Using a Hybrid Optimal Interpolation-Ensemble Kalman Filter for the Canadian Precipitation Analysis

Dikraa Khedhaouiria[1], Stéphane Bélair[1], Vincent Fortin[1], Guy Roy[2], and Franck Lespinas[2]

[1]Meteorological Research Division, Environment and Climate Change Canada, Dorval, QC, Canada
[2]Meteorological Service of Canada, Environment and Climate Change Canada, Dorval, QC, Canada

Correspondence: Dikraa Khedhaouiria (dikraa.khedhaouiria@ec.gc.ca)

Abstract. Several data assimilation (DA) approaches exist to generate consistent and continuous precipitation fields valuable for hydrometeorological applications and land data assimilation. Usually, DAs are based on either static or dynamic approaches. Static methods rely on deterministic forecasts to estimate background error covariance matrices, while dynamic ones use ensemble forecasts. Associating the two methods is known as hybrid DA and has proven beneficial for different applications as it combines the advantages of both approaches. The present study intends to explore hybrid DA for the 6-hour Canadian Precipitation Analysis (CaPA). Based on optimal interpolations (OI), CaPA blends forecasts and observations from surface stations and ground-based radar datasets to provide precipitation fields over the North American domain. The application of hybrid DA to CaPA consisted of finding the optimal linear combination between i) an OI based on the Regional Deterministic Prediction System (RDPS) and ii) an Ensemble Kalman Filter (EnKF) based on the 20-member Regional Ensemble Prediction System (REPS). The results confirmed the known effectiveness of the hybrid approach when low-density observation networks are assimilated. Indeed, the experiments conducted for the summer without radar datasets and for the winter (characterized by very few observations in CaPA) showed that attributing a relatively high weight of the EnKF (50 and 70% for summer and winter, respectively) gave better analysis skills and a reduction of false alarms than the OI method. A deterioration of the moderate to high-intensity precipitation bias was, however, observed during summer. Reducing to 30% the weight attributed to the EnKF permitted alleviating the bias deterioration while improving skill compared to the OI-based CaPA.

1 Introduction

Errors inherent in the observed data, the parameterization of subgrid processes, initial conditions, model physics, to name a few, combined with the chaotic nature of the atmosphere (Lorenz, 1963), lead to uncertain NWP forecasts, especially for precipitation fields (Ebert, 2001). Today, most meteorological centers have developed their operational ensemble prediction systems (EPS) to consider these uncertainties (Buizza, 2019). EPSs provide, for the same time and location, a set of forecasts obtained by introducing perturbations at different modeling stages (e.g., initial and boundary conditions, Buizza, 2019).

EPSs deliver probabilistic meteorological information that is also valuable for data-assimilation (DA) systems, which will be the focus of this paper for the particular case of the Canadian Precipitation Analysis (CaPA, Mahfouf et al., 2007; Fortin et al., 2015). The CaPA system produces gridded precipitation fields based on numerical weather prediction (NWP) forecasts

25 adjusted with observed precipitation (ground stations and radars) using optimal interpolation (OI) assimilation methods (Fortin et al., 2015; Lespinas et al., 2015; Fortin et al., 2018). Using EPSs in CaPA would allow the estimate of the 'errors of the day' from short-term ensemble forecasts through the computation of flow-dependent background errors covariances (Kalnay et al., 2007). As opposed to a static assimilation approach, weights attributed to the background and the observations would be potentially better distributed (Wang et al., 2007). However, DA systems using EPSs such as Ensemble Kalman Filter (EnKF,

30 Evensen, 2003) often require large ensembles to avoid sampling errors and the associated rank problem of covariance matrices (Houtekamer and Zhang, 2016). In contrast, DA systems using deterministic forecasts such as standard variational approaches are more computationally efficient (Wang et al., 2008a; Houtekamer and Mitchell, 1998). Nevertheless, these DAs make hypotheses on the homogeneity, the isotropy, and the invariance of error covariance matrices. To combine desirable aspects of both DA scheme families (Houtekamer and Zhang, 2016), linear combinations of the static and the dynamic error covariance

35 matrices, known as hybrid approaches, have been developed (Hamill and Snyder, 2000; Wang et al., 2008a; Kleist and Ide, 2015).

Hybrid assimilation schemes have been thoroughly and successfully tested for many applications. Hamill and Snyder (2000) combined a three-dimensional variational (3DVar) and EnKF DA schemes for atmospheric assimilation for the United States of America (USA) domain. The authors showed improvements over 3DVar alone, especially in data-poor networks and, to a lesser

40 extent, in denser networks. For the same domain, Wang et al. (2008a) also showed that analyses generated with the same hybrid approach (3DVar+EnKF) produced 12-hour forecasts that were more accurate than 3DVar and confirmed the efficiency of the method in regions with low observation density. Other studies worked on different hybridization variants, Wang et al. (2008b) combined the Kalman Ensemble Transform Filter (ETKF) and 3DVar for the DA system over the USA using the Weather Research and Forecasting Model (WRF). The authors drew the same conclusions regarding the density of observations. In

45 ocean forecasting, Counillon et al. (2009) combined optimal interpolation (OI) and an EnKF method. They demonstrated a reduction in forecast errors using hybrid covariances for small dynamic ensembles (10 members) relatively to the OI or EnKF.

The successful combination of two different DA schemes for producing better analysis and forecasts has led various meteorological centers to use this approach for operational products such as Houtekamer et al. (2019) for the Canadian Centre for Meteorological and Environmental Prediction (CCMEP) of Environment and Climate Change Canada (ECCC); Bonavita et al.

50 (2015) for the European Centre for Medium-Range Weather Forecasts (ECMWF) and Penny et al. (2015) for ocean forecasting at the National Centers for Environmental Prediction (NCEP). Several other studies have shown the relevance of hybrid approaches for DA, but a full review of their uses is beyond this research scope.

At CCMEP, the availability of the Regional Ensemble Precipitation System (REPS) encompassing the North American domain allows the investigation of hybrid approaches for analyses at a higher resolution than those currently existing (see Houtekamer et al., 20

55 The 20-member REPS product is available at an approximately 10-km grid spacing and has been operational since summer 2019. The purpose of this study is to explore the potential improvements brought by hybrid approaches to the operational 6-hour CaPA, which is ; solely based on an OI where the background field is the Regional Deterministic Prediction System (RDPS, Caron et al., 2015). Various factors motivated the exploration of the well-established and documented hybrid DA methods. First of all, the REPS is fully operational and has the same domain and resolution as CaPA and, therefore, can be used

as a background field without any interpolation. Second, the hybrid approaches may positively impact the several data-poor areas (e.g., northern Canada) of the CaPA domain where the OI has its limitations. Thirdly, to our knowledge, no other studies have explored this approach for precipitation analyses. Finally, the OI's specificity in CaPA makes it interesting to explore in the context of hybrid DA approaches. Indeed, conclusions from several studies on hybrid DA methods were generally obtained when combining static and dynamic DA approaches (Wang et al., 2007; Counillon et al., 2009). However, the OI in CaPA is not strictly speaking static as observations, and forecasting errors are updated for each analysis time using variographic analysis ~~(for more details see Fortin et al., 2015)~~ (Fortin et al., 2015) as opposed to static errors estimated with the climatology.

The rest of the article is organized as follows. ~~Section ?? introduces the~~ Sections 2 and 3 introduce the datasets and the hybrid analysis scheme ~~, while~~ methodology, respectively. Section 4 describes the method to select the optimal weighting for the hybrid approach~~. Section 2 and 5 present respectively the datasets and~~, while section 5 presents the experimental design followed by the verification strategy in Section 6. The results of the experiments are made available in Section 7, and the conclusion is given in Section 8.

2 Datasets

2.1 Model description

The operational 20-member REPS (version 3.0.0, ECCC, 2019) and the Regional Deterministic Prediction System (RDPS, Caron et al., 20 the same configuration. The domain covers the North American continent (Figure 1) with a horizontal grid spacing of 0.09° ($\sim$ 10-km) and 84 vertical levels. The RDPS/REPS generate 72-h forecasts four times per day at 00, 06, 12, and 18 UTC. Uncertainty in the REPS is represented by perturbed initial and lateral boundary conditions (ICs and LBCs) and stochastically perturbed physics tendencies (SPPT, Charron et al., 2010) but with the same physical parameterization for all members. The atmospheric ICs derive from a 20-member interpolated Global Ensemble Prediction System (GEPS, Charron et al., 2010; Houtekamer et al. perturbations centered around the RDPS initial analysis. Every hour, the LBCs are also provided by the GEPS.

2.2 Observations

The 6-hour analysis assimilates precipitation from surface stations from Canadian and the contiguous United States (US) networks, some of which are operational only during the warm season. Thirty-three and ~~the discussions are given in Section 8~~31 C-band radar QPEs, covering the US and Canada (mostly located along the US border), are also assimilated. Since 2017, new Canadian dual-polarization Doppler radars have been progressively replacing their C-band counterpart. They contribute to retrieving information on a broader range of meteorological events than standard C-band radars (Kollias et al., 2020), and would ultimately enable better analysis. Both types of observation follow an extensive quality control described in Section 3.2.

3 ~~Hybrid assimilation approach for the precipitation analysis~~

2.1 Stage IV precipitation

90   In addition to verification at station locations (see Section 6.1), the 6-hour analyses were also compared to 6-h Stage IV (ST4) analysis from the National Centers for Environmental Prediction (NCEP, Lin and Mitchell, 2005). The objective is to allow verification against spatially-continuous precipitation fields. ST4 is a mosaic of regional multi-sensor (gauges and WSR-88D radars) analysis that is designed differently from CaPA and ensures a certain degree of independence during the verification. The ST4 domain covers the contiguous United States (CONUS), but for known limitations in the western domain

95   (see Nelson et al., 2016, for more details), only the CONUS east of 105°W was used for verification purposes. The ST4 native horizontal spacing of ~ 4.7 km was interpolated to the coarser RDPS grid using circular filtering (see Fig. 3.a in Jacques et al., 2018) to allow verifications on a common grid. ST4 will not provide a picture of performance over the entire RDPS domain, as Canada and the southern part of the RDPS are not covered. However, ST4 is a valuable dataset that could help compare the results obtained with the different configurations of the hybrid approach.

100   ## 3 Methodology

3.1 Hybrid assimilation approach for the precipitation analysis

Similarly to what was proposed by Hamill and Snyder (2000), the ~~the~~ background field error covariance matrix using the hybrid approach ($\widetilde{\mathrm{P}}^{\mathrm{b}}$) is a weighted sum of the OI-based ($\mathrm{P^b}_{\mathrm{OI}}$) and dynamically-based ($\mathrm{P^b}_d$) matrices such as:

$$\widetilde{\mathrm{P}}^{\mathrm{b}} = (1-\beta)\mathrm{P^b}_{\mathrm{OI}} + \beta\mathrm{P^b}_{\mathrm{d}}, \tag{1}$$

105   where the parameter β is comprised between 0 and 1, ensuring that the total background error covariances are conserved (Wang et al., 2008a). Thus, when β=0, the analysis is solely based on the OI, while when β=1 the background errors are ruled by the REPS. The β selection methodology is further ~~detailled~~ detailed in Section 4.

It is worth mentioning that the precipitation from both the observation and background model (y) are primarily Box-Cox transformed such as:

110    $$x = \begin{cases} \lambda^{-1}\left[y^\lambda - 1\right], & \text{if } \lambda > 0, \\ \log(y), & \text{if } \lambda = 0, \end{cases} \tag{2}$$

and where in CaPA configuration $\lambda = 1/3$. This pre-processing strategy has been used as it provided more skillful analyses (Fortin et al., 2015; Lespinas et al., 2015). The final analyses are obtained from the back-transformation, followed by corrections of the biases induced by this transformation (Evans, 2013).

An exponential and isotropic model is assumed for the spatial correlation function of the background field errors (Fortin

115   et al., 2015). The i-th row and j-th column of $\mathrm{P^b}_{\mathrm{OI}}$ writes as:

$$\mathrm{p^b_{OI}}(i,j) = \sigma^2_{\mathrm{OI}} \exp\left(-\delta(i,j)/l_{\mathrm{OI};}\right) \tag{3}$$

where σ_{OI}^2, $\delta(i,j)$, and l_{OI} are, respectively, the variance of the background errors, the Euclidean distance between locations i and j and the correlation length. ~~These parameters are estimated using~~ In CaPA algorithm, parameters σ_{OI}^2, l_{OI} but also σ_o^2 – the variance errors of the observation – needed to build the observation error covariance matrix (explained further below), are estimated before the analysis as such using a variographic analysis of the innovations~~, $Z = d - Hx_f$. The~~. The innovations are classically defined as $Z = d - Hx_f$, where d, x_f and H correspond respectively to the measurements, the forecasts and the observation operator~~,~~; which is here the nearest neighbour interpolation (Fortin et al., 2015).

The degree of spatial dependence of the innovations is described through a theoretical exponential function fitted on the empirical semivariogram (Cressie, 2015) and is defined as:

$$\gamma_{i,j} = \frac{1}{2}\left\langle |z_i - z_j|^2 \right\rangle = \begin{cases} \sigma_o^2 + \sigma_{\mathrm{OI}}^2(1 - \exp\left(-\delta(i,j)/l_{\mathrm{OI}}\right)), & \text{if } \delta(i,j) > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

where z_i, z_j and σ_o^2 ~~corresponds to~~ correspond, respectively to the innovations at locations i and j and to variance errors of the observation. As opposed to other DA approaches where the error matrix is static, the variographic analysis is computed for each analysis time step and therefore allows for time varying elements in $\mathrm{P}^a_{\mathrm{OI}}$ $\mathrm{P}^b_{\mathrm{OI}}$.

The covariance matrix $\mathrm{P}^b_{\mathrm{d}}$ $\mathrm{P}^b_{\mathrm{d}}$ depicts the flow-dependant errors estimated from the REPS and is defined as:

$$\mathrm{P}^b_{\mathrm{d}} = \frac{1}{N-1}\frac{1}{N-2}\mathrm{A}'\mathrm{A}'^{\mathrm{T}}, \tag{5}$$

where $\mathrm{A}' \in \mathbb{R}^{m \times N}$ denotes the anomalies estimated from the N-member ensemble at m grid points. The superscript T corresponds to a matrix transpose. To avoid underestimation of the variance of the background errors due to the limited size of the dynamical ensemble (Houtekamer and Mitchell, 1998), the anomaly computation follows Hamill and Snyder (2000) suggestions and writes as:

$$\mathrm{A}'_{:,j[:,j]} = \mathrm{X}_{f[:,j]} - \overline{\mathrm{X}_f}_{[\![1,N]\!]\backslash j}, \tag{6}$$

where ~~$\mathrm{A}'_{:,j}$ is the~~ the subscript $[:,j]$ refers to the j-th column of ~~A' and $\overline{\mathrm{A}}_{[\![1,N]\!]\backslash j}$ is~~ a given matrix and X_f is $(m \times N)$ precipitation ensemble forecast matrix. The m lengh vector $\overline{\mathrm{X}_f}_{[\![1,N]\!]\backslash j}$ corresponds to the average across the N members without the j-th member.

The analysis is performed sequentially for each grid-cell, for which the error matrices are estimated using up to 16 neighbors per observation type to speed up the computation (Fortin et al., 2015). Therefore, when assimilating both surface observations and radar Quantitative Precipitation Estimates (QPEs), $\widetilde{\mathrm{P}}^b$ is a matrix of size up to 32×32. In cases where no observation within a radius of 500 km is encountered, the analysis process is accelerated by setting the grid-cell value equal to the background field.

The equations for the analysis estimates are solved as:

$$x_a = x_f + \mathrm{W} \cdot (d - \mathrm{H}x_f),$$
$$\mathrm{W} = \widetilde{\mathrm{P}}^b\mathrm{H}^{\mathrm{T}} \cdot \left(\mathrm{H}\widetilde{\mathrm{P}}^b\mathrm{H}^{\mathrm{T}} + \mathrm{R}\right)^{-1}, \tag{7}$$

where R and W are, respectively, the observation error covariance matrix (see details in Fortin et al., 2015) and the weight vector.

The variances of the error made by estimating the precipitation analysis is obtained from the covariance structure and the weights as follows (Fortin et al., 2015):

$$\sigma_a^2 = \sigma_b^2 - \mathrm{W}^{\mathrm{T}} \cdot \widetilde{\mathrm{P}}^{\mathrm{b}} \mathrm{H}^{\mathrm{T}} \tag{8}$$

For the hybrid approach, the variance of background errors is defined as:

$$\sigma_b^2 = (1 - \beta)\sigma_{\mathrm{OI}}^2 + \beta\sigma_{b,\mathrm{REPS}}^2, \tag{9}$$

where σ_{OI}^2 is estimated with variogram modelling presented in ~~equation 4~~ Eq. (4). In contrast, $\sigma_{b,\mathrm{REPS}}^2$ has the advantage of being directly estimated from the REPS and defines for a given grid-cell location, s_o, ~~such as:~~ as:

$$\sigma_{b,\mathrm{REPS}}^2 (s_o) = \frac{1}{N-1} \frac{1}{N-2} \sum_{i=1}^{N} \left(\boldsymbol{x}_{s_o s_{o,i}}{}^{(\mathrm{REPS})} - \overline{x}_{s_o s_{o[[1,N]]\backslash i}} \right) \left(\boldsymbol{x}_{s_o s_{o,i}}{}^{(\mathrm{REPS})} - \overline{x}_{s_o s_{o[[1,N]]\backslash i}} \right))^{\mathrm{T}}, \tag{10}$$

where $\boldsymbol{x}_{s_o}^{(\mathrm{REPS})}$ is the N-length vector of REPS precipitation at location s_o and $\overline{\boldsymbol{x}}_{s_o}$ $\overline{x}_{s_o[[1,N]]}$ is the ensemble mean at the same location ~~.~~ as defined in Eq. (5).

In addition to the estimate of the analysis error already provided by the OI, CaPA provides a spatially and temporally varying index that describes the confidence granted to the precipitation analysis. This index is based on the assumption that the most trustful data is observation. For a given grid-cell and valid time, the confidence index of the analysis (CFIA) has been defined as ~~one (1)~~ minus the ratio of the error variances of the analysis, σ_a^2, to the background, σ_b^2, such as:

$$\mathrm{CFIA} = 1 - \frac{\sigma_a^2}{\sigma_b^2} \tag{11}$$

The CFIA ranges from 0 to 1. CFIAs close to 1 (0) depict high contributions of the observation (background) into the analysis estimates. CFIAs close to 0 occurs when no observations are assimilated in the vicinity of a given grid-cell. Therefore, this index is convenient for users desiring to select only grid-cells influenced by observations. Using the hybrid approach in CaPA enables to account for flow-dependent errors. It would therefore have impacts on CFIA estimates further discussed in Section 7.4.

3.2 Quality Control of the observations during the assimilation

Observed surface precipitation and radar QPEs undergo an extensive quality control (QC) process to remove untrustworthy data. The QC is automatically done for each assimilation valid time, 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. A temporal QC is performed to identify persistent problems that occur over a given period, such as a station that reports no or too much precipitation over a long time. Rejecting or keeping a surface station is based on the statistical distribution of the differences

between the observation and the analysis obtained over the most recent cases (see details in Section 2.b. in Lespinas et al., 2015).

175 A different quality control, hereafter referred to as spatial QC, is carried out to identify surface stations that have very different precipitation than those in the immediate vicinity. For this purpose, an analysis is estimated at a site s_k $(x_{s_k}^{(a)})$ using neighboring stations in a leave-one-out (LOO) approach. The observation at the same site $(x_{s_k}^{(o)})$ is rejected or, is said invalid, if the following is not fulfilled:

$$\left| x_{s_k}^{(o)} - x_{s_k}^{(a)} \right| < \text{tol} \cdot \sqrt{(\sigma_o^2 + \sigma_a^2)}, \tag{12}$$

180 where tol is a tolerance factor set equal to 4 for the operational CaPA. Both observed precipitation and analysis estimates of Eq. (12) undergone the Box-Cox transformation beforehand (see Eq. 2). The size of the neighborhood depends on the background field correlation length which varies for each analysis and therefore adapts with seasons (Lespinas et al., 2015). This approach helps avoid rejecting very localized summer precipitation events. An additional QC is also applied during the cold season, with the rejection of radar QPEs and surface observations during windy condition (Rasmussen et al., 2012).

185 Several other QC are applied to precipitation input datasets, but their extensive description is beyond the scope of this manuscript (see Fortin et al., 2015; Lespinas et al., 2015, for further information). Figure 1 illustrates the stations and radars that passed the QC for summer.

In CaPA, hybrid DA approaches can impact the quality control of observations. Indeed, changing the past analysis values (used in the temporal QC) and the standard deviation of the analysis error (used in the spatial QC) can induce differences in

190 the number of assimilated stations. The results showed slight changes in the number of observations assimilated when using hybrid approaches and are therefore not detailed in this study.

4 Selection of the weighting factor: β

Analysis experiments using $\beta \in [0, 1]$ with a 0.1 step are conducted for each valid time to identify the β value that provides the most accurate precipitation analysis. At least two different options are possible to select the most suitable β. First, the

195 analyses are conducted for a given season, and the β that minimizes an objective function is selected and stored for future uses. The second option consists of choosing the most suitable β for each valid time, implying a dynamical selection of the β. While the first option is computationally inexpensive as βs are estimated once for each season, it suggest that this parameter remains constant over years. The second option is more flexible but needs independent data from those which relied upon for the precipitation verification step, which is required to evaluate the CaPA system (Section 6).

200 To understand the choice that has been made for either option, the CaPA configuration must be presented first. CaPA generates two types of precipitation analysis. The first is conducted only at surface station locations in a leave-one-out (LOO) framework (hereafter, CaPA-LOO). This computationally inexpensive analysis is both a quality control step to reject suspicious precipitation observations (see Section 2.b in Lespinas et al., 2015, and Section 2.2 below) and a valuable dataset for verification purposes in a cross-validation manner. The second type of analysis is performed for each grid-cell of the domain.

A dynamical selection of β for each analysis valid time would have been the preferred option as it would have not required manual updates during, for example, major operational upgrades of the different systems (CaPA, REPS, or RDPS). For this purpose, preliminary tests using random samplings of the CaPA-LOO datasets have been done. For each valid time, a training sample was employed for the β selection, while the testing sample was kept for objective verification. The training and the testing sample sizes had to be large enough to, respectively, capture the optimal β value over the domain and to realize consistent precipitation verification. Different sample sizes were tested with up to 40% of the initial CaPA-LOO dataset to estimate the optimal β. However, the results were too noisy due to sampling effects, especially during winter where the station density is relatively low. For this reason, the first option to select β value was preferred. The normalized root mean square error (NRMSE) averaged over a selected period and based on the CaPA-LOO datasets was chosen to identify optimal β. The normalization was realized to lessen the influence of the total precipitation amount during a given analysis time step (Bachmann et al., 2019). For more reliability, the NRMSE is computed only at ~~SYNOP~~ surface synoptic observations (hereafter SYNOP) and manual SYNOP stations during the summer and the winter, respectively, and is defined as follow:

$$
\text{NRMSE} = \frac{\sqrt{\sum_{k=1}^{M} \left(a_k^{\text{LOO}} - o_k\right)^2}}{\sqrt{\sum_{k=1}^{M} \left(a_k^{\text{LOO}} + o_k\right)^2}}, \tag{13}
$$

where M, a^{LOO} and o correspond, respectively, to the number of cases during a given valid time, the LOO analysis, and the observations. The NRMSEs of zero (0) illustrate perfect analyses.

5 ~~Datasets~~

4.1 ~~Model description~~

~~The operational 20-member REPS (version 3.0.0, ECCC, 2019) and the Regional Deterministic Prediction System (RDPS, Caron et al., 20 the same configuration. The domain covers the North American continent (Figure 1) with a horizontal grid spacing of 0.09° (~ 10-km) and 84 vertical levels. The RDPS/REPS generate 72-h forecasts four times per day at 00, 06, 12, and 18 UTC. Uncertainty in the REPS is represented by perturbed initial and lateral boundary conditions (ICs and LBCs) and stochastically perturbed physics tendencies (SPPT, Charron et al., 2010) but with the same physical parameterization for all members. The atmospheric ICs derives from a 20-member interpolated Global Ensemble Prediction System (GEPS, Charron et al., 2010; Houtekamer et a perturbations centered around the RDPS initial analysis. Every hour, the LBCs are also provided by the GEPS.~~

4.1 ~~Observations~~

~~The 6-hour analysis assimilates precipitation from surface stations from Canadian and the contiguous United States (US) networks, some of which are operational only during the warm season. Thirty-three and 31 C-band radar QPEs, covering the US and Canada (mostly located along the US border), are also assimilated. New Canadian dual-polarization Doppler radars have been progressively added to the observation database. They contribute to retrieving information on a broader range of meteorological events than standard C-band radars (Kollias et al., 2020), and would ultimately enable better analysis.~~

Observed precipitation and radar QPEs undergo an extensive quality control (QC) process to remove untrustworthy data. The QC is automatically done for each assimilation valid time, 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). A first temporal QC is performed to identify persistent problems that occur over a given period, such as a station that reports no or too much precipitation over a long time. Rejecting or keeping a surface station is based on the statistical distribution of the differences between the observation and the analysis obtained over the most recent cases (see details in Section 2.b. in Lespinas et al., 2015). A second quality control, hereafter referred to as spatial QC, is carried out to identify surface stations that have very different precipitation than those in the immediate vicinity. For this purpose, an analysis is estimated at a site s_k using neighboring stations in a LOO approach. The observation is rejected or, is said invalid, if:

$$\left| x_{s_k}^{(\text{OBS})} - x_{s_k}^{(\text{CaPA})} \right| < \text{tol} \cdot \sqrt{(\sigma_o^2 + \sigma_a^2)},$$

where tol is a tolerance factor set equal to 4 for the operational CaPA, and obs and CaPA superscripts refer to, respectively, the observation and the analysis in the transformed space. The size of the neighborhood depends on the background field correlation length which varies for each analysis and therefore adapts with seasons (Lespinas et al., 2015). This approach helps avoid rejecting very localized summer precipitation events. An additional QC is also applied during the cold season, with the rejection of radar QPEs and surface observations during windy condition (Rasmussen et al., 2012). Figure 1 illustrates the stations and radars that passed the QC for summer.

In CaPA, hybrid DA approaches can impact the quality control of observations. Indeed, changing the past analysis values (used in the temporal QC) and the standard deviation of the analysis error (used in the spatial QC) can induce differences in the number of assimilated stations. The results showed slight changes in the number of observations assimilated when using hybrid approaches and are therefore not detailed in this study.

4.1 Stage IV precipitation

In addition to the LOO verification at station locations, the 6-hour analyses were also compared to 6-h Stage IV (ST4) analysis from the National Centers for Environmental Prediction (NCEP, Lin and Mitchell, 2005). The objective is to allow verification against seamless precipitation fields. ST4 is a mosaic of regional multi-sensor (gauges and WSR-88D radars) analysis that is designed differently from CaPA and ensures a certain degree of independence during the verification. The ST4 domain covers the contiguous United States (CONUS), but for robustness reasons, only the CONUS east of 105W was used for verification purposes (e.g., Nelson et al., 2016; Schwartz, 2019). The ST4 native horizontal spacing of ∼ 4.7 was interpolated to the coarser RDPS grid using circular filtering (see Fig 3.a in Jacques et al., 2018) to allow verifications on a common grid. ST4 will not provide a picture of performance over the entire RDPS domain, as Canada and the southern part of the RDPS are not covered. However, ST4 is a valuable dataset that could help compare the results obtained with the different β values in the hybrid approach.

5   Experimental set-up

Two sets of two-month experiments were conducted: i) one over July to August 2019 (hereafter summer), and ii) one over January to February 2020 (hereafter winter). For each 6-hour valid time, 11 sets of analyses were produced with the different β values presented in Section 4.

270   Summer and winter experiments are expected to differ due to both the background model's distinct seasonal performances in representing precipitation and the significantly smaller assimilated datasets in winter. The usefulness of a hybrid configuration has been demonstrated when low-density observation networks are assimilated (Hamill and Snyder, 2000; Wang et al., 2008a, b). Thus, to verify this point without being strongly influenced by seasonal effects, an additional experiment was conducted without the assimilation of radar QPEs. The experiment without radar was only conducted during the summer as no
275   impact is expected during the cold season, during which radar QPEs are not assimilated.

6   Verification strategy

6.1   Comparisons against ~~observation~~ observations from surface stations

Based on the 2x2 contingency table for binary events (Table 1), four different metrics, commonly used for precipitation objective evaluation, are computed for the different configurations of the LOO analysis. First, the frequency bias index (FBI) which compares the frequency of events in the analyses to those in the observations. Second, the equitable threat score (ETS) which
280   assess the agreement between the analysis and the observations. Following Table 1 annotations, FBI and ETS express as:

$$\text{FBI} = (a + b).\,(a + c)^{-1},$$
$$\text{ETS} = \frac{a - h_r}{a + b + c - h_r},$$

(14)

where $h_r = (a + b)\,(a + c)$ illustrates the hits expected by chance and make the ETS less sensitive to the climatological frequency of precipitation events. The probability of detection (POD) and the false alarm ratio (FAR) are the other two metrics
285   and define as:

$$\text{POD} = a \cdot (a + c)^{-1},$$
$$\text{FAR} = b \cdot (a + b)^{-1},$$

(15)

FBI-1 – which is equivalent to the normalized difference between false alarms and missed events – is preferred in the following as positive (negative) values imply positive (negative) bias (Lespinas et al., 2015). FBIs-1 and FARs of zero (0) are optimal, while ideal PODs and ETSs are one (1). For the four metrics, binary events define as 6-hour accumulations meeting or exceed-
290   ing selected thresholds, here 0.2, 1.0, 5.0 and 10.0 mm. Higher thresholds are not shown because the samples were too small, leading to overly noisy scores.

Statistical differences between the scores using $\beta = 0.0$ (hereafter the reference experiment) and $\beta > 0.0$ were assessed using a stationary block bootstrapping with a 95% confidence level (Brown et al., 2012). The bootstrapping implementation specific to CaPA is detailed in Lespinas et al. (2015).

6.2 Comparisons against Stage IV

Two different types of comparison against ST4 were conducted. First, aggregated areal coverages of 6-h accumulated precipitation meeting or exceeding selected accumulation thresholds were assessed to evaluate how precipitations are distributed in ST4 and CaPA.

Second, the Fraction Skill Score (FSS) was computed to assess the analysis skill in spatially placing precipitation events (Roberts and Lean, 2008; Schwartz et al., 2009). A selected threshold is first applied to each grid-cell of CaPA and ST4 to define the occurrences of precipitation events for a given valid time and for each grid-cell. Then, the ~~fractions of grid-cells above the threshold(probabilities~~fractional values at the ith grid cells (probability of precipitation above a selected threshold) in a ~~pre-selected neighborhood (for example, a square of~~ preselected neighborhood (e.g., 30 km square) are estimated in CaPA ($f_{a(i)}$) ~~are calculated for CaPA (f_a)~~ and ST4 (~~f_o~~)$f_{o(i)}$), respectively. The FSS compares the differences of fractions to the largest possible fraction difference and expresses as:

$$\text{FSS} = 1 - \frac{\frac{1}{N_y}\sum_{i=1}^{N_y}\left(\mathrm{f}_{a}(i) - \mathrm{f}_{o(i)}\right)^2 \quad \frac{1}{N_y}\sum_{i=1}^{N_y}\left(\mathrm{f}_{a(i)} - \mathrm{f}_{o(i)}\right)^2}{\frac{1}{N_y}\left[\sum_{i=1}^{n}\mathrm{f}_{a(i)}^2 + \sum_{i=1}^{n}\mathrm{f}_{o(i)}^2\right] \quad \frac{1}{N_y}\left[\sum_{i=1}^{N_y}\mathrm{f}_{a(i)}^2 + \sum_{i=1}^{N_y}\mathrm{f}_{o(i)}^2\right]}, \tag{16}$$

where N_y is the number of grid-cells on the verification domain. The FSS is averaged over the period and was calculated for squares of 20, 30 and 50 km. Values close to 1 of the FSS are optimal, while 0 indicates no skills. By construction, as the size of the neighbourhood increases, FSSs also increase because overlaps of precipitation events in the observed and analysis datasets are more likely.

7 Results and discussion

7.1 Selection of the optimal β value

Figure 2 shows the NRMSE estimated at the SYNOP stations for the different β values and in the LOO framework. The presence of a minimum at $\beta > 0.0$ for the three experiments, in summer with and without radar QPEs and in winter, demonstrates the usefulness of the hybrid approach.

Comparing the two summer experiments (Fig. 2.a and 2.b), it can be first noticed that, as expected, the addition of radar QPEs generally reduces the NRMSE and thus improves the performance of the analysis. It also decreases the added value of the hybrid approach. NRMSE values for β ranging from 0.0 to 0.4 were ~~indeed~~ very similar for the summer experiment assimilating radar QPEs (Fig. 2.b), ~~with a minimum obtained at~~ where the optimal β equal to 0.4 corresponds to an NRMSE reduction of 2.2% compared to $\beta = 0.0$. On the other hand, the experiment without radar QPEs showed ~~larger~~ more significant

variability in NRMSEs for the different βs. The optimal value of 0.5 reduces the NRMSEs by 3.6% when compared to the experiment using $\beta=0.0$. These results suggest that when the density of assimilated observations is lower, the hybrid approach brings more added value and is thus consistent with the literature (Hamill and Snyder, 2000; Wang et al., 2008a). During the summer, the use of $\beta > 0.6$ (0.9) with (without) radar assimilation reduced the NRMSE values and therefore deteriorated the analyses compared to the reference analysis ($\beta = 0.0$).

Interestingly, the winter experiments illustrated a different pattern. According to the NRMSEs (Fig. 2.c), the analysis improved when βs increased and reached a minimum at $\beta = 0.7$, leading to a 7.5% reduction in the NRMSE values. The analysis deteriorated for even higher βs but was not worse than the reference experiment, meaning that in that case, the dynamic approach was more suited than the static one. The low density of assimilated observations during the winter solid precipitation compared to liquid phase precipitation season may partly explain this result and is consistent with the experiment with and without radars. Another point that could explain this performance is the higher winter forecast skill of both the RDPS and the REPS that allow more accurate background fields and a better specification of the $\widetilde{P}^b$ matrix.

Both the seasons and the quantity of assimilated observations seem to have an influence on the optimal value of β. Summer 2019, with and without radar QPEs, and winter 2020 have optimal β that is equal to 0.5, 0.4, and 0.7, respectively. For all three experiments, the hybrid approach showed its relevance as it overcame both the static ($\beta = 0.0$) and the dynamic configuration ($\beta = 1.0$). Further evaluation metrics to verify that these optimal values contribute to improve different aspects of the 6-hour precipitation distribution are presented in the following section.

7.2 Contingency table verification

Metrics based on the contingency table FBI-1, ETS, POD and FAR (Section 6.1) were calculated for the three experiments and for all β values. Attention was paid to the experiments with the optimal β according to the NRMSE presented in the previous section (see Fig. 2). However, experiments with different β values were also investigated because they seemed to provide a good compromise, improving some metrics while deteriorating very little the others. Results with β values that degraded too much the reference analysis are not presented.

Figure 3.a illustrates the metrics for the summer without the radar QPEs for the optimal $\beta = 0.5$ and $\beta = 1.0$, both compared to $\beta = 0.0$. Filled markers indicate no significant differences at the 95% confidence level between the two experiments for a given threshold. The 6-h precipitation analysis with $\beta = 0.5$, displayed a significant increase in skill (at the 95% confidence level) as shown by the ETS and a decrease of the false alarms (FAR) for all the selected thresholds. The POD was slightly deteriorated, especially for small precipitation events. As illustrated by the FBI-1, the selection of $\beta = 0.5$ led to generally lower precipitation amounts than with $\beta = 0.0$. The impact was positive for small precipitation events (thresholds of 0.2 and 1.0 mm), but it tended to smooth out higher intensity events. Interestingly, using a completely dynamic configuration, $\beta = 1.0$, showed improved performances compared to $\beta = 0.0$ and $\beta = 0.5$, but only for precipitation events above 0.2 and 1.0 mm and except POD, which degraded for all thresholds. However, looking at precipitation events of higher intensity with $\beta = 1.0$, e.g., 5 and 10 mm, did not necessarily degrade FBI-1 or the ETS scores compared to $\beta = 0.0$ but were not as well represented as when using $\beta = 0.5$. Looking at other

β values, $\beta = 0.3$ (Fig. 3.b) seemed to be a good compromise between skill and bias of the precipitation analysis. Indeed, the ETS and FAR remained improved compared to the reference experiment, and the deterioration of the POD was less important than with $\beta = 0.5$, while the degradation of FBI-1 for heavy precipitation was acceptable.

The results for the analysis during the same season but with the assimilation of the radar QPEs is showed in Figures 3.c and 3.d. Similarly to what was observed for the NRMSE values, all metrics were generally improved when assimilating the radar QPEs but the differences between the reference and the optimal $\beta = 0.4$ were less pronounced. The analysis with $\beta = 0.4$ showed significantly reduced FARs for all thresholds. However, the improvement in ETS when compared to the $\beta = 0.0$ was significant only for small thresholds (0.2 and 1.0 mm). For higher thresholds, the skill was slightly improved but was not significant at the 95% confidence level. Similarly to the experiment without radar QPEs (Fig. 3a. and 3.b.), the POD slightly ~~was slightly~~ deteriorated, and the FBI-1 reduced for small precipitation events but increased for events of medium to high intensity. The dynamical approach, $\beta = 1.0$, showed a different pattern when assimilating radar datasets. Almost all scores and thresholds were significantly degraded compared to $\beta = 0.0$ and $\beta = 0.4$. The only improvement regards precipitation events above 0.2 mm, but the degradation for other thresholds is too substantial. The use of a $\beta = 0.3$ seemed again to be a good compromise regarding the bias by reducing the analysis smoothness when compared to $\beta = 0.4$, while preventing the POD deterioration.

Finally, Figure ~~5~~4.a illustrates the same metrics during the winter and compares the $\beta = 0.7$ and 1.0 to the reference experiment. ~~The ETS,~~ With $\beta = 0.7$, the ETS was significantly improved, and the false alarms at the 95% confidence level were reduced. Fewer precipitation events were generated for all selected thresholds with the analysis using $\beta = 0.7$ than with $\beta = 0.0$. Again, this improves the performance for 6-hour precipitation greater than 0.2 and 1.0 mm, but not for accumulations greater than 2.0 mm. However, the degradation of FBI-1 for high-intensity precipitation was much less pronounced than in summer, especially for such a high β value. The probability of detecting events (POD) greater than 0.2 mm was significantly reduced, but was increased for heavy events precipitation (> 10 mm). In between performances were obtained for experiment with $\beta = 1.0$, the latter did no better than when using $\beta = 0.7$ but were not that different for several thresholds and scores (e.g., FBI-1 for precipitation above 0.2 mm in Fig. 3.a).

For the sake of comparison with the summer season, contingency table metrics using $\beta = 0.3$ are also displayed in Fig. ~~5~~4.b. It appeared that this value also provides improved metrics even though it was not identified as optimal. Indeed, FBI-1 was still, but to a lesser extent, reduced for precipitation above 0.2 mm compared to the reference experiment. It is interesting to note that for other thresholds, the degradation of FBI-1 was less pronounced and was not significant in most cases. Similarly, the POD degradation for the 0.2 mm threshold was less pronounced. The ETS and FAR were still significantly improved for all thresholds relative to the reference experiment, but to a lesser extent than with $\beta = 0.7$.

In light of these results, ~~it appears that the use of the~~ the optimal β value identified through the use of NRMSE ~~did indeed show an improvement in skills and a reduction in~~ showed improved performances compared to static ($\beta = 0.0$) and dynamic ($\beta = 1.0$) configurations. Improved skills and reduced false alarm rates were indeed obtained when using the optimal β for both summer and winter experiments. However, the frequency of moderate to ~~high intensity~~ high-intensity 6-hour precipitation was reduced, especially in summer. The latter can be detrimental to very localized high-intensity precipitation events. The

390 examination of the same scores for other values of β highlighted that ~~the use of~~ $\beta = 0.3$ offered a good compromise for skill gain while not overly damaging the frequency bias for high-intensity precipitation events for both seasons. For these reasons, $\beta = 0.3$ was maintained as optimal in the sense of a compromise and is used for the additional checks presented below.

7.3 Verification against ST4

The impact of radar QPE assimilation in the hybrid approach has been demonstrated in the previous sections. Therefore, to
395 simplify the discussion, the comparison of CaPA analyses with ST4 is only performed for the experiments that assimilated the ground-based radars (i.e., excluding the summer without radar QPE experiment). This point is also justified by the fact that the operational CaPA is integrated with the radar QPEs and that the final objective is to compare with this system. Thus, in the following, the summer experiments refer to the one that assimilates the radar QPEs.

Figure ~~6~~ 5 shows the accumulated precipitation for ST4 and CaPA using $\beta = 0.3$ and the differences between these two
400 datasets. More spatial patterns and higher summer precipitation amounts were observed in ST4 than in the CaPA using $\beta = 0.3$ (Fig. ~~6~~5.a). The higher native resolution of ST4 and the limited ability of the CaPA background model to predict convective precipitation, especially at the 10 km grid spacing, may explain such results.

The winter (Fig~~6.~~. 5.b), generally characterized by spatially larger precipitation events, showed a better agreement between CaPA and ST4. Indeed, ST4 and CaPA with $\beta = 0.3$ had similar spatial patterns with both over- and underestimated accumu-
405 lations depending on the location, but to a lesser extent than in summer. These results can be interpreted both by the greater forecasting capacity of the background model in winter and by the expected lower impact of the horizontal resolution during this season. Accumulated precipitation, using either $\beta = 0.0$ or $\beta = 0.3$, was compared to ST4 at the grid point scale to identify the best parameter (not shown for conciseness). It was found, that without spatial consistency, some locations (grid cells) were better represented with $\beta = 0.3$ while for other locations $\beta = 0.0$ was more suitable. Therefore, moving from $\beta = 0.0$ to 0.3
410 values have a minimal impact on seasonal precipitation accumulations.

Figures ~~7~~6.a and ~~7~~6.b depict the fraction of grid-cells above selected 6-h thresholds averaged for the summer and winter, respectively. During summer, the differences between the configurations $\beta = 0.0$ and $\beta = 0.3$ were very small (< 2 %), as shown by the very close golden and blue curves. Consistent with results on the FBI-1 (Fig. 3), the fraction of precipitation events above 0.2 mm generated in the analysis using $\beta = 0.0$ or 0.3 were slightly higher than in ST4 for both season. During
415 summer, the fraction of grid points greater than 1, 2 and 5 ~~mm.6h^{-1}~~ mm.6 h^{-1} were systematically underestimated by 2 to 3% in CaPA compared to ST4 and these fractions appear to contribute importantly to the underestimation of ST4 precipitation totals (Fig. 6). Winter was similar (Fig~~7.~~. 6.b) as fractions of grid-cells for the 6-h precipitation analysis were generally overestimated in the analysis. However, this overestimation was relatively smaller ($< 1\%$) than summer and became almost negligible for 5 and 10 mm.6h^{-1} thresholds.
420 The FSSs for the 11 analyses differing by β values and using ST4 as a reference were calculated for the 0.2 and 1.0 ~~mm.6h^{-1}~~ mm 6h^{-1} thresholds with neighborhood length scales of 20, 30, and 50 km (Fig. ~~??~~7). For a given neighborhood length scale and for the 0.2 mm threshold during the summer (Fig. ~~??~~7.a), the FSS increased when β went from 0.0 to 0.5 and then decreased for higher βs, thus are consistent with the results obtained with NRMSEs. Similarly, during the winter and for the

same threshold, FSSs increased from when β went from 0.0 to 0.7 and decreased thereafter (Fig. ~~??~~7.c). The gain associated with the use of $\beta > 0.0$ was less marked for the 1.0 mm threshold, especially for the summer (Fig. ~~??~~7.b). These results are consistent with those obtained with scores based on the contingency tables for the summer experiment. For both seasons (Fig. ~~??~~7.b and ~~??~~7.d), the use of $\beta = 0.3$ did not, however, deteriorate the FSSs compared to those of the reference experiment and even showed a slight improvement during the winter.

7.4 Confidence index

In the OI-based CaPA, CFIA fields are characterized by circular structures as shown for two winter cases in Figure 8.b. Small dots of high values are centered around the surface stations and large disks represent radar footprints. These particular structures are a direct consequence of the assumptions on the error isotropy.

Using the hybrid approach in CaPA enables to account for flow-dependent errors, allowing CFIA spatial distributions to be linked to meteorological situations and thus lessen the influence of the aforementioned assumptions. However, in this configuration, the variances of the background field errors are also spatially variable, in contrast to the spatially constant value used for operational CaPA (~~eq~~Eqs. 4 and 9). Therefore, this makes the interpretability of CFIA, in its current definition (~~eq~~Eq. 11), less straightforward. To illustrate this point two different meteorological situations were selected and are illustrated in Figures 8 (a,d). The 6-h precipitation fields for January 5, 2020 (Fig. 8.a) and January 18, 2020 (Fig. 8.d) valid at 12UTC, were both characterized by large synoptic events but at different locations. Increasing the contribution of the REPS in the analysis computation, $\beta = 0.7$ in Fig. 8 (c,f), leads to CFIAs following meteorological spatial distributions. Generally, CFIAs tended to be higher at places with precipitation than when using $\beta = 0.0$ as shown in the ~~eastearn~~ eastern part of the domain for the January ~~18~~ 18th case (Fig. 8.f). Inversely, locations with no precipitation in the background field tended to show small CFIA values. This last result is consistent with the calculation of P^b_d (Eq. 5), which, when the ensemble members have no or very little precipitation, will tend to a zero matrix. Nonetheless, CFIA remained close to zero above the Atlantic ocean for the case of January 5 (Fig. 8.c) despite the synoptic event that occurred in that location. This is explained by the current CaPA computation framework~~, where gridcells are set~~. No matter the value of β, analysis precipitation at a given grid-cell is equal to the background when no observations ~~to be assimilated~~ are available in the vicinity. The latter lead to error variances of the analysis (σ_a^2) being close to those in the background (σ_b^2), and by construction CFIA will tend towards small values (Eq. 11). As a result, the current CFIA (~~eq~~Eq. 11) revealed two limitations for the hybrid assimilation approach: (i) users interested in ~~gridcells~~ grid-cells with high CFIAs may indirectly favor locations with precipitation which in turn can lead to biased interpretation and (ii) low CFIA values may have two different interpretations. To overcome the first limitation, a field representing the density of assimilated observations for each ~~gridcell~~ grid-cell will come along with the CaPA precipitation fields. In addition, to ease CFIA interpretation, the reference, which is currently σ_b^2 in ~~eq. 11~~Eq. (11), will be changed for a temporal climatology of the background field errors. To do this, the denominator in the ~~equation 11~~Eq. (11) will be replaced by the time average of σ_b^2, calculated for each grid cell using an exponential filter that gives more weight to the most recent cases. Different tests made on CFIA estimates are beyond the scope of this paper and will be discussed in a different study.

8   Conclusions

Based on optimal interpolation, the Canadian Precipitation Analysis, currently operational at CCMEP, is used for different applications (see Fortin et al., 2018, for further details). The recently available operational 20-member REPS with a ~ 10 km grid-spacing naturally led to examine whether it was possible to improve the estimation of background error covariances used by the current CaPA OI. From this perspective, the selected approach is a so-called hybrid approach (Hamill and Snyder, 2000; Counillon et al., 2009) and it consists of a weighting of two covariance matrices, one obtained by the OI and the other from an ensemble Kalman filter based on the REPS. The primary motivation is to overcome some simplifying assumptions about the structure of background errors when using the OI. Indeed, the operational CaPA uses a spatial covariance structure of background errors assumed isotropic and is modeled as an exponential function. The second motivation is the documented positive impact of this approach for domains with low observation density, which is the case for some regions of the CaPA domain (e.g., northern Canada) and during the winter season when few stations are assimilated.

In practice, this consists of choosing an optimal weighting between two covariance matrices of background field errors, one from the OI and the other from REPS, through a parameter – β – which varies between 0 and 1. The experiments were conducted for 6-hour precipitation accumulations and over two seasons, summer 2019 and winter 2020. To verify the impact of the amount of assimilated observations, an additional experiment without radar QPEs' assimilation was conducted for summer 2019. This additional experiment was not performed for winter as the assimilated radar QPEs during the solid precipitation season are anyway negligible in CaPA (Fortin et al., 2015).

Results have shown that first, the analysis with the assimilation of the radar QPEs limits the positive impacts of the increase of the weighting towards the REPS (i.e., high βs). Indeed, with and without the radar QPEs assimilation, and according to NRMSE estimates, summer had optimal values of 0.4 and 0.5, respectively. Winter precipitation showed higher optimal values of β, around 0.7, meaning that a higher weighting in favor of the REPS is preferable during that season. This result confirmed the known positive impact of the hybrid method in the configuration of a low station density. The second reason is the higher forecasting skill of the background model (RDPS) and the REPS during the winter season, leading to a better error specification. Interestingly, these optimal values obtained when comparing precipitation from surface stations to the analysis (in a leave-one-out framework) were consistent with those obtained using the Fractions Skill Scores, using a relatively independent observation dataset and on a smaller domain, the Stage IV analysis (Lin and Mitchell, 2005).

The use of the hybrid approach with the optimal β values for each experiment increased the analysis skills of both the operational CaPA ($\beta = 0.0$) and a purely dynamic approach ($\beta = 1.0$). This result is especially true for low-intensity precipitation events, for which the false alarm rate has been significantly reduced. In summer, the frequency bias score was ~~reduced~~ improved for small precipitation events (up to 1 mm) but degraded for higher thresholds. High-intensity precipitation events were indeed smoothed when using the optimal β value. This result was also observed for the winter but to a lesser extent. Using a trade-off value of $\beta = 0.3$ for the summer experiments reduced the impact of smoothing high-intensity events while allowing for significant skill improvement. According to the winter 2020 experiment results, a higher β would have been more suitable. However, in practice, this means choosing a date in the CaPA configuration for which β must be modified. A more

extended test period and further experiments would be needed to distinguish whether the parameter changes should be made during the regular North American domain seasons or during precipitation phase changes (approximately in May and October). In the meantime, other tests, using $\beta = 0.3$ during the winter, have provided acceptable results in skill and bias while being conservative.

495   Different flavors of the hybrid approach have also been tested but have not been developed in this article as they did not contribute positively or significantly to the selected metrics. As suggested in Counillon et al. (2009) study, an inflation factor ranging from 1.0 to 2.0 was applied to alleviate the possible lack of spread in the REPS but the experiments showed no or negative impacts on the verification metrics. Another hybrid configuration was also conducted to mitigate the impact of high-intensity events smoothing during summer precipitation. For that test, the hybrid approach was activated only if at least half of

500   the REPS members had strictly positive precipitation; otherwise, β was set to zero. The resulting scores showed higher biases for low-intensity precipitation events than the current approach and no bias reduction for high-intensity events.

     In addition to the work planned for a better specification of the CFIA (Section 7.4), future developments should focus on improving the current hybrid approach by accounting for the observation density ~~observations~~ when selecting the optimal β values (i.e., allow for spatial variability of β). Therefore, the hybrid system could benefit places without observations while not

505   degrading the analysis at locations with a high density of observations. This task will be facilitated by the availability of a new field informing on the amount of data assimilated (Section 7.4). Other works should explore the possibility of extending the current approach to create an ensemble version of the precipitation analysis. The REPS members, with prior bias corrections, could be used to create the analysis background field members. The hybrid approach could also help create a global version of CaPA using the GDPS and the GEPS (Houtekamer et al., 2014), respectively, as a background model and an ensemble system.

510   A global version of CaPA has already been identified as useful for land surface data assimilation at CCMEP. In that context, the hybrid approach would be especially relevant for the many regions of the world that do not benefit from high-density precipitation observations.

Appendix A: List of acronyms

CaPA	Canadian Precipitation Analysis
CFIA	ConFidence Index of the Analysis
DA	Data Assimilation
ECCC	Environement and Climate Change Canada
EnKF	Ensemble Kalman Filter
EPS	Ensemble Prediction System
GEPS	Global Ensemble Forecast System
ETS	Equitable Threat Score
FAR	False Alarm Rate
FBI	Frequency Bias Index
FSS	Fraction Skill Score
LOO	Leave-one-out
NRMSE	Normalized Root Mean Squared Error
NWP	Numerical Weather Prediction
OI	Optimal Interpolation
POD	Probability of Detection
QC	Quality Control
QPE	Quantitative Precipitation Estimates
RDPS	Regional Deterministic Prediction System
REPS	Regional Ensemble Prediction System
ST4	Stage IV
SYNOP	Synoptic stations

Data availability. REPS and CaPA (regional configuration) are both available as operational systems of ECCC at the following locations, respectively, https://dd.meteo.gc.ca/ensemble/reps/10km/grib2 and https://dd.meteo.gc.ca/analysis/precip/rdpa/grib2/polar_stereographic. The output data from this study (CaPA using hybrid methodology for the assimilation) have been archived and are available upon request to the corresponding author.

Author contributions. The main contributions from each co-author are as follows: under the supervision of SB and VF, DK proposed the methodological, the experimental design, and the preparation of the paper. GR contributed to the adaptation of the CaPA data assimilation code to integrate the hybrid modeling. VF, SB, and FL contributed to interpreting results and paper preparation.

Competing interests. The authors declare that they have no conflict of interest.

References

Bachmann, K., Keil, C., and Weissmann, M.: Impact of radar data assimilation and orography on predictability of deep convection, Quarterly Journal of the Royal Meteorological Society, 145, 117–130, https://doi.org/https://doi.org/10.1002/qj.3412, 2019.

525 Bonavita, M., Hamrud, M., and Isaksen, L.: EnKF and Hybrid Gain Ensemble Data Assimilation. Part II: EnKF and Hybrid Gain Results, Monthly Weather Review, 143, 4865–4882, https://doi.org/10.1175/MWR-D-15-0071.1, 2015.

Brown, J. D., Seo, D.-J., and Du, J.: Verification of Precipitation Forecasts from NCEP?s Short-Range Ensemble Forecast (SREF) System with Reference to Ensemble Streamflow Prediction Using Lumped Hydrologic Models, Journal of Hydrometeorology, 13, 808 – 836, https://doi.org/10.1175/JHM-D-11-036.1, 2012.

530 Buizza, R.: Ensemble forecasting and the need for calibration, in: Statistical Postprocessing of Ensemble Forecasts, pp. 15–48, Elsevier, 2019.

Caron, J.-F., Milewski, T., Buehner, M., Fillion, L., Reszka, M., Macpherson, S., and St-James, J.: Implementation of Deterministic Weather Forecasting Systems Based on Ensemble–Variational Data Assimilation at Environment Canada. Part II: The Regional System, Monthly Weather Review, 143, 2560–2580, 2015.

535 Charron, M., Pellerin, G., Spacek, L., Houtekamer, P., Gagnon, N., Mitchell, H. L., and Michelin, L.: Toward random sampling of model error in the Canadian ensemble prediction system, Monthly Weather Review, 138, 1877–1901, 2010.

Counillon, F., Sakov, P., and Bertino, L.: Application of a hybrid EnKF-OI to ocean forecasting, Ocean Science, 5, 389–401, https://doi.org/10.5194/os-5-389-2009, 2009.

Cressie, N.: Statistics for spatial data, John Wiley & Sons, 2015.

540 Ebert, E.: Ability of a poor man's ensemble to predict the probability and distribution of precipitation, Monthly Weather Review, 129, 2461–2480, 2001.

ECCC: Regional Ensemble Prediction System (REPS) Version 3.0.0 Summary of changes with respect to version 2.4.0 and validation, Tech. rep., Dorval, Qc, Canada, https://collaboration.cmc.ec.gc.ca/cmc/cmoi/product_guide/docs/lib/technote_reps-300_20190703_e.pdf, 2019.

Evans, A. M.: Investigation of enhancements to two fundamental com- ponents of the statistical interpolation method used by the Canadian

545 Precipitation Analysis (CaPA), Master's thesis, University of Manitoba,Winnipeg, Manitoba, 2013.

Evensen, G.: The ensemble Kalman filter: Theoretical formulation and practical implementation, Ocean dynamics, 53, 343–367, https://doi.org/10.1007/s10236-003-0036-9, 2003.

Fortin, V., Roy, G., Donaldson, N., and Mahidjiba, A.: Assimilation of radar quantitative precipitation estimations in the Canadian Precipitation Analysis (CaPA), Journal of Hydrology, 531, 296 – 307, https://doi.org/https://doi.org/10.1016/j.jhydrol.2015.08.003, 2015.

550 Fortin, V., Roy, G., Stadnyk, T., Koenig, K., Gasset, N., and Mahidjiba, A.: Ten Years of Science Based on the Canadian Precipitation Analysis: A CaPA System Overview and Literature Review, Atmosphere-Ocean, 56, 178–196, https://doi.org/10.1080/07055900.2018.1474728, 2018.

Hamill, T. M. and Snyder, C.: A hybrid ensemble Kalman filter–3D variational analysis scheme, Monthly Weather Review, 128, 2905–2919, https://doi.org/https://doi.org/10.1175/1520-0493(2000)128<2905:AHEKFV>2.0.CO;2, 2000.

555 Houtekamer, P. and Zhang, F.: Review of the Ensemble Kalman Filter for atmospheric data assimilation, Monthly Weather Review, 144, 4489–4532, https://doi.org/https://doi.org/10.1175/MWR-D-15-0440.1, 2016.

Houtekamer, P. L. and Mitchell, H. L.: Data assimilation using an Ensemble Kalman Filter technique, Monthly Weather Review, 126, 796–811, https://doi.org/https://doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2, 1998.

Houtekamer, P. L., Deng, X., Mitchell, H. L., Baek, S.-J., and Gagnon, N.: Higher Resolution in an Operational Ensemble Kalman Filter, Monthly Weather Review, 142, 1143 – 1162, https://doi.org/10.1175/MWR-D-13-00138.1, 2014.

Houtekamer, P. L., Buehner, M., and De La Chevrotière, M.: Using the hybrid gain algorithm to sample data assimilation uncertainty, Quarterly Journal of the Royal Meteorological Society, 145, 35–56, https://doi.org/https://doi.org/10.1002/qj.3426, 2019.

Jacques, D., Michelson, D., Caron, J. F., , and Fillion, L.: Latent heat nudging in the Canadian Regional Deterministic Prediction System, Monthly Weather Review, 146, 3995–4014, https://doi.org/10.1175/MWR-D-18-0118.1, 2018.

Kalnay, E., Li, H., Miyoshi, T., Yang, S.-C., and Ballabrera-Poy, J.: 4D-Var or ensemble Kalman filter?, Tellus A, 59, 758–773, https://doi.org/10.1111/j.1600-0870.2007.00261.x, 2007.

Kleist, D. T. and Ide, K.: An OSSE-Based Evaluation of Hybrid Variational–Ensemble Data Assimilation for the NCEP GFS. Part I: System Description and 3D-Hybrid Results, Monthly Weather Review, 143, 433–451, https://doi.org/10.1175/MWR-D-13-00351.1, 2015.

Kollias, P., Bharadwaj, N., Clothiaux, E., Lamer, K., Oue, M., Hardin, J., Isom, B., Lindenmaier, I., Matthews, A., Luke, E., et al.: The ARM Radar Network: At the Leading-edge of Cloud and Precipitation Observations, Bulletin of the American Meteorological Society, 101, E588 – E607, https://doi.org/10.1175/BAMS-D-18-0288.1, 2020.

Lespinas, F., Fortin, V., Roy, G., Rasmussen, P., and Stadnyk, T.: Performance Evaluation of the Canadian Precipitation Analysis (CaPA), Journal of Hydrometeorology, 16, 2045–2064, https://doi.org/10.1175/JHM-D-14-0191.1, 2015.

Lin, Y. and Mitchell, K.: The NCEP Stage II/IV hourly precipitation analyses: development and applications, in: Preprints of the 19th Conference on Hydrology, American Meteorological Society, San Diego, CA, 9-13 January 2005, p. Paper 1.2, 2005.

Lorenz, E. N.: Deterministic nonperiodic flow, Journal of the Atmospheric Sciences, 20, 130–141, 1963.

Mahfouf, J., Brasnett, B., and Gagnon, S.: A Canadian Precipitation Analysis (CaPA) project: Description and preliminary results, Atmosphere-Ocean, 45, 1–17, https://doi.org/10.3137/ao.v450101, 2007.

Nelson, B. R., Prat, O. P., Seo, D. J., and Habib, E.: Assessment and implications of NCEP stage IV quantitative precipitation estimates for product intercomparisons, Weather and Forecasting, 31, 371–394, https://doi.org/10.1175/WAF-D-14-00112.1, 2016.

Penny, S. G., Behringer, D. W., Carton, J. A., and Kalnay, E.: A Hybrid Global Ocean Data Assimilation System at NCEP, Monthly Weather Review, 143, 4660–4677, https://doi.org/10.1175/MWR-D-14-00376.1, 2015.

Rasmussen, R., Baker, B., Kochendorfer, J., Meyers, T., Landolt, S., Fischer, A. P., Black, J., Thériault, J. M., Kucera, P., Gochis, D., et al.: How well are we measuring snow: The NOAA/FAA/NCAR winter precipitation test bed, Bulletin of the American Meteorological Society, 93, 811–829, 2012.

Roberts, N. M. and Lean, H. W.: Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events, Monthly Weather Review, 136, 78 – 97, https://doi.org/10.1175/2007MWR2123.1, 2008.

Schwartz, C. S.: Medium-range convection-allowing ensemble forecasts with a variable-resolution global model, Monthly Weather Review, 147, 2997–3023, https://doi.org/10.1175/MWR-D-18-0452.1, 2019.

Schwartz, C. S., Kain, J. S., Weiss, S. J., Xue, M., Bright, D. R., Kong, F., Thomas, K. W., Levit, J. J., and Coniglio, M. C.: Next-Day Convection-Allowing WRF Model Guidance: A Second Look at 2-km versus 4-km Grid Spacing, Monthly Weather Review, 137, 3351 – 3372, https://doi.org/10.1175/2009MWR2924.1, 2009.

Wang, X., Hamill, T. M., Whitaker, J. S., and Bishop, C. H.: A Comparison of Hybrid Ensemble Transform Kalman Filter–Optimum Interpolation and Ensemble Square Root Filter Analysis Schemes, Monthly Weather Review, 135, 1055–1076, https://doi.org/10.1175/MWR3307.1, 2007.

Wang, X., Barker, D. M., Snyder, C., and Hamill, T. M.: A hybrid ETKF–3DVAR data assimilation scheme for the WRF model. Part I: Observing system simulation experiment, Monthly Weather Review, 136, 5116–5131, https://doi.org/https://doi.org/10.1175/2008MWR2444.1, 2008a.

Wang, X., Barker, D. M., Snyder, C., and Hamill, T. M.: A Hybrid ETKF–3DVAR Data Assimilation Scheme for the WRF Model. Part II: Real Observation Experiments, Monthly Weather Review, 136, 5132–5147, https://doi.org/10.1175/2008MWR2445.1, 2008b.

600

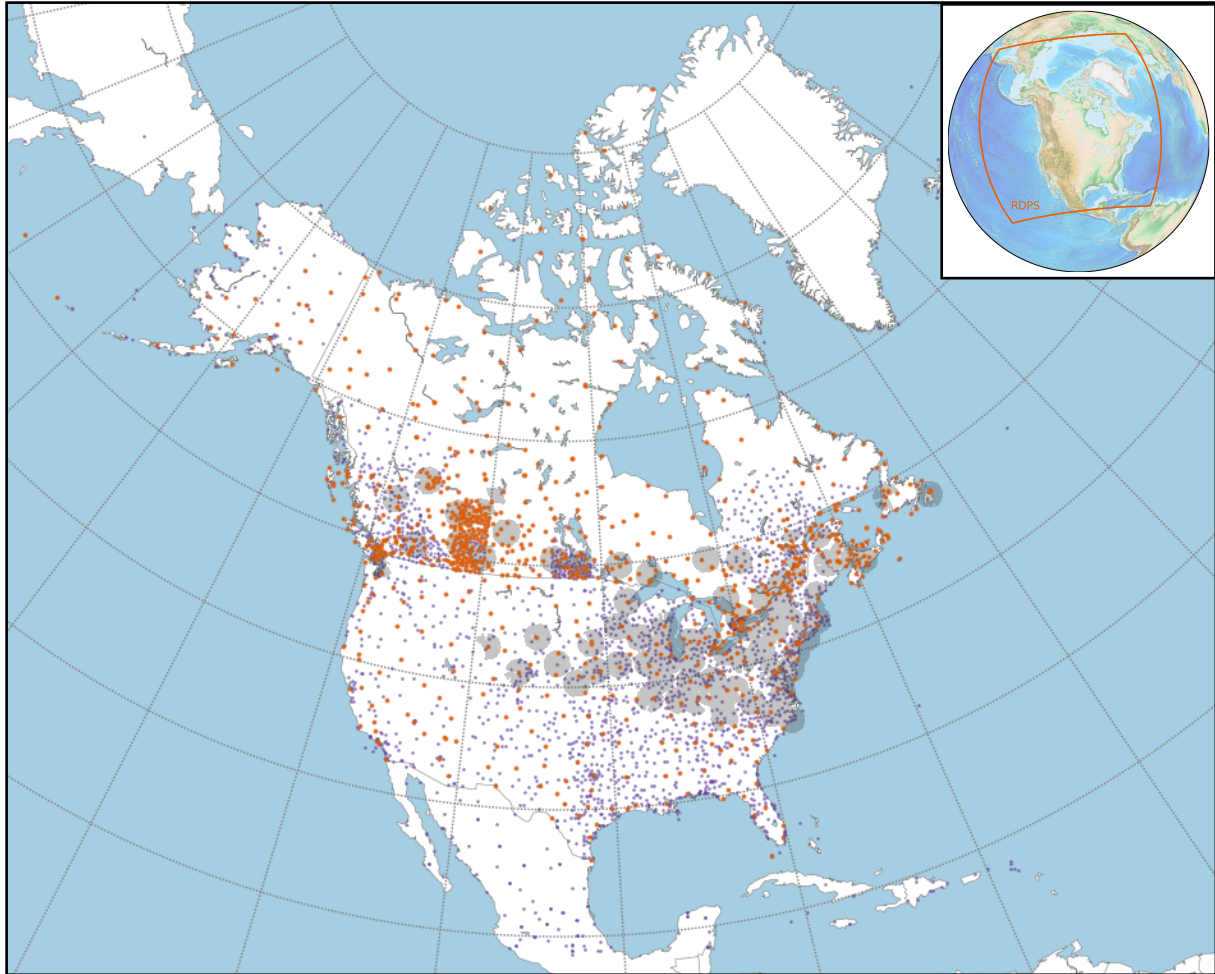



Figure 1. Blue and orange dots represent, respectively, the various precipitation networks and the SYNOP stations assimilated in summer 2019. Shaded gray areas illustrate the assimilated portion of radar beams. The top right panel illustrates the analysis domain.

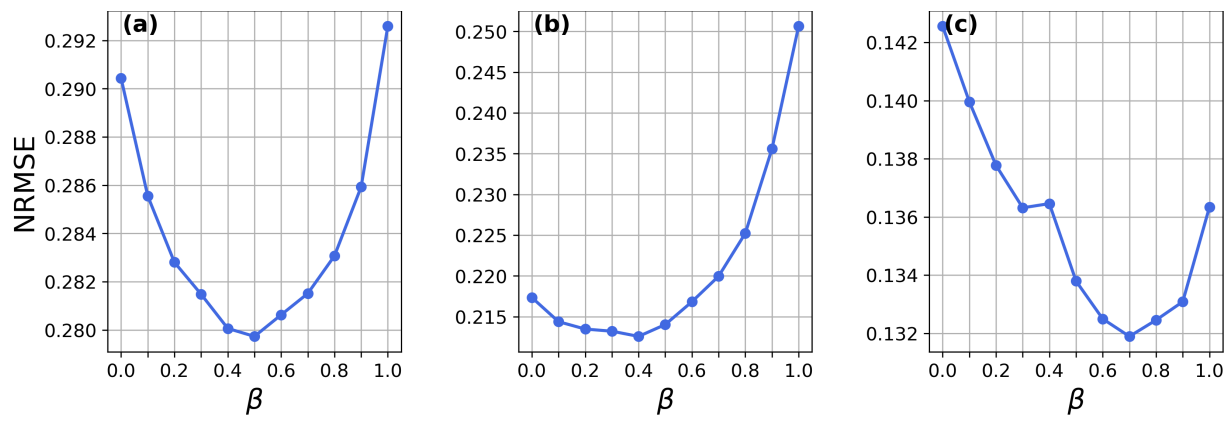
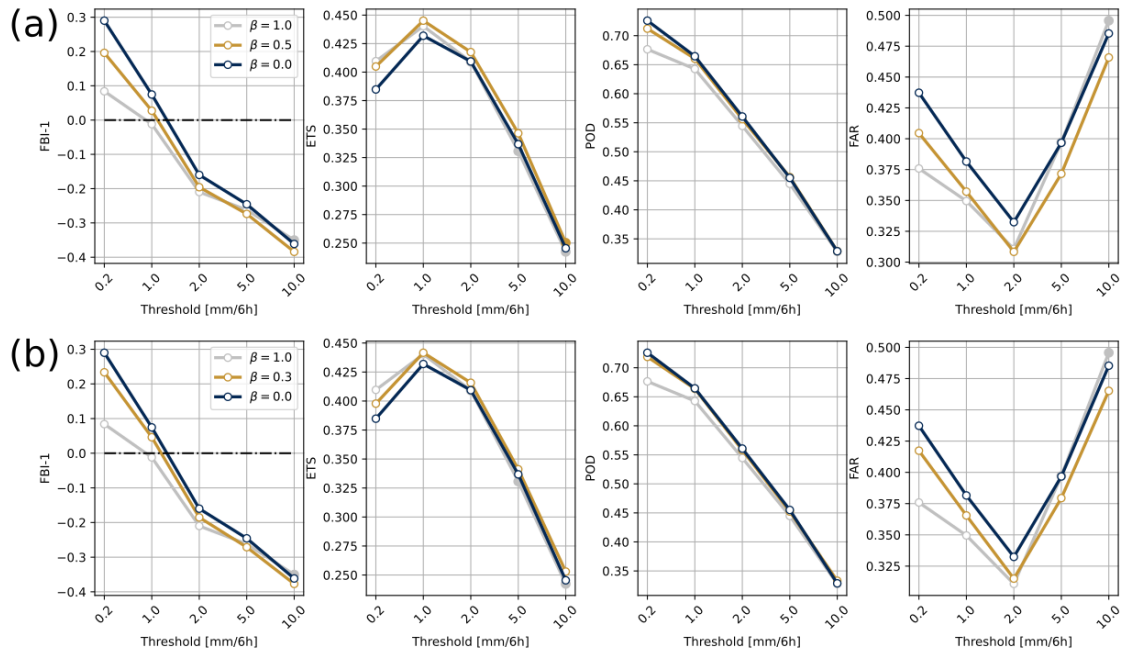



Figure 2. NRMSE averaged over the summer period without (a) and with (b) the assimilation of radar QPEs and over the winter (c) for the different β values.

Summer without radars

Summer with radars

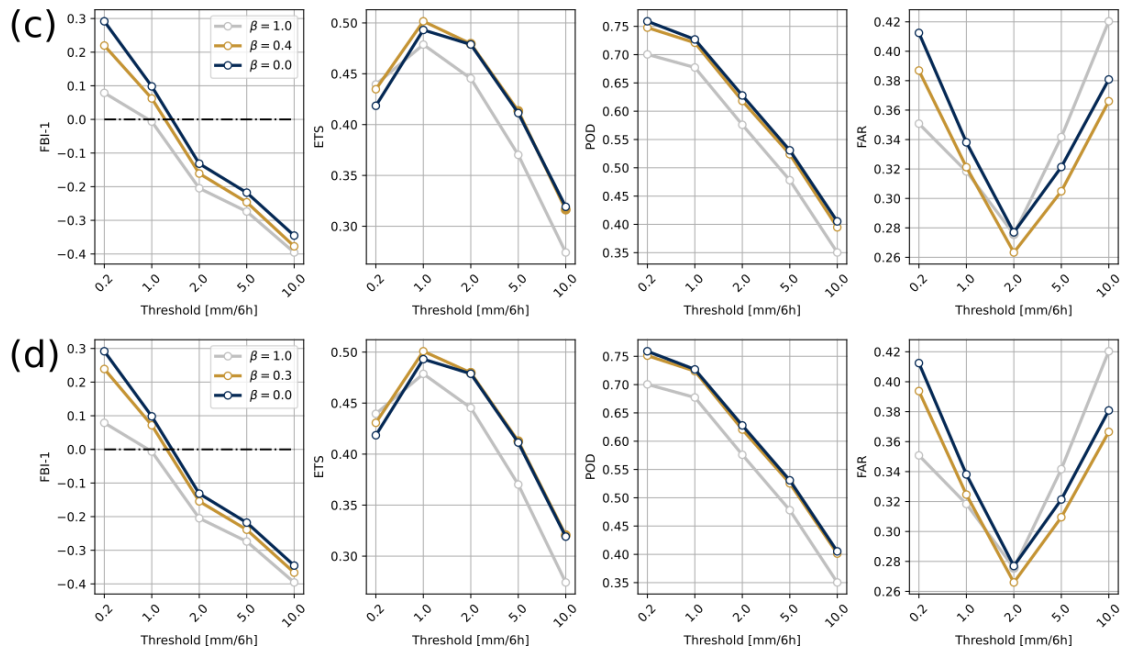



Figure 3. FBI-1, ETS, POD and FAR across the whole domain for summer experiment without radar QPEs for precipitation analysis with $\beta = 0.0$ (dark blue line), $\beta = 1.0$ (grey line), $\beta = 0.5$ (yellow ~~line~~ lines in (a)) and $\beta = 0.3$ (yellow ~~line~~ lines in (b)). Same figures but for the summer experiment with the assimilation radar QPEs with $\beta = 1.0$ (grey lines), $\beta = 0.4$ (yellow ~~line~~ lines in (c)) and $\beta = 0.3$ (yellow ~~line~~ lines in (d) ~~both~~) all three compared to the reference experiment when $\beta = 0.0$ (blue ~~line~~ lines). Filled markers indicate no significant differences at the 95% confidence level between the reference experiment $\beta = 0.0$ and $\beta = 1.0$, $\beta = 0.5$, $\beta = 0.4$ or $\beta = 0.3$ experiments.

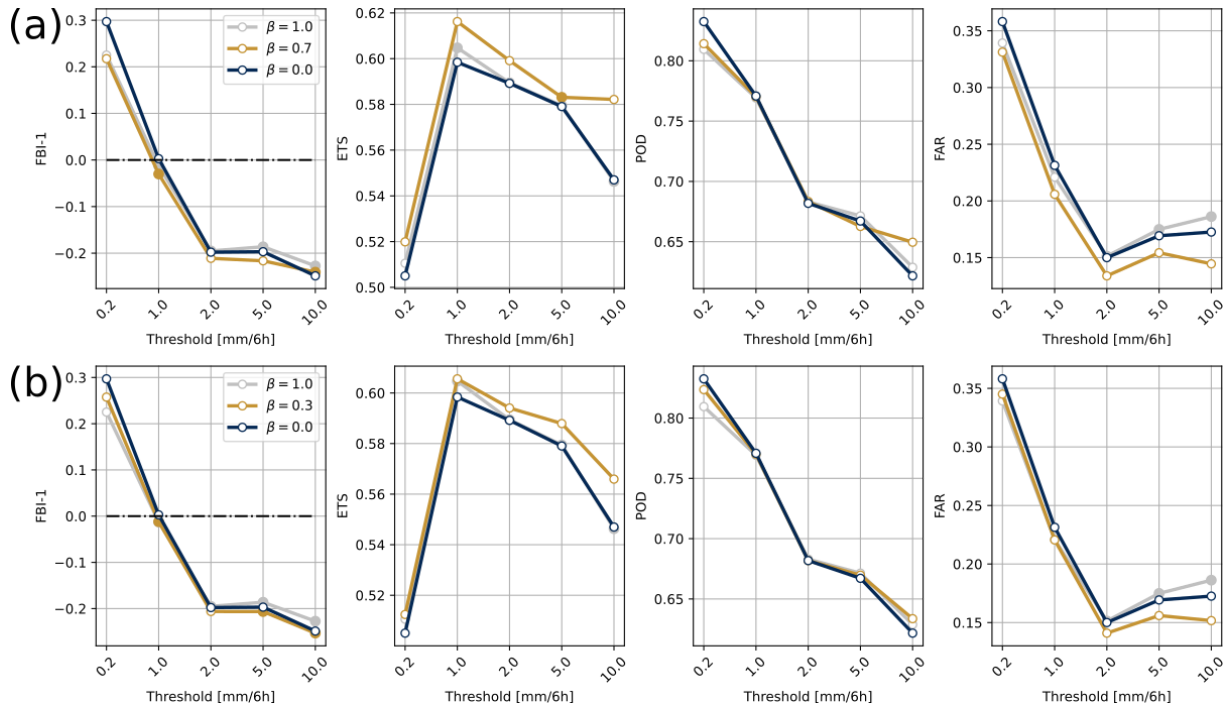



Figure 4. Same as Fig. ~~3~~ 3 but for the winter experiment and with $\beta = 0.7$ (top panel)~~and~~, $\beta = 0.3$ (bottom panel) and $\beta = 1.0$ (two panels), ~~both~~ all three compared to the reference experiment ~~when~~ ($\beta = 0.0$).

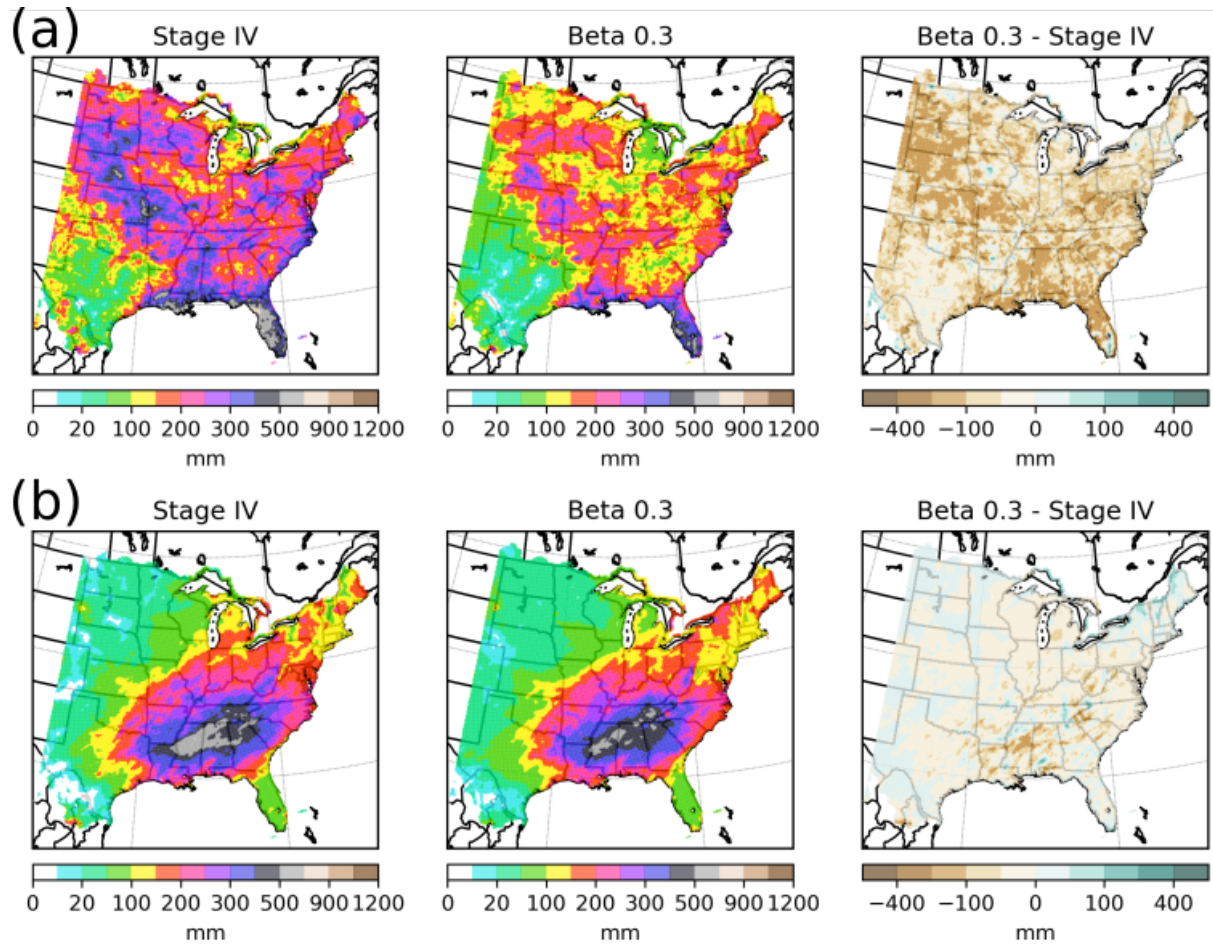



Figure 5. (a) From left to right, 6-h precipitation accumulated over the summer for ST4, for CaPA with $\beta = 0.3$, and the difference between the two; (b) same as (a) but for the winter experiment.

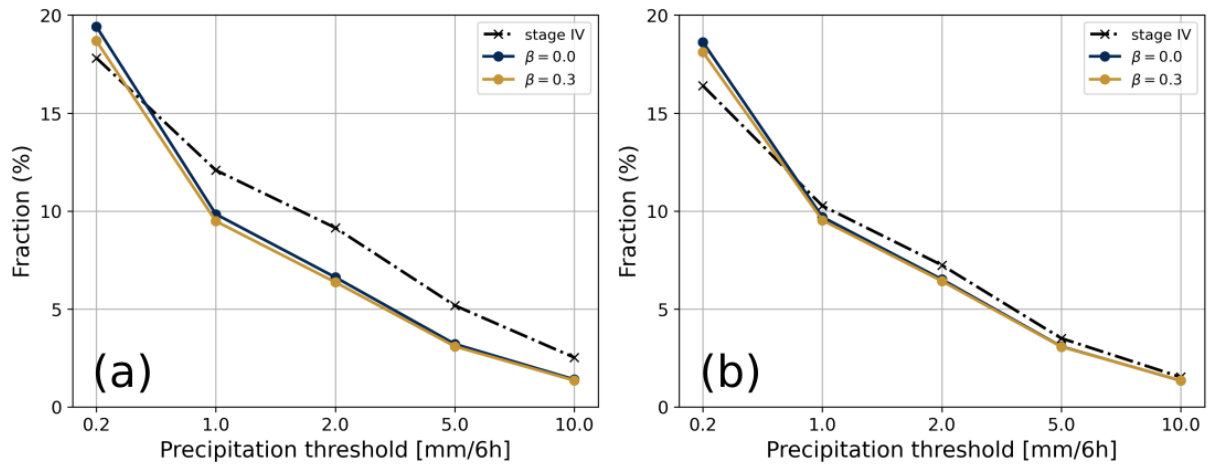



Figure 6. Fractional areal coverage (%) of 6-hour accumulated precipitation meeting or exceeding selected thresholds (x-axis) during the summer (a) and the winter (b); for ST4 (dashed black line) and CaPA using the configuration with $\beta = 0.0$ (blue line) and 0.3 (yellow line) over their common domain (Fig. 65).

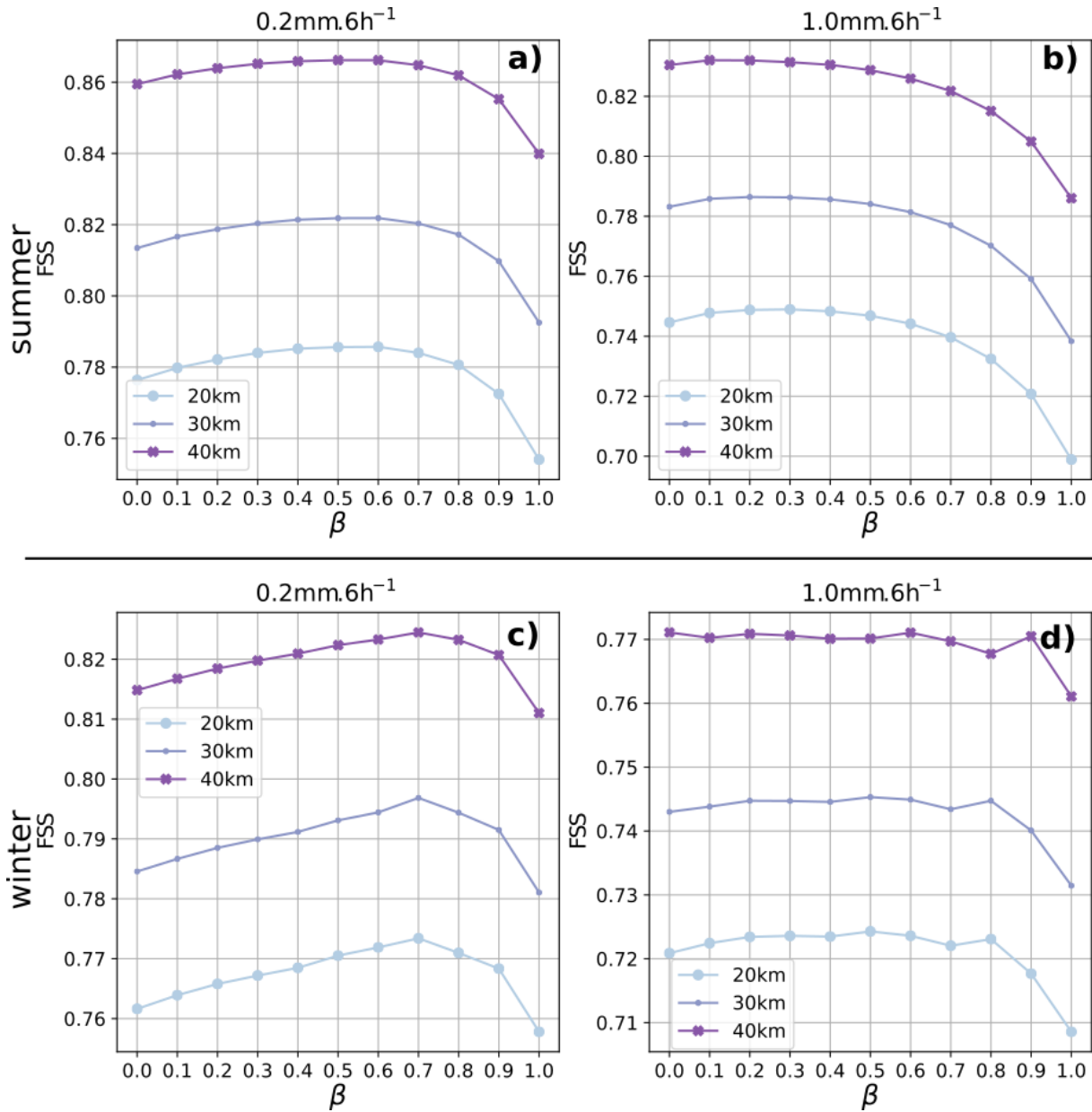



Figure 7. Fraction skill score (FSS) as a function of the different β values for 6-hour precipitation meeting or exceeding 0.2 mm (a,c) and 1.0 mm (b,d) thresholds for summer (first row) and winter (second row); using ST4 precipitation as reference. Each curve represents different neighboring length scale (20, 30 and 50 km).

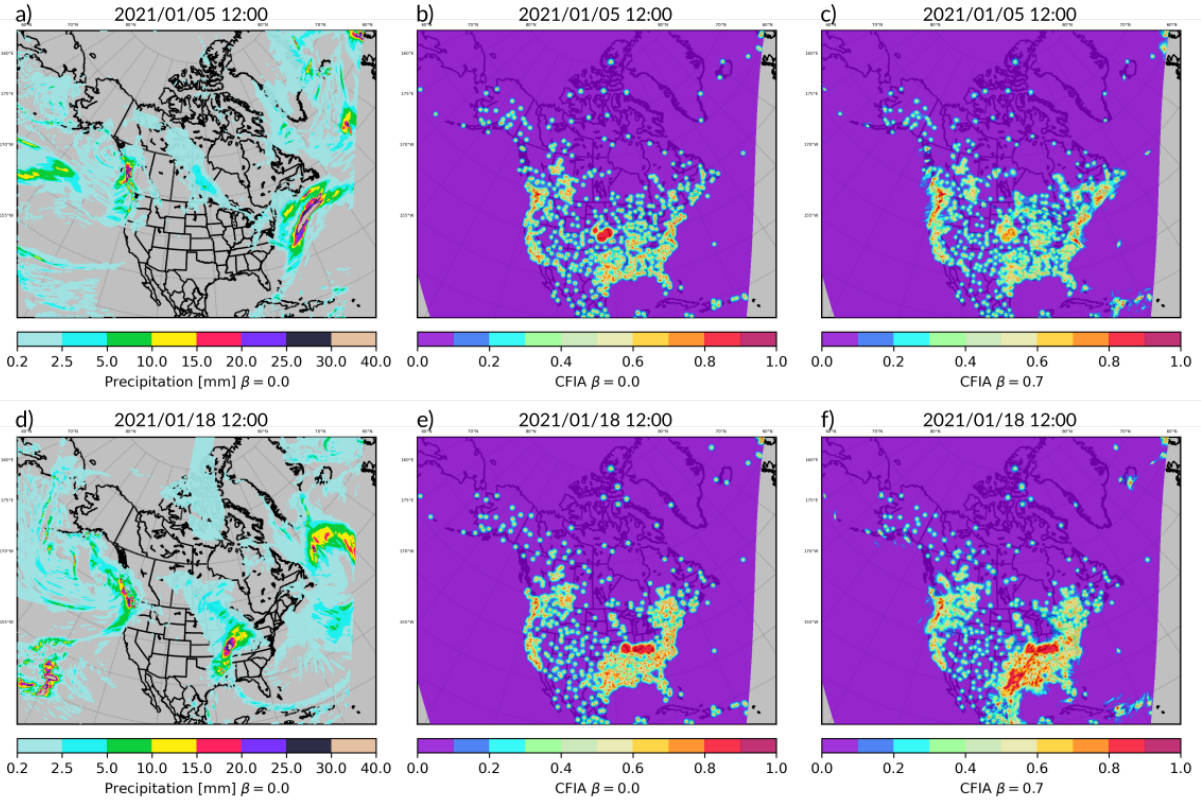



Figure 8. Precipitation field (a,d) from CaPA and associated CFIA values (b,e) for the 5 January 2020 /01/05 12:00 and 2020/01/18 12:00 January 2020, both valid timeat 12UTC. Panel (c) and (f) shows the CFIA for the same date but using the hybrid approach with high contribution of the REPS, here with $\beta = 0.7$.

Table 1. The 2×2 Contingency table for binary events where a, b, c and d corresponds to the frequency of hits, false alarm, misses and correct negatives.

	Observed event Yes	Observed event No	Total
Analysis event Yes	a (hit)	b (false alarm)	a+b
Analysis event No	c (miss)	d (correct negative)	c+d
Total	a+c	b+d	1