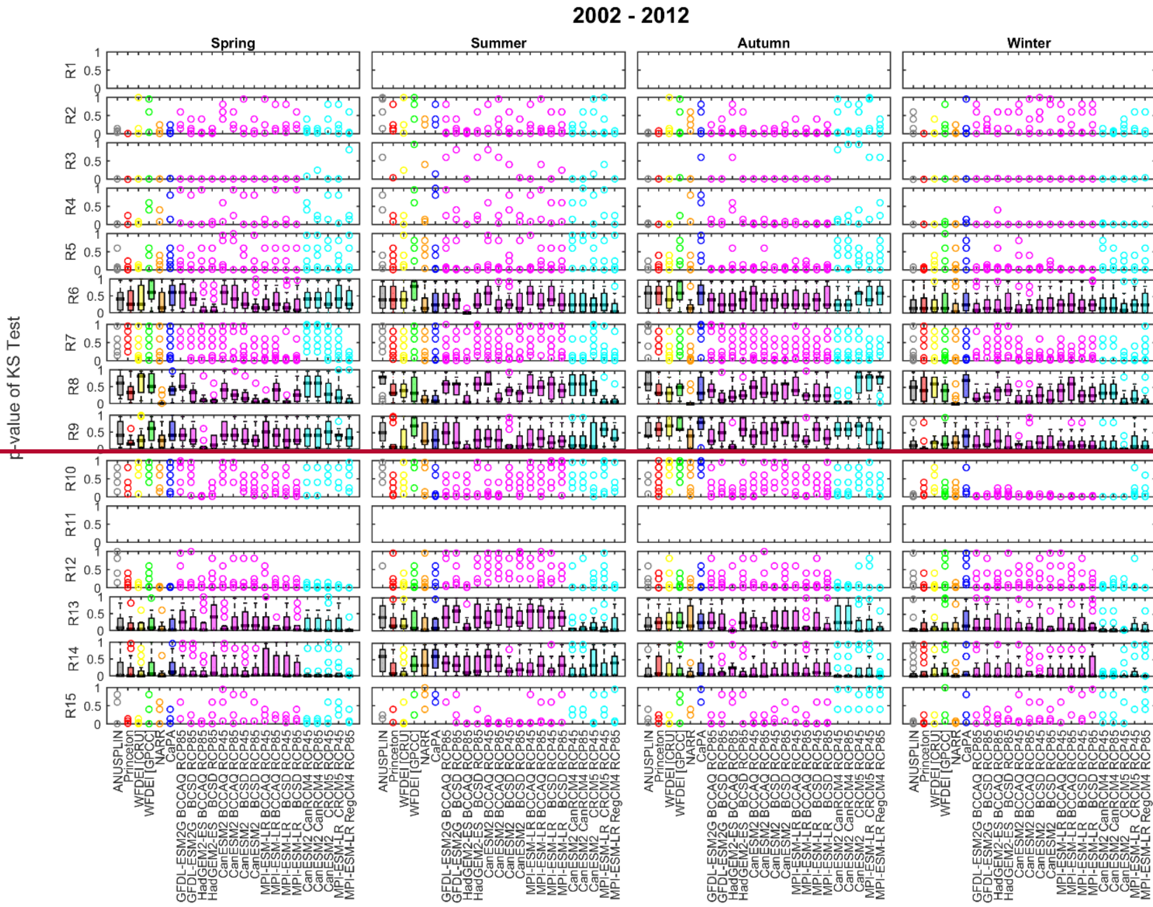


Figure 3. Distributions of p-value of the K-S test in the 15 ecozones in four seasons for the period of 1979 to 2012 (long-term comparison without CaPA). Note that the numbers of precipitation-gauge stations in each ecozone are different (see Table 4). Each hollow circle represents one p-value of the 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.


2002 - 2012
Spring
Summer
Autumn
Winter
p-value of KS Test
R1
R2
R3
R4
R5
R6
R7
R8
R9
R10
R11
R12
R13
R14
R15
ANUSPLIN
Princeton
WFDEI [CRU]
WFDEI [GPCC]
NARR
CaPA
GFDL-ESM2G BCCAQ RCP85
GFDL-ESM2G BCSD RCP85
HadGEM2-ES BCCAQ RCP85
HadGEM2-ES BCSD RCP85
CanESM2 BCCAQ RCP45
CanESM2 BCCAQ RCP85
CanESM2 BCSD RCP45
CanESM2 BCSD RCP85
MPI-ESM-LR BCCAQ RCP45
MPI-ESM-LR BCCAQ RCP85
MPI-ESM-LR BCSD RCP45
MPI-ESM-LR BCSD RCP85
CanESM2 CanRCM4 RCP45
CanESM2 CanRCM4 RCP85
CanESM2 CRCM5 RCP45
MPI-ESM-LR CRCM5 RCP45
MPI-ESM-LR RegCM4 RCP85

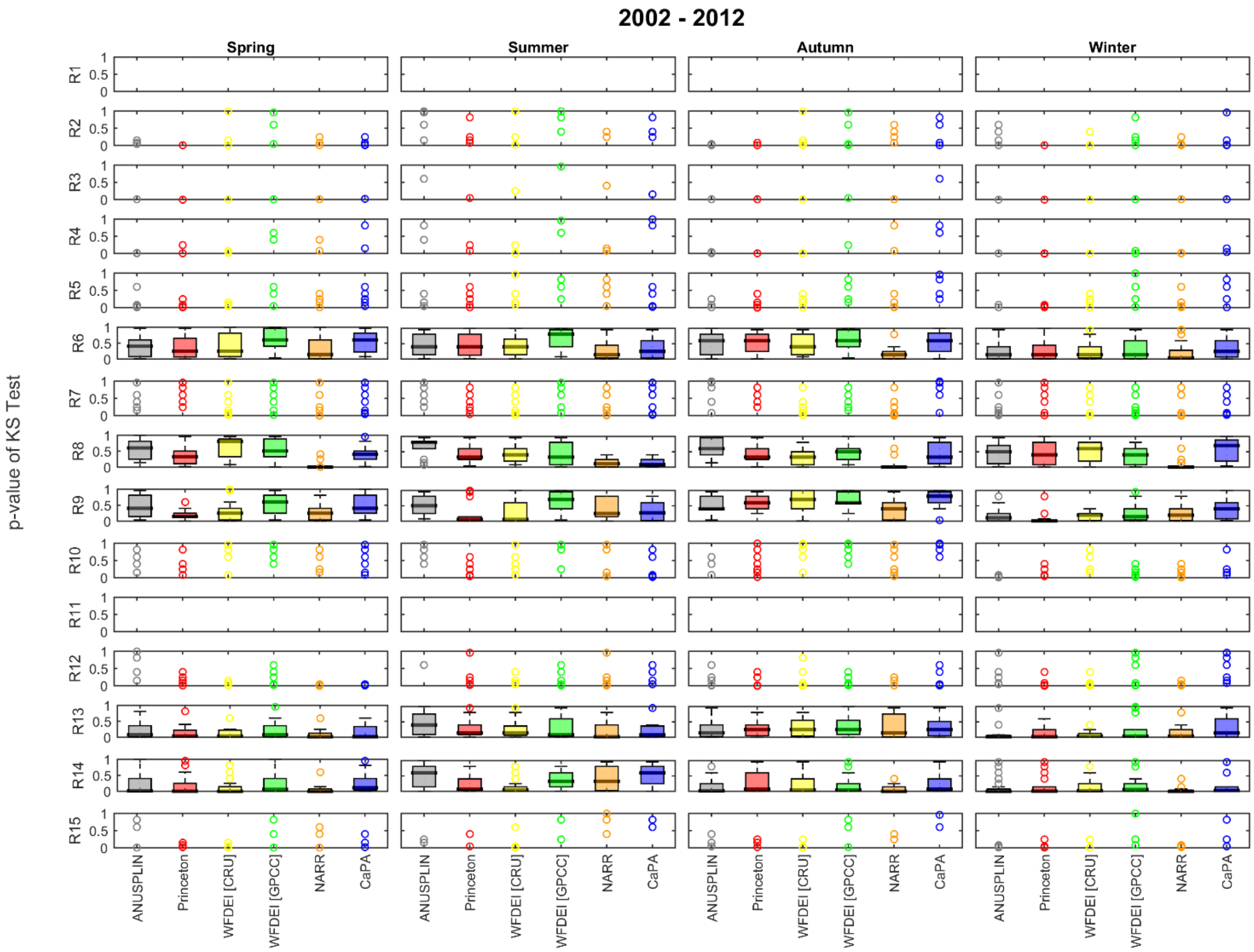


Figure 4. Distributions of p-value of the K-S test in the 15 ecozones in four seasons for the period of 2002 to 2012 (short-term comparison with the inclusion of CaPA). Note that the numbers of precipitation-gauge stations in each ecozone are different (see Table 4). Each hollow circle represents one 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.

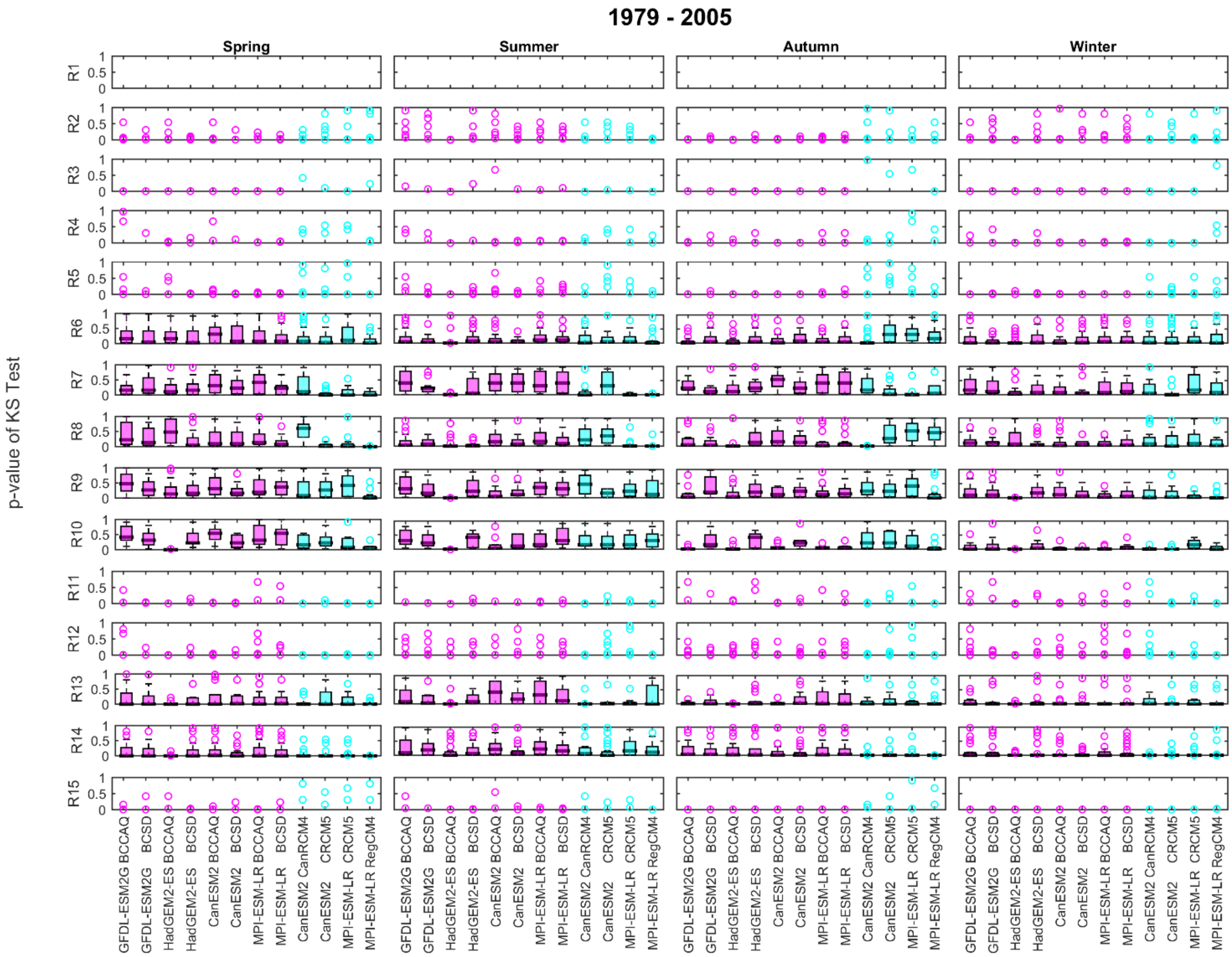


Figure 5. Distributions of p-value of the K-S test in the 15 ecozones in four seasons for the period of 1979 to 2005 (long-term comparison of PCIC and NA-CORDEX). Note that the numbers of precipitation-gauge stations in each ecozone are different (see Table 4). Each hollow circle represents one p-value 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.

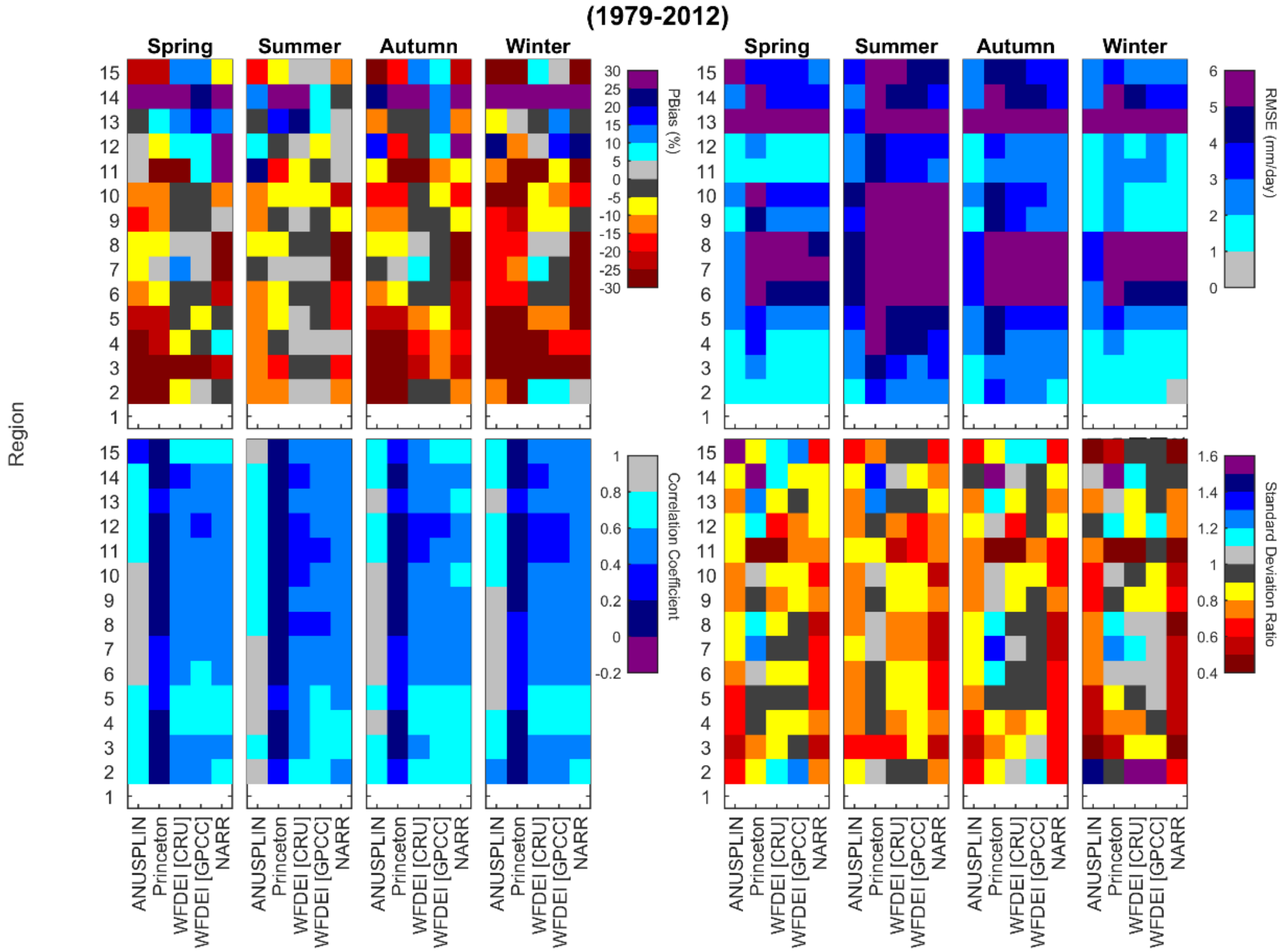


Figure 6. Portrait diagram showing the accuracy ($PBias$) (top left), magnitude of the errors ($RMSE$) (top right), strength and direction of relationship between gridded products and precipitation-gauge stations (r) (bottom left), and amplitude of the variations (σ_G/σ_R) (bottom right) of each type of gridded precipitaiton products when evaluating against the precipitation-gauge station data in each ecozone (Region 1 to 15) in four seasons for the time period of 1979 to 2012. Each column indicates one gridded precipitation product and each row represents one ecozone with numerical code corresponding to region shown in Fig. 1. White indicates that no data are available due to no precipitation-gauge stations exisiting in that region.

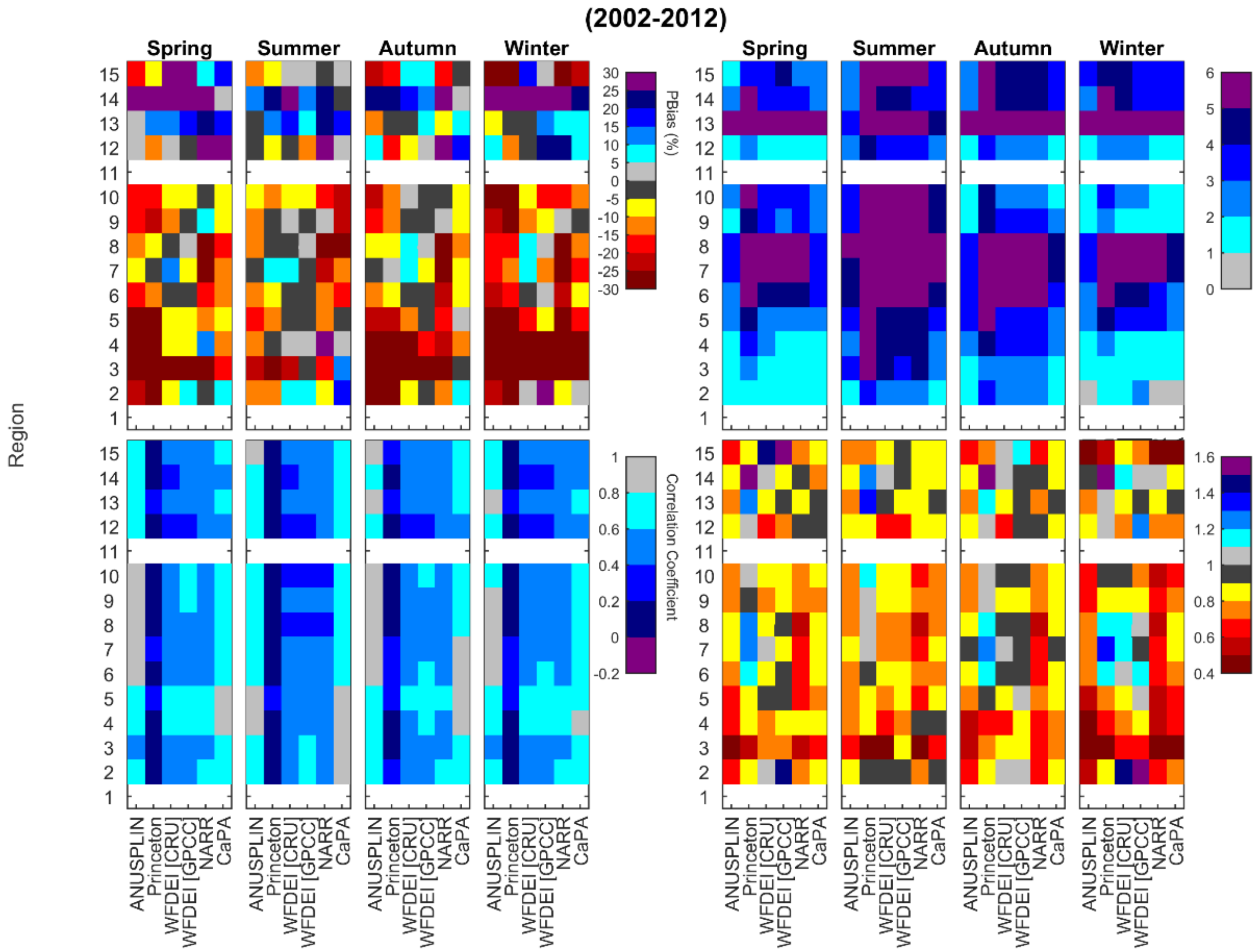


Figure 7. Portrait diagram showing the accuracy ($PBias$) (top left), magnitude of the errors ($RMSE$) (top right), strength and direction of relationship between gridded products and precipitation-gauge stations (r) (bottom left), and amplitude of the variations (σ_G/σ_R) (bottom right) of each type of gridded precipitaiton products when evaluating against the precipitation-gauge station data in each ecozone (Region 1 to 15) in four seasons for the time period of 2002 to 2012. Each column indicates one gridded precipitation product and each row represents one ecozone with numerical code corresponding to region shown in Fig. 1. White indicates that no data are available due to no precipitation-gauge stations exisiting in that region.

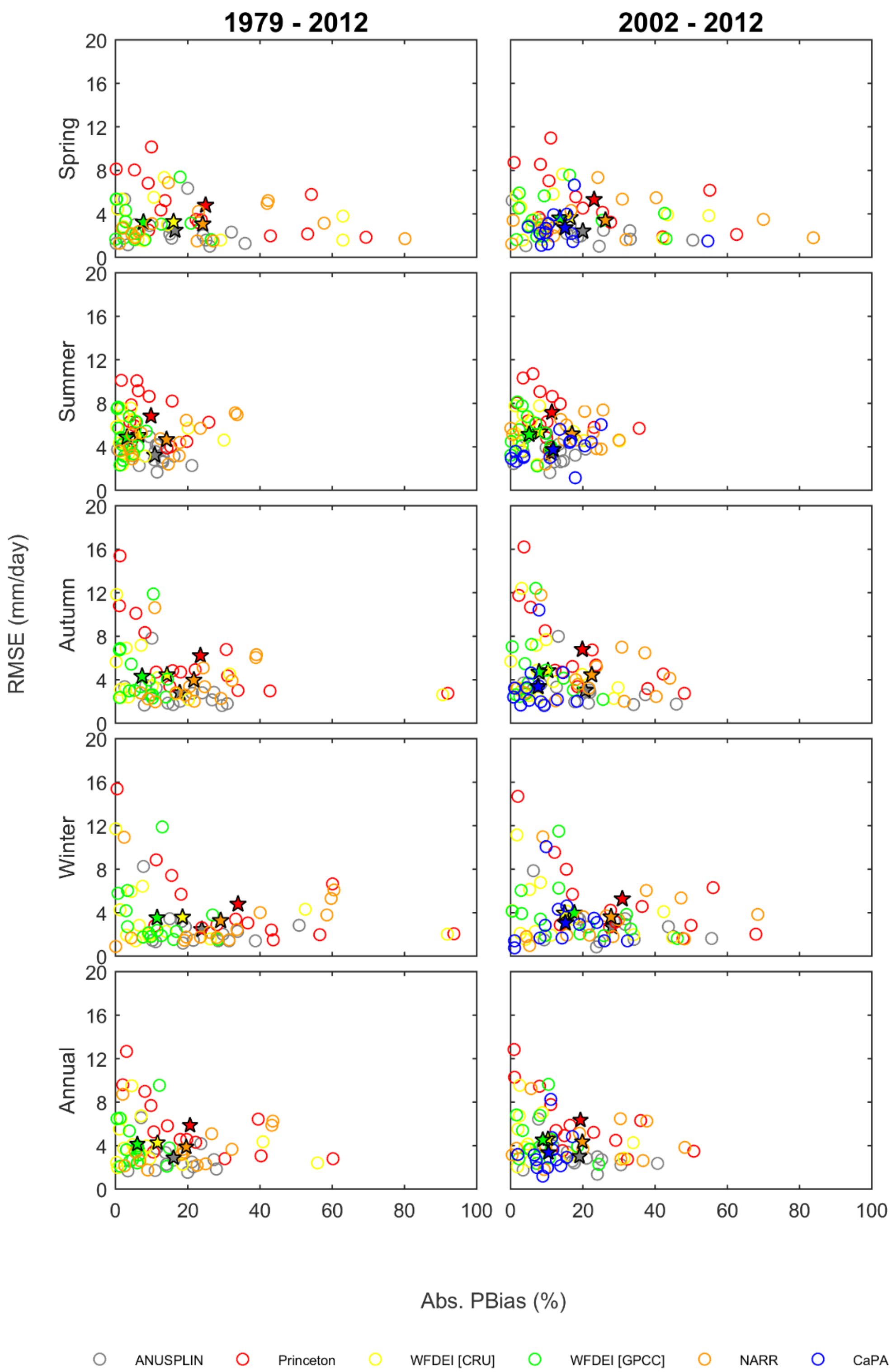


Figure 8. Scatter plots showing absolute $PBias$ (x-axis) versus $RMSE$ (y-axis) of each precipitation dataset in four seasons and the entire year for the period of 1979 to 2012 (left panel) and 2002 to 2012 (right panel). Each hollow circle represents one ecozone and the solid stars indicate the overall average across ecozones.

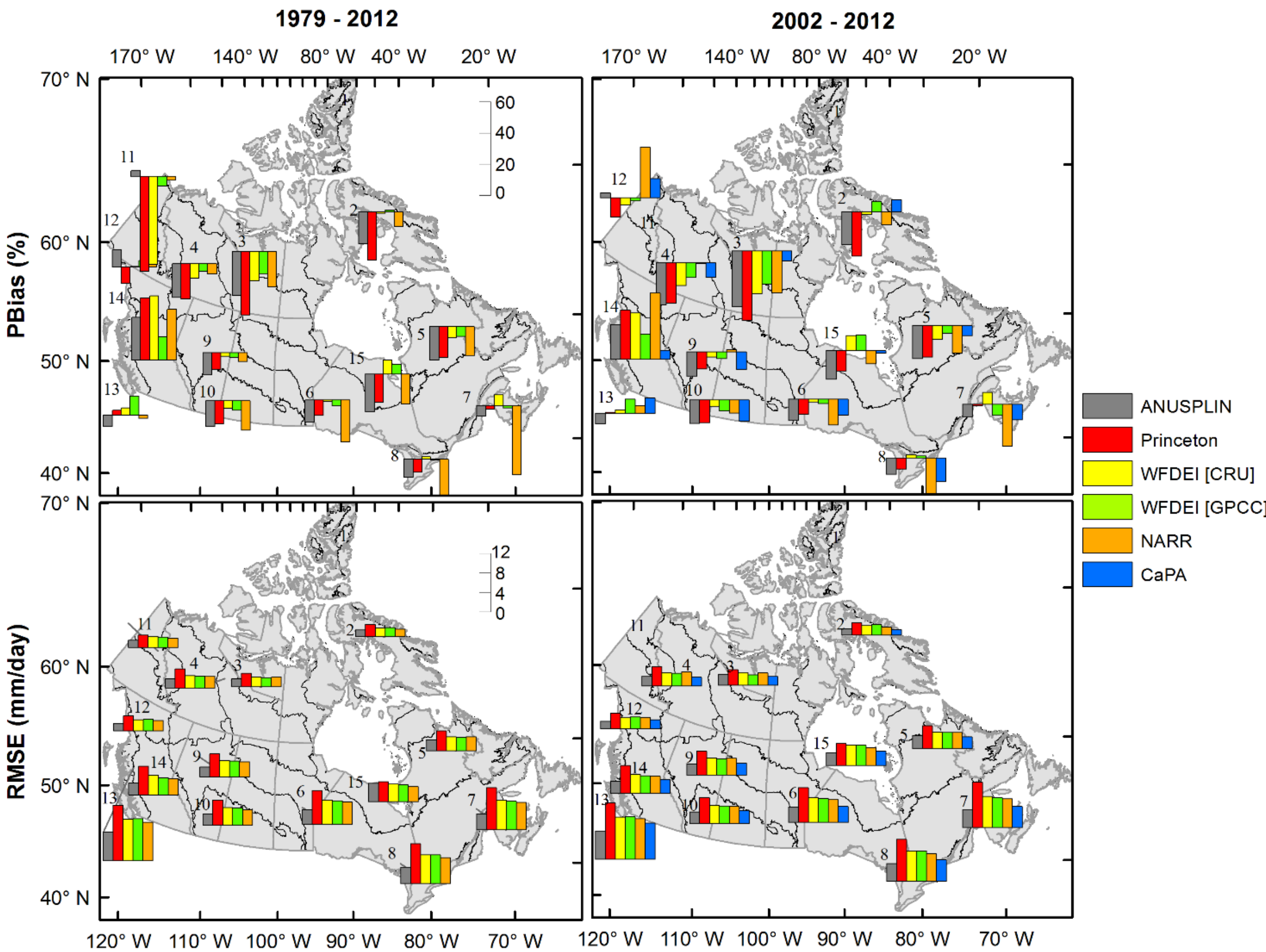


Figure 9. Bar graphs showing the annual accuracy ($PBias$) (first row) and magnitude of the errors ($RMSE$) (second row) of each precipitation dataset for the period of 1979 to 2012 (left panel) and 2002 to 2012 (right panel) in different ecozones. The white bar shows the scale of the bars with number beside it indicating the value of the bar.