Response to reviewer #1

Dear Anonymous Referee #1,

We want to thank referee #1 for the review and the opportunity to improve our paper. We hope we have adequately answered all the reviewer' comments addressed in the following with a point-by-point response in *italic*. Sentences that we suggest for addition or modification to the revised version of the manuscript are indicated in *italic blue*.
Best regards,

Dikraa Khedhaouiria on the behalf of all co-authors

Highlights

The paper presents an extension of the classical 'static' data assimilation to incorporate ensemble forcast and by such allowing to releveive some restricting constraints on the shape of the covariance matrix of the errors terms in the classical DA approach.This approach is interesting, and make use nicely of some recent developments in DA.

General comments

The paper is rather easy to read and is well structured, altough I got quite lost in all the acrynonims, maybe a list of them could be appreciated.

- *We thank the reviewer for this positive comment. About the acronyms, we suggest an Appendix that will list all the acronyms present in the manuscript (see page 10 of this document for the appendix).*

As I am not directly an expert of the domain, I do not know what are the models, so it took me some time to figure out what how it is constructed. It may worth introducing the whole model in section 2 (analysis and observations models).

- *We think that it is essential that non-expert in data assimilation approaches can follow the thread of the article. For this reason, we intend to restructure the article so that current Section 4, which describes the datasets, to be inserted as section 2. The current Section 2, which presents the assimilation approaches, will be Section 3, and so on for the remaining sections. However, as the observation's quality control described in lines 176-195 needs some aspects of the assimilation to be defined first, it will be moved as a subsection of the new Section 3. Therefore, we propose the following structure for the next version of the manuscript:*

  2. Datasets

  2.1 Model Description

  2.2 Observations

  2.3 Stage IV precipitation

  3. Methodology

  3.1 Hybrid assimilation approach for the precipitation analysis

  3.2 Quality Control of the observations during the assimilation

  Except for the section number, the other sections will not be restructured. Finally, it is worth mentioning that with this new plan, small rearrangements of the reviewed manuscript were needed:

  - *Lines 68-71 currently as: "Section 2 introduces the hybrid analysis scheme, while Section 3 describes the method to select the optimal weighting for the hybrid approach. Section 4 and 5 present respectively the datasets and experimental design followed by the verification strategy in Section 6. The results of the experiments are made available in Section 7 and the conclusion and the discussions are given in Section 8." will be changed for "Sections 2 and 3 introduce the datasets and the hybrid analysis scheme methodology, respectively. Section 4 describes the method to select the optimal weighting for the hybrid approach, 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 and the discussions are given in Section 8."*

  - *Reference to LOO in lines 197-198 will be removed as it is not yet introduced at this point of the proposed version of the manuscript. Lines 197-198 currently as: "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)." will be changed for "In addition to verification at station locations (see Section 3.2 below), 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)".*
  - *As β was also not yet introduced in the proposed Section 2.3, the sentence in lines 205-206 "However, ST4 is a valuable dataset that could help compare the results obtained with the different β values in the hybrid approach." will be modified such as: "However, ST4 is a valuable dataset that could help compare the results obtained with the different configurations of the hybrid approach."*

In Eq 6, the authors use the approach of Hamill and Snyder to estimate the variance of the background errors, but it seems that they do not do the same for the hydrid approach (Eq 10), I may be wrong, or may the authors comment on that ?

- *We apologize for the unclear explanation, and we thank the reviewer for this comment for two reasons. First, it helped us catch errors in equations (5) and (10). The denominator should have been $N-2$ instead of $N-1$ (see equation 5 in Hamill and Snyder 2000). It is essential to mention that this typo is not present in the analysis computation; therefore, no changes are expected in the results. Equation (5) will be corrected for:*

$$P^b{}_d = \frac{1}{N-2} A' A'^T$$

Second, we did use in equation (10) the anomaly computation as explained in equation (6)[Hamill and Snyder 2000], and we agree that the way we explained the meaning of $\overline{x}_{s_o}$ can be misleading. For these reasons, we intend to change the following (line 120-121 p.5) "[...]:

$$\sigma^2_{b,REPS}(s_o) = \frac{1}{N-1}\left(x^{(REPS)}_{s_o} - \overline{x}_{s_o}\right)\left(x^{(REPS)}_{s_o} - \overline{x}_{s_o}\right)^T, \quad (1)$$

where $x^{(REPS)}_{s_o}$ is the N-length vector of REPS precipitation at location s_o and $\overline{x}_{s_o}$ is the ensemble mean at the same location." for "[...]:

$$\sigma^2_{b,REPS}(s_o) = \frac{1}{N-2}\sum_{i=1}^{N}\left(x^{(REPS)}_{s_o,i} - \overline{x}_{s_o[\![1,N]\!]\setminus i}\right)\left(x^{(REPS)}_{s_o} - \overline{x}_{s_o[\![1,N]\!]\setminus i}\right)^T, \quad (2)$$

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

Concerning the results I am a bit surprised that the performance curve un beta (figure 2-c) goes up between .3 and .4, is a sampling artifact ? Maybe the authors should add some comment on this fact or provide some estimation of the variability of the NRMSE.

- *Figure 2.c illustrates the NRMSE values that help compare the analysis, for different βs, during the winter against observation at surface stations in a leave-one-framework. The reviewer is right to mention the peculiarities present for $\beta = 0.3$ and 0.4. We carefully checked, and the difference between the two NRMSEs is around 0.1% and is indeed a sampling effect.*

In the same objective of better understanding the gain linked to the assimilation of the data in the model, would it be possible to compute the metrics (FBI-I, ETS, POD and FAR) when β=1 ?

- *We thank reviewer #1 for this comment also raised by reviewer #2. We will comment and add results when β=1 at different places in the new version of the manuscript. The following items list all the modification suggestions:*
  - *Figures 3 and 4 will additionally display the metric (FBI-I, ETS, POD, FAR) values when $\beta = 1$ (grey curve). The new version of these figures and their captions are provided at the end of this document (pages 11-12). We want to draw the reviewer's attention to Figure 3.d. An error occurred while merging different figures. In the current manuscript version, Figure 3.d is the same as Figure 4.b, which is wrong. We will correct Figure 3 accordingly. As shown on page 11 of this document, the results are much more consistent with the obtained NRMSE values (Figure 2.b). Indeed, using $\beta = 0.3$ and $\beta = 0.4$ during summer, with the assimilation of radar QPEs, lead to similar outputs in verification metrics.*
  - *In section 7.1, we intend to add the following sentence: "For all three experiments, the hybrid approach showed its relevance as it overcame both the static ($\beta = 0.0$) and the dynamic configuration ($\beta = 1.0$)."*
  - *We suggest to modify in section 7.2 lines 281-286:*
    * *" Figure 3.a illustrates the metrics for the summer without the radar QPEs for the optimal $\beta = 0.5$ compared to $\beta = 0.0$, where filled markers indicate no significant differences at the 95% confidence level between the two experiments for a given threshold. The 6-h precipitation analysis displayed a significant increase of 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 when looking at the small precipitation events. As illustrated by the FBI-1, the selection of $\beta = 0.5$ led to generally lower precipitation amounts than with the use of $\beta = 0.0$. 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.'*
    * *for "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 $\beta = 0.5$."*
  - *We also propose to add right after the sentence "the POD slightly deteriorated, and the FBI-1 reduced for small precipitation events but increased for events of medium to high intensity." (line 296-297) the followings "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 $.2$ mm, but the degradation for other thresholds is too substantial."*

- *To comment of the winter results, we will change:*
  - *lines 300-306: "Finally, Figure 4.a illustrates the same metrics during the winter and compares the $\beta = 0.7$ to the reference experiment. 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)."*
  - *for "Finally, Figure 4.a illustrates the same metrics during the winter and compares the $\beta = 0.7$ and 1.0 to the reference experiment. 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 that 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 (Fig 3.a)."*
- *Finally, we intend to change lines 313-314 "In light of these results, it appears that the use of the optimal β value identified through the use of NRMSE did indeed show an improvement in skills and a reduction in false alarm rates for both summer and winter experiments." for "In light of these results, the optimal β value identified through the use of NRMSE 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."*

In the metrics, the authors point out the values that are not significantly different, maybe they could also plot the variability (errors bars, or boxplots ?)

- *We thank the reviewer for this interesting comment. The statistical significance of the difference between scores of different configurations regards Figures 3 and 4, for a total of 24 subplots. In the revised version of the manuscript, each subplot will contain the results for $\beta = 1.0$ (see reply to the previous comment), leading to three curves, and adding the variability information for each of this curve would, therefore, highly burden Figure 3 and 4, without adding much information. We hope it is okay if we do not make any modifications regarding the variability in the scores.*

Specific issues

- Eq 1, the model could be presented first, so that we know what $\widetilde{\mathrm{P}}^{\mathrm{b}}$ corresponds to. In particular, it could be useful to have the size of the matrices

- *The introduction of the forecasting models before presenting the hybrid approach will be addressed in the revised version of the manuscript. This topic is thoroughly discussed in the General Comments section above. $\widetilde{P}^b$ is a covariance matrix and is therefore by definition a square matrix between each pair of elements of a given random vector. In the context of data assimilation, the covariance matrix of the background error is computed at each grid-cell of the domain using the neighboring grid-cells where observations are collocated. If there are M locations with observations in the vicinity of a grid-cell, the background values (i.e., the REPS and the RDPS) at those locations will be used to build a $M \times M$ covariance matrix. In CaPA, no matter the configuration, we use a maximum of 16 points per type of observation in the vicinity when building covariance matrices. Past testing experiments led to this choice and helped speed up the analysis computation while ensuring a good quality product. Moreover, the matrice size is already discussed in Lines 105-109. We added information on the matrix size when confusion was possible, for example, in equations 5 and 6. For these reasons, we suggest not adding more information regarding the size of the different matrices, and we hope the reviewer will understand this point.*

- Eq 5 the prime notation is introduced a bit too early I guess, and P^{α}_{OI} (L96) is not yet defined. Similarly, $A_{:,j}$ (eq 6) is not defined.

- *We are sorry for the confusion here. We will change the P^a_{OI} for P^b_{OI}. P^a_{OI} corresponds to the analysis error covariance matrix when the analysis is solely based on the OI approach. However, it is not explicitly defined in the manuscript. We intend to keep the prime notation for the anomaly (perturbation) matrix as in equation (5) as it is not the conventional way to compute it. However, the reviewer is correct about the lack of clarity in Equation 5, which also has an error (see our previous comments in the General comments section). Indeed, the denominator should be $(N-2)$ and not $(N-1)$. For these reasons, we propose to change the following:*

  - **line 96-104**: *"The covariance matrix ${P^b}_d$ depicts the flow-dependant errors estimated from the REPS and is defined as:*

$$ {P^b}_d = \frac{1}{N-1} A' {A'}^T, \tag{3} $$

    where $A' \in R^{m \times N}$ 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:

$$ {A'}_{:,j} = A_{:,j} - \overline{A}_{[\![1,N]\!]\setminus j}, \tag{4} $$

    where ${A'}_{:,j}$ is the j-th column of A' and $\overline{A}_{[\![1,N]\!]\setminus j}$ is the average across the N members without the j-th member."

  - *by: "The covariance matrix ${P^b}_d$ depicts the flow-dependant errors estimated from the REPS and is defined as:*

$$ {P^b}_d = \frac{1}{N-2} A' {A'}^T, \tag{5} $$

where $A' \in R^{m \times N}$ 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:

$$A'_{[:,j]} = X_{f\,[:,j]} - \overline{X_f}_{[\![1,N]\!]\setminus j}, \tag{6}$$

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

- L155: I understand that SYNOP is a netwrok of stations, but the term is not introduced earlier.

- *The reviewer is totally right, SYNOP stands for synoptic stations and the acronym was not defined. We will modify*
  - **line 155**: *"For more reliability, the NRMSE is computed only at SYNOP and manual SYNOP stations during the summer and [...]"*
  - *by: "For more reliability, the NRMSE is computed only at surface synoptic observations (hereafter SYNOP) and manual SYNOP stations during the summer and [...]"*

  Moreover, the suggested Appendix A (see page of the present document) will also define SYNOP among other acronyms.

- L174: 'progressively' if quite unprecise, if the authors think it is worth mentioning, I think it should be precised;

- *We are sorry for the lack of precision on this point. The sentence in line 174 : "New Canadian dual-polarization Doppler radars have been progressively added to the observation database." refers to the ongoing replacement of C-band to S-band radars initiated in 2017 (see https://www.canada.ca/en/environment-climate-change/services/weather-general-tools-resources/radar-overview/modernizing-network.html). We propose to change line 174 for "Since 2017, new Canadian dual-polarization Doppler radars have been progressively replacing their C-band counterpart."*

- L187: 'in the transformed space', does it refer to the box-cox transformed data ?

- *Yes, it does. We use this comment to point out that, thanks to reviewer #2, an observation is rejected if eq. (13) is* **not** *satisfied and not the other way around as specified in the current version of the manuscript. Therefore, we propose the following modification to help better understand the CaPA quality control and correct the misinterpretation of eq. 13. :*
  - *line 183-187: "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:*

  $$\left| x_{s_k}^{(OBS)} - x_{s_k}^{(CaPA)} \right| < tol \cdot \sqrt{(\sigma_o^2 + \sigma_a^2)}, \tag{13}$$

  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"

– *by: For this purpose, an analysis is estimated at a site s_k ($x_{s_k}^{(a)}$) using neighboring stations in a LOO approach. The observation at the same site ($x_{s_k}^{(o)}$) is rejected or, is said invalid, if the following is not fulfilled:*

$$\left|x_{s_k}^{(o)} - x_{s_k}^{(a)}\right| < tol \cdot \sqrt{(\sigma_o^2 + \sigma_a^2)}, \tag{13}$$

where tol is a tolerance factor set equal to 4 for the operational CaPA. Both observed precipitation and analysis estimates of equation (13) undergone the Box-Cox transformation beforehand (see eq. 2).

- L201: what are the 'robustness reasons' ?

- *We are sorry to have taken the reader's knowledge of Stage IV product performance for granted. Stage IV results from an advanced blending of surface stations and radar datasets, but this product still suffers from bias and has some limitations, especially in the western CONUS domain. Nelson et al. (2016) study provides a thorough picture of the strengths and limitations of Stage IV. The lack of reliability of radar data in complex terrain is one reason to discard the western CONUS domain. It is a common practice when Stage IV is used as a reference for verifications. A quick search in Google Scholar for papers quoting Nelson et al. (2016) article provides an idea of how common it is to not use the western part of the CONUS domain in Stage IV. We, therefore, suggest replacing the following:*
  - *lines* **200-202***: "The ST4 domain covers the contiguous United States (CONUS), but for robustness reasons, only the CONUS east of 105W was used for verification purposes [...]"*
  - *by: "The ST4 domain covers the contiguous United States (CONUS), but for known limitations in the western domain (see Nelson et al. 2016 for more details), only the CONUS east of 105W was used for verification purposes [...]"*

- Eq 16: the sum in the denominator are between 1 and N_y ?

- *We thank the reviewer for catching these typos. Reviewer #2 also highlighted that a bracket is not at the right place in the numerator. Therefore, we will modify equation 16, initially written as:*
  - **line 246***: "*

$$FSS=1-\frac{\frac{1}{N_y}\sum_{i=1}^{N_y}\left(f_a(i)-f_{o(i)}\right)^2}{\frac{1}{N_y}\left[\sum_{i=1}^{n}f_{a(i)}^2+\sum_{i=1}^{n}f_{o(i)}^2\right]} \tag{16}$$

"

  - *for "*

$$FSS=1-\frac{\frac{1}{N_y}\sum_{i=1}^{N_y}\left(f_{a(i)}-f_{o(i)}\right)^2}{\frac{1}{N_y}\left[\sum_{i=1}^{N_y}f_{a(i)}^2+\sum_{i=1}^{N_y}f_{o(i)}^2\right]} \tag{16}$$

"

- L374: the "two different interpretations" should be precised, they are not clear (at least to me)

- *We are sorry for not being precise enough about the difficulty of interpreting low CFIA values when using the hybrid approach. Section 7.4 discusses the limitations of using the CFIA field in its definition (eq. 11). Presently, for a given grid-cell, if there are no stations or radars in the vicinity, the analysis is set equal to the background field (lines 109-110), therefore in equation 11, σ_a^2 will tend to be equal to σ_b^2, and CFIA will tend towards low values. This way of computing CFIA is* *true no matter the value of β. Increasing the contribution of the REPS, i.e., increasing β, leads to CFIAs following meteorological spatial distributions, which is a nice feature. However, when no precipitation is observed in ensemble forecasts, the anomaly matrix tends towards a null matrix. Thus the matrix P_d^b (eq. 5), tends towards zero. Here again, σ_a^2 as defined in equation 9 will tend to be equal to σ_b^2 and again, CFIA values will be small. Therefore, the user may ask if the small CFIA values he/she is reading reflect the absence of precipitation or the absence of assimilated observations. To better illustrate this point, we intend to add after line 369:*

- *"Generally, CFIAs tended to be higher at places with precipitation than when using $\beta = 0.0$, as shown in the eastern part of the domain for the January 18 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_d^b (eq. 5), which, when the ensemble members have no or very little precipitation, will tend to a zero matrix."*

We also suggest to modify line 370-371 (p.13):

- *"This is explained by the current CaPA computation framework, where gridcells are set equal to the background when no observations are available in the vicinity."*
- *for "'This is explained by the current CaPA computation framework. 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)."*

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

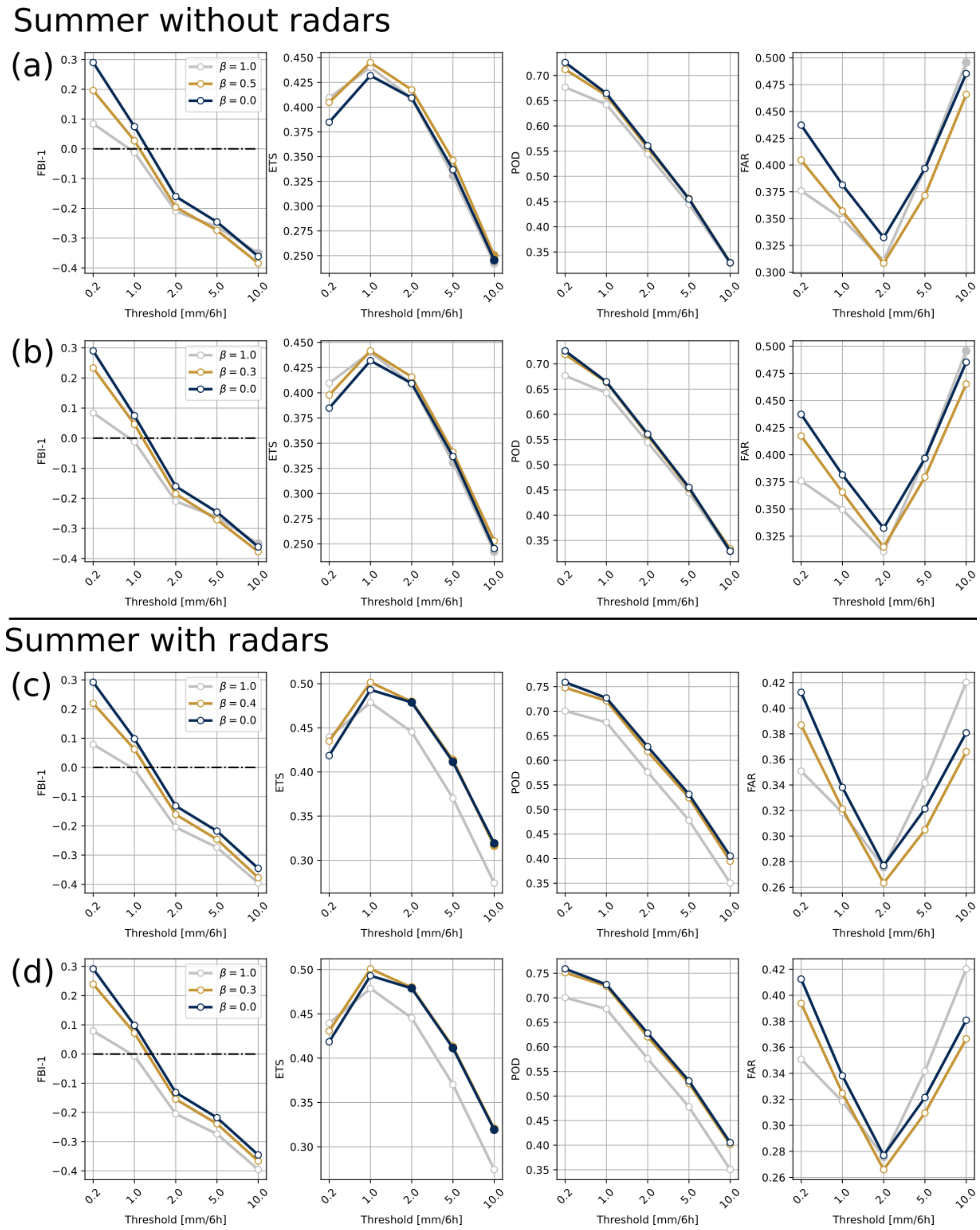

Figure 3. FBI-1, ETS, POD and FAR across the whole domain for summer experiment without radar QPEs for precipitation analysis with $\beta = 0.0$ (dark blue line), $\beta = 1.0$ (grey line), $\beta = 0.5$ (yellow line in a) and $\beta = 0.3$ (yellow line in b). Same figures but for the summer experiment with the assimilation radar QPEs with $\beta = 1.0$ (grey line), $\beta = 0.4$ (yellow line c) and $\beta = 0.3$ (yellow line d) all three compared to the reference experiment when $\beta = 0.0$ (blue line). 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.

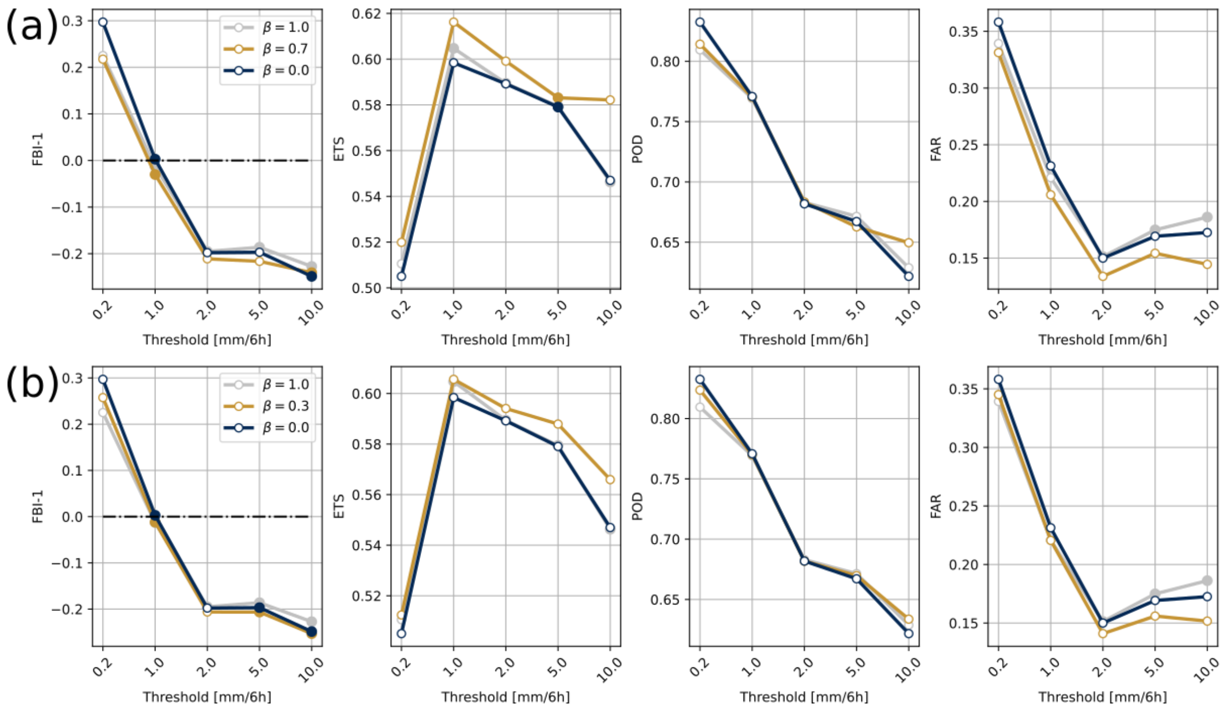

Figure 4. Same as Fig. 3 but for the winter experiment and with $\beta = 0.7$ (top panel), $\beta = 0.3$ (bottom panel) and $\beta = 1.0$ (two panels), all three compared to the reference experiment ($\beta = 0.0$).