

An algorithm for blending multiple satellite precipitation estimates with in situ precipitation measurements in Canada

Achan Lin¹ and Xiaolan L. Wang¹

Received 9 June 2011; revised 27 August 2011; accepted 29 August 2011; published 11 November 2011.

[1] This study proposes an algorithm for blending multiple satellite precipitation estimates (SPEs) with in situ gauge precipitation measurements in Canada. Depending on the number of gauge stations in the target area, the algorithm employs gauge data alone or blends gauge data with the corresponding SPEs that have been corrected for biases using a novel bias removal procedure developed in this study. The performance of this algorithm is evaluated in terms of root-mean-square error (RMSE), frequency bias index, and Pierce skill score, using 10 year gauge data from southwestern Canada where there are enough valid gauge stations to be split into a training data set and an evaluation data set. Sensitivity of the algorithm to gauge density is assessed by using five training data sets representing sparse to moderate gauge densities. The results show that, in comparison with the SPEs and a kriging analysis of gauge data, the blended analysis has the smallest RMSE and is least biased and most skillful in all seasons, and that the lower the gauge density, the more superior the blended analysis is. When gauge density is low, kriging analysis of gauge data is worse than bias-corrected SPEs. The unadjusted SPEs are the worst by all measures considered, which indicate a need for a proper correction of biases in the SPEs. The blending algorithm is promising for producing a more realistic gridded precipitation, especially for gauge sparse regions, such as northern Canada. A blended analysis of monthly precipitation is produced and compared with several existing precipitation analyses.

Citation: Lin, A., and X. L. Wang (2011), An algorithm for blending multiple satellite precipitation estimates with in situ precipitation measurements in Canada, *J. Geophys. Res.*, 116, D21111, doi:10.1029/2011JD016359.

1. Introduction

[2] Precipitation is a key variable for specifying the state of the climate system. Quantifying the amount of precipitation and assessing its changes both spatially and temporally are critically important for the assessment and understanding of climate change and its impacts on the environment, ecosystem, and human society.

[3] Traditionally, precipitation amount is measured with precipitation gauges. Gauge measurements provide the most reliable point observations of precipitation, also with good representation of the temporal variability of precipitation. However, precipitation gauge networks are often too sparse to represent the very high spatial variability of precipitation. This problem is particularly severe in northern Canada where two gauge stations are typically 500~700 km apart (could be over 1100 km in northern Quebec). The large geographical gaps in northern Canada make it difficult to generate a realistic gridded precipitation data set using gauge precipitation data alone. Gridded precipitation data

sets derived solely from spatial interpolation of gauge measurements suffer greatly from inadequate density of precipitation gauges. Yet, there is an urgent need for a high-quality gridded precipitation database that can well represent both the temporal and spatial variations of precipitation over Canada, for validating GCM/RCM simulations and for assessing climate change and its impacts, among many other applications.

[4] Satellite observations of infrared and microwave radiance have been used successfully to retrieve precipitation information. The satellite based precipitation estimates/data sets have been attracting more and more attention; they provide better representation of the precipitation field (the spatial variability), although they also contain non-negligible errors because of the indirect nature of the relationship between observations and precipitation, and because of the inadequate sampling and algorithm imperfections. Combining satellite precipitation estimates that have better spatial coverage with in situ gauge data that have better temporal coverage has become the most promising approach to produce a gridded precipitation data set that can well represent both the temporal and spatial variations of precipitation over regions of insufficient gauge density.

[5] Many merging techniques and products of different time and spatial scales and coverages have been developed in the past decades. Among them, the product of the Global

¹Climate Research Division, Science and Technology Branch, Environment Canada, Toronto, Ontario, Canada.

Precipitation Climatology Project (GPCP) [Huffman *et al.*, 1995, 1997; Adler *et al.*, 2003] and the Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP) [Xie and Arkin, 1997] have global coverage with a long time span and are widely recognized. The GPCP combined precipitation data were developed and computed by the NASA/Goddard Space Flight Center's Laboratory for Atmospheres as a contribution to the GEWEX Global Precipitation Climatology Project. As stated in Adler *et al.* [2003], GPCP is "a merged analysis that incorporates precipitation estimates from low-orbit satellite microwave data, geosynchronous-orbit satellite infrared data, and surface rain gauge observations." GPCP products, which are available online at <http://precip.gsfc.nasa.gov/>, include (1) monthly, 2.5° merged analysis for the period from 1979 to date [Adler *et al.*, 2003]; (2) pentad (5 d), 2.5° merged analysis for 1979–present [Xie *et al.*, 2003]; and (3) daily, 1° merged analysis for 1997–present [Huffman *et al.*, 2001]. We use the GPCP monthly analysis version 2.1 (http://precip.gsfc.nasa.gov/gpcp_v2.1_comb_new.html) for comparison later in this study. Similarly, the CMAP merged data sets also combine rainfall estimates from a variety of satellite and in situ data sources, including pentad and monthly analysis of global precipitation on a $2.5^\circ \times 2.5^\circ$ latitude-longitude grid for the period extending back to 1979. In addition, simulations of precipitation from Numerical Weather Prediction models are also used to fill in gaps at high latitudes in the enhanced version of the CMAP data set. The CMAP data are available online at <ftp://ftp.cpc.ncep.noaa.gov/precip/cmap>.

[6] An intercomparison study [Yin *et al.*, 2004] shows that the GPCP and CMAP products "are in very good agreement in describing the large-scale precipitation spatial patterns and temporal variations," though "there are still some significant differences between them that appear to be related to the input data and/or the method of merging the satellite and gauge data. The most obvious is the differences in the tropical oceans and the high-latitude land areas."

[7] This study proposes an algorithm for producing a gridded monthly precipitation data set of half-degree resolution by blending multiple satellite precipitation estimates with in situ gauge data for Canada. Although the focus of this study is restricted to Canada (40°N – 90°N , 50°W – 150°W), we believe that the algorithm is applicable for producing a global blended analysis of precipitation with little adjustment.

[8] The remainder of the paper is arranged as follows. Section 2 gives a brief description and preprocessing procedure of the data sets used in this study. Section 3 details the blending algorithm, with an assessment of its performance presented in section 4. Section 5 compares the resulting blended analysis with the GPCP and CMAP data sets, and with the Canadian gridded precipitation (CANGRD) data set, a Canadian gridded precipitation product. Finally, we present some concluding remarks in section 6.

2. Data Sets and Preprocessing Procedures

[9] This study focuses on monthly time scale. In addition to the in situ gauge precipitation data for Canada, two major types of satellite precipitation estimates (SPEs) are used in this study. How we preprocess these data sets is described in the subsections below.

2.1. In Situ Gauge Data and Virtual Observation Values

[10] The gauge precipitation data used in this study are monthly mean precipitation totals for 1979–2003, as derived by Hutchinson *et al.* [2009] from daily precipitation amounts (after quality control) in the digital archive of Environment Canada. The monthly precipitation data include correction for trace precipitation and distributing an accumulated amount to each day in the period of accumulation whenever possible and necessary (e.g., when the period goes from one calendar month to the next). The data processing and quality control procedure are detailed by Hutchinson *et al.* [2009]. Unlike satellite precipitation estimates, which are snapshot area quantities and are usually expressed as monthly or daily or subdaily mean precipitation rates, gauge precipitation measurements are accumulated point quantities (usually daily amounts recorded at stations). For each station, the daily amounts in a month are first aggregated to a monthly total amount, which is then divided by the number of days in the month to obtain the monthly mean precipitation rate (unit: mm per day). However, these monthly mean precipitation rates are still point values. We need a gauge precipitation data set that is comparable to the SPEs, with a value representing the average rate of precipitation over each designated grid box (which is a $0.5^\circ \times 0.5^\circ$ grid box in this study). That is, we need to aggregate all the point values available within a designated grid box, to produce a grid box average value.

[11] Since the density of gauge stations varies from region to region, many of the designated $0.5^\circ \times 0.5^\circ$ grid boxes have no valid gauge station (i.e., station of valid precipitation observation for the month in question) in it. In this study, a virtual observation grid box/location (VOGB) refers to a grid box/point for which there is at least one valid gauge station within the respective grid box. The number of VOGBs may vary from month to month, depending on the number and locations of valid gauge stations for the month. As an example, the map of valid gauge stations for September 2003 is shown in Figure 1a, and the corresponding VOGBs is shown in Figure 1b.

[12] Having defined the VOGBs, we take the inverse-distance-weighted (IDW) average of all valid gauge data within a VOGB (or an expanded VOGB if necessary, see below) as the virtual observation value at the VOGB (representing the VOGB average). If there are less than four valid gauge stations (note that there is at least one) within the target VOGB, the IDW average is calculated using up to 12 nearest valid stations within a 75 km radius from the VOGB center, or within a 125 km radius if the number of valid stations within the 75 km radius is still less than four. If the number of valid stations within the 125 km radius is still less than four, the IDW average is calculated using the available, less than four valid stations (could be just one station). Remember that a $0.5^\circ \times 0.5^\circ$ grid box is not a VOGB if there is no valid gauge station in it.

[13] Following the above procedure, a set of virtual observation values (denoted as O, with one value for each VOGB) is produced for each calendar month in the period of analysis. The number of valid gauge stations used to derive a virtual observation value varies from location to location within a month, and from month to month. In July 2003, for

a. Stations of gauge data for September 2003



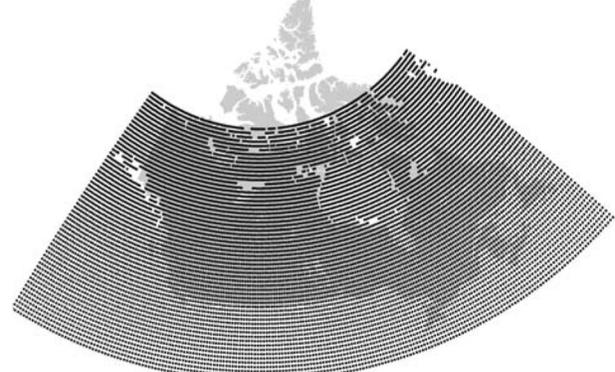
b. Virtual observation locations for September 2003



c. GPROF data coverage for January 2000



d. GPROF data coverage for August 2000



e. UMORA data coverage for January 2000



f. UMORA data coverage for August 2000

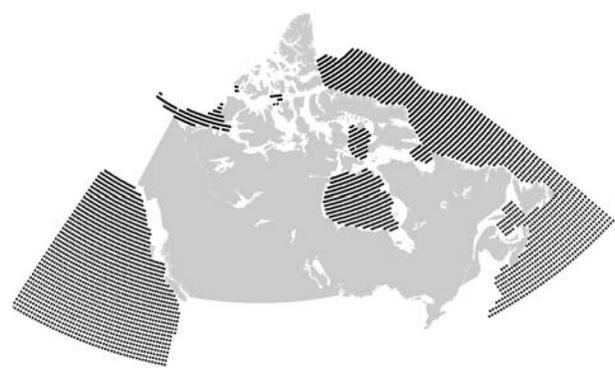


Figure 1. An example map of (a) the stations of gauge precipitation (monthly total) data, (b) the corresponding virtual observation locations (for September 2003), (c, d) the Goddard Profiling Algorithm (GPROF) coverage for January and August 2000, and (e, f) the Unified Microwave Ocean Retrieval Algorithm (UMORA) coverage for January and August 2000.

example, 9.4% of the virtual observation values were derived using 4~18 valid gauge stations within the $0.5^\circ \times 0.5^\circ$ grid box; the majority (85.6%) of them were produced using 1~3 valid gauge stations within the $0.5^\circ \times 0.5^\circ$ grid box, along with other stations (up to a total of 12 stations) in the area of up to 125 km radius from the VOGB center; and 5% of them

were set as the only gauge value within 125 km radius from the VOGB center (mostly in northern Canada; in this case the only gauge is also within the VOGB). Since the grid box is small ($0.5^\circ \times 0.5^\circ$), especially in the north, it is reasonable to take the single gauge/point value to represent the grid box average for a small number of VOGBs in northern Canada.

Table 1. Satellite Precipitation Estimates Used in This Study^a

Algorithm	Input Data	Spatial Scale	Time Scale	Area Coverage	Time Coverage
TOVS	HIRS, MSU	$1.0^\circ \times 1.0^\circ$	daily/monthly	global	Jan 1979 to Dec 2004
AIRS	AIRS, AMSU	$1.0^\circ \times 1.0^\circ$	daily/monthly	global	Sep 2002 to present
GPROF	SSM/I	$0.5^\circ \times 0.5^\circ$	orbit segments	70°N–70°S	Jul 1987 to present
UMORA	SSM/I	$0.25^\circ \times 0.25^\circ$	daily/monthly	global ocean	Jul 1987 to present

^aTOVS, Television and Infrared Observation Satellite Operational Vertical Sounder (J. Susskind, personal communication, 2010); HIRS, High-Resolution Infrared Radiation Sounder (NOAA); MSU, Microwave Sounding Unit (NOAA); AIRS, Atmospheric Infrared Sounder (Aqua), AIRX3SPM, AIRS/Aqua Level 3 monthly support physical retrieval product (AIRS + AMSU); AMSU, Advanced Microwave Sounding Unit (Aqua); GPROF, Goddard Profiling Algorithm version 6.0 (http://rain.atmos.colostate.edu/RAINMAP/gprof_description.html); SSM/I, Special Sensor Microwave Imager (DMSP); UMORA, Unified Microwave Ocean Retrieval Algorithm (<http://www.ssmi.com/>).

Note that single isolated station does not fill all the grid boxes within the 125 km radius, because a $0.5^\circ \times 0.5^\circ$ grid-box is not a VOGB if there is no valid gauge station within this grid box (a VOGB must have at least one gauge station within the $0.5^\circ \times 0.5^\circ$ grid box) and virtual observation values are calculated for VOGBs only.

[14] Just like SPEs, the virtual observation values represent the grid box average precipitation rates. For each month, we also interpolate the virtual observation values (O) onto the designated grid by ordinary kriging, obtaining the virtual observation field $K(O)$ for later use. The data, which are on a latitude-longitude grid, were first projected onto an oblique stereographic coordinate system centered at (110°W, 60°N) before being used in the kriging. Then, we use the ArcGIS geostatistical analyst to do the kriging with a spherical semivariogram model.

2.2. Satellite Precipitation Estimates

[15] Several data sets of satellite precipitation estimates (SPEs) exist today, but few have data for high-latitude areas. Following works of GPCP, CMAP and *Sapiano et al.* [2008], we choose to use the following satellite based precipitation estimates: Television and Infrared Observation Satellite Operational Vertical Sounder (TOVS) based precipitation estimates, Atmospheric Infrared Sounder (AIRS) based precipitation estimates, and two Special Sensor Microwave/Image (SSM/I) based precipitation estimates: Goddard Profiling Algorithm (GPROF) [*Kummerow et al.*, 2001] and Unified Microwave Ocean Retrieval Algorithm (UMORA) [*Hilburn and Wentz*, 2008; *Wentz and Spencer*, 1998; *Wentz*, 1997]. The spatial and temporal resolution and coverage of these data sets are summarized in Table 1. A brief description of these data sets is given below.

2.2.1. TOVS and AIRS Based Satellite Precipitation Estimates

[16] The TOVS instruments were operated aboard the NOAA series of polar orbiting NIMBUS satellites. They delivered vertical profiles of temperature, humidity and liquid water content. The TOVS-based precipitation product was produced by the Satellite Research Team under the direction of Joel Susskind, at NASA Goddard Space Flight Centers Laboratory for Atmospheres. It is available in 1.0° resolution for the period from 1979 to 2003 by contacting Joel Susskind.

[17] Data from the TOVS instruments are processed using a multiple regression relationship between collocated rain gauge measurements and several TOVS based parameters that relate to cloud volume, cloud top pressure, fractional cloud cover and relative humidity profile. The relationships

vary seasonally and latitudinally; separate relationships are developed for ocean and land (see the work of *Susskind et al.* [1997] for more details about the TOVS-based data sets).

[18] From 2004 onward, the TOVS data sets are replaced with AIRS-based product AIRSX3SPM [*Susskind et al.*, 2011], which are available by contacting Joel Susskind. The processing technique of AIRS-based data follows the same principal of TOVS-based data sets.

[19] The TOVS and AIRS data sets together form a continuous global record of precipitation for the period from 1979 to present, which we use in this study. The TOVS/AIRS-based product of SPEs is particularly useful over regions where in situ measurements are sparse and is one of the few available sources for the high latitudes.

[20] The TOVS and AIRS precipitation estimates we have downloaded are on a global $1.0^\circ \times 1.0^\circ$ latitude-longitude grid. We interpolated the data to the designated $0.5^\circ \times 0.5^\circ$ latitude-longitude grid by inverse-distance-weighted averaging of the values at the four nearest grid points, while keeping the average value for the original $1.0^\circ \times 1.0^\circ$ grid box unchanged.

2.2.2. SSM/I-Based Satellite Precipitation Estimates

[21] The SSM/I data are recorded by the polar orbiting satellites of the Defense Meteorological Satellite Program (<http://www.ncdc.noaa.gov/oa/rsads/ssmi/gridded/>). SSM/I is one of the major instruments used for inferring precipitation from passive microwave signals. The techniques are usually based on radiative transfer: emission and scattering estimates. The emission estimation technique directly measures the emission from hydrometeors in the atmosphere, which requires a homogeneous background and is therefore unsuitable over land and coast. The scattering technique is based on the scattered radiation from ice particles in clouds and works over any background (ocean or land).

[22] In this study, we use the GPROF SSM/I data, which are scattering estimates over land and combined scattering/emission estimates over ocean, and the UMORA SSM/I data, which are emission estimates only over ocean. A comparison of the GPROF data set with the UMORA data set can be found in the work of *Hilburn and Wentz* [2008]. The GPROF data set is described in details by *Kummerow et al.* [2001] and the UMORA data set by *Hilburn and Wentz* [2008], *Wentz and Spencer* [1998], and *Wentz* [1997]. We only give a brief summary of these data sets below.

[23] The GPROF data set was downloaded from the NASA Goddard Space Flight Center. At the original sub-daily timings, the data set is on a $0.5^\circ \times 0.5^\circ$ grid for the following periods: July 1987 to November 1987, January

1988 to February 1990, and January 1991 to present. These data are accumulated to produce daily, then monthly, means on a $0.5^\circ \times 0.5^\circ$ latitude-longitude grid. Following the approach of *Sapiano et al.* [2008], we discarded a monthly average for which the number of ambiguous pixels exceeds 20%, with ambiguous pixels being defined as pixels that are possibly contaminated with snow/ice (they are a combination of channel values which are sometimes associated with snow/ice [*Sapiano et al.*, 2008]). Though a valuable product, GPROF SSM/I based applications have limited spatial coverage. For precipitation estimates, it is limited to latitudes ranging approximately from 70°N to 70°S in summer, and from 55°N to 55°S in winter, presumably because of the disturbance of the precipitation signal by surface snow or ice cover. Its coverage changes month by month, depending on the snow/ice coverage. Figures 1c and 1d show an example of its coverage in winter and summer.

[24] The monthly UMORA data set, downloaded from the Remote Sensing System, is on a $0.25^\circ \times 0.25^\circ$ latitude-longitude grid. We averaged them to the designated $0.5^\circ \times 0.5^\circ$ latitude-longitude grid. The data are available only for ocean areas for the period from July 1987 to present.

[25] Presumably because of contamination by snow/ice, the SSM/I data contain isolated unrealistically high figures. For example, in February 2003, within the latitude range of 40°N to 90°N and longitude range of 50°W to 150°W , the maximum GPROF value is 41.993 mm per day, the maximum UMORA value is 131.2 mm d^{-1} , but the regional maximum value calculated from gauge measurements in Canada is only 7.81 mm d^{-1} . First, we need to remove unrealistically high values in each month, using a cutoff value. We tried with different cutoff values: $(G_{\max} + C)$ with $C = 2, 4, 6, 8, 10$ (and also $\sqrt{G_{\max}^2 + 144}$), where G_{\max} is the regional maximum gauge precipitation (the maximum value over Canada; unit: mm per day) in that month. We found that the weights for the SSM/I data (and our results) are not sensitive to all these choices of cutoff value. Thus, we chose to use $(G_{\max} + 4)$, a moderate cutoff value, which retains all the satellite estimates within a reasonable range. All SSM/I values that are higher than the corresponding cutoff value are rejected, i.e., set to missing. The rejection rate varies from month to month. It is usually below 1%, and some months have no values rejected. The monthly average rejection rate over the period 1987–2007 is 0.30% for the GPROF data set, and 0.32% for the UMORA data set. Note that the rejection rate calculated this way is usually higher than what it would be if the regional maximum gauge value were obtained from the entire region covered by the satellite data ($40^\circ\text{N}\sim 90^\circ\text{N}$, $150^\circ\text{W}\sim 50^\circ\text{W}$), rather than just gauge measurements in Canada. However, this will have little effect on our blended precipitation product, which will only cover Canadian region.

[26] The two SSM/I data sets also include occasional 0 monthly mean precipitation rates for an unrealistically high number of grid points, which are also difficult to process using our algorithm which involves ratios between the SPEs and virtual observations values. Thus, we set the minimum value in the GPROF and UMORA data sets to a quarter of the minimum virtual observation value in the precipitation field for that month, or a quarter of $0.00001 \text{ mm d}^{-1}$ if the minimum virtual observation value is smaller than $0.00001 \text{ mm d}^{-1}$.

[27] Then, the two SSM/I based data sets, i.e., the GPROF data set that covers the region from $70^\circ\text{N}\sim 70^\circ\text{S}$ and the UMORA data set that covers global oceans are combined by simple averaging, producing a single global SSM/I data set of monthly mean precipitation rates on the designated $0.5^\circ \times 0.5^\circ$ latitude-longitude grid for later use in blending. Note that the UMORA data set has very limited contribution to this study because of its ocean-only coverage: It has data for the Hudson Bay and the Gulf of St. Lawrence areas only in the ice-free season (Figures 1e and 1f); besides, it only serves as sort of boundary conditions for the Canadian coastal areas.

3. The Blending Algorithm

[28] This section describes an algorithm we propose for blending multiple satellite precipitation estimates (mSPEs) with in situ gauge precipitation data for Canada. As detailed next in the subsections, the algorithm consists of three steps: (1) Combine the mSPEs into one single data set of SPEs, obtaining the initial estimates, SPE0, for Canada. (2) Quantify and remove biases in the SPE0 against the corresponding virtual observations. (3) Blend the bias-corrected SPEs with the corresponding virtual observations to produce a blended analysis of monthly precipitation field.

3.1. Combining Multiple Satellite Precipitation Estimates

[29] An observational error is defined as the difference between an observed value and its true value. Let σ_k^2 denote the spatial variance of errors (simply referred to as error variance) in the k -th data set D_k (for $k = 1, 2, \dots, M$). The combination of M data sets that is optimal in terms of error variance (i.e., has the smallest error variance) is obtained as follows:

$$D_c = \sum_{k=1}^M (w_k * D_k), \text{ where weights } w_k = \frac{1/\sigma_k^2}{\sum_{k=1}^M (1/\sigma_k^2)} \quad (1)$$

Simply put, the combined value is a weighted average with weights being inversely proportional to the error variance of the corresponding data set. Here, an error refers to the difference between a virtual observation value (obtained using the procedure described in section 2.1) and an SPE for the same grid box (a virtual observation grid box).

[30] We use (1) to combine the SSM/I data set and the TOVS/AIRS data set (both result from the preprocessing described in section 2.2). However, for each data set, we use a fixed weight, the 1987–2007 average of the weights for each calendar month that are estimated using the monthly error variances of the data set for each month in the period. Namely, the weight for each calendar month does not change from year to year; for each data set, we have 12 weights, one for each calendar month. The error variances were calculated from the data for the region south of 55°N only, because most of the SSM/I data are missing in the region north of 55°N in the cold season and north of 70°N in the warm season (using all available data gives very similar results). The resulting weights for the SSM/I data are listed in Table 2, which range from 0.44 in January to 0.14 in May; subtracting each of these weights from 1 gives the

Table 2. The 1987–2007 Mean Error Variances of the TOVS/AIRS and SSM/I Data Sets, and the 1987–2007 Mean Weights for the SSM/I Data for Each Calendar Month^a

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
TOVS/AIRS	12.04	7.22	4.74	2.60	1.86	1.82	2.10	1.49	2.50	4.16	7.62	6.66
SSM/I	14.65	12.88	9.21	5.53	5.82	6.60	6.77	5.89	4.36	4.96	9.43	10.92
Weights for SSM/I	0.44	0.36	0.32	0.24	0.14	0.16	0.17	0.18	0.29	0.41	0.43	0.39

^aNote that a mean weight here is the average of the weights calculated for each month over the period 1987–2007; it would be slightly different if calculated from the mean error variances.

corresponding weight for the TOVS/AIRS data set (because $M = 2$ here). We use these weights for all grid points in the region south of 55°N . For the grid points between 55°N and 70°N , the weights for the SSM/I data are reduced linearly from the values at 55°N to zero at 70°N , because very few SSM/I data are available at 55°N and higher latitudes in the cold season and at 70°N and higher latitudes in the warm season. When a SSM/I SPE is not available, the “combined” SPE equals the TOVS/AIRS SPE. The resulting combined data set is referred to as the initial satellite precipitation estimates, denoted as SPE0. This data set contains large systematic biases. Section 3.2 describes how we estimate and remove these biases.

[31] Table 2 shows that both the SSM/I and TOVS/AIRS data sets have much larger error variances in winter than in summer, although the TOVS/AIRS data always have a smaller error variance than do the SSM/I data for the same month, regardless of season. However, winter/summer error variance ratio is much larger for the TOVS/AIRS than for the SSM/I data; thus, the error variance ratios between SSM/I and TOVS/AIRS are larger in summer than in winter, and hence the weights for SSM/I data are smaller in summer than in winter (Table 2).

[32] For each calendar month, the weights in (1) are invariant over the region south of 55°N , which is a simplification that has room for improvement. The linear reduction of the weights for SSM/I data from 55°N to 70°N is also a simplification that might have a little room for improvement (note that the weights are not used when a SSM/I SPE is not available; so the effects of the linear reduction should be small). The error variance of SPEs could change from region to region, depending on the ground (e.g., canopy and topography) and precipitation characteristics. We could divide the analysis region into a few subregions of similar error variances and derive season- and region-dependent weights for combining the SPEs (including one or more subregions between 55°N and 70°N). This will be implemented in a future version of our algorithm.

3.2. Removing Systematic Biases in the SPE0 Fields

[33] Here, a bias refers to the difference between a virtual observation value and an SPE for the same virtual observation location. In calculating the error variance of the SPE data sets, we have noticed large biases in the SPEs, which remain in the SPE0 data set. For example, the spatial average (over Canada) of the 10 year (1994–2003) mean precipitation rates is 2.1783 for the virtual observation fields, but only 1.7698 for the SPE0 fields. In general, the SPE0 data set underestimates precipitation amount in the region south of 60°N , while it overestimates in the region north of 60°N . Note that the SPE0 for the south is truly a mix of the SSM/I and TOVS/AIRS data; but for the north it is almost entirely from the

TOVS/AIRS data. It is essential that we remove these systematic biases prior to any application of this data set. The bias removal is particularly important for grid boxes with no gauge data, because in this case the value in the blended analysis depends mainly on the SPE (see section 3.3).

[34] There are several ways to remove biases in SPEs. For example, *Xie and Xiong* [2011] remove biases in daily SPEs by matching their probability density function (PDF) with the PDF of the collocated daily gauge analysis, correcting precipitation-intensity-dependent biases on daily time scale. *Adler et al.* [2003] use ratios of the spatial mean precipitation rates (averaged over a moving window centered at a gauge location) between gauge values and SPEs to quantify the systematic biases in SPEs. In the latter case, the ratio may vary over time. *Sapiano et al.* [2008] use noise to signal variance ratios. Having tried several bias removal procedures (including that of *Adler et al.* [2003]), we develop the following bias removal procedure, which turns out to be most effective (leading to less biased and more skillful blended analysis).

[35] Similar to the procedure of *Adler et al.* [2003], we also use time varying ratios between virtual observation values and SPEs. For each individual month t in the period analyzed, we estimate the ratio field R_{ij} (where i is the designated grid point index) and use it as a multiplication factor to remove systematic biases in the SPE0 field. That is, a bias-corrected SPE field (SPE1) is obtained as:

$$S_{ii}^1 = R_{ii} * S_{ii}^0. \quad (2)$$

The ratio field R_{ii} is derived in the following three steps:

[36] 1. Calculate the ratio between the virtual observation value O_{ij} and the SPE0 value S_{ij}^0 at each virtual observation location; namely, $R_{ij}^0 = O_{ij}/S_{ij}^0$ for month t at virtual observation location j .

[37] 2. Limit the ratios R_{ij}^0 within the interval [0.1, 4.0]:

$$R_{ij}^a = \begin{cases} 4.0 & \text{if } R_{ij}^0 \geq 4.0 \\ R_{ij}^0 & \text{if } 0.1 \leq R_{ij}^0 < 4.0 \\ 0.1 & \text{if } R_{ij}^0 < 0.1 \end{cases} \quad (3)$$

These limits are necessary to avoid the unrealistically high and low ratios, because ratios are very sensitive to data errors.

[38] 3. Use ordinary kriging to interpolate the adjusted ratios R_{ij}^a to each grid point i on the designated $0.5^\circ \times 0.5^\circ$ grid, creating the ratio field R_{ii} . This is necessary because the ratios R_{ij}^a have values at only the virtual observation locations (not each grid point on the designated grid).

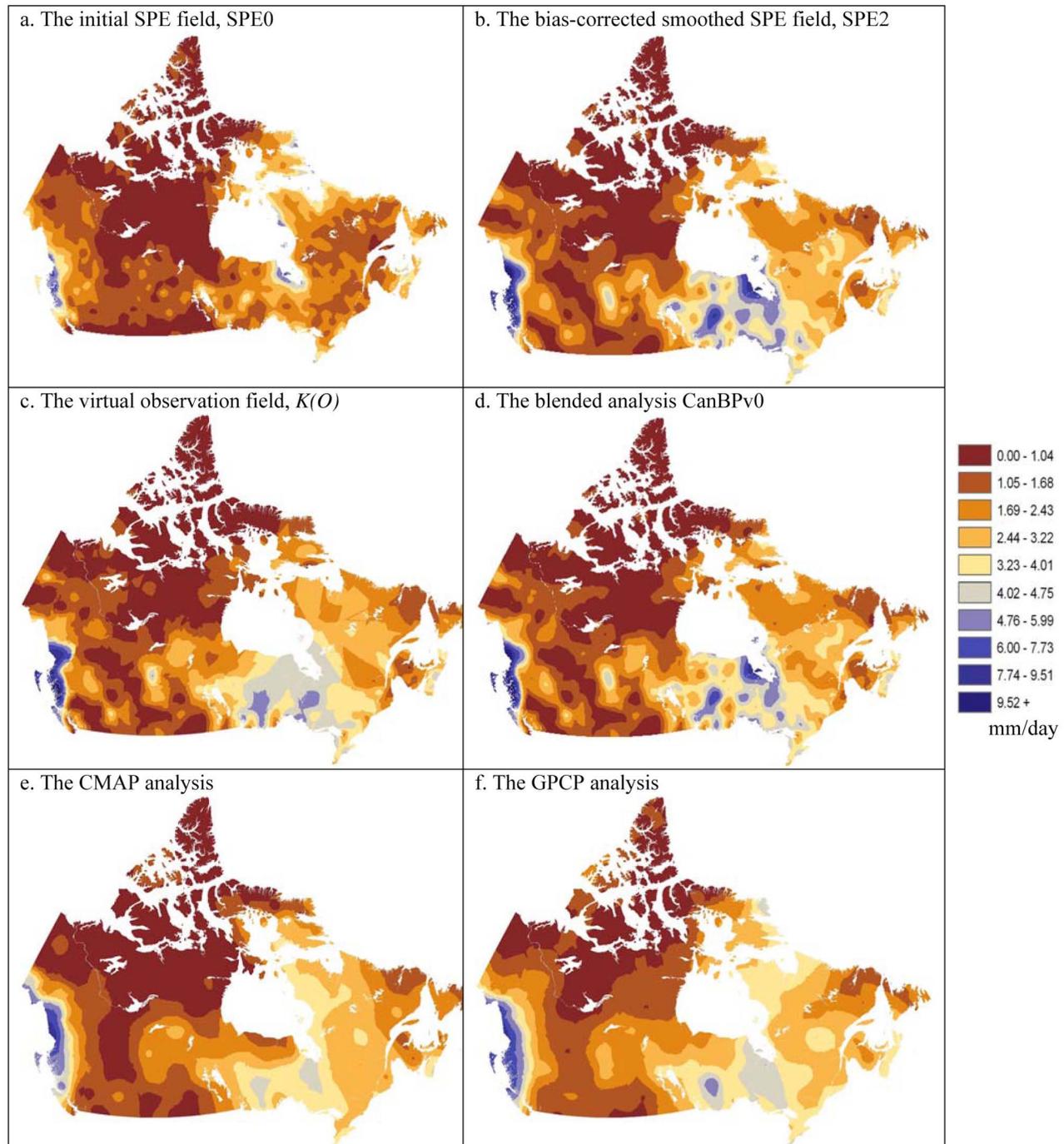


Figure 2. An example of the initial and bias-corrected satellite precipitation estimates (SPE) fields, in comparison with the virtual observation field (from ordinary kriging), the blended analysis Canadian Blended Precipitation version 0 (CanBPv0), Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP), and Global Precipitation Climatology Project (GPCP) (for September 2003). Note that the Canadian waters (e.g., the Hudson Bay) are masked off in these plots, but they are included in the output of the blended analysis.

[39] Note that large data errors can remain in a SPE1 field, because of the use of the ratio limits in (3). We can reduce the effect of large errors by making the SPE1 field smoother. To this end, we simply average the SPE1 values (on the $0.5^\circ \times 0.5^\circ$ grid) over a circle of 75 km radius centered at the target grid point, obtaining the SPE2 (we have tried with

50 km, 75 km, and 100 km and found that 75 km works best, giving an effect grid resolution of about 1.5° latitude by 2.5° longitude at 55°N latitude). This is the final version of the bias-corrected SPEs, which will be used to produce a blended precipitation analysis. As shown in Figures 2a–2c, the SPE2 field retains the main spatial pattern of the SPE0

field, but quantitatively closer to the corresponding virtual observation field $K(O)$. The bias-correction procedure actually anchors the major pattern of the SPE field to the gauge observation values, which is what we desire.

[40] It is worth pointing out that we have also tried with a bias correction algorithm similar to what was used to produce the GPCP data set by *Adler et al.* [2003]. We tried the GPCP bias correction procedure with some modifications. For example, instead of using a moving window that consists of a fixed number of latitude-longitude grid boxes (thus its size reduces as the latitude increases), we use a moving circle (of 125 km radius) whose area is basically latitude invariant. This is particularly important for Canada, because of its large latitude range and its very low gauge density in the north. The results (not shown) are not as good as the results obtained from using the bias correction procedure described above, in terms of the bias and skill measures discussed later in section 4.

3.3. Blending Satellite Precipitation Estimates With Virtual Observations

[41] This section describes the final step to blend the bias-corrected SPEs, i.e., the SPE2 data set obtained in section 3.2, with the corresponding virtual observation field $K(O)$ (see section 2.1), to obtain a blended analysis of precipitation. The algorithm is inspired by the conditional merging by kriging [*Sinclair and Pegram*, 2005].

[42] Let B denote the blended analysis. The gist of the blending algorithm can be expressed as:

$$B = \lambda * K(O) + (1 - \lambda) * [S + K(O - S_{VOL})],$$

$$\text{with } \lambda = \begin{cases} 1 & \text{if } m \geq 8 \\ m/8 & \text{if } m < 8 \end{cases} \quad (4)$$

where S denotes the SPE2 values on the whole designated $0.5^\circ \times 0.5^\circ$ grid, S_{VOL} denotes the S values at virtual observation locations, $K(O - S_{VOL})$ denotes the residual field (detailed below), and m is the number of valid gauge stations within 75 km radius from the target grid point. The residual field $K(O - S_{VOL})$ is obtained by interpolating, through ordinary kriging, the differences between the SPE2 values at the virtual observation locations, S_{VOL} , and the corresponding virtual observation values, O , onto the designated $0.5^\circ \times 0.5^\circ$ grid. When $\lambda = 0$ (which is true for all designated grid points that have no valid gauge station within 75 km radius), the blended analysis $B = [S + K(O - S_{VOL})]$, which is, in a loose sense, a statistical interpolation

combination of the kriging and the statistical interpolation. Note that setting the weighting coefficients to be inversely proportional to the error variance of the individual input fields would help reduce the error variance. However, the error variance of the individual input fields has been implicitly used in the bias correction process using (2). Thus, λ is simply set to be a linear function of number of gauges inside the target grid box.

[43] Our blending algorithm involves kriging. An accurate definition of error structure of the input fields involved is critical for success of kriging. However, we do not define the error structure of the input fields directly. Our approach is to evaluate indirectly the performance of kriging through rigorous evaluation of the performance of our blending algorithm, as detailed in section 4 below. The results indicate that our algorithm has satisfactory performance (see section 4).

[44] *Xie and Arkin* [1995] point out that at least five gauges are needed (in the region they studied) to compute a $2.5^\circ \times 2.5^\circ$ grid box average of monthly rainfall with a relative root mean square error of 10% or less. In this study, we calculate $0.5^\circ \times 0.5^\circ$ grid box averages of monthly mean precipitation rates using all gauge stations within 75 km radius from the target grid point; and the blended value equals the $K(O)$ value (kriging analysis of gauge data) if there are 8 or more stations within the 75 km radius. The threshold number 8 is chosen after trying with a range (4 to 16) of trial threshold numbers of stations. We find that it is unnecessary (and better not) to use the satellite estimates (which are generally of lower accuracy) when there is sufficient number of valid gauge stations within 75 km radius of the target grid point to represent the grid box average, and that 8 appears to be the best choice for the data analyzed in this study. Notice that the grid box averages are calculated by kriging, not by simply averaging all gauge data within the grid box. Thus, stations outside the grid box also play a significant role.

[45] Note that formula (4) involves addition and subtraction. It could generate negative values, which are nonsense for precipitation amount. To avoid negative values in the resulting blended analysis, we replace the term $[S + K(O - S_{VOL})]$, which might be negative, with the following term:

$$\begin{cases} \min(S + K(O - S_{VOL}), 4*S) & \text{if } K(O - S_{VOL}) \geq 0 \\ \max(S + K(O - S_{VOL}), 0.5*S) & \text{if } K(O - S_{VOL}) < 0 \end{cases} \quad (5)$$

Thus, the blended analysis is eventually obtained as

$$B = \begin{cases} \lambda * K(O) + (1 - \lambda) * \min[S + K(O - S_{VOL}), 4*S] & \text{if } K(O - S_{VOL}) \geq 0 \\ \lambda * K(O) + (1 - \lambda) * \max[S + K(O - S_{VOL}), 0.5*S] & \text{if } K(O - S_{VOL}) < 0 \end{cases} \quad (6)$$

of the virtual observations (O) that uses the corresponding satellite estimates (S) as the background field. When $\lambda = 1$, i.e., when there are eight or more valid gauge stations within 75 km radius from the target grid point, $B = K(O)$; the term $[S + K(O - S_{VOL})]$ and hence the satellite estimates (S) are not used; the blended value is obtained from the gauge data alone by kriging. When $0 < \lambda < 1$, the blended analysis is a

The 4 and 0.5 in (5) and (6) are empirical numbers that were determined by trial and error method: the 4 was chosen from trial values of 2, 3, 4, 5, and 6, and the 0.5, from trial values of 0.1, 0.25, and 0.5; all these trials have similar results, because these parameters are used only in a very small number of cases. In a loose sense, the second term in (6) can still be regarded as a statistical interpolation of the virtual

observations O with the corresponding satellite estimates S as the background field.

[46] The blending algorithm, as formulated in (6), is used to produce a blended precipitation data set for Canada, which is compared with other existing precipitation analysis in section 5. Its performance is assessed in section 4.

4. Evaluation of the Algorithm

[47] In order to assess the performance of the blending algorithm formulated in (6), we need to have two non-overlapping data sets. One is referred to as a training set, which is to be used to correct biases in the SPEs and to be blended with the bias-corrected SPEs to obtain a blended analysis. The other is referred to as an evaluation set, which is to be used to represent the true precipitation field against which to validate the blended and kriging analyses as well as the bias-corrected SPEs. We need a gauge network that has sufficient gauge stations for splitting into a training set and an evaluation set. The gauge network in southwestern Canada (48°N–62°N, 90°W–140°W) for the period 1994–2003 can serve for this purpose. For the 10 year period from 1994 to 2003, the total number of gauge stations in this test region ranges from 873 to 1258, though most of the gauge stations are still located in its south part (e.g., Figure 1a). Also, the number of stations with valid precipitation data is always smaller in winter than in summer.

[48] In order to assess the sensitivity of the blending algorithm to the density of gauge stations, we form five training sets, which consist of 125, 100, 70, 40, and 20 stations, respectively. Each smaller training set is a subset of the larger training set on the next level. The training sets are chosen carefully so that they have valid data for each and every month in the 10 year period (1994–2003) and they have a relatively even coverage over the test region (Figures 3a, 3c, and 3e). In other words, the five training sets remain the same for each month in the 10 year period. For each training set, all stations that are not included in the training set and have valid data for the month in question form the evaluation set for that month. Thus, the smaller the training set, the bigger the evaluation data set (see Figures 3b, 3d, and 3f); also, while the training set of stations does not change over time, the stations for evaluation often varies from month to month because the total number of valid stations often varies from month to month. The gauge density of the smallest training set, shown in Figure 3a, is similar to the gauge density in northern Canada (north of 60°N). Note that the evaluation set is always much larger than the training set; i.e., the true precipitation field is well represented, while the blended analysis is produced using a very limited number of gauge data. In other words, the evaluation settings are strict for the blending algorithm; they have good representation of the hardest situation in northern Canada, but they are much harder than the reality in southern Canada (even the 125 station training set has a much lower gauge density than the one existing in southern Canada).

[49] For each month, both the training set and the evaluation set of observations are aggregated to the designated $0.5^\circ \times 0.5^\circ$ grid points in the test region, using the procedure detailed in section 2.1. The resulting virtual observations of the training set are used to correct biases in the corresponding SPEs and are then blended with the bias-corrected

SPEs to produce a blended analysis for the month in question. This blended analysis is then evaluated against the virtual observations of the evaluation data set for that month, at the virtual observation locations of the evaluation data set.

[50] The following measures of performance are used in this section: mean error (ME), root-mean-square error (RMSE), pattern correlation score (PCS), frequency bias index (FBI), and Pierce skill score (PSS). The definitions of these measures are described in the Appendix. The PSS, also known as the true skill score, usually has a value between 0 and 1 inclusive; it is truly equitable, while the commonly used equitable threat score is not equitable [Hogan *et al.*, 2010]. The higher the PSS value, the higher skill the forecast/analysis has. For an unbiased forecast the FBI value is 1; and an FBI value greater (smaller) than unity indicates that the forecast/analysis overestimates (underestimates) the quantity in question. We also define a measure of the FBI bias as the standard deviation from unity of the FBI values, i.e., the FBI_{std} in the Appendix.

[51] The 10 year (1994–2003) mean values of the MEs, RMSEs, and PCSs of the blended analysis based on each of the five training sets, and of the corresponding initial SPEs (SPE0), the bias-corrected SPEs (SPE1), the bias-corrected smoothed SPEs (SPE2), and the kriging result of the virtual observations of the training set, are reported in Table 3. For the blended analyses that are based on the 20 and 100 station training sets, the box-and-whisker plots of the MEs are shown in Figure 4, and the time series of the RMSEs in Figure 5.

[52] As shown in Table 3, the SPE2 is much better than the SPE0 by all measures: its MEs are only about 42%~1% of the MEs of the SPE0 (in absolute values), and its RMSEs, about 81%~66%; it also has a much better pattern correlation with the evaluation data set. Also, it is important to point out that in terms of RMSEs and PCSs (Table 3), the SPE2 fields are superior to the SPE1 fields, and also superior to the kriging analysis when the gauge density is low to moderate, like the five training sets analyzed in this study. All these suggest that our bias correction procedure is necessary and successful.

[53] The SPE0 data set has the largest MEs in absolute value among the four data sets in comparison, showing a negative bias for all training data sets (Table 3 and Figure 4). It also has much larger RMSEs than the other data sets, especially when the training set is larger (Table 3; because biases in the SPEs can be better corrected when more gauge data are available for estimating the biases). Actually, the SPE0 is the worst among the five data sets by all measures (see Table 3) and hence will not be discussed further hereafter.

[54] The blended analysis is generally better than the kriging analysis, especially when the gauge density is low (Table 3). When the gauge density is low (like the 40 station set or lower), both the SPE2 and the blended analysis are better than the kriging analysis (Table 3 and Figure 5a). Overall, the blended analysis is the best among the five data sets (Table 3 and Figure 4); it has the smallest RMSEs for the majority of the 120 months in the test period (Figure 5) and the highest PCSs with the evaluation data set.

[55] To further evaluate the algorithm, for each of the four seasons, the 10 year (1994–2003) mean values of the FBI and PSS are shown as a function of the threshold q in

a. The 20-station training set (~6 gauges per 10^6 km²)

b. The 20-station test set for September 2003

c. The 70-station training set (~21 gauges per 10^6 km²)

d. The 70-station test set for September 2003

e. The 100-station training set (~30 gauges per 10^6 km²)

f. The 100-station test set for September 2003



Figure 3. (a, c, e) Three training sets of stations used in this study (the 40 and 125 station training sets are not shown) and (b, d, f) the corresponding evaluation sets for September 2003. The numbers in parentheses are the gauge station density of the training set in question.

Table 3. Ten Year, 1994–2003, Mean Values of the MEs, RMSEs, and PCs of the SPE0, the SPE1, the SPE2, the Kriging of the Virtual Observations of the Training Sets, and the Blended Analysis Based on Each of the Five Training Sets^a

	Stations				
	20	40	70	100	125
Mean precipitation rates (mm d^{-1})	1.975342	1.743527	1.873111	1.843574	1.863117
	<i>10 Year Mean MEs (mm d^{-1})</i>				
SPE0	-0.3033	-0.302	-0.2979	-0.3048	-0.3038
SPE1	0.0207	-0.1021	0.0748	0.0173	0.0338
SPE2	0.0257	-0.0991	0.0848	0.0288	0.0443
Kriging	0.0855	-0.147	-0.0169	-0.0591	-0.0437
Blended	0.0161	-0.1283	0.0414	-0.0082	-0.0028
	<i>10 Year Mean RMSEs (mm d^{-1})</i>				
SPE0	1.5814	1.5881	1.5891	1.6076	1.6078
SPE1	1.3652	1.3631	1.2870	1.2142	1.2154
SPE2	1.2787	1.2855	1.1583	1.1061	1.0838
Kriging	1.5192	1.4614	1.2171	1.1057	1.0859
Blended	1.2788	1.2533	1.152	1.0711	1.0678
	<i>10 Year Mean PCs</i>				
SPE0	0.7061	0.7065	0.7039	0.7047	0.7055
SPE1	0.7372	0.7481	0.7860	0.8041	0.8128
SPE2	0.7716	0.7814	0.8159	0.8321	0.8401
Kriging	0.6707	0.7292	0.8014	0.836	0.8417
Blended	0.7666	0.7848	0.8198	0.841	0.8458

^aIn order to help judge the relative importance of MEs and RMSEs, the regional means of the 10 year mean precipitation rates of the evaluation data set (observations) are also listed in the second row.

Figures 6 and 7 for the blended analyses that are based on the 20 and 100 station training sets, in comparison with the corresponding kriging analysis and SPE2. The seasons are defined as follows: spring (March–May), summer (June–August), autumn (September–November), and winter (December–February). The 10 year average values of FBI_{std} for each of the four seasons are reported in Table 4.

[56] The results in Table 4 and Figure 6 indicate that, overall, the blended analysis is least biased in all four seasons

among the three data sets; and the kriging analysis is most biased. The FBI_{std} values of the blended analysis are about 44% to 102% (or 79% on average; exceeding 96% only in the 40 station training set in summer) of their SPE2's counterpart, and about 29% to 81% (or 59% on average) of those of the kriging analysis (Table 4). The frequency bias decreases notably as the gauge density increases from the 40 station set to the 70 station set (Figure 6 and Table 4). It is smaller in the cold season than in the warm season (largest in summer, smallest in autumn; Table 4).

[57] Figure 6 also shows that all three data sets tend to overestimate the lower quantiles but underestimate the higher quantiles of precipitation in each season (e.g., underestimate summer precipitation when precipitation rates are greater than 2.0 mm d^{-1}). The smaller the training data set (i.e., the lower the gauge density), the more severe are the overestimates and underestimates (Figure 6). The FBI range corresponding to the 20 station training set is about three times as wide as that of the 100 station training set in winter (Figure 6). For the 100 station training set (the gauge density is about 30 stations per 10^6 km^2), the biases are very small for the entire range of precipitation rate in autumn and winter (Figures 6g and 6h).

[58] As shown in Figure 7, the blended analysis is most skillful among the three data sets, with the skill improvement for the whole range of precipitation rates (i.e., threshold values q). Note that all the three data sets are more skillful for moderate precipitation than for small or large precipitation (each curve in Figure 7 drops at both ends). The skill increases as the gauge density increases (compare left and right panels in Figure 7). When the gauge density is very low (like the 20 station set), the kriging analysis is the worst, notably worse than the SPE2 (Figures 7a–7d). For moderate gauge density (like the 100 station set), the kriging analysis is still not better than the SPE2 (Figure 7e–7h).

[59] We notice that the algorithm performs better in the cold season than in the warm season. That is something we would like to further investigate and improve.

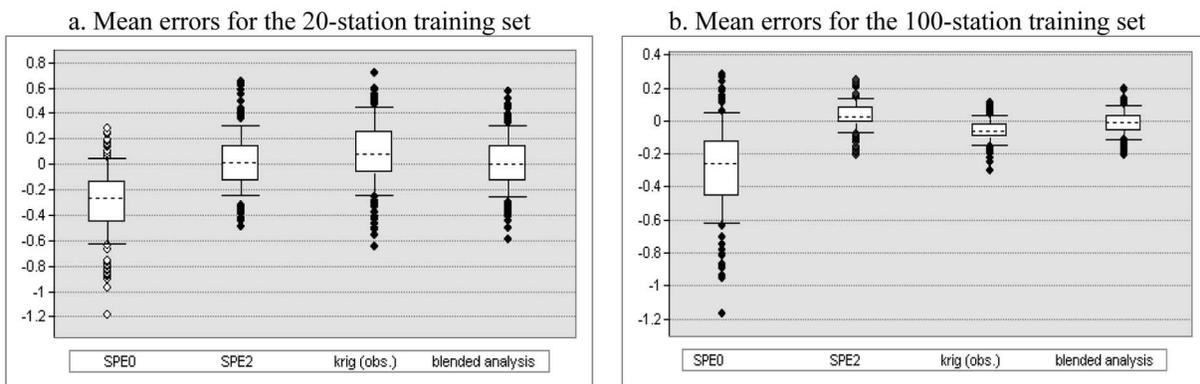


Figure 4. The box-and-whisker plots of the mean errors (mm d^{-1}) in the blended analysis based on (a) the 20 station training set and (b) the 100 station training set. Here, the errors refer to the differences between the blended analysis values and the corresponding virtual observation values of the evaluation set; the mean refers to the average over all virtual observation locations of the evaluation set. The evaluation period is 1994–2003 (10 years).

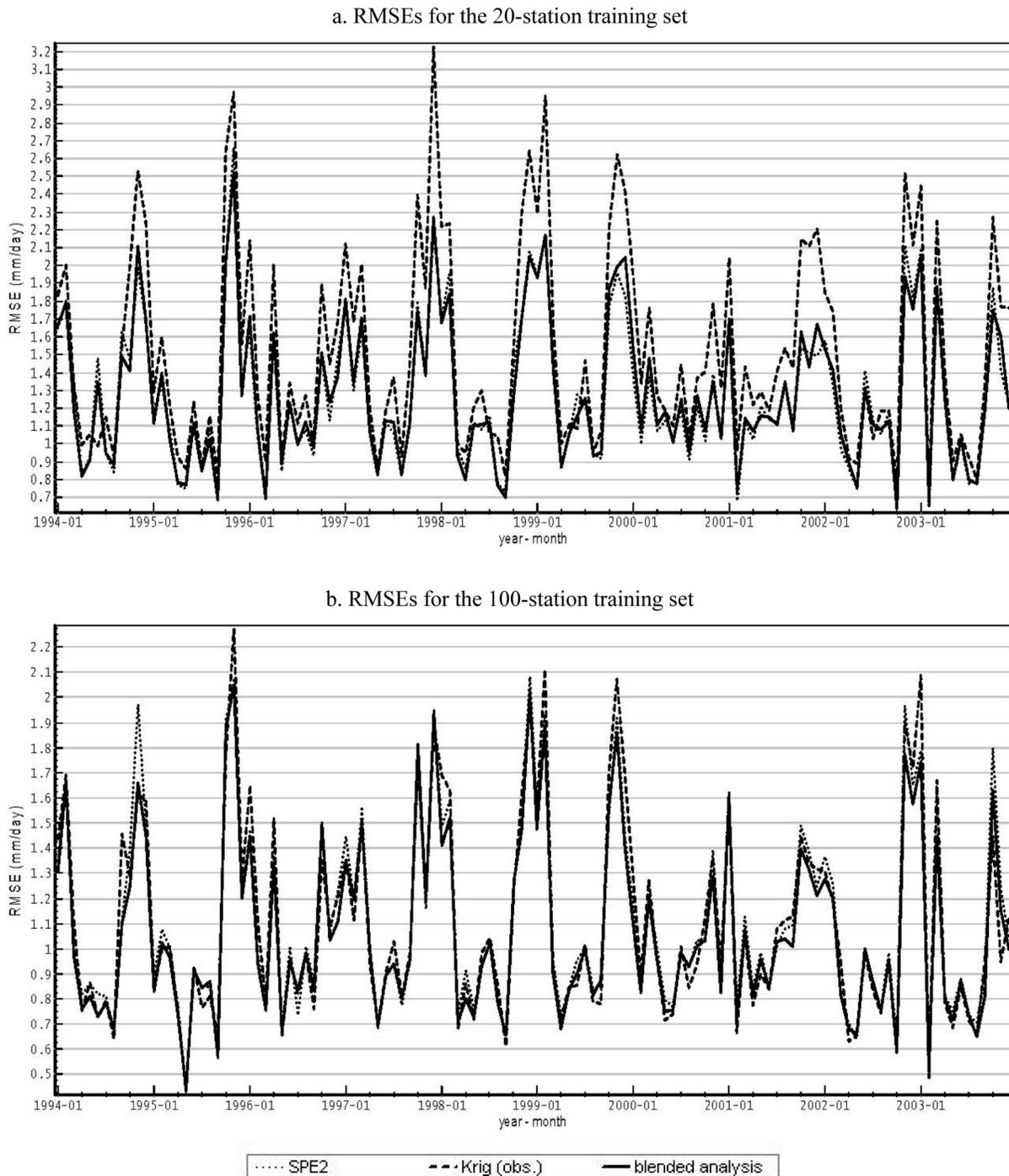


Figure 5. The time series of root-mean-square errors (RMSEs) for (a) the 20 station training set and the (b) 100 station training set. The evaluation period is 1994–2003 (10 years).

[60] It is important to stress that, for each month, the data in the evaluation set are not included in the training set; namely, the evaluation data set is independent from the training data set. The evaluation data set was not used in any way to produce the SPE2, or the blended analysis, or the kriging analysis being evaluated in this section.

[61] In summary, the blended analysis is least biased and most skillful in presenting the spatial patterns of precipita-

tion. It is better than either the bias-corrected SPEs or the kriging of available observations alone, although it still tends to underestimate larger precipitation and overestimate smaller precipitation in each season (i.e., underestimate the width of the distribution of precipitation amount). The unadjusted SPEs contain large biases and should not be used without a proper correction for the biases. The benefit of

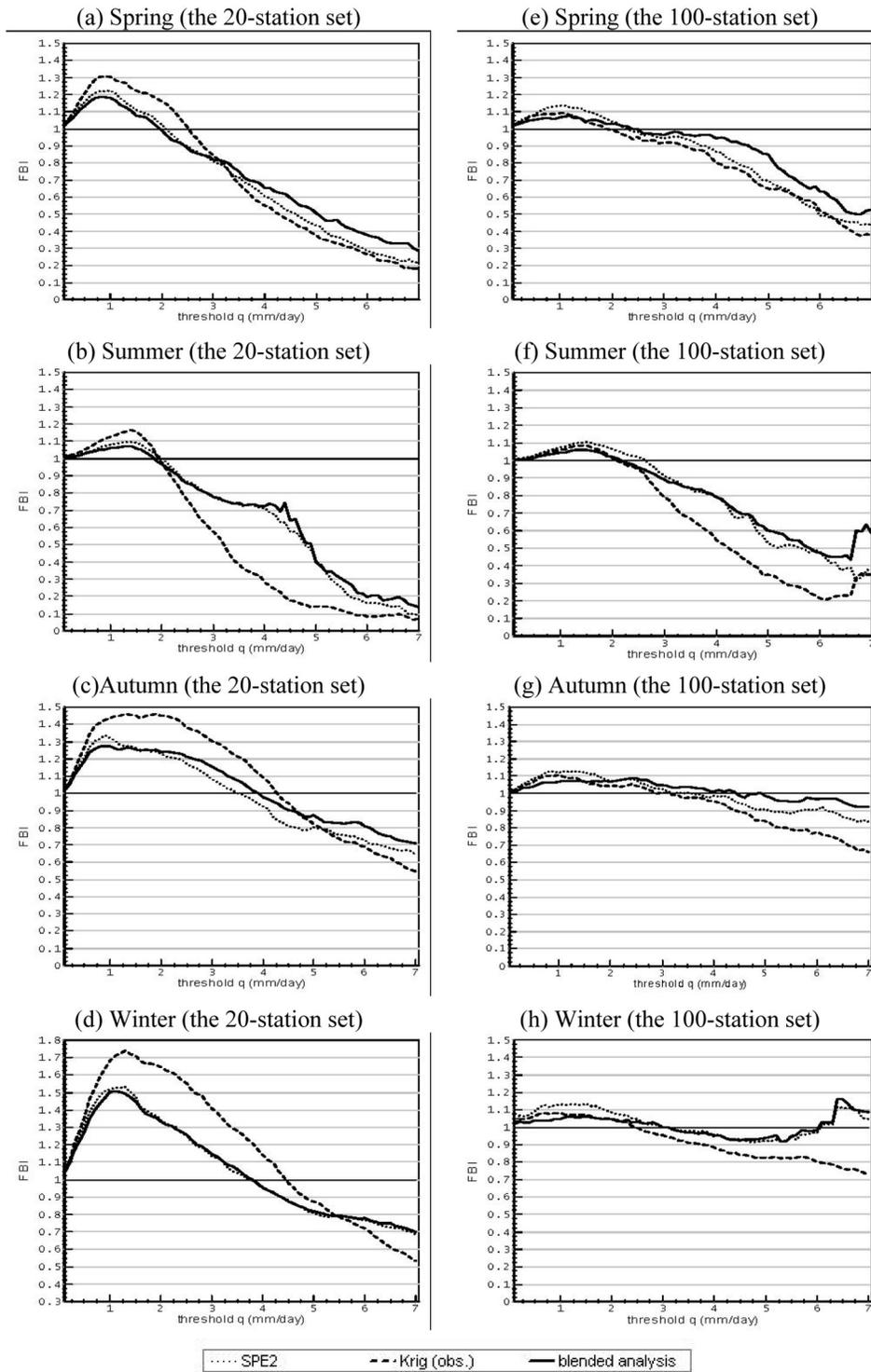


Figure 6. The 10 year (1994–2003) mean frequency bias indices for each of the four seasons, for (a–d) the 20 station training set and (e–h) the 100 station training set.

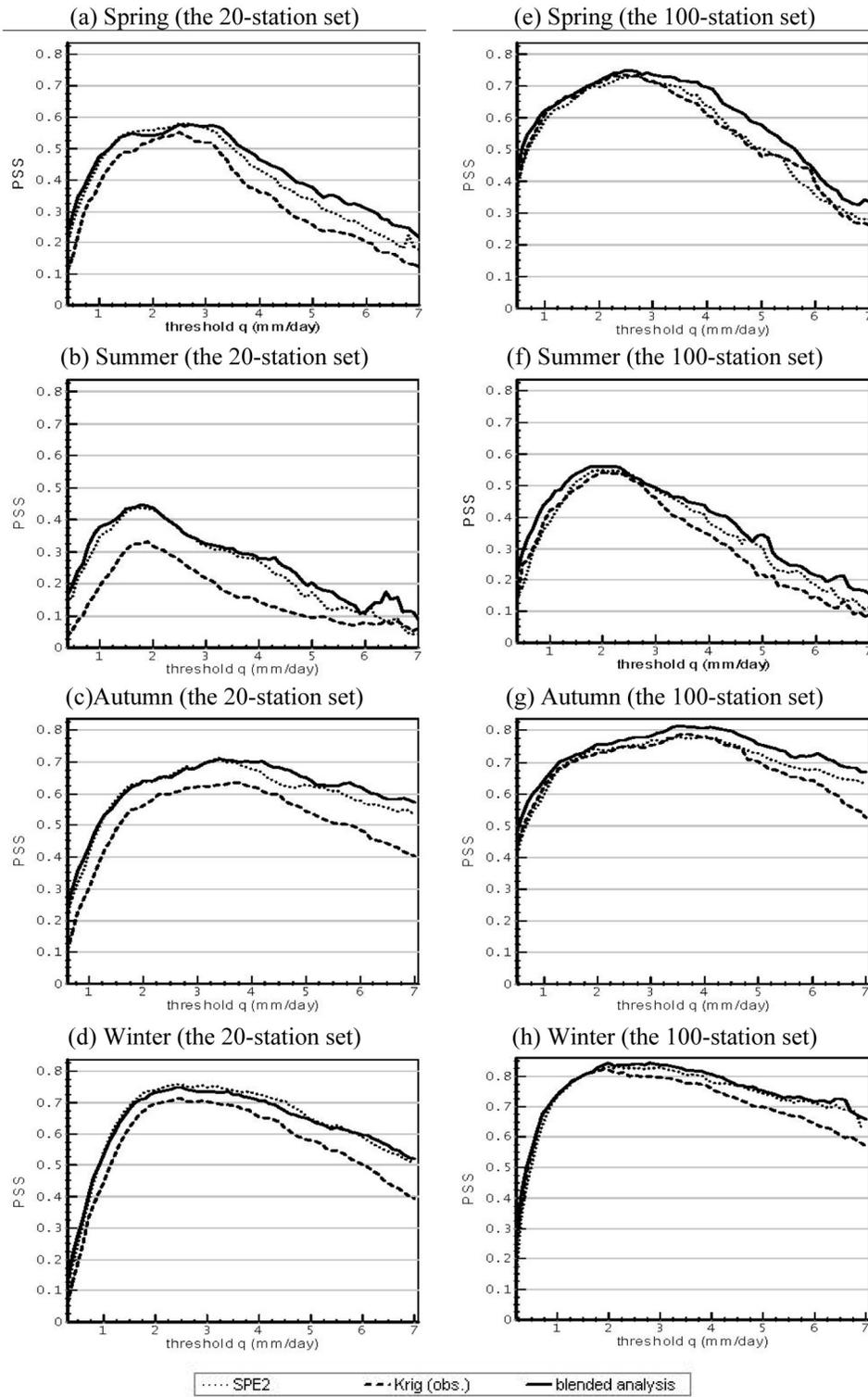


Figure 7. The 10 year (1994–2003) mean Pierce skill scores for each of the four seasons for (a–d) the 20 station training set and (e–h) the 100 station training set.

Table 4. Ten Year, 1994–2003, Average Values of the Root-Mean-Square of FBI Biases, i.e., the FBI_{std} Defined in Appendix A, of the Blended Analysis Based on Each of the Five Training Sets, and of the Corresponding SPE2 and the Kriging of the Virtual Observations of the Training Sets

	Stations				
	20	40	70	100	125
<i>Spring</i>					
SPE2	0.4459	0.5187	0.1842	0.2772	0.2143
Kriging	0.4808	0.5986	0.2477	0.2917	0.2283
Blended	0.3898	0.4386	0.1334	0.2123	0.1593
<i>Summer</i>					
SPE2	0.481	0.4135	0.342	0.336	0.3601
Kriging	0.6282	0.6172	0.4734	0.4654	0.4446
Blended	0.4537	0.4231	0.285	0.2908	0.3057
<i>Autumn</i>					
SPE2	0.2224	0.1934	0.076	0.0908	0.081
Kriging	0.3221	0.291	0.1279	0.148	0.1227
Blended	0.1926	0.129	0.0654	0.0519	0.036
<i>Winter</i>					
SPE2	0.2776	0.2599	0.111	0.0782	0.0797
Kriging	0.4221	0.3461	0.1547	0.1355	0.1064
Blended	0.2656	0.2124	0.0719	0.058	0.0676

combining bias-corrected SPEs with gauge data to produce a blended analysis of precipitation is clearly not negligible.

5. Comparison With Other Precipitation Analyses

[62] In this section, we compare our blended precipitation analysis, denoted as CanBPv0 (Canadian Blended Precipitation version 0), with three existing gridded precipitation products: Canadian gridded precipitation data set ((CANGRD), from Environment Canada (E. Milewska and R. Whitewood, Tools for monitoring areal trends in Canada, submitted to *Atmosphere-Ocean*, 2011), the GPCP monthly analysis version 2.1 (http://precip.gsfc.nasa.gov/gpcp_v2.1_comb_new.html), the CMAP monthly analysis (rain-1 [Xie and Arkin, 1996, 1997]). The period for comparison is 1994–2003.

[63] The CanBPv0 data set is based on all available gauge data in Canada for each target month; it is on a $0.5^\circ \times 0.5^\circ$ latitude-longitude grid over Canada. The CANGRD data set is based on gauge monthly precipitation data for 400+ long-term stations, which have been adjusted for wetting losses, gauge under catch, and station joining [Mekis and Hogg, 1999; Milewska and Whitewood, submitted manuscript, 2011]; it is on an oblique stereographic projection grid system with 50 km spatial resolution. Both the GPCP and CMAP data sets are products from merging gauge data with multiple SPEs; they are on a $2.5^\circ \times 2.5^\circ$ latitude-longitude grid. Numerical weather prediction model simulations of precipitation are also used in the CMAP rain-1 data set [Xie and Arkin, 1996, 1997]. For comparison purpose, the CanBPv0 and CANGRD data sets were first converted to the $2.5^\circ \times 2.5^\circ$ latitude-longitude grid of the GPCP and CMAP data sets for the Canadian region ($40^\circ\text{N}\sim 90^\circ\text{N}$, $150^\circ\text{W}\sim 50^\circ\text{W}$), shown in Figure 8.

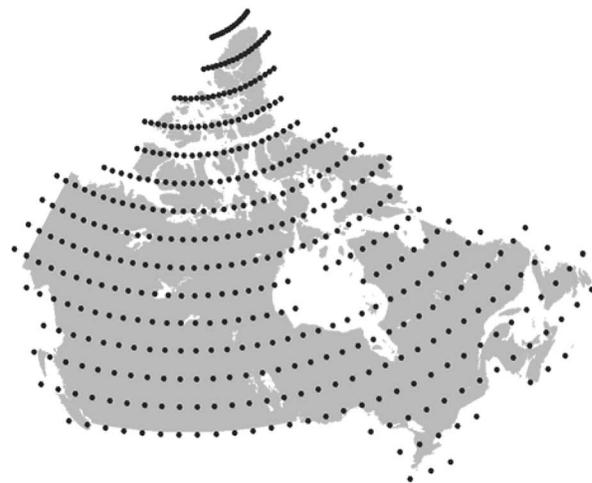


Figure 8. The $2.5^\circ \times 2.5^\circ$ latitude-longitude grid points at which the blended analysis is compared with the GPCP, CMAP, and Canadian gridded precipitation (CANGRD) data sets.

[64] The comparison is done on locations where all the four data sets have valid values, in terms of the regional average and standard deviation, as well as the pattern correlation score (PCS) between the CanBPv0 analysis and the other data sets. The comparison was done for southern and northern Canada (south and north of 60°N), separately. The 10 year mean values of the regional average, standard deviation (SD), and PCS are reported in Table 5.

[65] The box-and-whisker plots of the regional averages and spatial standard deviations of monthly mean precipitation rates are shown in Figure 9, for southern and northern Canada separately. In southern Canada, CanBPv0 is closest to GPCP in terms of both the regional means and spatial variability, while CANGRD and CMAP data sets have smaller regional means and slightly lower spatial variability in comparison with CanBPv0 and GPCP (Table 5 and Figures 9a and 9c). CanBPv0 has the highest spatial variability among all four data sets (Table 5). In northern Canada, CanBPv0 is closest to CANGRD in terms of both regional mean and spatial variability, while GPCP and CMAP data sets have greater regional means and higher spatial variability than the other data sets.

Table 5. Ten Year, 1994–2003, Mean Values of the Regional Average and Standard Deviation of the Five Data Sets in Comparison, and of the Pattern Correlation Score Between the CanBPv0 Analysis and the Other Four Data Sets^a

	CanBPv0	GPCP	CMAP	CANGRD
<i>Southern Canada</i>				
Average	2.3486	2.3612	2.0603	2.2216
SD	1.7715	1.7026	1.4272	1.5893
PCS	1.0000	0.8613	0.7977	0.8939
<i>Northern Canada</i>				
Average	0.6641	0.8392	0.7478	0.6410
SD	0.4582	0.6732	0.6190	0.3992
PCS	1.0000	0.6309	0.4479	0.6738

^aSee section 5. SD, standard deviation.

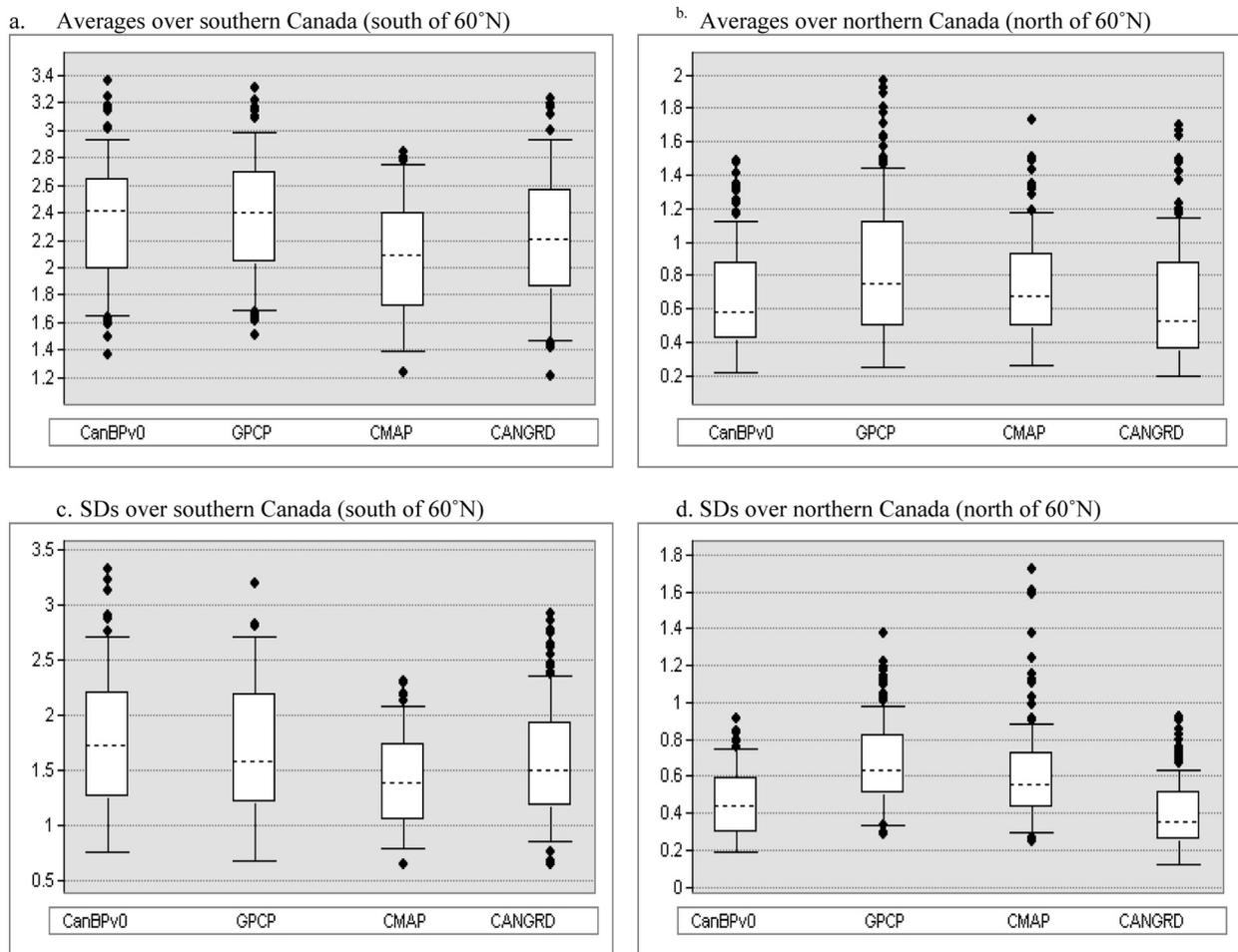


Figure 9. (a, b) The box-and-whisker plots of the regional averages and (c, d) spatial standard deviations (SDs) of monthly mean precipitation rate (mm d^{-1}) for southern Canada (Figures 9a and 9c) and northern Canada (Figures 9b and 9d), as derived from the four data sets. The period for comparison is 1994–2003 (10 years).

[66] Time series of pattern correlation scores between CanBPv0 and the other three data sets (GPCP, CMAP, and CANGRD) are shown in Figure 10 for northern and southern Canada, separately. Figures 2d–2f show an example of the blended analysis (for September 2003), as well as the corresponding CMAP and GPCP analysis.

[67] In terms of PCS, CanBPv0 is best correlated with CANGRD, second best with GPCP, in both southern and northern Canada (Table 5 and Figure 10). The CMAP data set has much lower PCSs with CanBPv0, especially in northern Canada, which should be expected since TOVS-based estimates are not included in the CMAP data set. In general, the correlations are higher in southern Canada than in northern Canada (Figure 10 and Table 5). This is because there are much more gauge data available in southern Canada and hence more gauge data have been “shared” by the four precipitation products in comparison.

[68] In summary, the CanBPv0 analysis is more comparable to the CANGRD and GPCP data sets than to the CMAP data set, both in southern and northern Canada. The relatively lower precipitation amount in the CMAP data set

is due to the fact that the CMAP data set does not include wind under-catch correction while the other data sets do.

6. Concluding Remarks

[69] We have developed an algorithm for blending multiple satellite precipitation estimates (SPEs) with in situ gauge precipitation measurements in Canada. The algorithm employs gauge data alone when there are at least eight valid gauge data within 75 km radius from the target grid point; otherwise, it blends the available gauge data with the SPEs that have been corrected for systematic biases, where the biases were estimated using a newly developed bias correction procedure and the corresponding gauge data.

[70] We have assessed the performance of this algorithm in terms of RMSE, frequency bias index, and Pierce skill score, using gauge data from Southwestern Canada where there are enough valid gauge stations to be split into a training data set and an evaluation data set. We have also assessed the sensitivity of the algorithm to gauge density by using five training data sets that represent sparse to moderate gauge densities. The validation results show that, in com-

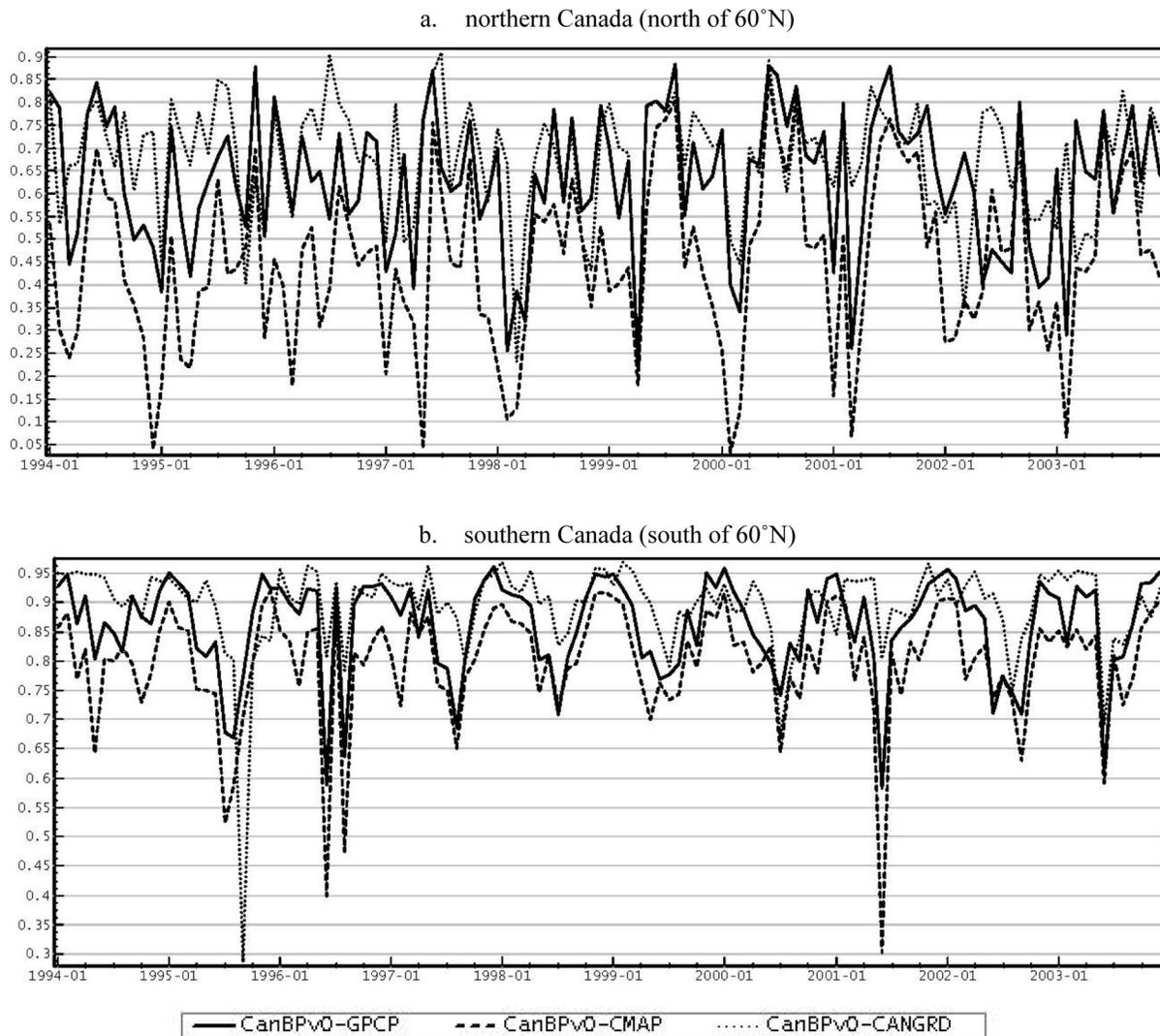


Figure 10. Time series of pattern correlation score (PCS) between the CanBPv0 and the GPCP, CMAP, and CANGRD data set for (a) northern Canada and (b) southern Canada. The time is shown as year/month.

parison with the SPEs and a kriging analysis of gauge data alone, the blended analysis CanBPv0 has the smallest RMSE and is least biased and most skillful in representing the spatial pattern of precipitation in all seasons, and that the lower the gauge density, the more superior is the blended analysis. When gauge density is low, kriging analysis is worse than bias-corrected SPEs alone. The unadjusted SPEs are the worst by all measures of performance used, which indicate a need for a proper correction of biases in the SPEs. The blending algorithm is a promising algorithm for producing a more realistic gridded precipitation, especially for gauge sparse regions, such as northern Canada.

[71] Ordinary kriging analysis is heavily used in our algorithm. Ordinary kriging is a method that is often associated with acronym BLUE for “Best Linear Unbiased Estimator.” It is “linear” in the sense that the estimates are weighted linear combinations of the available data; it is “unbiased” in the sense that it tries to have the mean of prediction error equal to zero; it is “best” in the sense that it

aims at minimizing the variance of the prediction errors. Kriging is a highly regarded interpolation technique in many fields. In a recent comparison study for six interpolation methods for climate data [Hofstra *et al.*, 2008], kriging is selected as the best performing method overall.

[72] Kriging alone is not perfect. Generally speaking, kriging is the BLUE when the data are normally distributed (Gaussian). When the data are not normally distributed, it is still the best linear predictor, but there may very well be better nonlinear predictors that give a much better result. Precipitation data is generally not normally distributed. Transformations (e.g., logarithm or Box Cox transformation) are commonly used to bring the data closer to normal. In mountainous areas, ordinary kriging of the actual gauge value is often not desirable. However, if we could detrend (deseasonalize) the data (when a reasonable trend is present), the residuals are usually much closer to normal than are the original gauge data. That is one of the reasons that kriging usually performs better on a residual field than on

Table A1. Contingency Table

	Observed $\geq q$ (Yes)	Observed $< q$ (No)
Forecast $\geq q$ (Yes)	a (hits)	b (false alarms)
Forecast $< q$ (No)	c (misses)	d (correct negatives)

the corresponding original field [Chen *et al.*, 2002]. One of our ongoing projects is to estimate monthly mean precipitation rates in mountainous areas using gauge data and Digital Elevation Model, which we will report in a separate study. We will try to improve our kriging analysis by accounting for geographic complexity, especially in mountainous area in the next version of this product by integrating these two projects.

[73] We are now also in the process of preparing a more accurate, more complete data set of gauge precipitation measurements in Canada, by investigating various flagged values, correcting for wetting losses and gauge under catch, accounting for trace precipitation (which accounts for a large portion of annual precipitation in high-latitude regions and hence is very important for Canada), etc.. We plan to use the algorithm developed in this study along with a more accurate gauge data set for a longer period (including the after 2003 period) to produce a new version of blended monthly precipitation analysis for Canada in the near future.

Appendix A: Performance Measures

[74] Let $O = \{o_1, o_2, \dots, o_m\}$ denote the observation set, $F = \{f_1, f_2, \dots, f_m\}$ denote the forecast (analysis) set, where m is the number of locations with observation and forecast. Further, let \bar{o} and \bar{f} be the spatial mean of O and F , respectively. The mean error (ME) and the root-mean-square error (RMSE) of the forecast are defined respectively as

$$ME = \frac{1}{m} \sum_{k=1}^m (o_k - f_k) \quad (\text{A1})$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{k=1}^m (o_k - f_k)^2} \quad (\text{A2})$$

The pattern correlation score (PCS) between F and O is defined as:

$$PCS = \frac{\sum_{k=1}^m (o_k - \bar{o})(f_k - \bar{f})}{\sqrt{\sum_{k=1}^m (o_k - \bar{o})^2 \sum_{k=1}^m (f_k - \bar{f})^2}} \quad (\text{A3})$$

[75] To assess objectively the skill of a sequence of n Yes/No forecasts (e.g., the precipitation rate is not less than q mm per day), we first obtain a contingency table (see Table A1). Here, the sum of the counts of hits, false alarms, misses, and correct negatives is equal to the total number of forecasts: $a + b + c + d = n$. Then, the Pierce skill score (PSS) and

the frequency bias index (FBI) for the forecast of whether or not the forecast $\geq q$ are defined respectively as:

$$PSS(q) = \frac{a}{a+c} - \frac{b}{b+d} \quad (\text{A4})$$

$$FBI(q) = \frac{a+b}{a+c} \quad (\text{A5})$$

The PSS is also known as the true skill score and is truly equitable [Hogan *et al.*, 2010]. The higher the PSS value, the higher skill the forecast has. For an unbiased forecast the FBI value is 1; and an FBI value greater (smaller) than unity indicates that the forecast overestimates (underestimates) the quantity in question. Since the FBI is a function of the threshold q , we also define the following as a measure of the mean FBI bias:

$$FBI_{std} = \sqrt{\frac{1}{m_q} \sum_{j=1}^{m_q} [FBI(q_j) - 1]^2} \quad (\text{A6})$$

where m_q is the number of thresholds used in the evaluation.

[76] **Acknowledgments.** The authors wish to thank Joel Susskind for his help on obtaining TOVS and AIRS based satellite precipitation estimates and Mathew R. P. Sapiano for his help and suggestions on the preprocessing of SSM/I data. The authors are grateful to Ewa Milewska and Normand Bussi eres of Environment Canada for their help/suggestion during the course of this study. Ewa Milewska and Vincent Fortin of Environment Canada, as well as the three anonymous reviewers are acknowledged for their useful comments/suggestions on an earlier version of this manuscript.

References

- Adler, R. F., et al. (2003) Thev 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.
- Chen, M., P. Xie, and J. E. Janowiak (2002), Global land precipitation: A 50-yr monthly analysis based on gauge observations, *J. Hydrometeorol.*, *3*, 249–266, doi:10.1175/1525-7541(2002)003<0249:GLPAYM>2.0.CO;2.
- Hilburn, K. A., and F. J. Wentz (2008), Intercalibrated passive microwave rain products from the Unified Microwave Ocean Retrieval Algorithm (UMORA), *J. Appl. Meteorol.*, *47*(3), 778–794, doi:10.1175/2007JAMC1635.1.
- Hofstra, N., M. Haylock, M. New, P. Jones, and C. Frei (2008), Comparison of six methods for the interpolation of daily, European climate data, *J. Geophys. Res.*, *113*, D21110, doi:10.1029/2008JD010100.
- Hogan, R. J., C. A. T. Ferro, I. T. Jolliffe, and D. B. Stephenson (2010), Equitability revisited: Why the “equitable threat score” is not equitable, *Weather Forecast.*, *25*(2), 710–726, doi:10.1175/2009WAF2222350.1.
- Huffman, G. J. (1997), The Global Precipitation Climatology Project monthly mean precipitation dataset, *WMO/TD 808*, 37 pp., World Meteorol. Organ., Geneva, Switzerland.
- Huffman, G. J., R. F. Adler, B. Rudolf, U. Schneider, and P. R. Keehn (1995), Global precipitation estimates based on a technique for combining satellite-based estimates, rain gauge analysis, and NWP model precipitation information, *J. Clim.*, *8*, 1284–1295, doi:10.1175/1520-0442(1995)008<1284:GPBOA>2.0.CO;2.
- Huffman, G. J., R. F. Adler, M. Morrissey, D. T. Bolvin, S. Curtis, R. Joyce, B. McGavock, and J. Susskind (2001), Global precipitation at one-degree daily resolution from multi-satellite observations, *J. Hydrometeorol.*, *2*(1), 36–50, doi:10.1175/1525-7541(2001)002<0036:GPAODD>2.0.CO;2.
- Hutchinson, M. F., D. W. McKenney, K. Lawrence, J. H. Pedlar, R. F. Hopkinson, W. Milewska, and P. Papadopol (2009), Development and testing of Canada-wide interpolated spatial models of daily minimum-maximum temperature and precipitation for 1961–2003, *J. Appl. Meteorol. Climatol.*, *48*, 725–741, doi:10.1175/2008JAMC1979.1.
- Kummerow, C., Y. Hong, W. S. Olson, S. Yang, R. F. Adler, J. McCollum, R. Ferraro, G. Petty, D-B. Shin, and T. T. Wilheit, (2001), The evolution

- of the Goddard Profiling Algorithm (GPROF) for rainfall estimation from passive microwave sensors, *J. Appl. Meteorol.*, *40*, 1801–1820, doi:10.1175/1520-0450(2001)040<1801:TEOTGP>2.0.CO;2.
- Mekis, É., and W. Hogg (1999), Rehabilitation and analysis of Canadian daily precipitation time series, *Atmos. Ocean*, *37*(1), 53–85, doi:10.1080/07055900.1999.9649621.
- Sapiano, M. R. P., T. M. Smith, and P. A. Arkin (2008), A new merged analysis of precipitation utilizing satellite and reanalysis data, *J. Geophys. Res.*, *113*, D22103, doi:10.1029/2008JD010310.
- Sinclair, S., and G. Pegram (2005), Combining radar and rain gauge rainfall estimates using conditional merging, *Atmos. Sci. Lett.*, *6*(1), 19–22, doi:10.1002/asl.85.
- Susskind, J., P. Piraino, L. Rokke, L. Iredell, and A. Mehta (1997), Characteristics of the TOVS Pathfinder Path A dataset, *Bull. Am. Meteorol. Soc.*, *78*, 1449–1472, doi:10.1175/1520-0477(1997)078<1449:COTTPP>2.0.CO;2.
- Susskind, J., J. Blaisdell, L. Iredell, and F. Keita (2011), Improved temperature sounding and quality control methodology using AIRS/AMSU data: The AIRS Science Team Version 5 Retrieval Algorithm, *IEEE Trans. Geosci. Remote Sens.*, *49*, 883–907, doi:10.1109/TGRS.2010.2070508.
- Wentz, F. J. (1997), A well calibrated ocean algorithm for special sensor microwave/imager, *J. Geophys. Res.*, *102*(C4), 8703–8718, doi:10.1029/96JC01751.
- Wentz, F. J., and W. Spencer (1998), SSM/I rain retrievals within a unified all-weather ocean algorithm, *J. Atmos. Sci.*, *55*, 1613–1627, doi:10.1175/1520-0469(1998)055<1613:SIRRWA>2.0.CO;2.
- Xie, P., and P. A. Arkin (1995), 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.
- Xie, P., and P. A. Arkin (1996), Analyses of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions, *J. Clim.*, *9*, 840–858, doi:10.1175/1520-0442(1996)009<0840:AOGMPU>2.0.CO;2.
- Xie, P., and P. A. Arkin (1997), Global precipitation: A 17 year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs, *Bull. Am. Meteorol. Soc.*, *78*, 2539–2558.
- Xie, P., J. E. Janowiak, P. A. Arkin, R. Adler, A. Gruber, R. Ferraro, G. J. Huffman and S. Curtis (2003), GPCP pentad precipitation analyses: An experimental dataset based on gauge observations and satellite estimates, *J. Clim.*, *16*, 2197–2214.
- Xie, P., and A.-Y. Xiong (2011), A conceptual model for constructing high-resolution gauge-satellite merged analyses of precipitation, *J. Geophys. Res.*, doi:10.1029/2011JD016118, in press.
- Yin, X., A. Gruber, and P. Arkin (2004), Comparison of the GPCP and CMAP merged gauge-satellite monthly precipitation products for the period 1979–2001, *J. Hydrometeorol.*, *5*, 1207–1222, doi:10.1175/JHM-392.1.

A. Lin and X. L. Wang, Climate Research Division, Science and Technology Branch, Environment Canada, 4905 Dufferin St., Toronto, ON M3H 5T4, Canada. (xiaolan.wang@ec.gc.ca)