Responses to Reviewer 2 comments on Manuscript HESS-2016-511

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

Authors: Jefferson Wong et al

Manuscript No: hess-2016-511

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

Response to Reviewer 2

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 will be reduced and the spatiotemporal scales of evaluation will be included in the revised abstract. The following shows the revised abstract, with deleted materials being crossed out by drawing a line through them (and revised sentences being coloured in red):*

>> *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 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 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~~. Results showed that most of the products were relatively skillful in central Canada. However, they tended to overestimate precipitation amounts on the west coast and underestimate on the east and especially 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 from 2002. All the products 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 from this study can be used as a guidance for 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 will be addressed and the GPM mission will be mentioned in the revised manuscript. Accordingly, the corresponding references will also be added. The following shows the revised discussion of satellite-based estimates in the original manuscript [P4:L7-14], 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 will address this comment by providing better explanation of the calculation and revising the wordings in the equation for better clarity in the revised manuscript. The following shows the revised Sections 4.1, with deleted materials being crossed out by drawing a line through them and revised sentences being coloured in red:

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 and the ground observations. ~~at 5 % significance level (α = 0.05) to support the~~ The null hypothesis (H_0) for this test is that the two datasets came from same population. Monthly total precipitation data were used and aggregated for each season because the existence of numerous zero values in the daily precipitation data might 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 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), which is shown by Eq. (1).

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

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

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 most 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 [P27:L12-14] of the original manuscript 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) [P22:L23] 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. We will enlarge the figures as much as possible and provide 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. The revised figures are shown as follows:*

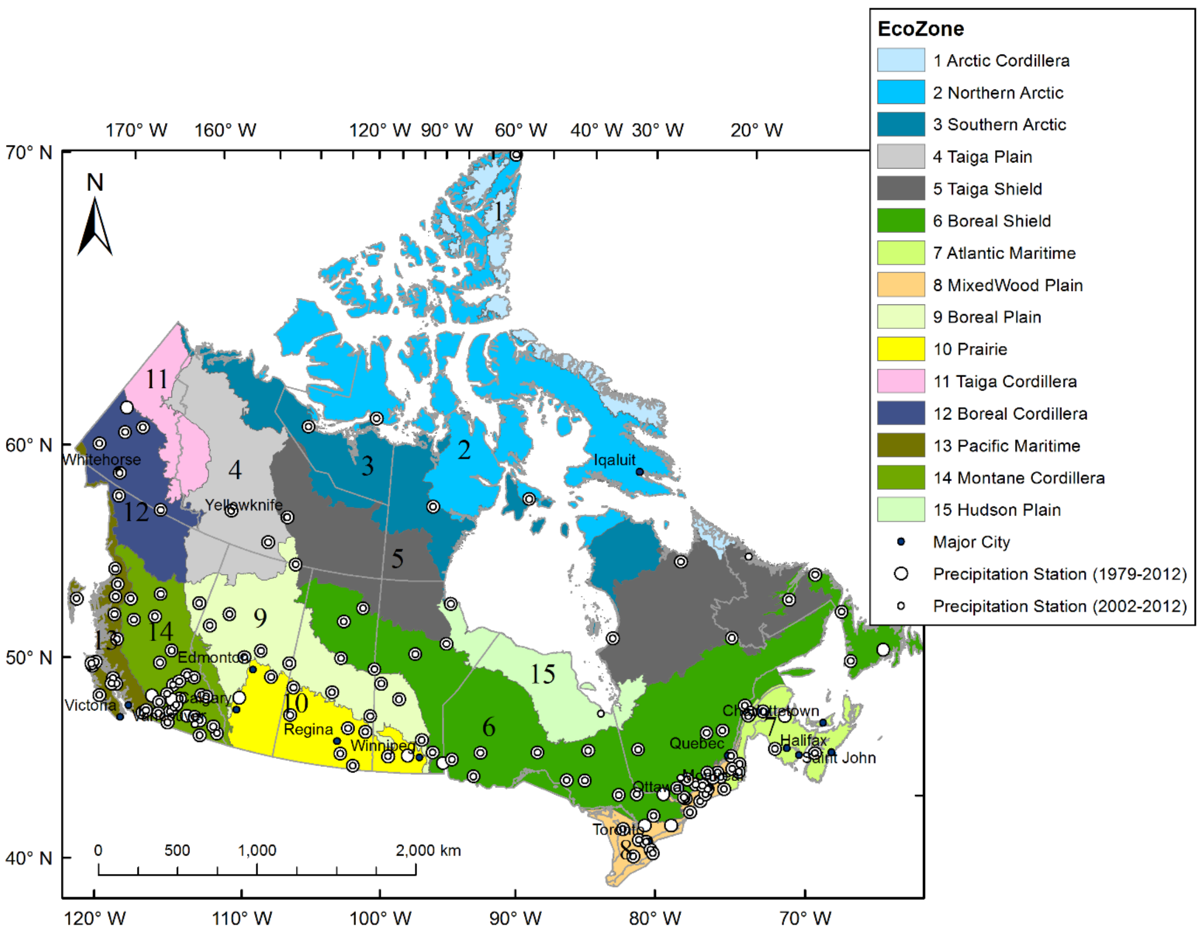

Figure 1. 15 terrestrial ecozones of Canada with numerical codes indicating Region from 1 Arctic Cordillera to 15 Hudson Plain. Big (a total of 145) and small (a total of 137) white dots are the extracted precipitation-gauge stations from the Canadian adjusted and homogenized precipitation datasets of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

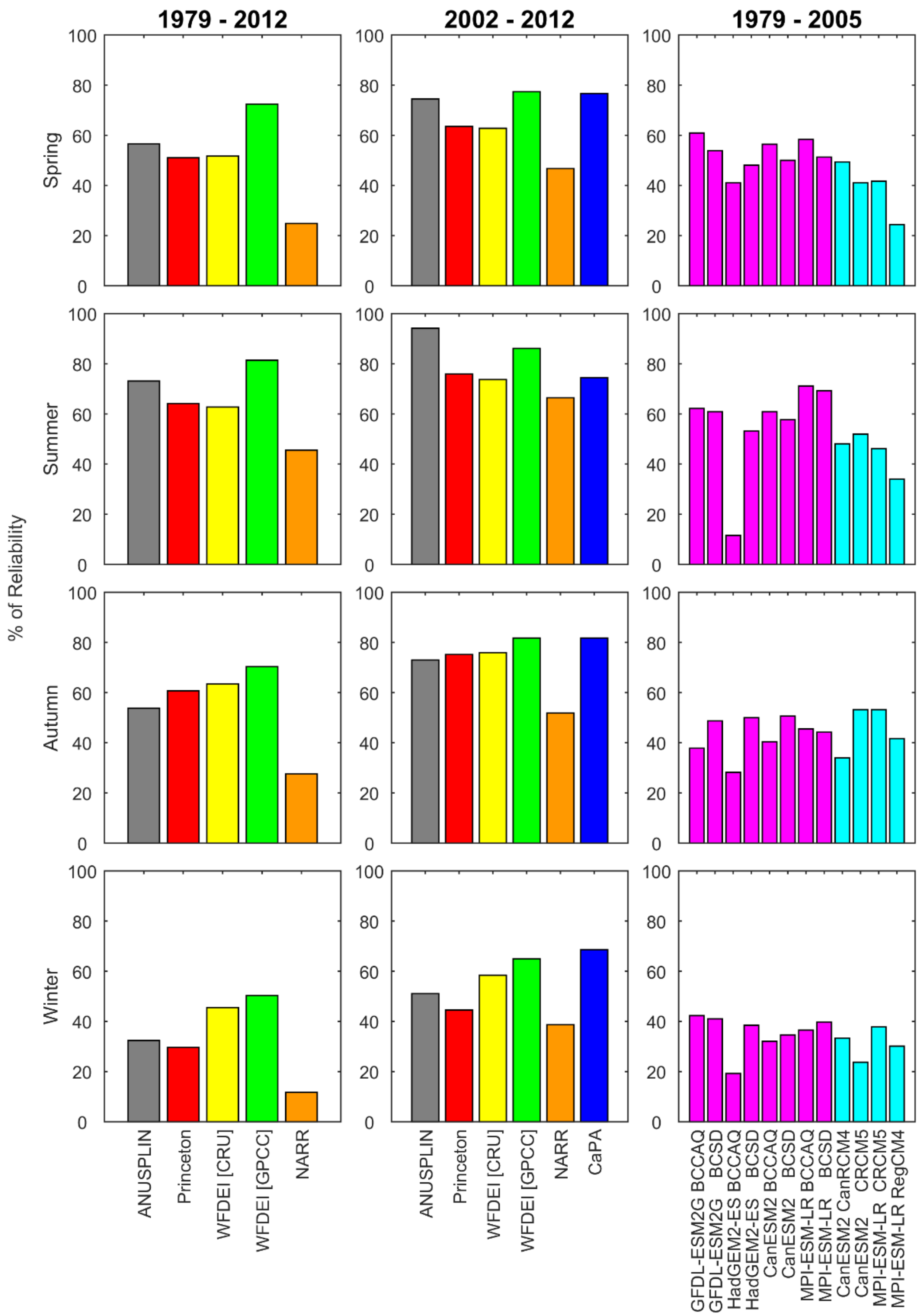

Figure 2. The percentage of reliability, calculated by the Eq. (1), of each precipitation dataset in four seasons for the period of 1979 to 2012 (left panel) and 2002 to 2012 (right panel) across Canada. The higher the percentage, the more reliable the precipitation dataset. Different colours represent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

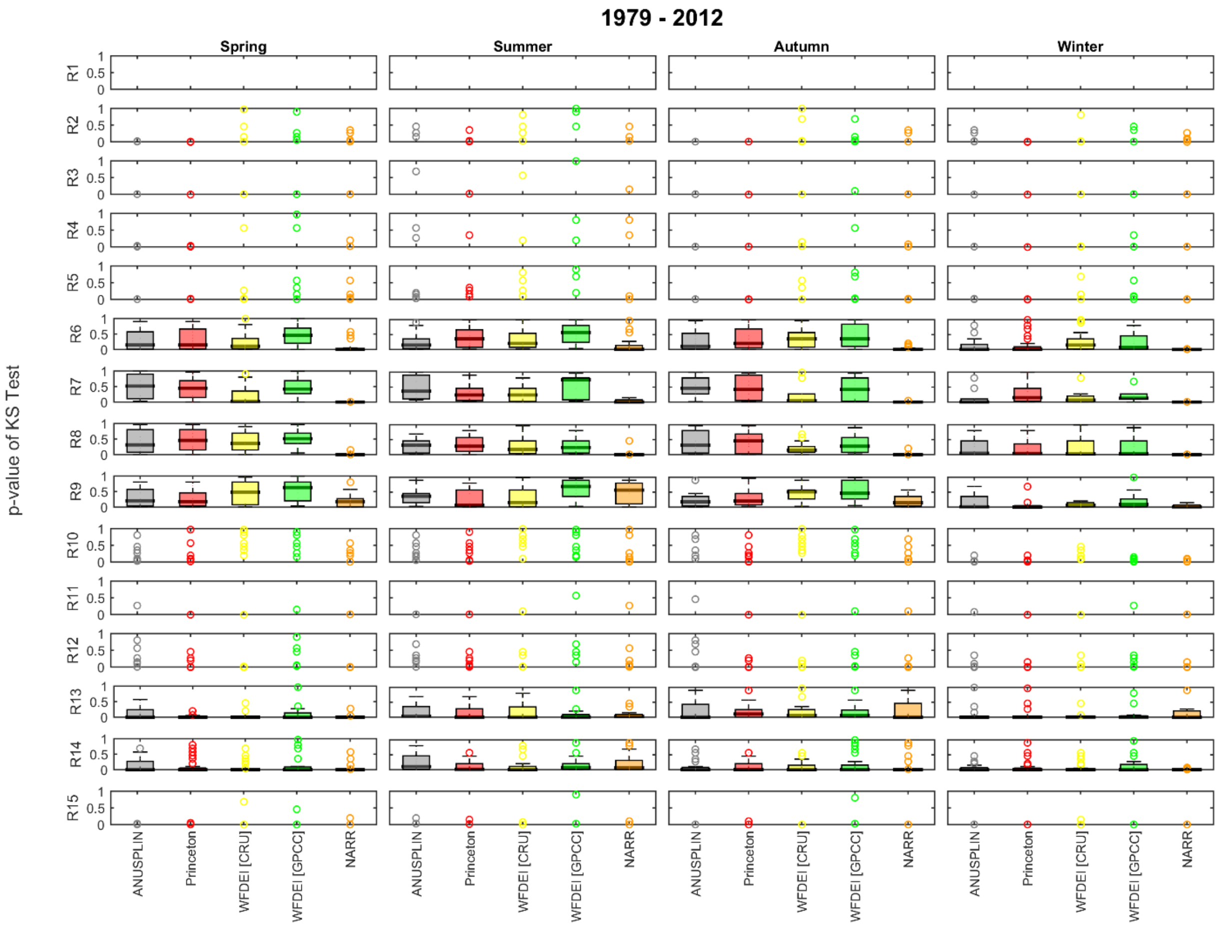

Figure 3. Distributions of p-value of the K-S test in the 15 ecozones in four seasons for the period of 1979 to 2012 (long-term comparison without CaPA). Note that the numbers of precipitation-gauge stations in each ecozone are different (see Table 4). Each hollow circle represents one p-value of the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

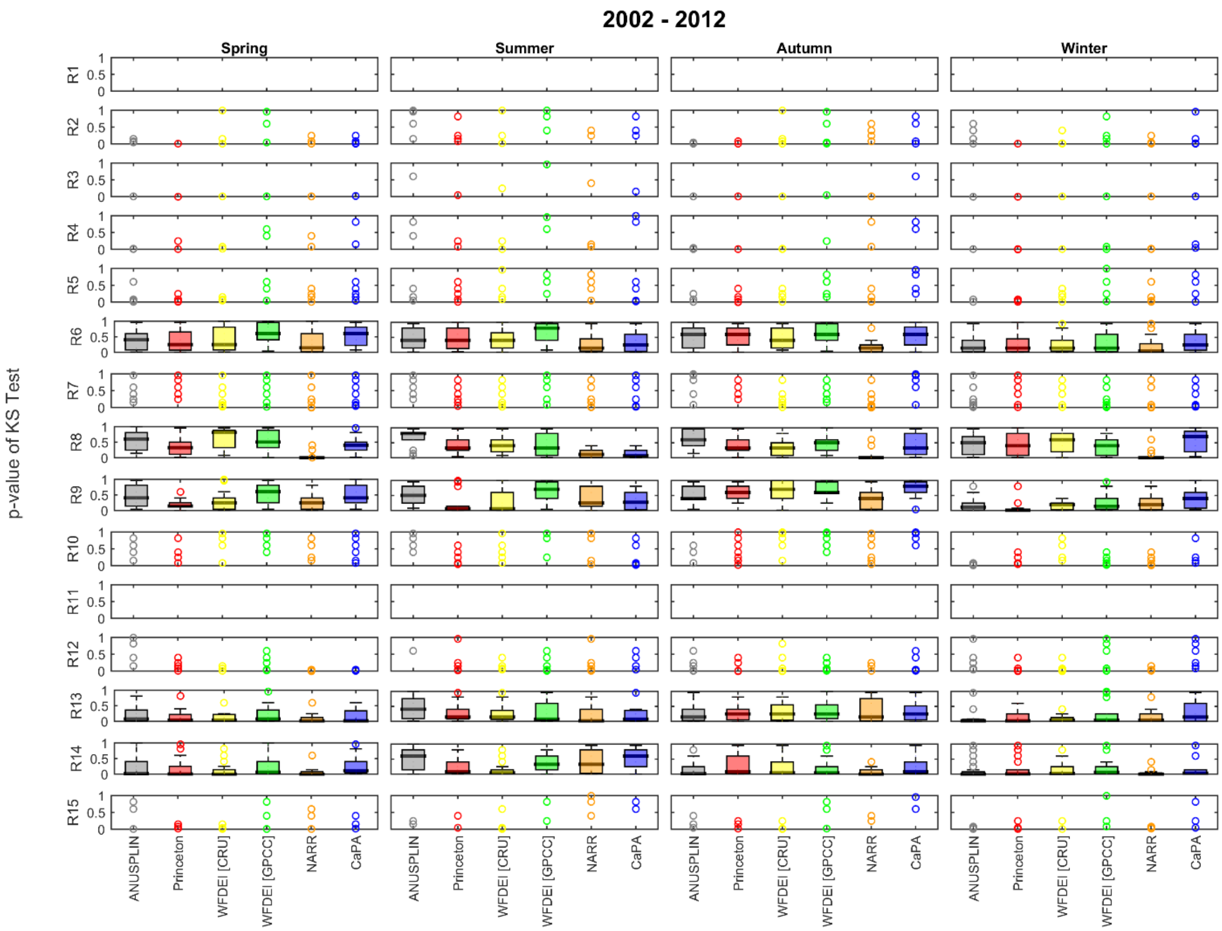

Figure 4. Distributions of p-value of the K-S test in the 15 ecozones in four seasons for the period of 2002 to 2012 (short-term comparison with the inclusion of CaPA). Note that the numbers of precipitation-gauge stations in each ecozone are different (see Table 4). Each hollow circle represents one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

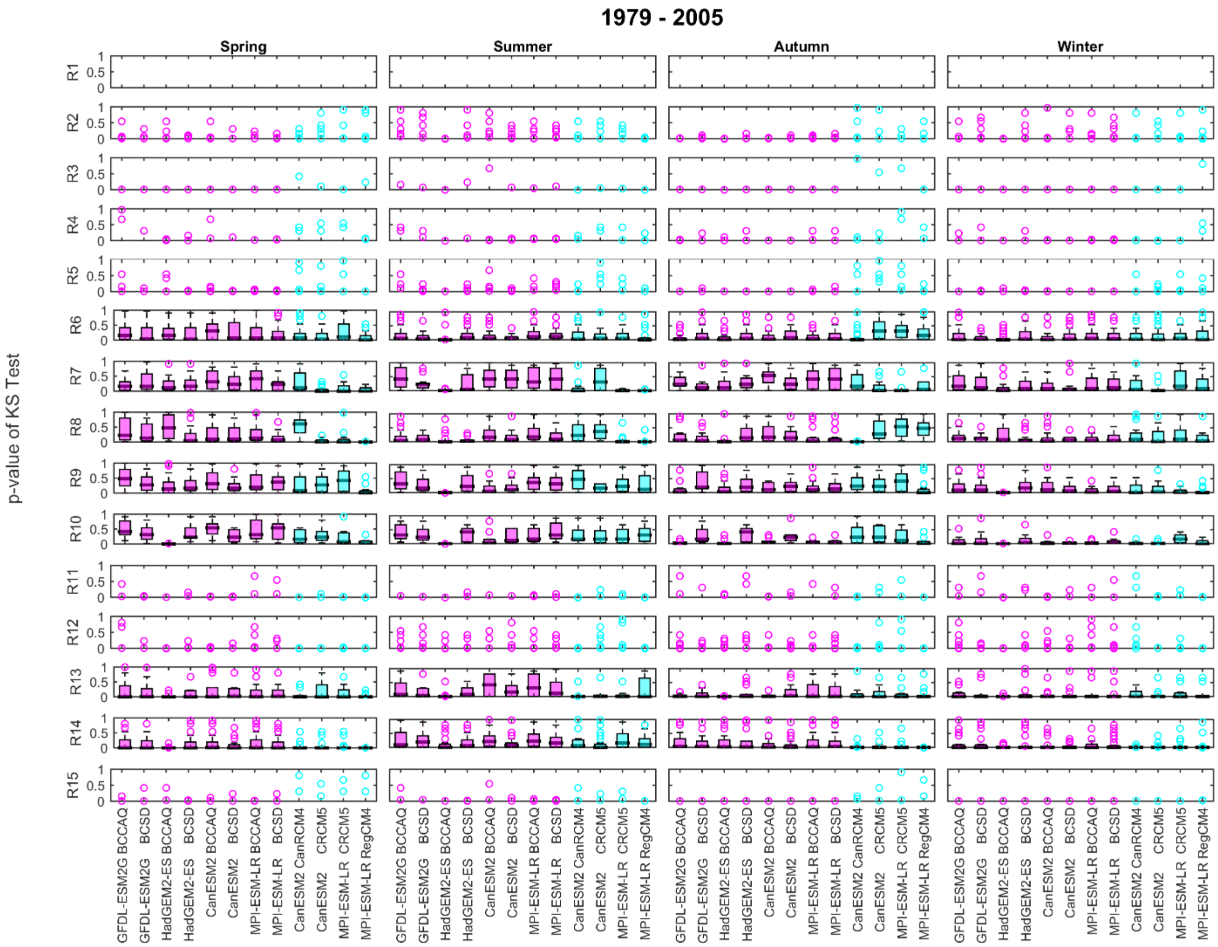

Figure 5. Distributions of p-value of the K-S test in the 15 ecozones in four seasons for the period of 1979 to 2005 (long-term comparison of PCIC and NA-CORDEX). Note that the numbers of precipitation-gauge stations in each ecozone are different (see Table 4). Each hollow circle represents one p-value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

References:

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.

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.

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.

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.

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.