

High-Resolution (2.5km) Ensemble Precipitation Analysis across Canada

DIKRA KHEDHAOUIRIA, STÉPHANE BÉLAIR, AND VINCENT FORTIN

Meteorological Research Division, Environment and Climate Change Canada, Dorval, Quebec, Canada

GUY ROY AND FRANCK LESPINAS

Meteorological Service of Canada, Environment and Climate Change Canada, Dorval, Quebec, Canada

(Manuscript received 3 December 2019, in final form 1 May 2020)

ABSTRACT

Consistent and continuous fields provided by precipitation analyses are valuable for hydrometeorological applications and land data assimilation modeling, among others. Providing uncertainty estimates is a logical step in the analysis development, and a consistent approach to reach this objective is the production of an ensemble analysis. In the present study, a 6-h High-Resolution Ensemble Precipitation Analysis (HREPA) was developed for the domain covering Canada and the northern part of the contiguous United States. The data assimilation system is the same as the Canadian Precipitation Analysis (CaPA) and is based on optimal interpolation (OI). Precipitation from the Canadian national 2.5-km atmospheric prediction system constitutes the background field of the analysis, while at-site records and radar quantitative precipitation estimates (QPE) compose the observation datasets. By using stochastic perturbations, multiple observations and background field random realizations were generated to subsequently feed the data assimilation system and provide 24 HREPA members plus one control run. Based on one summer and one winter experiment, HREPA capabilities in terms of bias and skill were verified against at-site observations for different climatic regions. The results indicated HREPA's reliability and skill for almost all types of precipitation events in winter, and for precipitation of medium intensity in summer. For both seasons, HREPA displayed resolution and sharpness. The overall good performance of HREPA and the lack of ensemble precipitation analysis (PA) at such spatiotemporal resolution in the literature motivate further investigations on transitional seasons and more advanced perturbation approaches.


1. Introduction


Accurate spatiotemporal characterization of precipitation fields is essential (Serinaldi and Kilsby 2014) for many applications such as hydrological modeling, impact studies, and agricultural impact assessments, which require spatially and temporally continuous precipitation inputs (Lamb et al. 2016). Continuous input series improve the ability to simulate infiltration, runoff, and

all major hydrological processes to provide better risk assessments.

Consistent with that goal, gridded precipitation has been developed, often based on satellite radiances (Kidd and Levizzani 2011), radar composites (Chandrasekar et al. 2012), mesoscale analysis (Joyce et al. 2004), interpolated station records (Hopkinson et al. 2011), reanalysis, and a blending of all or part of the preceding products (Fortin et al. 2018; Beck et al. 2019). Each of these datasets has its advantages, but they are all prone to a multitude of errors and uncertainties (for an extensive review, see Tapiador et al. 2017). When used as input for subsequent modeling (e.g., hydrology, land surface data assimilation), the uncertainties of the precipitation need to be estimated (Pappenberger and Beven 2006).

The current study focuses on uncertainties of gridded precipitation analysis (PA) by using ensembles. Recent research on ensemble meteorological fields (Buizza 2019; Duan et al. 2019; Raut et al. 2018, among others)

 Denotes content that is immediately available upon publication as open access.

 Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JHM-D-19-0282.s1>.

Corresponding author: Dikra Khedhaouiria, dikraa.khedhaouiria@canada.ca

DOI: 10.1175/JHM-D-19-0282.1

For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy \(www.ametsoc.org/PUBSReuseLicenses\)](https://www.ametsoc.org/PUBSReuseLicenses).

has revealed an increasing interest in the uncertainty assessment and an awareness of the limitation in using single deterministic products, especially during extreme events (Roundy et al. 2019; Krzysztofowicz 2001). So far, no other studies have developed PA ensembles at such spatiotemporal resolution (2.5-km grid spacing) for a large domain. The few gridded datasets providing an ensemble of precipitation realizations are daily products with coarser spatial resolutions (>10 km) and covering more extended periods (>30 years). Over the United States, Newman et al. (2015) extended the Clark and Slater (2006) study and proposed ensembles of daily precipitation at approximately 12-km resolution by interpolating site observations. Aalto et al. (2016) employed comprehensive interpolation approaches of seven climate variables observed at climate stations, including daily precipitation, to provide 10-km gridded products and their uncertainties over Finland. Europe-wide 100-member daily precipitation and temperature datasets were proposed by Cornes et al. (2018) on a 25-km grid resolution, solely based on interpolation from station observations. All of these gridded precipitation products, suited for climate studies, are highly dependent on the station density. As such, they are more likely to be inaccurate at places with sparse networks. PA has the advantage of relying on the dynamic modeling of atmospheric processes at those places (Fortin et al. 2015).

The general objective of the current study is to present an ensemble of high-resolution 6-h PA based on stochastic perturbations to provide uncertainty estimates. With this aim, an experimental probabilistic version of the Canadian Precipitation Analysis (CaPA) developed at Environment and Climate Change Canada (ECCC) is evaluated regarding quality attributes (e.g., reliability, skill; Murphy 1993) for events of interest. The added value of an ensemble version of the high-resolution CaPA was assessed through 24 members for the 2018 summer (July–August) and the 2018/19 winter (December–February).

The following section provides an overview of the ensemble PA development at ECCC, followed by a summary of the CaPA assimilation scheme in section 3. Section 4 describes the process used to create the ensembles and the experimental setup. The assessment method of the ensemble is provided in section 5, while the associated results are illustrated in section 6. Finally, section 7 provides a discussion and conclusions.

2. Overview of ensemble precipitation analysis development at ECCC

Over the North American domain, a useful alternative dataset to the scattered ground-based observations

is the CaPA (Mahfouf et al. 2007). CaPA is a gridded near-real-time quantitative precipitation estimates (QPE) developed by ECCC (Fortin et al. 2015; Lespinas et al. 2015). CaPA has been widely used as a reference for both research projects (see Fortin et al. 2018, for an extensive review) and operational uses. The CaPA system produces gridded precipitation fields based on numerical weather prediction (NWP, hereafter the background field) adjusted at each analysis step with observed precipitation (ground stations and radars) using optimal interpolation (OI) assimilation methods (Fortin et al. 2018). At ECCC, the latest CaPA version is currently available on a 10- and 2.5-km grid, both running at 6- and 24-h time resolution.

To provide ensemble PA, uncertainties in the components of the analysis need to be identified. The main sources of uncertainties in CaPA are those related to the surface observations (discussed in section 4 below) and to the background field. Currently, the High-Resolution Regional Deterministic Prediction System (HRDPS; Milbrandt et al. 2016), integrated on a national domain with 2.5-km grid spacing, provides the background field for CaPA. The uncertainty estimate of numerical prediction systems is ongoing research at ECCC and has led to the production of the Regional Ensemble Prediction System (REPS), operationally running across the North American domain with 10-km grid spacing since the summer of 2019. The direct use of the REPS members, generated with perturbed initial and boundary conditions and physical tendencies, as background fields is part of future research for the ensemble PA production, as the REPS resolution is presently too coarse for the targeted high resolution (2.5-km grid spacing).

Based on these considerations, this study is aimed toward the generation of ensembles based on stochastic methods, especially since such a CaPA ensemble already exists at ECCC. This ensemble is currently integrated as an internal module of the operational Canadian Land Data Assimilation System (CaLDAS; Carrera et al. 2015). One of the CaLDAS strengths is to use as forcing an ensemble PA to feed its ensemble Kalman filter (EnKF) land data assimilation system. Carrera et al. (2015) demonstrated that the use of PA, which benefits from the optimal combination of observations and past forecasts, resulted in better superficial soil moisture estimates and more accurate root-zone soil moisture analyses. Currently, an embedded CaPA within CaLDAS runs as many times as the number of desired members by assimilating perturbed background fields and observations. However, CaLDAS uses an outdated version of CaPA [see Table 1 in Fortin et al. (2018) for different CaPA versions] as regular updates

within the CaLDAS environment would imply important technical changes.

Despite very positive results in the current CaLDAS when simulating surface variables (e.g., soil moisture; [Carrera et al. 2015](#)), the newest version of CaPA is expected to substantially enhance those results. Indeed, the latest CaPA additionally assimilates radar QPE, includes more observation networks, and undergoes an enhanced quality control process ([Fortin et al. 2018](#)). The positive impact on CaPA performances of radar QPEs combined with a denser surface network was already demonstrated in [Fortin et al. \(2015\)](#) and [Lespinas et al. \(2015\)](#) studies. These results were confirmed by additional experiments (not shown for conciseness) that consisted in comparing performances of CaPA within CaLDAS to the most recent operational CaPA.

Improving CaLDAS is of high importance as many subsequent systems are highly dependent on its outputs. For example, the ECCC's Water Cycle Prediction System (WCPS; [Durnford et al. 2018](#)) would benefit from better CaLDAS runoff to improve modeling of river streamflows. The PA ensemble generated within CaLDAS is also of high interest for other users, even though this PA ensemble has never been available nor archived as no specific needs were identified in the past. An external ensemble that runs independently from CaLDAS with up-to-date observations and version is therefore proposed here, as the High-Resolution Ensemble Precipitation Analysis (HREPA), and would represent a more optimal solution for CaLDAS and other users. In this context, modifying CaLDAS will not require changes in its assimilation scheme as only CaLDAS forcings will be adjusted to account for an updated ensemble of PA, HREPA. This new CaLDAS configuration is currently running at ECCC and would allow a clear quantification of the impact of using HREPA on CaLDAS analyses.

Although the perturbation methodology (see [section 4](#) below) is similar to the one used in CaPA within CaLDAS, the present work differs from [Carrera et al. \(2015\)](#) study on two main points. First, precipitation from radars are added in the assimilation scheme and are perturbed for the ensemble generation, and second, it includes an in-depth verification of the ensemble regarding precipitation.

3. Canadian Precipitation Analysis system

The precipitation datasets involved in CaPA, namely, the observations and the background fields ([section 3a](#)), and the assimilation approach ([section 3b](#)) are briefly described below. For more specific details on CaPA, the

reader is referred to [Mahfouf et al. \(2007\)](#), [Lespinas et al. \(2015\)](#), and [Fortin et al. \(2018\)](#) studies.

a. Precipitation datasets

CaPA assimilates surface observations from 1) the synoptic (SYNOP) and the ECCC network with around 300 manual and 750 automatic stations; 2) the meteorological terminal aviation routine weather report (METAR) network with around 1200 sites that report precipitation occurrences; 3) the Réseau Météorologique Coopératif du Québec; Québec's provincial network, with 120 stations; 4) Ontario Ministry of Natural Resources and Forestry network with 60 automatic stations; and 5) the British Columbia Forest network with 228 automatic stations available during the summer. Additional networks reporting daily precipitation accumulation are also available ([Boluwade et al. 2017](#)) but are not appropriate for the 6-h temporal resolution. Thirty-three and 31 C-band radar QPEs, covering the United States and Canada (mostly located along the U.S. border), respectively, are assimilated in CaPA at 2.5 km. The new Canadian S-band radars are also being progressively added to the CaPA observation database and would ultimately enable better analysis.

Observed precipitation and radar QPE undergo an extensive quality control process to remove untrustworthy data. The QC is done automatically before each assimilation time step, leading to a time-varying number of assimilated observations, as described in [Lespinas et al. \(2015\)](#) (for surface observations) and in [Fortin et al. \(2015\)](#) (for radar QPEs). The impact of QC is substantial during the cold season. Solid precipitation is indeed challenging to measure at gauging sites due to large undercatch during windy conditions ([Rasmussen et al. 2012](#)). The radar QPEs are discarded as specific adjustments necessary for snow are not currently undertaken for CaPA (e.g., [Sivasubramaniam et al. 2018](#)), mainly due to the lack of reliable ground observations. The QC process eliminates around 75% of observed data in winter.

The 6-h CaPA background field for a given valid time corresponds to the 6–12-h HRDPS forecasts of precipitation accumulations, initialized 12 h before the valid time. For example, the analysis valid on 1 January 2020 at 1800 UTC uses the 6–12-h time window from the forecast initiated on 1 January 2020 at 0600 UTC. The HRDPS is developed at ECCC and is a limited area model (LAM) of the Global Environment Multiscale (GEM) atmospheric model ([Mailhot et al. 2006](#)). HRDPS uses land surface initial conditions from CaLDAS based on a coupled 2.5-km GEM 6-h cycle. It produces 48-h forecasts on a national domain at a 2.5-km grid spacing with initial and boundary

conditions provided by the Regional Deterministic Prediction System (RDPS; Vaillancourt et al. 2012). Complete information on HRDPS is provided in Milbrandt et al. (2016).

b. Optimal interpolation

In CaPA, observations O and background fields B are combined using the OI method to produce the precipitation analyses A . The OI is conducted on cubic root transforms $[\phi(\cdot)]$ of precipitation datasets following:

$$\phi(A) = \phi(B) + \sum_{k=1}^K W_k [\phi(O_k) - \phi(B_k)], \quad (1)$$

where subscript k denotes the k neighboring observation points, and W_k refers to the weight applied to $[\phi(O_k) - \phi(B_k)]$, called the innovation (Daley 1991). The OI is based on the spatial interpolation of innovations by using kriging approaches to estimate the error statistics of the surface observations, the radar QPEs, and the background fields. Based on these error statistics and imposing error analysis variances to be minimum, the OI weights are estimated as described in appendix A. The final analyses are obtained from back transformation with Eq. (1), followed by corrections to the biases induced by the cubic root transformation (Fortin et al. 2018; Evans 2013). More details on the interpolation scheme and its underlying hypotheses can be found in Mahfouf et al. (2007); Fortin et al. (2018), while technical specifications (e.g., configuration of the HRDPS) are available in ECCC (2018) guide.

4. Ensemble generation and experimental setup

Precipitation ensembles provide multiple probable outcomes of the same events (Liehti and Zappa 2019) and help quantify associated uncertainties, which is the goal of this study. A stochastic perturbation approach (section 4a) was selected to generate such an ensemble for CaPA and was applied for a specific experimental setup (section 4b).

a. Perturbation approach

Ensemble generation of the PA consists of adding stochastic perturbations in the analysis input components, which are known to have a certain degree of uncertainties. The CaPA uncertainties mainly result from errors in 1) the background field (e.g., parameterization of subgrid processes in the NWP models; Rummukainen 2016), 2) the observed datasets (e.g., inaccurate representation of light and extreme precipitation by tipping buckets; Tapiador et al. 2017), and 3) the radar composite (e.g., beam blocking, attenuation, superrefraction,

uncertain radar reflectivity–rain rate relationship; Villarini et al. 2008);

Following the method used in CaLDAS (Carrera et al. 2015), only phasing errors were considered when perturbing the background field. Accordingly, a mean displacement error of 0 km with a standard deviation of 25 km was randomly applied to the 6–12-h background field, independently on both the latitude and longitude directions.

Uncertainties for measurements from gauging sites and for radar QPEs are assumed to be on the magnitude of the precipitation events. Random perturbations are therefore sampled from a zero-mean Gaussian distribution with variance equal to σ_o^2 (σ_R^2) and are added to surface observation (radar QPEs) to create an ensemble of 24 equiprobable surface observation networks (radar QPEs). The σ_o^2 and σ_R^2 correspond to gauge and radar variance errors, and both are obtained by variographic analysis employed during the OI process (Fortin et al. 2015). The same perturbation was spatially applied within a radar beam, thus assuming a perfect spatial autocorrelation.

The spatiotemporal structure of perturbed surface observations and radar QPEs is not explicitly modeled, and two main reasons support this choice. First, HREPA sensitivity tests to the different perturbations (not shown for conciseness) only show a small impact on the PA quality when precipitation from surface observation and radar were not perturbed (see also Table 3 in Carrera et al. 2015). Also, the current approach of perturbing the background field maintains the spatiotemporal precipitation structure from the control realization, and this for all the perturbed members. More sophisticated approaches to perturb surface observations and radar QPEs could possibly improve HREPA (see section 7), but are beyond the scope of this study.

Random perturbations of observed, radar, and the background field precipitation are realized independently in the transformed space and are used later on to run CaPA OI (see section 3b). An approach such as the ensemble OI (EnOI) could also have enabled HREPA generation in a more comprehensive way (Evensen 2003; Counillon and Bertino 2009). The EnOI method allows the creation of background field ensembles through sampling from a long-term model integration. Therefore, EnOI provides background covariance estimates in a parsimonious approach (Evensen 2003). However, the current background field, corresponding to HRDPS forecasts, is operational only since December 2017, which is a relatively short period for precipitation. Moreover, HRDPS is continuously evolving, meaning that inhomogeneities may arise and in turn may be problematic for proper sampling.

At this stage and following the CaLDAS approach, 24 random PA members are generated. The choice of the ensemble size is dictated by the sensitivity analysis conducted in Carrera et al. (2015) study. The authors used the same perturbation approach but with an outdated CaPA version. With precipitation ensembles of 6, 12, 24, 36, and 48 members, (Carrera et al. 2015) demonstrated that the impact on different scores and soil moisture variability did not change beyond 24 members. A deterministic CaPA member using original observation and background field without any perturbations was additionally generated for the same period and is referred to as the control member. Finally, negative precipitation values obtained after the back transformation were set to zero.

b. Experimental setup

The summer 2018 and winter 2018/19, from 30 June to 31 August 2018 and from 1 December 2018 to 28 February 2019 (hereafter summer and winter), are respectively the two seasons used for HREPA evaluation across the 2.5-km high-resolution domain. Despite a high number of stations used during the assimilation process, only 551 SYNOP (176 manual SYNOP) stations during summer (winter) are examined to increase reliability in the HREPA performance assessment.

5. Ensemble evaluation framework

The 6-h precipitation ensemble members evaluation focuses on two points: (i) the performance of each of the members with deterministic scores typically used for CaPA verification (Fortin et al. 2018; Lespinas et al. 2015) and (ii) the performance of the joint distribution of the ensemble and the observations (hereafter probabilistic verification, Murphy 1993). The first type provides an overview of how the members compare to each other and observations (section 5a), while the second one allows for the comparison of different event outcomes (the members) with the single associated observation (section 5b).

a. Deterministic verification

The deterministic evaluation was here conducted to provide information on the bias and skill of each member using the frequency bias index (FBI) and the equitable threat scores (ETS), respectively. Both metrics are based on the 2×2 contingency table for binary events, where an event is defined equal to 1 when precipitation above a selected threshold occur and 0 otherwise. To study different precipitation events, thresholds of the 0th, 20th, 50th, 70th, 80th, 95th, and 99th percentiles are used based on the nonzero observed precipitation

distribution. In the following, a nonzero precipitation is defined as a 6-h precipitation accumulation above 0.2 mm. Such a threshold is necessary because of frequent low-intensity precipitation from the background field.

The FBI and ETS are defined as

$$\begin{aligned} \text{FBI} &= \frac{h + f_a}{h + m}, \\ \text{ETS} &= \frac{h - h_r}{h + m + f_a - h_r}, \end{aligned} \quad (2)$$

in which h , m , f_a , c are the hits, the misses, the false alarms, and the correct negatives, respectively; h_r corresponds to random hits and expresses as follows $h_r = (h + m)(h + f_a)/(h + m + f + c)$. The FBI compares the frequency of events in the analyses to those in the observations. The ETS assesses the proportion of hits (events issued by the analyses did occur) for a given precipitation event, while accounting for hits expected by chance. For practical reasons, FBI-1—which is equivalent to the normalized difference between false alarms and missed events—is preferred in the following as positive (negative) values implies positive (negative) bias. FBI-1 ranges from -1 to $+\infty$ with perfect score of 0, whereas ETS ranges from $-1/3$ to 1 with perfect score of 1.

b. Joint-distribution verification

The joint probability distribution of HREPA f and observed precipitation events x — $\Pr(x, f)$ —is also investigated. Different features of the ensemble quality aspects are illustrated through the relationship between ensemble members and one observation (Bradley et al. 2016). Several scores are identified to represent different attributes of the ensemble, and are based on the decomposition of $\Pr(x, f)$ into conditional, $\Pr(x|f)$ and $\Pr(f|x)$, and marginal distributions, $\Pr(x)$ and $\Pr(f)$ (Murphy and Winkler 1987; Bradley et al. 2016). The 6-h precipitation time series are discretized into binary events using the same thresholds presented above. The observed event x takes the value of 1 or 0 depending on its occurrence or nonoccurrence. In HREPA, binary events are defined by their relative frequency of occurrence f in the 24-member ensemble and vary from 0 to 1.

Reliability is one of the most desirable attributes of an ensemble. It characterizes the relation between the observed frequency of an event and its forecast probability. Reliability is therefore a property of $\Pr(x|f)$ and $\Pr(f)$. An ensemble is said to be reliable when both probabilities agree. Combined with a given level of reliability, sharpness is another valuable attribute in ensemble evaluation and is especially useful in decision-making (Bradley et al. 2016; Gneiting et al. 2007).

Sharpness provides information on the ensemble variability without considering the observations. The histogram of the marginal $\Pr(f)$ gives insight on the ensemble sharpness, where a U-shape histogram indicates that probabilities cluster around 0 and 1 and suggest the sharpness of the ensemble. On the other hand, a bell shape or flat histogram indicates no sharpness. Another relevant attribute is resolution, which informs on how much the ensemble deviates from the observed climatology, with implications on the ensemble's ability to resolve different outcomes.

The attribute diagram is here selected to illustrate in one figure the reliability, the sharpness, and the resolution of the ensemble. For each event (e.g., precipitation ≥ 0.2 mm), the attribute diagram illustrates (i) $\Pr(x|f)$ (y axis) against $\Pr(f)$ (x axis) curve which for a reliable ensemble would lie on the diagonal, (ii) the histogram of $\Pr(f)$ as an inset in the diagram to illustrate the sharpness, and (iii) the variation around observed climatology line for the resolution. Following [Bröcker and Smith \(2007\)](#), a resampling approach is used to assign consistency bars to the observed relative frequencies. The estimated 5%–95% consistency bars are plotted around the mean of $\Pr(x|f)$ for each forecast probability bin and illustrate the expected sampling error. The idea is to graphically evaluate the reliability of an ensemble while considering the fluctuation of the observed relative frequencies.

The skill of an ensemble is an attribute that informs on the improvement of the studied system relative to a reference. To this end, the Brier skill score (BSS), based on the Brier score (BS), was used here and is expressed as follows:

$$\text{BSS} = 1 - \frac{\text{BS}}{\text{BS}_{\text{clim}}}, \quad (3)$$

where the reference is the observed climatology (clim). The BS (see [appendix B](#) for details) represents an averaged squared error between x and f events and therefore illustrates the correspondence between the pairs of observed occurrences and associated frequencies in the ensemble. A skillful system is represented by positive BSS values when compared to the climatology, with 1 as a perfect score. Negative BSS values correspond to a system worse than the climatology.

The ensemble ability to discriminate dichotomous events is assessed through the receiver operating characteristics (ROC) curve. The ROC curve is based on contingency tables for probabilistic forecasts and relates the hit rate (HR, y axis) and false alarm rate (FAR, x axis) as the probability threshold varies. For each

precipitation event (e.g., precipitation ≥ 0.2 mm), ensemble probability bins (e.g., $[0, 0.1[$, \dots , $[0.9, 1.0]$) are first defined and are used as thresholds. The HR and the FAR are estimated using the predefined bins. The HR corresponds to the proportion of occurrences, meaning that both observed and forecasted precipitation are in the same category. The FAR is defined as the proportion of nonoccurrences (i.e., when precipitation in the analysis is not in the observed category). The integrated area under the ROC curve (AUC) summarizes the HREPA performance regarding the discrimination attribute. AUC is equal to 0.5 when HR is equal to FAR and therefore indicates that an ensemble is without skill, whereas AUC values close to 1 point to a skillful ensemble.

c. Regional cross-validation framework

The large domain with its contrasting regions results in high spatial variability of the event frequency, which in turn may mislead the interpretation of several scores ([Hamill and Whitaker 2006](#)). Accordingly, the verification was done both across the domain and for seven different climatic regions adapted from the Bukovsky classification ([Bukovsky 2012](#)). [Figure 1](#) presents the spatial distribution of the selected stations within the climatic regions. The original [Bukovsky \(2012\)](#) classification was adapted by pooling some regions, either for low station density reasons (e.g., northern regions) or because scores were similar between regions (not shown for conciseness).

Additionally, a leave-one-out cross-validation (LOOCV) method was applied to characterize the predictive capacity of HREPA. Despite high autocorrelations between validation sites ([Panthou et al. 2012](#)) and potential contrasting interpretations of regional results due to different station densities (e.g., north versus the southeast), this approach has the advantage to be simple and already implemented in the operational CaPA. Indeed, CaPA provides precipitation analyses and LOOCV estimates for each analysis time step and observation site. In this context, CaPA assimilates all stations minus one as a training sample for the calibration and the remaining station for validation. This step is repeated across all stations. The LOOCV for HREPA was conducted in a similar way across the 24 members and the control run.

6. Results and discussion

Scores presented below were estimated across different regions. Precipitation series from all sites of the domain or of a specific region were pooled together for the calculations of the different deterministic and

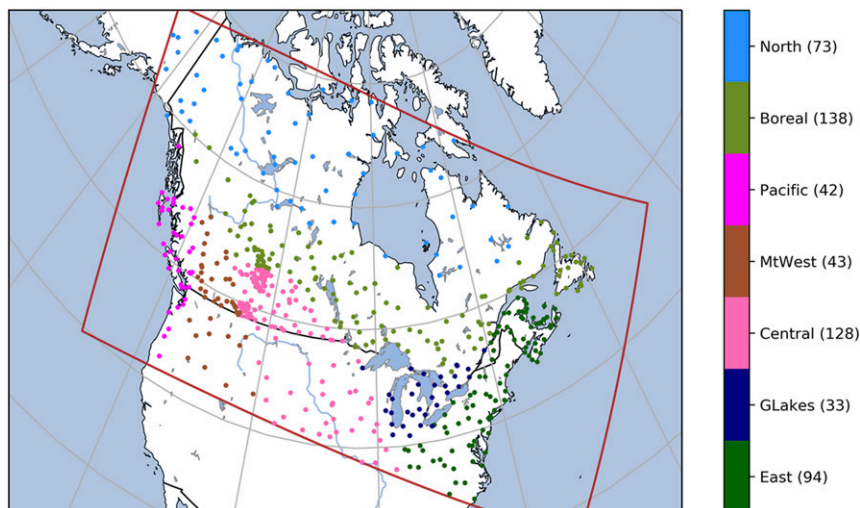


FIG. 1. Precipitation SYNOP stations used for verification and the analysis domain (dark red frame). The different climatic regions based on the Bukovsky (2012) classification are, from north to south and west to east: North, Boreal, Pacific, MtWest, Central, GLakes, and East. The number of stations within each region is indicated in parentheses.

probabilistic scores. Only the validation results of the LOOCV framework are presented in this section.

a. Frequency-based deterministic evaluation

The FBI-1 results for the summer experiment (Fig. 2) show that, for most regions, the 95% interval estimated from the 24 FBI-1 values exhibit relatively small spread with the control member lying within that interval. The mountainous (Pacific and MtWest) and, to a lesser extent, northern regions, differed and showed FBI-1 of the control member outside of this same interval. This counterintuitive result may be explained by the ranking of the control member precipitation distribution among the 24 distributions from the members. Indeed, the nonzero precipitation distribution of the control member lies within the distributions of the perturbed members for both seasons and all regions (not shown for conciseness). However, the 6-h precipitation amounts from the control member ranged as one of the smallest among the other members, especially for high quantiles. This last point leads to frequencies of events throughout the random members larger than in the control run and therefore implies FBI-1 values that are shifted toward higher values.

HREPA exhibits small biases across the domain and for most precipitation events (bottom-right panel, Fig. 2) with FBI-1 values below 0.2. The lowest bias (close to 0) is obtained for precipitation above the 70th percentile (around 2.5 mm). Differences in the FBI-1 values arise when evaluating the different regions but with generally the same pattern. Indeed, the FBI-1

reveals positive biases for the high- to middle-frequency events (corresponding to 0.2–2 mm and up to 3 mm for some regions), notably higher for the Great Lakes (GLakes) region, and to a lesser extent for the East and Central regions. In contrast, negative biases are obtained for low-frequency events (>90th percentile). These results reveal that too many small precipitation events are simulated by the members when compared to surface observations and inversely not enough precipitation events of high intensity.

The challenging representation of the summer precipitation regime in the background field may partly explain this underestimation for high-intensity precipitation events. These discrepancies may also be explained by the representativeness errors (Gervais et al. 2014) that contribute to this underestimation as areal precipitation in HREPA was here compared to point-scale observations. All of these results were consistent with those obtained in Fortin et al. (2015) and Lespinas et al. (2015) studies during a different summer and with other CaPA versions. Northern regions (North and Boreal) display different patterns with the FBI-1 oscillating slightly above and below 0. For these regions, precipitation regimes are less prone to convection and are therefore easier to represent in the background field.

The winter experiment, illustrated in Fig. 3, displays distinct results with systematically fewer events in the PA than in the observations as represented by the negative FBI-1 values for the control member. An evaluation of the background field (without any assimilation) against surface observations (not shown for conciseness)

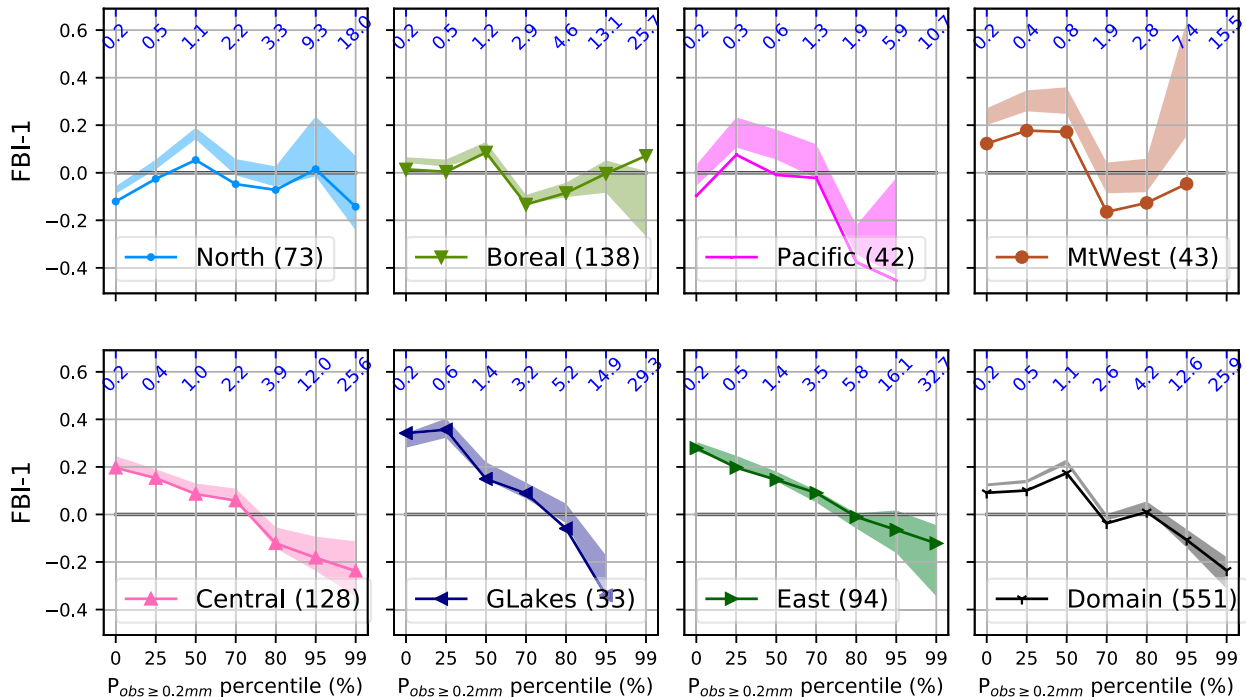


FIG. 2. FBI-1 for binary events defined for different percentiles of the nonzero precipitation ($P_{\text{obs} \geq 0.2 \text{ mm}}$) across the Bukovsky regions and the domain for the summer experiment. The top x axis with blue labels indicates the corresponding percentile value in millimeters. Solid lines and the associated shaded area correspond to, respectively, FBI-1 for the control member and the 95% interval of the 24 FBI-1 estimated from the perturbed members. Positive (negative) FBI-1 indicates positive (negative) bias in the frequency of precipitation events. The number of stations within each region is depicted in parentheses at the bottom of each panel.

exhibits systematic large negative frequency biases in the background field, which are significantly reduced with the assimilation of observations. For the same reasons, FBI-1 values from the control member were also outside the 95% intervals of the 24 FBI-1 values. However, the small number of stations for many regions during the winter made it challenging to evaluate.

The ETS values, during the summer experiment and for all regions (Fig. 4), exhibits increasing HREPA skills from high-frequency (0th–50th percentile) to middle-frequency events (50th–70th percentile), in contrast with decreasing skills for low-frequency events (>70th percentiles). ETS results are therefore consistent with those obtained for the FBI-1 (Fig. 2). Regional performance ranks are also consistent with those of FBI-1 results, where eastern regions displayed the best performance (ETS values up to 0.6 for the GLakes). As expected, the control member performed better than the perturbed members by being in the upper limit of the narrow 95% interval estimated from the 24 ETS values.

Figure 5 illustrates ETS values during the winter experiment. Despite fewer stations for that experiment, it is interesting to note that ETS values are generally larger than in summer (Fig. 4), implying that HREPA is more skillful in winter. This result was expected as the

background field is usually more skillful in predicting winter precipitation, generally associated with larger-scale meteorological fronts rather than summer convection. For many regions, ETS values displayed small variability among the different thresholds suggesting similar precipitation regime driving these events (unlike the different regimes occurring in summer). Finally, similarly to the FBI-1 performance during the winter, the control member is outside the narrow 95% interval of the 24 ETS values and is systematically more skillful than the perturbed members.

b. Probabilistic assessment

Figure 6 illustrates attribute diagrams across selected regions for some precipitation events. For conciseness, results for the remaining regions [Boreal, Mountain West (MtWest), and Central] are provided in Fig. S2 in the online supplemental material. The HREPA displays some reliability but was overconfident about precipitation occurrences (precipitation $\geq 0.2 \text{ mm}$) during the summer experiment (left panel, Fig. 6). HREPA overestimated the observed frequencies, illustrated by a slope shallower than 1, implying biases in the HREPA probability of precipitation when compared to observations at stations. The East and GLakes (not shown)

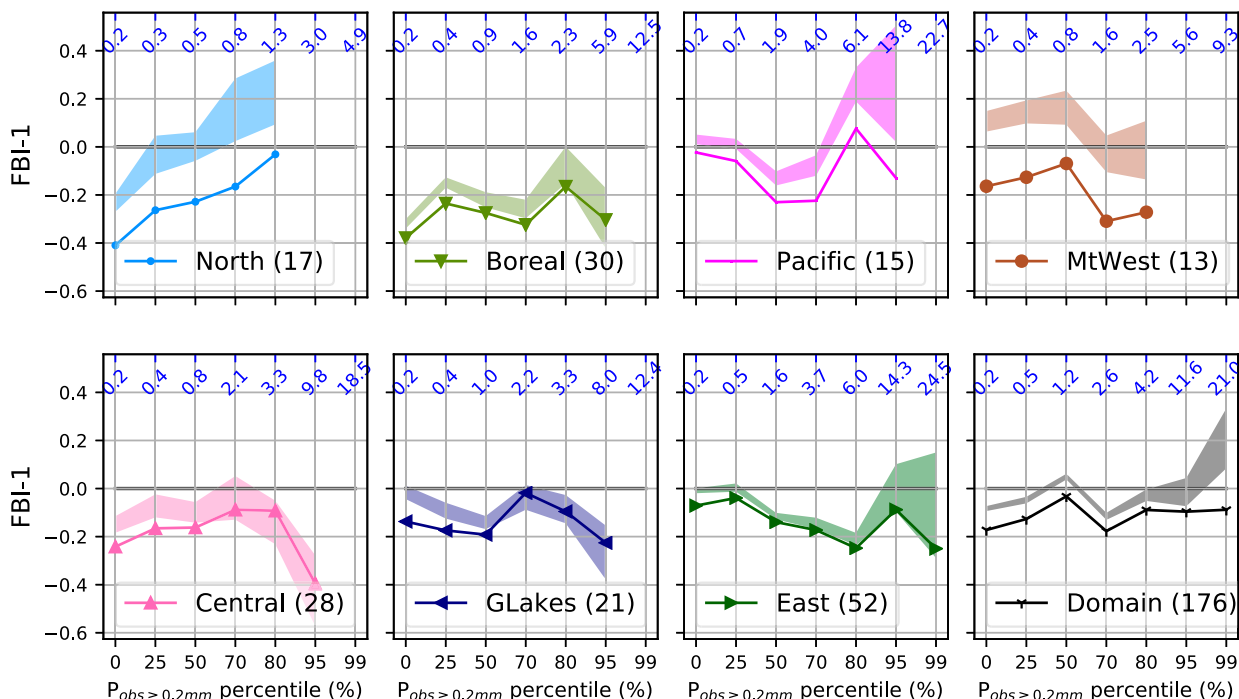


FIG. 3. As in Fig. 2, but for the winter experiment.

regions exhibit the highest conditional bias while the North, Pacific, and Boreal (Fig. S2) regions were the most reliable regarding the precipitation occurrence ($P \geq 0.2$ mm).

As illustrated in the middle and right panels of Fig. 6, the HREPA reliability increases for precipitation events of larger intensity. The slopes in the reliability diagram are steeper and closer to one, and HREPA frequencies lie within the consistency bars. Similarly to the FBI-1 performance in summer, the HREPA reliability increased up to events above the 70th–90th percentiles of the nonzero precipitation distribution. The sample size of events above the 95th percentile is too small to properly estimate the reliability diagram.

The HREPA sharpness is also illustrated in Fig. 6 as the histogram of HREPA probabilities $p(f)$. The shape of the histogram suggests that the 0–0.1 probability category bins, and to a lesser extent, the 0.9–1.0 category bin, is used much more often than others, revealing sharpness of the ensemble. Finally, HREPA shows high resolution as a wide range of frequency of observations is close to HREPA probabilities for all regions and events.

The reliability of the ensemble is slightly different during the winter experiment, as illustrated by Fig. 7 and Fig. S3. HREPA is highly reliable for the precipitation occurrence (left panel, Fig. 7) for almost all regions, as shown by the observed relative frequencies close to

the diagonal. In agreement with deterministic results, HREPA performs better in winter than in summer, except for the North and the MtWest regions (Fig. S3), where HREPA was generally overconfident. The Boreal region (Fig. S3) stayed underconfident but for probabilities below 0.7. Greater reliability of the ensemble is obtained for precipitation events above the median (middle panel, Fig. 7) across the domain. Although reliability seemed altered at the regional scales (e.g., ensemble being underconfident for the Boreal, Pacific, GLakes, and the East regions), the observed relative frequencies still lie within the consistency bars. The ensemble system is therefore still reliable considering sampling uncertainties in the observation, which are higher than during summer due to a smaller number of surface observations.

For precipitation events of larger intensities (right panel, Fig. 7), HREPA probabilities underestimate the observed frequencies for most regions but still generally remained within the consistency intervals. The Pacific and Boreal regions showed the lowest conditional bias, while the East region displayed the highest underestimations. Similarly to the summer experiment, the ensemble exhibits resolution and sharpness throughout the different regions and for all precipitation events.

The HREPA skill relative to the climatology was estimated with the BSS and is presented in Fig. 8. BSS values are positive across all regions for almost all

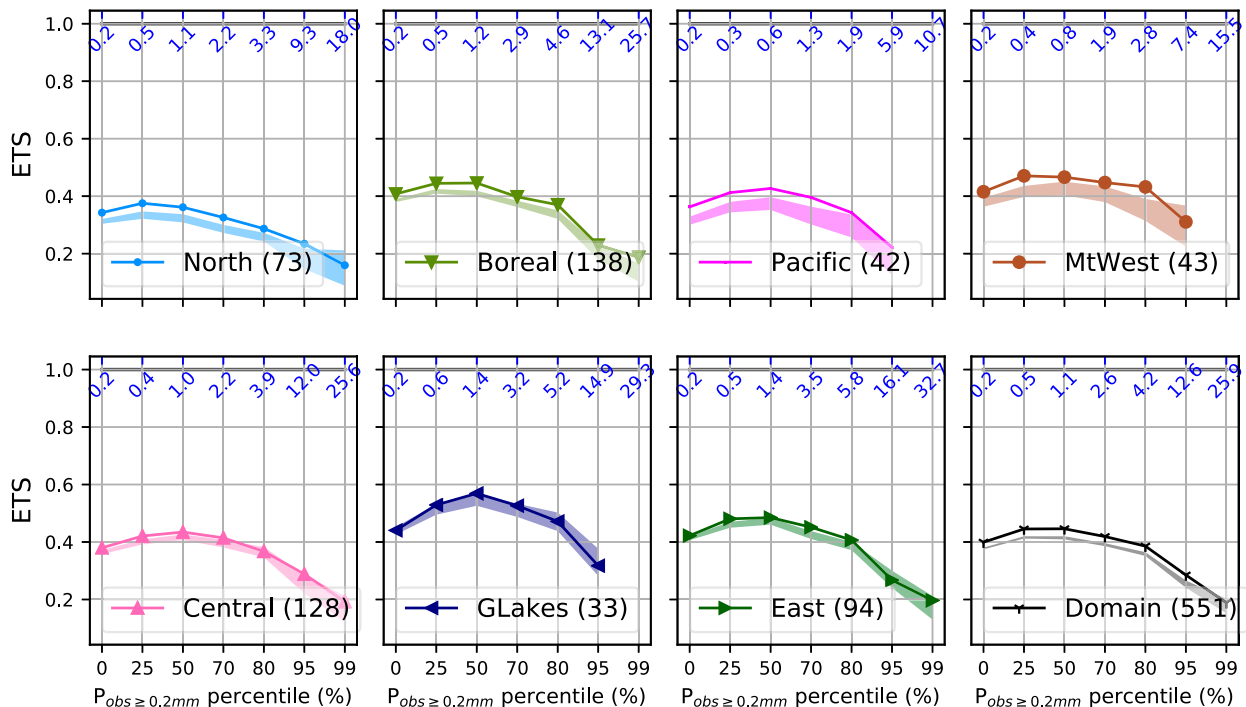


FIG. 4. ETS for binary events defined for different percentiles of the nonzero precipitation ($P_{\text{obs} \geq 0.2\text{mm}}$) across the Bukovsky regions and the domain for the summer experiment. The top x axis with the blue labels indicates the corresponding percentile value in millimeters. Solid lines and the associated shaded area correspond to, respectively, ETS for the control member and the 95% interval of the 24 ETS estimated from the perturbed members. ETS close to one corresponds to an ideal score.

precipitation events, meaning that HREPA outperforms forecasts based on climatology for both seasons. BSS shows summer performance in line with previous results in this study. HREPA skills (relative to the climatology) increase from events above the 0th percentile (0.2 mm) to events above the 70th–80th percentiles and decrease for events of larger intensities. Eastern regions (East and GLakes) display better predictive skills (up to 0.6) than other regions. The precipitation regime impacted the HREPA forecast skill at the seasonal scale for most regions with winter outperforming summer results. In such cases, the more likely synoptic systems occurring in winter, characterized by long duration and weak intensity with a relatively large spatial extent, are known to be better represented by the background field. BSS values were generally larger in winter, especially for the East region (BSS values above 0.7 for many events). Northern (North and Boreal) and MtWest regions show small differences between seasons.

The HREPA exhibit good discrimination between events and nonevents, as shown in Fig. 9. The area under the ROC curves, AUC, is indeed systematically over 0.75 for all events, regions, and seasons. The HREPA discrimination performance, when increasing precipitation thresholds, is quite similar to the other scores.

The AUC increases from the lowest threshold (events ≥ 0.2 mm) up to precipitation above the 50th–70th percentile and decreased for events of higher intensity. Winter shows better performances than summer for most regions and at the domain scale. AUC values are generally above 0.9 during winter for most precipitation events and exceeding 0.95 for eastern regions (GLakes, Central, and East). The North, Boreal, and MtWest regions show a different pattern with AUC being slightly higher during the summer. Similar to other scores, northern regions display the worst performance. It seems, however, reasonable to assume that HREPA should not perform as well over the northern regions where the volume of assimilated data is smaller than for the southern regions.

7. Summary

The Canadian Precipitation Analysis (CaPA) is extensively used at ECCC within its land data assimilating system (CaLDAS; Carrera et al. 2015) and more generally for several hydrometeorological applications (see Fortin et al. 2018). This study aims to present an upgrade of CaPA by providing uncertainty estimates through an ensemble of 6-h PA at ~ 2.5 -km grid spacing. By using

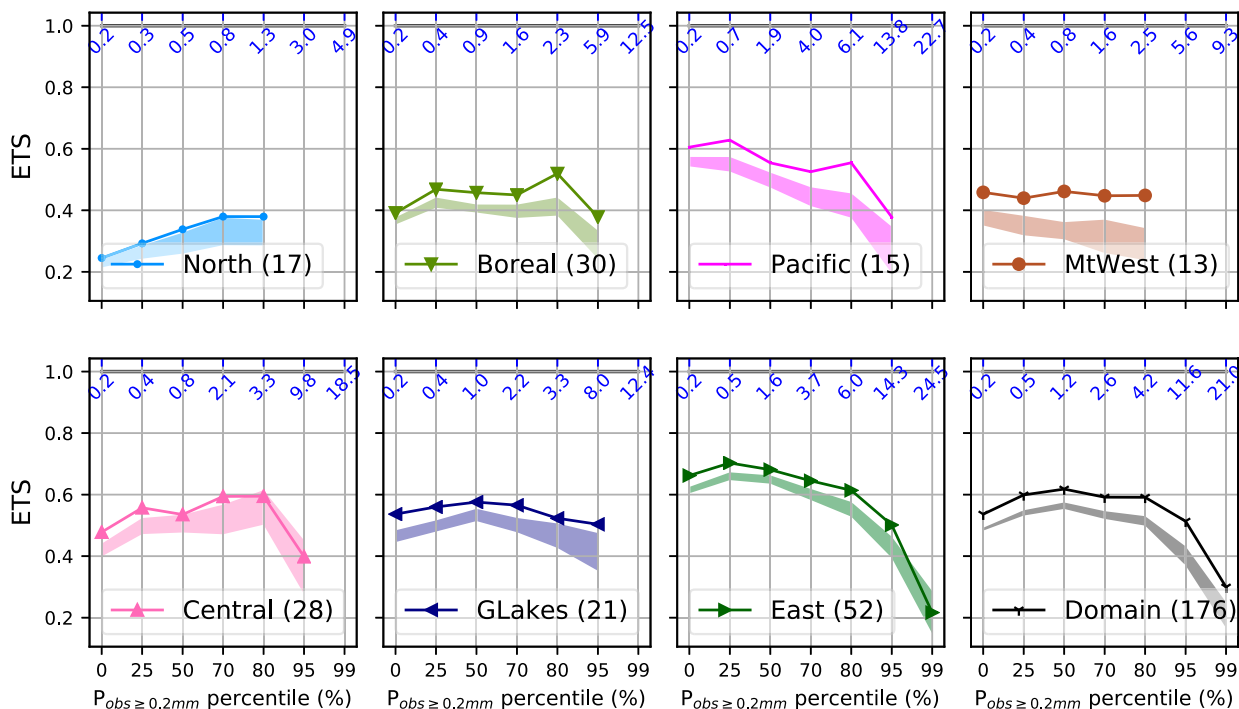


FIG. 5. As in Fig. 4, but for the winter experiment.

stochastic approaches, the three main components of the CaPA analysis, namely, (i) the at-site observations, (ii) the radar composites, and (iii) the background field, were randomly perturbed and used within a classical OI assimilation approach (Mahfouf et al. 2007; Fortin et al. 2018) to generate 24 realizations of the analysis plus one control analysis. Evaluation of this ensemble was conducted in summer (July–August 2018) and winter (December 2018–February 2019) in a cross-validation framework.

During the summer experiment, HREPA performances are impacted by the too frequent light precipitation [$< \sim 1 \text{ mm (6 h)}^{-1}$] in the background field relatively to at-site observations. When considering all precipitation events ($\geq 0.2 \text{ mm}$), the frequency biases (FBI-1) are positive across all HREPA members, the skill (ETS and BSS) are relatively low, and the ensemble is overconfident (reliability diagram). For medium to heavy precipitation, the ETS, reliability, and BSS metrics is increased, and the frequency bias is reduced across the members.

Inversely, the too few summer extreme events in the background field lead to negative frequencies bias and a decrease of the reliability and skill of the ensemble for this event category. This result can be partly explained by the scale mismatch between at-site and HREPA precipitation (representativeness errors). Only northern regions have small sensitivity to precipitation event

type, especially for bias-related scores. The frequency bias of the members oscillated around the zero value, and the ensemble reliability remained suitable for the different events. Differences among regions can, however, be observed. Eastern regions display the best predictive skills, while the northern regions were the least skillful for different metrics (ETS, BSS, and AUC). The low station density and the lack of radars in northern regions may partially explain the later result.

For the winter experiment, HREPA systematically detects less precipitation events than observed, resulting in negative FBI-1 values for all members regardless of the event types. This result is directly linked to the negative frequency bias found in the background field during that particular winter, which is significantly reduced after the assimilation of surface observations and radar QPEs. In addition, the attribute diagram shows underestimated observed relative frequencies but with some reliability when considering sampling uncertainties in the observations.

As expected, HREPA is more skillful (ETS, BSS, and AUC) in winter than in summer conditions. Winter synoptic systems (i.e., systems of long duration and low intensity with relatively large spatial extent) tend to be better represented in the background field. It is difficult to draw general conclusions regarding HREPA winter performances due to the low station density (barely 30% of what is used in summer).

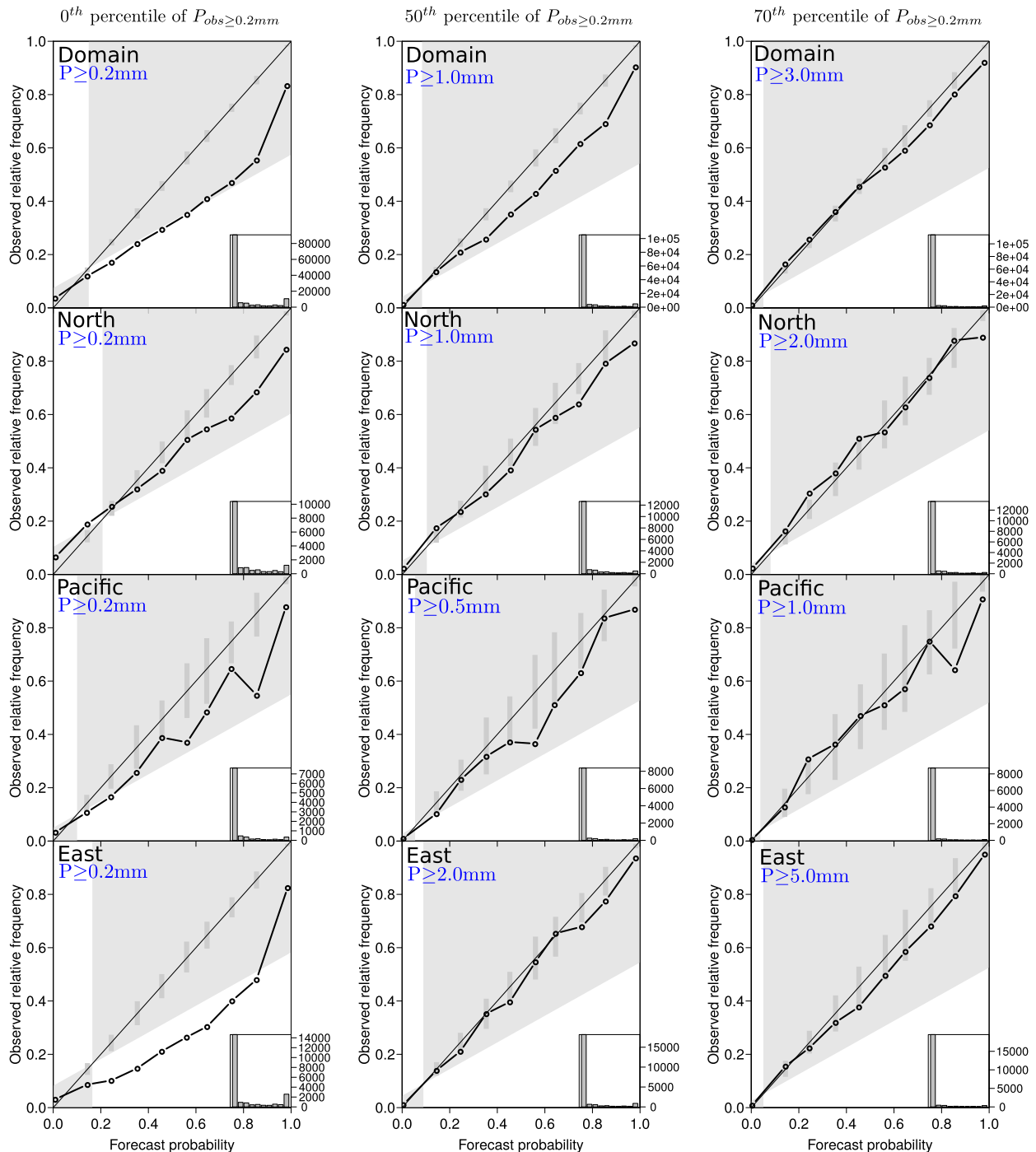


FIG. 6. Attribute diagrams for the precipitation above the (left) 0th, (center) 50th, and (right) 70th percentiles for the entire domain and for the North, Pacific, and East regions during the summer experiment. The gray lines illustrate the consistency bars with 95% confidence level. HREPA probabilities falling in the shaded gray area contribute to skill in reference to sample climatology. Sharpness diagrams are presented in the right corners of the plots.

Despite these limitations, both in the verification process and in the datasets, HREPA was demonstrated to be a reliable ensemble for different precipitation types and was shown to have resolution and sharpness for both

seasons. Extending the verification for transitioning seasons (spring and autumn) would add robustness on HREPA performance evaluation and is part of the next steps before HREPA become operational at ECCC.

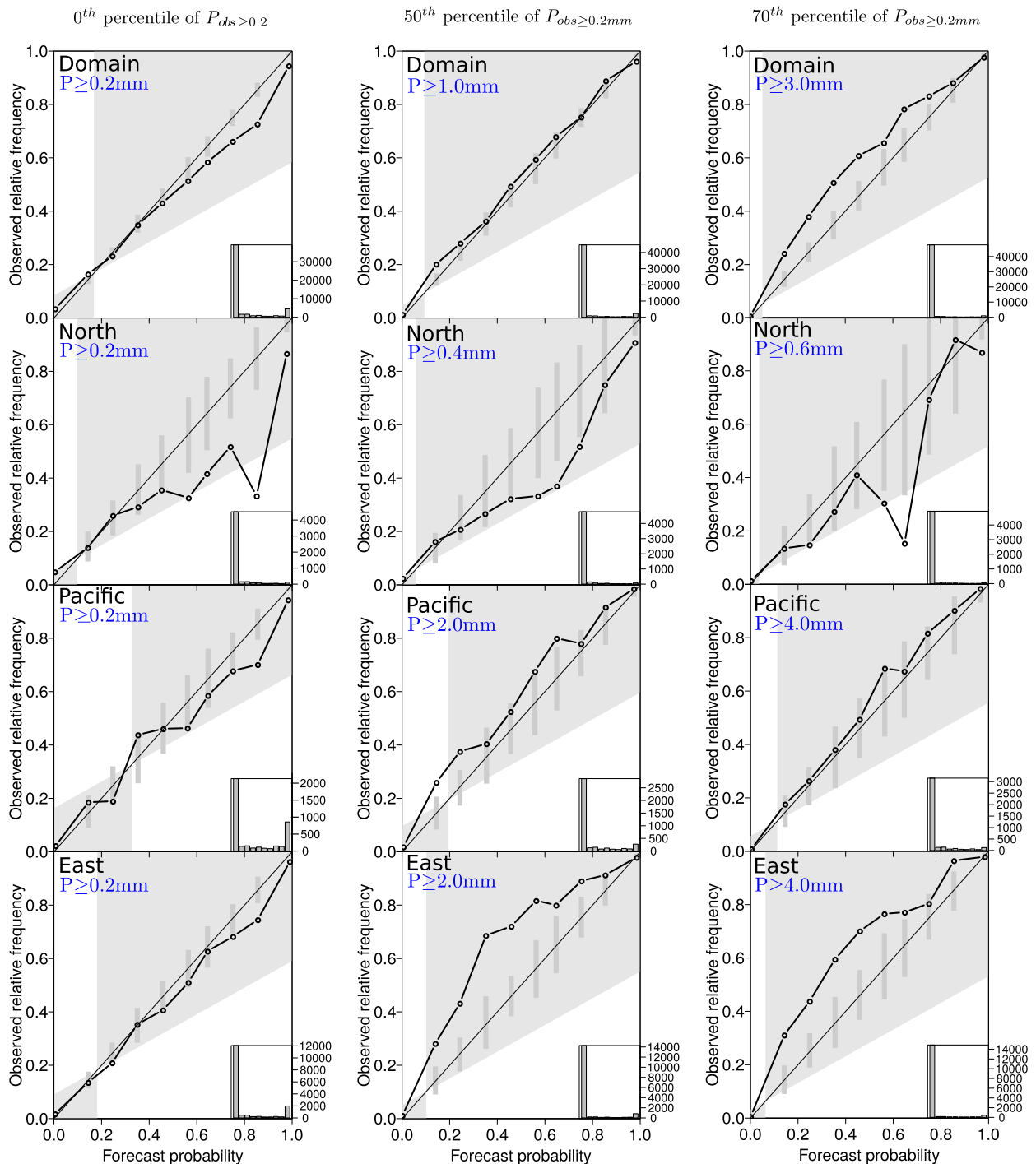


FIG. 7. As in Fig. 6, but for the winter experiment.

Projected improvements in the deterministic precipitation analysis (CaPA), such as the assimilation of satellite-based retrievals (Boluwade et al. 2017) combined with radar QPEs, the correction of the known bias in winter observations and the addition of new surface networks will potentially impact HREPA performance

positively. The observed biases and skill of CaPA, and hence HREPA, are linked to the background field quality. The ongoing research on improving precipitation forecasts from atmospheric models is expected to help reduce biases and increase the reliability of the ensemble. Furthermore, a fundamental issue contributing to this

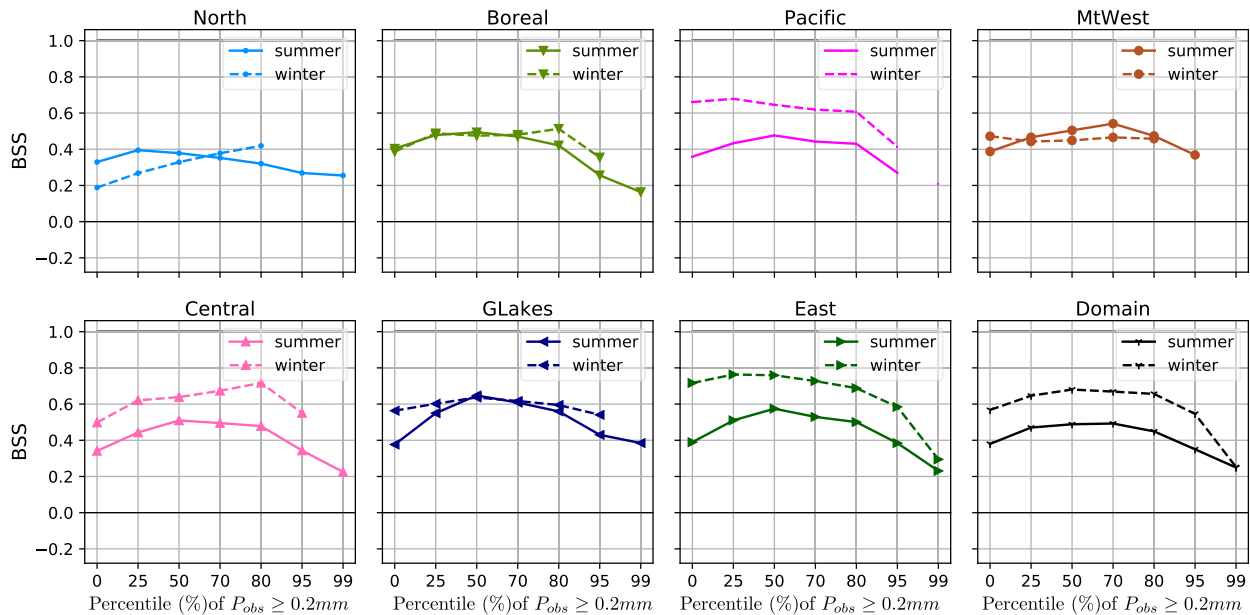


FIG. 8. BSS for nonzero precipitation above selected thresholds (x axis) across the domain and climatic regions during the summer (plain line) and winter (dashed line) experiments. BSS values above (below) 0 indicate a more (less) skillful ensemble than the climatological mean.

apparent bias and lack of skill for specific precipitation events, is the verification of HREPA (area average over a grid box) against point observations, especially pronounced for complex terrains (Lundquist et al. 2019). Using a different variable such as the soil

humidity or runoff from a land data assimilation system, such as CaLDAS, forced by HREPA could allow us to partly overcome this shortcoming and, at the same time, alleviate some of the leave-one-out cross-validation limitations.

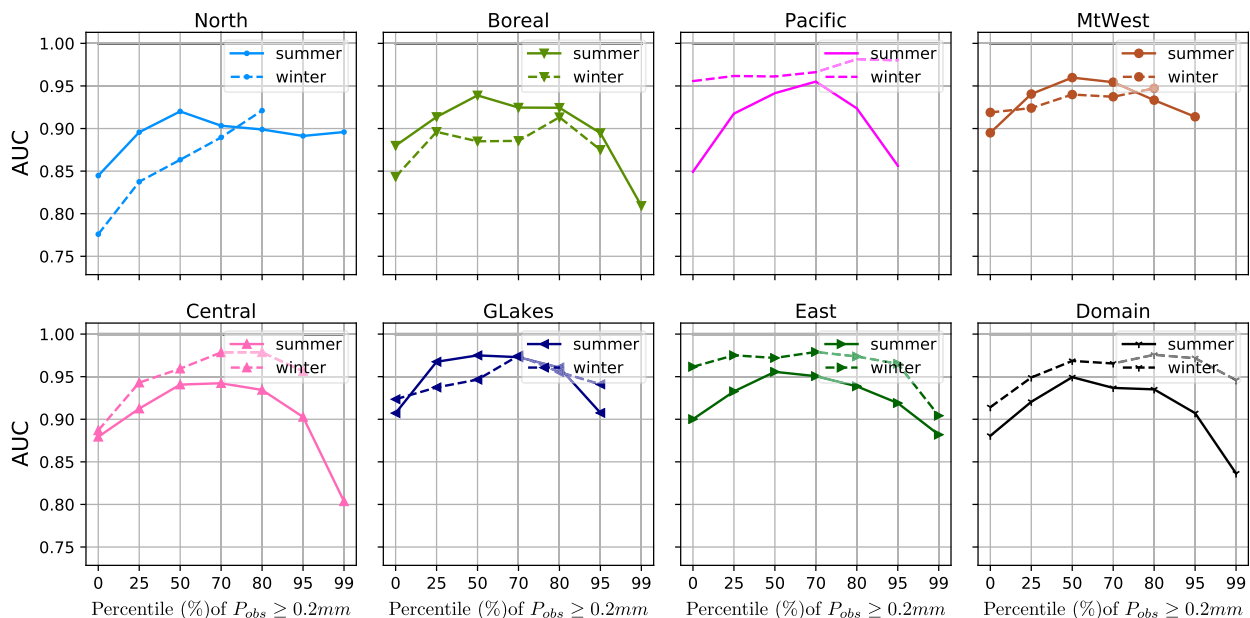


FIG. 9. Area under ROC curves (AUC) for nonzero precipitation above selected thresholds (x axis) across the domain and climatic regions during the summer (plain line) and winter (dashed line) experiments. AUC values close to 1 indicate that HREPA is able to discriminate between event and nonevents, while values around 0.5 correspond to a lack of discrimination capacity.

Beyond the improvements on the deterministic PA, one interesting research avenue is related to the use of perturbed precipitation background fields from the Regional Ensemble Prediction System (REPS) developed at ECCC. The REPS has the advantage to be physically consistent, meaning that spatial uncertainties in, for instance, mountainous areas are different from those in the Great Lakes region, and uncertainties in convection-type precipitation systems are different from synoptic systems. The added value of a physically based versus randomly based perturbations in the background fields on the PA would then be achieved. The challenge of the coarser resolution (~ 10 km versus ~ 2.5 km) will have to be addressed first. Hybrid approaches combining high-resolution deterministic BF to lower-resolution ensemble of background field is currently under study to provide flow-dependent background error covariances (Counillon et al. 2009). HREPA could also benefit from a more complex perturbation approach to represent uncertainties in the observations. Modeling the spatial correlation of radar errors to be used for random generation of radar QPEs (Cecinati et al. 2017) is one example of an approach to be tested.

APPENDIX A

Optimal Interpolation: Assimilation Scheme

In a manner equivalent to Eq. (1), and in the transformed space, the analysis x_A at a given point s_0 is given by

$$x_A(s_0) = x_B(s_0) + \mathbf{w}^T \mathbf{d}, \quad (\text{A1})$$

where x_A , x_B are the analysis and the background field, respectively. The vector \mathbf{w} corresponds to the weight vector to be estimated. Finally, the innovations \mathbf{d} at different points $S = \{s_1, s_1, \dots, s_n\}/s_0$ are written as

$$\mathbf{d} = \begin{bmatrix} x_O(s_1) - x_B(s_1) \\ \vdots \\ x_O(s_n) - x_B(s_n) \end{bmatrix} \quad (\text{A2})$$

The observations x_O and the background fields are assumed to be unbiased. For minimizing the analysis variance errors, it can be shown that the vector of weight must be equal to

$$\mathbf{w} = \mathbf{b}(\mathbf{R} + \mathbf{B})^{-1}, \quad (\text{A3})$$

where \mathbf{b} is the $n \times 1$ covariance vector of background errors at s_0 and sites S . The terms \mathbf{R} and \mathbf{B} are both $n \times n$ matrices and correspond to covariance matrices of the

observation and background errors at points S , respectively. Additional hypotheses are made on \mathbf{b} and on the elements of \mathbf{R} and \mathbf{B} matrices (Fortin et al. 2015), such as

- \mathbf{R} is a block diagonal matrix:

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}^{(1)} & 0 \\ 0 & \mathbf{R}^{(2)} \end{bmatrix}, \quad (\text{A4})$$

where $\mathbf{R}^{(1)} = \sigma_O^2 \mathbb{I}$, σ_O^2 corresponds to the error observation variances (no correlation between sites is assumed), while $\mathbf{R}^{(2)}$ is the covariance matrix of radar observation errors. Each element of $\mathbf{R}^{(2)}$ is assumed to be equal to $r_{ij}^{(2)} = \sigma_{rd}^2 \exp[-d_{rd}(i, j)/l_{rd}]$; where σ_{rd} , d_{rd} , l_{rd} are the error variances, correlation length, and distances between grid cell i and j of the radar composite, respectively.

- $B_{ij} = \sigma_B^2 \exp(-d_{ij}/l_B)$, where σ_B^2 , l_B , and d_{ij} are the error background variances, the correlation length, and distance between the site s_i and s_j , respectively.
- $b_{ij} = \sigma_B^2 \exp(-d_{ij}/l_B)$, where σ_B^2 , l_B , and d_{Oj} are the error background variances, the correlation length, and distance between the site s_O and s_j , respectively.

A crucial step is the estimation of the different error statistics to completely determine the OI weights and therefore the analysis. Currently, the error statistics (σ_B^2 and σ_O^2) are calculated by using kriging of the innovations. An experimental variogram of the innovations is calculated at each analysis time step and consist in relating the averaged squared deviation of the innovation to the distances (d) between two sites. Using the least squares method, the following theoretical variogram function is fitted to the empirical variogram:

$$\gamma(d)/2 = \begin{cases} \sigma_O^2 + \sigma_B^2 \left(1 - e^{-\frac{d}{l_B}}\right), & \text{if } d > 0, \\ 0, & \text{if } d = 0. \end{cases} \quad (\text{A5})$$

Similarly, interpolation of the differences between the observation and radar QPE is realized to estimate σ_{rd}^2 and l_{rd} . The errors statistics are therefore completely determined.

APPENDIX B

Calculation of Different Probabilistic Scores

a. Brier score and Brier skill scores

The Brier score (BS) assesses the accuracy of probability forecast for binary events (Wilks 2011) and is calculated as

$$BS = \frac{1}{N} \sum_{i=1}^N (x_i - f_i)^2, \quad (B1)$$

where x_i , f_i , and N denote, respectively, the observed binary event ($x_i = 1$ if the event occurs and 0 otherwise), the probability forecast of the same event, and the number of pairs of forecast and observation. Smaller BS values are desirable for an accurate forecast.

The Brier skill score (BSS) evaluates the forecast relative accuracy and allows comparisons among regions and seasons. A reference forecast is selected, usually the climatology, and the BS is normalized as follows:

$$BSS = 1 - \frac{BS}{BS_{\text{clim}}}, \quad (B2)$$

with

$$BS_{\text{clim}} = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2, \quad (B3)$$

where \bar{x} corresponds to the observed climatological probability. A (negative) positive BSS indicates that the forecast (underperform) outperform the climatology.

REFERENCES

- Aalto, J., P. Pirinen, and K. Jylhä, 2016: New gridded daily climatology of Finland: Permutation-based uncertainty estimates and temporal trends in climate. *J. Geophys. Res. Atmos.*, **121**, 3807–3823, <https://doi.org/10.1002/2015JD024651>.
- Beck, H. E., and Coauthors, 2019: Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. *Hydrol. Earth Syst. Sci.*, **23**, 207–224, <https://doi.org/10.5194/hess-23-207-2019>.
- Boluwade, A., T. Stadnyk, V. Fortin, and G. Roy, 2017: Assimilation of precipitation estimates from the integrated multisatellite retrievals for GPM (IMERG, early run) in the Canadian Precipitation Analysis (CaPA). *J. Hydrol. Reg. Stud.*, **14**, 10–22, <https://doi.org/10.1016/j.ejrh.2017.10.005>.
- Bradley, A. A., J. Demargne, and K. J. Franz, 2016: Attributes of forecast quality. *Handbook of Hydrometeorological Ensemble Forecasting*, Q. Duan et al., Eds., Springer, 1–44.
- Bröcker, J., and L. A. Smith, 2007: Increasing the reliability of reliability diagrams. *Wea. Forecasting*, **22**, 651–661, <https://doi.org/10.1175/WAF993.1>.
- Buizza, R., 2019: Ensemble forecasting and the need for calibration. *Statistical Postprocessing of Ensemble Forecasts*, S. Vannitsem, D. Wilks, and J. Messner, Eds., Elsevier, 15–48.
- Bukovsky, M., 2012: Masks for the Bukovsky regionalization of North America. Regional Integrated Sciences Collective, Institute for Mathematics Applied to Geosciences, National Center for Atmospheric Research, accessed 5 June 2019, <http://www.narccap.ucar.edu/contrib/bukovsky/>.
- Carrera, M. L., S. Bélair, and B. Bilodeau, 2015: The Canadian Land Data Assimilation System (CaLDAS): Description and synthetic evaluation study. *J. Hydrometeor.*, **16**, 1293–1314, <https://doi.org/10.1175/JHM-D-14-0089.1>.
- Cecinati, F., M. A. Rico-Ramirez, G. B. Heuvelink, and D. Han, 2017: Representing radar rainfall uncertainty with ensembles based on a time-variant geostatistical error modelling approach. *J. Hydrol.*, **548**, 391–405, <https://doi.org/10.1016/j.jhydrol.2017.02.053>.
- Chandrasekar, V., Y. Wang, and H. Chen, 2012: The CASA quantitative precipitation estimation system: A five year validation study. *Nat. Hazards Earth Syst. Sci.*, **12**, 2811–2820, <https://doi.org/10.5194/nhess-12-2811-2012>.
- Clark, M. P., and A. G. Slater, 2006: Probabilistic quantitative precipitation estimation in complex terrain. *J. Hydrometeor.*, **7**, 3–22, <https://doi.org/10.1175/JHM474.1>.
- Cornes, R. C., G. van der Schrier, E. J. M. van den Besselaar, and P. D. Jones, 2018: An ensemble version of the E-OBS temperature and precipitation datasets. *J. Geophys. Res. Atmos.*, **123**, 9391–9409, <https://doi.org/10.1029/2017JD028200>.
- Counillon, F., and L. Bertino, 2009: Ensemble optimal interpolation: Multivariate properties in the Gulf of Mexico. *Tellus*, **61A**, 296–308, <https://doi.org/10.1111/j.1600-0870.2008.00383.x>.
- , P. Sakov, and L. Bertino, 2009: Application of a hybrid EnKF-OI to ocean forecasting. *Ocean Sci.*, **5**, 389–401, <https://doi.org/10.5194/os-5-389-2009>.
- Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press, 457 pp.
- Duan, Q., F. Pappenberger, J. Thielen, H. Cloke, and J. Schaake, 2019: *Handbook of Hydrometeorological Ensemble Forecasting*. Springer, 1528 pp.
- Durnford, D., and Coauthors, 2018: Toward an operational water cycle prediction system for the Great Lakes and St. Lawrence River. *Bull. Amer. Meteor. Soc.*, **99**, 521–546, <https://doi.org/10.1175/BAMS-D-16-0155.1>.
- ECCC, 2018: High Resolution Deterministic Precipitation Analysis System (CaPA-HRDPA): Implementation of version 4.0.0. Tech. Note, 20 pp., https://collaboration.cmc.ec.gc.ca/cmc/cmoin/product_guide/docs/lib/CAPA-HRDPA_4_1_0_Tech_note_e.pdf.
- Evans, A. M., 2013: Investigation of enhancements to two fundamental components of the statistical interpolation method used by the Canadian Precipitation Analysis (CaPA). M.S. thesis, Dept. of Civil Engineering, University of Manitoba, 307 pp., <https://hdl.handle.net/1993/22276>.
- Evensen, G., 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367, <https://doi.org/10.1007/s10236-003-0036-9>.
- Fortin, V., G. Roy, N. Donaldson, and A. Mahidjiba, 2015: Assimilation of radar quantitative precipitation estimations in the Canadian Precipitation Analysis (CaPA). *J. Hydrol.*, **531**, 296–307, <https://doi.org/10.1016/j.jhydrol.2015.08.003>.
- , —, T. Stadnyk, K. Koenig, N. Gasset, and A. Mahidjiba, 2018: Ten years of science based on the Canadian precipitation analysis: A CaPA system overview and literature review. *Atmos.–Ocean*, **56**, 178–196, <https://doi.org/10.1080/07055900.2018.1474728>.
- Gervais, M., J. R. Gyakum, E. Atallah, L. B. Tremblay, and R. B. Neale, 2014: How well are the distribution and extreme values of daily precipitation over North America represented in the community climate system model? A comparison to reanalysis, satellite, and gridded station data. *J. Climate*, **27**, 5219–5239, <https://doi.org/10.1175/JCLI-D-13-00320.1>.
- Gneiting, T., F. Balabdaoui, and A. E. Raftery, 2007: Probabilistic forecasts, calibration and sharpness. *J. Roy. Stat. Soc.*, **69B**, 243–268, <https://doi.org/10.1111/j.1467-9868.2007.00587.x>.
- Hamill, T. M., and J. S. Whitaker, 2006: Probabilistic quantitative precipitation forecasts based on reforecast analogs: Theory

- and application. *Mon. Wea. Rev.*, **134**, 3209–3229, <https://doi.org/10.1175/MWR3237.1>.
- Hopkinson, R. F., D. W. McKenney, E. J. Milewska, M. F. Hutchinson, P. Papadopol, and L. A. Vincent, 2011: Impact of aligning climatological day on gridding daily maximum-minimum temperature and precipitation over Canada. *J. Appl. Meteor. Climatol.*, **50**, 1654–1665, <https://doi.org/10.1175/2011JAMC2684.1>.
- Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. Xie, 2004: 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, [https://doi.org/10.1175/1525-7541\(2004\)005<0487:CAMTPG>2.0.CO;2](https://doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2).
- Kidd, C., and V. Levizzani, 2011: Status of satellite precipitation retrievals. *Hydrol. Earth Syst. Sci.*, **15**, 1109–1116, <https://doi.org/10.5194/hess-15-1109-2011>.
- Krzysztofowicz, R., 2001: The case for probabilistic forecasting in hydrology. *J. Hydrol.*, **249**, 2–9, [https://doi.org/10.1016/S0022-1694\(01\)00420-6](https://doi.org/10.1016/S0022-1694(01)00420-6).
- Lamb, R., D. Faulkner, P. Wass, and D. Cameron, 2016: Have applications of continuous rainfall–runoff simulation realized the vision for process-based flood frequency analysis? *Hydrol. Processes*, **30**, 2463–2481, <https://doi.org/10.1002/hyp.10882>.
- Lespinas, F., V. Fortin, G. Roy, P. Rasmussen, and T. Stadnyk, 2015: Performance evaluation of the Canadian Precipitation Analysis (CaPA). *J. Hydrometeorol.*, **16**, 2045–2064, <https://doi.org/10.1175/JHM-D-14-0191.1>.
- Liechti, K., and M. Zappa, 2019: Verification of short-range hydrological forecasts. *Statistical Postprocessing of Ensemble Forecasts*, S. Vannitsem, D. Wilks, and J. Messner, Eds., Elsevier, 954–974.
- Lundquist, J., M. Hughes, E. Gutmann, and S. Kapnick, 2019: Our skill in modeling mountain rain and snow is bypassing the skill of our observational networks. *Bull. Amer. Meteor. Soc.*, **100**, 2473–2490, <https://doi.org/10.1175/BAMS-D-19-0001.1>.
- Mahfouf, J.-F., B. Brasnett, and S. Gagnon, 2007: A Canadian precipitation analysis (CaPA) project: Description and preliminary results. *Atmos.–Ocean*, **45**, 1–17, <https://doi.org/10.3137/ao.v450101>.
- Mailhot, J., and Coauthors, 2006: The 15-km version of the Canadian regional forecast system. *Atmos.–Ocean*, **44**, 133–149, <https://doi.org/10.3137/ao.440202>.
- Milbrandt, J. A., S. Bélair, M. Faucher, M. Vallée, M. L. Carrera, and A. Glazer, 2016: The pan-Canadian high resolution (2.5 km) deterministic prediction system. *Wea. Forecasting*, **31**, 1791–1816, <https://doi.org/10.1175/WAF-D-16-0035.1>.
- Murphy, A. H., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, **8**, 281–293, [https://doi.org/10.1175/1520-0434\(1993\)008<0281:WIAGFA>2.0.CO;2](https://doi.org/10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2).
- , and R. L. Winkler, 1987: A general framework for forecast verification. *Mon. Wea. Rev.*, **115**, 1330–1338, [https://doi.org/10.1175/1520-0493\(1987\)115<1330:AGFFV>2.0.CO;2](https://doi.org/10.1175/1520-0493(1987)115<1330:AGFFV>2.0.CO;2).
- Newman, A. J., and Coauthors, 2015: Gridded ensemble precipitation and temperature estimates for the contiguous United States. *J. Hydrometeorol.*, **16**, 2481–2500, <https://doi.org/10.1175/JHM-D-15-0026.1>.
- Panthou, G., T. Vischel, T. Lebel, J. Blanchet, G. Quantin, and A. Ali, 2012: Extreme rainfall in West Africa: A regional modeling. *Water Resour. Res.*, **48**, W08501, <https://doi.org/10.1029/2012WR012052>.
- Pappenberger, F., and K. J. Beven, 2006: Ignorance is bliss: Or seven reasons not to use uncertainty analysis. *Water Resour. Res.*, **42**, W05302, <https://doi.org/10.1029/2005WR004820>.
- Rasmussen, R., and Coauthors, 2012: How well are we measuring snow: The NOAA/FAA/NCAR winter precipitation test bed. *Bull. Amer. Meteor. Soc.*, **93**, 811–829, <https://doi.org/10.1175/BAMS-D-11-00052.1>.
- Raut, B. A., A. W. Seed, M. J. Reeder, and C. Jakob, 2018: A multiplicative cascade model for high-resolution space-time downscaling of rainfall. *J. Geophys. Res. Atmos.*, **123**, 2050–2067, <https://doi.org/10.1002/2017JD027148>.
- Roundy, J. K., Q. Duan, and J. C. Schaake, 2019: Hydrological predictability, scales, and uncertainty issues. *Handbook of Hydrometeorological Ensemble Forecasting*, Q. Duan et al., Eds., Springer, 1–29.
- Rummukainen, M., 2016: Added value in regional climate modeling. *Wiley Interdiscip. Rev.: Climate Change*, **7**, 145–159, <https://doi.org/10.1002/WCC.378>.
- Serinaldi, F., and C. G. Kilsby, 2014: Simulating daily rainfall fields over large areas for collective risk estimation. *J. Hydrol.*, **512**, 285–302, <https://doi.org/10.1016/j.jhydrol.2014.02.043>.
- Sivasubramaniam, K., A. Sharma, and K. Alfredsen, 2018: Estimating radar precipitation in cold climates: The role of air temperature within a non-parametric framework. *Hydrol. Earth Syst. Sci.*, **22**, 6533–6546, <https://doi.org/10.5194/hess-22-6533-2018>.
- Tapiador, F., and Coauthors, 2017: Global precipitation measurements for validating climate models. *Atmos. Res.*, **197**, 1–20, <https://doi.org/10.1016/j.atmosres.2017.06.021>.
- Vaillancourt, P., and Coauthors, 2012: Improvements to the Regional Deterministic Prediction System (RDPS) from version 2.0. 0 to version 3.0. Canadian Meteorological Centre Tech. Note, 78 pp., http://collaboration.cmc.ec.gc.ca/cmc/cmci/product_guide/docs/lib/op_systems/doc_opchanges/technote_rdp300_20121003_e.pdf.
- Villarini, G., P. V. Mandapaka, W. F. Krajewski, and R. J. Moore, 2008: Rainfall and sampling uncertainties: A rain gauge perspective. *J. Geophys. Res.*, **113**, D11102, <https://doi.org/10.1029/2007JD009214>.
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. International Geophysics Series, Vol. 100, Academic Press, 704 pp.