Norwegian Meteorological Institute

METreport

No. 13/2022
ISSN 2387-4201
Meteorology

Addressing predictability in the AROME-Arctic ensemble using Stochastically Pertrubed Parameters (SPP) and Ensemble Data Assimilation (EDA)

Alertness project deliverable
Andrew Singleton

Photo: Eivind Støylen

Norwegian Meteorological Institute

METreport

Title	Date
Addressing predictability in the AROME-Arctic ensemble using Stochastically Pertrubed Parameters (SPP) and Ensemble Data Assimilation (EDA)	August 24, 2022
Section Meteorology	**Report no.** 13/2022
Author(s) Andrew Singleton	**Classification** ● Free ○ Restricted
Client(s) Norwegian Research Council	**Client's reference** Project number 280573 'Advanced models and weather prediction in the Arctic: enhanced capacity from observations and polar process representations (ALERTNESS)

Abstract

This report describes work done as part of the Alertness product to address uncertainties in the AROME-Arctic ensemble using Ensemble Data Assimilation (EDA) and Stochastically Perturbed Parameters (SPP). Forecasts using EDA and SPP were compared against reference forecasts using a more basic implementation of the AROME-Arctic ensemble. In general both SPP and EDA led to increased spread in the ensemble than the reference without impacting the errors in a negative way. However, SPP introduced some large biases, especially for 2m temperature and 10m wind speed, and EDA resulted in stringer biases for total cloud cover. It is recommended that SPP should not be taken into use until the biases are rectified, but EDA can be taken into use since overall it leads to a better representation of the uncertainties in the forecasts.

Keywords

Ensembles, Arctic, Verification, Weather Forecasting

Disciplinary signature

Responsible signature

Norwegian Meteorological Institute
Org.no 971274042
post@met.no
www.met.no / www.yr.no

Oslo
P.O. Box 43, Blindern
0313 Oslo, Norway
T. +47 22 96 30 00

Bergen
Allégaten 70
5007 Bergen, Norway
T. +47 55 23 66 00

Tromsø
P.O. Box 6314, Langnes
9293 Tromsø, Norway
T. +47 77 62 13 00

Contents

1 Introduction

Predictability in the atmosphere can be broken down into two distinct categories - intrinsic predictability and practical predictability. Intrinsic predictability is the predictability that could be achieved by a perfect model with perfect initial conditions (Lorenz 1969), and is essentially governed by the chaotic nature of the atmosphere. Practical predictability on the other hand is the predictability that is achievable by a numerical weather prediction (NWP) model that is limited by uncertainties in its initial conditions and modelling methodologies (Surcel et al. 2015). Ensemble prediction systems (EPS) have been developed to account for limitations to both intrinsic and practical predictability by producing many forecasts, in which initial conditions and model processes are perturbed, such that the uncertainty of a forecast can be estimated. These perturbations are often designed to represent the uncertainties in, for example, the observing systems from which the initial conditions are derived and the empirical relationships that govern the computation of physical processes. In this sense, the perturbations are used to take account of the practical predictability. However, the perturbations also model the intrinsic predictability of the atmosphere in that they will have a larger impact on the forecast in less predictable situations (Lorenz 1963).

Predictability has both spatial and temporal dimensions. For example, in a study of the spatial and temporal scales of predictability for MEPS, the operational implementation of the HarmonEPS model (Frogner et al. 2019a) for a domain over the Nordic countries, it was shown that the minimum predictable spatial scale for hourly accumulated precipitation grew from approximately 15 km in the first hours of the forecast, to 75 km by hour 12 and 100 km by hour 36 (Frogner et al. 2019b). Furthermore, it was found that MEPS was better able to account for uncertainties than the global model, IFSENS, from the European Centre for Medium-range Weather Forecasts (ECMWF) for a range of weather parameters.

In a previous study for the ALERTNESS project (Singleton and Grote 2020), it was found that the AROME-Arctic ensemble (AAEPS) better modelled uncertainties for a range of weather parameters than IFSENS, which is the current state of the art ensemble available for the Arctic region. This suggested that AAEPS is a promising solution for weather forecasting for the Arctic region around northern Norway, the Barents Sea and Svalbard. However, uncertainties in the initial conditions were modelled using the fairly unsophisticated SLAF method (Toth and Kalnay 1993) and no attempt was made to account for uncertainties in the modelling system itself. In this report, we document the impact of using the stochastically perturbed parameters (SPP) method (Ollinaho et al. 2017) to model uncertainties in the modelling system and ensemble data assimilation (EDA) (Bouttier et al. 2016) to better model uncertainties in the initial conditions.

SPP is a method that attempts to take account of uncertainties in physics parameterizations that are typically based on empirical relationships. The scheme adds spatially and temporally correlated stochastic noise to selected parameters in various parameterization schemes (Frogner et al. 2022). In this way uncertainties are accounted throughout the integration time of the model forecast. EDA on the other hand only takes account uncertainties in the initial conditions. Each ensemble member has its own analysis that is derived by perturbing both conventional (synop, ship, buoy, aircraft and radiosonde) observations, and radiances from satellites (Isaksen et al. 2010; Frogner et al. 2019a). Over the Arctic, the relative paucity of conventional observations means that satellite radiances are especially important for EDA.

2 Methods

In this section the methods used in this study are briefly described. In general, the same methods are used here as described in Singleton and Grote (2020) and where they differ is noted below. More detailed information is available in the cited works.

2.1 AROME-Arctic EPS (AAEPS)

As in Singleton and Grote (2020), the model used is based on the operational AROME-Arctic model (Müller et al. 2017) in an ensemble implementation (Frogner et al. 2019a) with 2.5 km horizontal resolution. It was found that 6 perturbed members achieved very similar results to the 10 perturbed members in Singleton and Grote (2020), so only 6 perturbed members are used in this study to save computational resources. Additionally, Singleton and Grote (2020) showed that the ensemble introduced biases in near surface temperature and humidity. This was due to small unbalances in the surface perturbations that grew with every data assimilation cycle. Switching off data assimilation for the perturbed members was found to alleviate this problem. Therefore surface data assimilation cycling is switched off for the perturbed ensemble members here - i.e. surface perturbations are added to the control analysis for each ensemble member. Forecasts from this basic setup are referred to as REF in this report.

The HarmonEPS model is updated in cycles and for this study cycle 40 of the model is used for the SPP runs, and cycle 43 is used for the EDA runs. Reference runs are done for both model cycles. For all experiments one long ensemble forecast (36 - 48h) is run per day initialized at 00 UTC with 3-hour cycling of the data assimilation system.

2.2 SPP

The SPP scheme in HarmonEPS (Frogner et al. 2022) follows that developed by ECMWF (Ollinaho et al. 2017). Perturbations to selected parameters in physics schemes related to cloud microphysics, radiation, convection and turbulence are initially drawn from a lognormal distribution with a prescribed mean and standard deviation for each parameter. The spatial and temporal correlations are controlled using the stochastic pattern generator (SPG: Tsyrulnikov and Gayfulin 2017). Following Frogner et al. (2022), the spatial correlation length scale is set to 200 km and the temporal correlation length scale is set to 12 hours. The same pattern is used for each model level in the vertical.

The parameters that are perturbed are summarized in Table 1. Note that the percentiles of the distributions are given as the ratio of the perturbed value to the unperturbed value. The results of REF and SPP are compared for periods during YOPP (Year of Polar Prediction) special observing periods (SOP) 1 (8 - 31 March 2018) and 2 (10 July - 1 August 2018).

Table 1: Summary of SPP perturbations. Type is the type of parameterization scheme that is impacted where CONV is convection, IM is ice microphysics, LM is liquid microphysics, RAD is radiation and TURB is turbulence; Parameter is the variable name in the HarmonEPS code; Control is the value for the control member; Std Dev, 5th %ile and 95th %ile are the standard deviation, 5th and 95th percentiles of the distribution from which the perturbations are sampled, expressed relative to the control value.

Type	Parameter	Description	Cntrl	Std Dev	5th %ile	95th %ile
CONV	CLDDPTHDP	Threshold for cloud thickness used in shallow or deep convection decision	4000	0.1	0.07	3.50
IM	ICE_CLD_WGT	Cloud ice content impact on cloud thickness	1.00	0.4	0.07	3.50
IM	ICENU	Ice nuclei concentration	1.00	0.7	0.03	31.60
LM	VSIGQSAT	Saturation limit sensitivity for condensation	0.03	0.4	0.07	3.50
LM	KGN_ACON	Kogan autoconversion speed	10.00	0.5	0.03	3.81
LM	KGN_SBGR	Kogan subgrid-scale (cloud fraction) sensitivity	0.50	0.2	0.31	2.24
RAD	RADGR	Graupel impact on radiation	0.50	0.3	0.15	2.93
RAD	RADSN	Snow impact on radiation	0.50	0.3	0.15	2.93
TURB	RFAC_TWO_COEF	Top entrainment efficiency	2.00	0.4	0.07	3.50
TURB	RZC_H	Stable conditions length scale	0.15	0.4	0.07	3.50
TURB	RZI_INF	Asymptotic free atmosphere length scale	100	0.6	0.01	3.84

2.3 EDA

EDA results in an ensemble of analyses from which to initialize each ensemble member of the forecast. The method used was initially designed to estimate error covariance matrices that are a part of variational data assimilation systems and is based on Isaksen et al. (2007). The data assimilation is done using 3DVAR (3-dimensional variational data assimilation) separately for the upper air and surface observations, with the perturbations added for each ensemble member such that each member gets its own analysis to start each model run from.

At the time of running the experiments, EDA was only functioning reliably for cycle 43 of the HarmonEPS model so it was decided to use this cycle. Furthermore, many more satellite observations became available after the YOPP SOPs used in Singleton and Grote (2020) and the SPP runs meaning that a later period was chosen to analyse the EDA performance (25 October - 7 November 2019). This period was chosen since it included a polar low event.

2.4 Verification metrics

The aim of this report is analyse the performance of SPP and EDA relative to a reference where these methods are not used. The goal of both methods is to introduce more spread to the forecast without increasing errors, or unbalancing the ensemble. For this reason, we concentrate on spread - skill comparisons whereby the spread of the ensemble (the square root of the mean variance) is compared with the the route mean square

error (RMSE) of the ensemble mean. For a well calibrated ensemble, the ensemble spread should be equal to the RMSE. In order to check the balance of the perturbations, biases of the individual ensemble members are inspected. Where there are imbalances the verification is stratified into ranges of the observed parameter to ascertain whether the imbalances are across the board or only at the extremes.

Verification is done for a range of near surface parameters (2m temperature, 2m specific humidity, 10m wind speed, mean sea level pressure, 6-hour accumulated precipitation and total cloud cover) at available land based weather stations; and for vertical profiles of temperature, specific humidity, wind speed and geopotential height for all available radiosonde launches. The location of the stations is shown in Fig. 1.

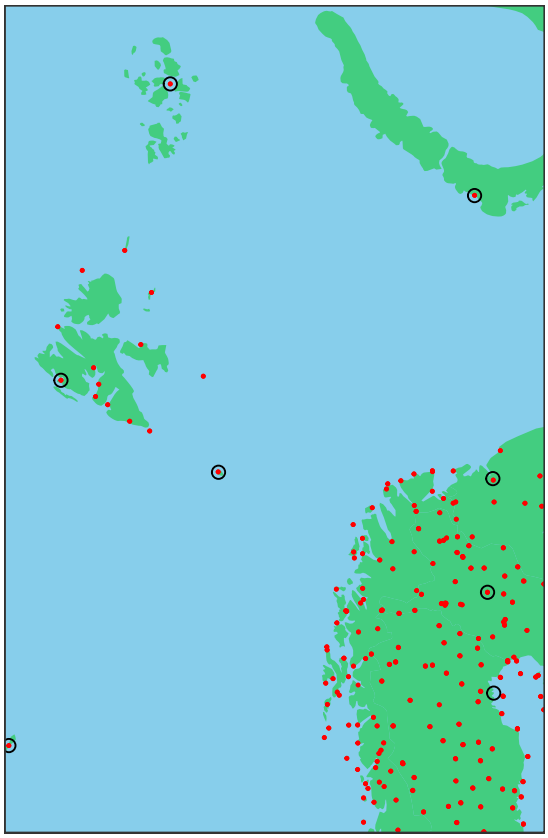

Figure 1: Location of observation stations used in the verification. Red dots denote surface stations and black circles are radiosonde launch locations.

3 Results

3.1 SPP

In this section, the results for the SPP experiments are reported for both SOP 1 (winter) and SOP 2 (summer) individually, and combined.

3.1.1 2m Temperature

Spread-skill plots for 2m temperature are shown in Fig. 2. For the two SOPs combined, the spread can be seen growing with lead time for both REF and SPP, with SPP having larger spread throughout the forecast. The RMSE for REF and SPP appears to be close to equal. For the individual SOPs a clear diurnal cycle in the spread is apparent, with SPP generating more spread during the night, which is particularly noticeable for SOP 1 when the nights are much longer.

While these results look promising for SPP, the bias of the individual members tells a slightly different story (Fig. 3). It is clear that SPP introduces a cooler bias into the perturbed members when compared with REF. While some of the problems with REF reported in Singleton and Grote (2020) (warmer bias in winter, cooler bias in summer) persist, switching off surface data assimilation cycling for the perturbed members has improved this situation. The cooling impact of SPP on the perturbed members appears to be stronger in the winter (SOP 1) than the summer (SOP 2).

Stratifying the biases for the perturbed members relative to the control member by observed 2m temperature suggests that the cooling due to SPP occurs for all observed 2m temperatures, but is weakest for temperatures between $2.5°C$ and $10°C$. The strongest cooling occurs between observed 2m temperatures of $-12.5°C - 0°C$ and $17.5°C - 25°C$.

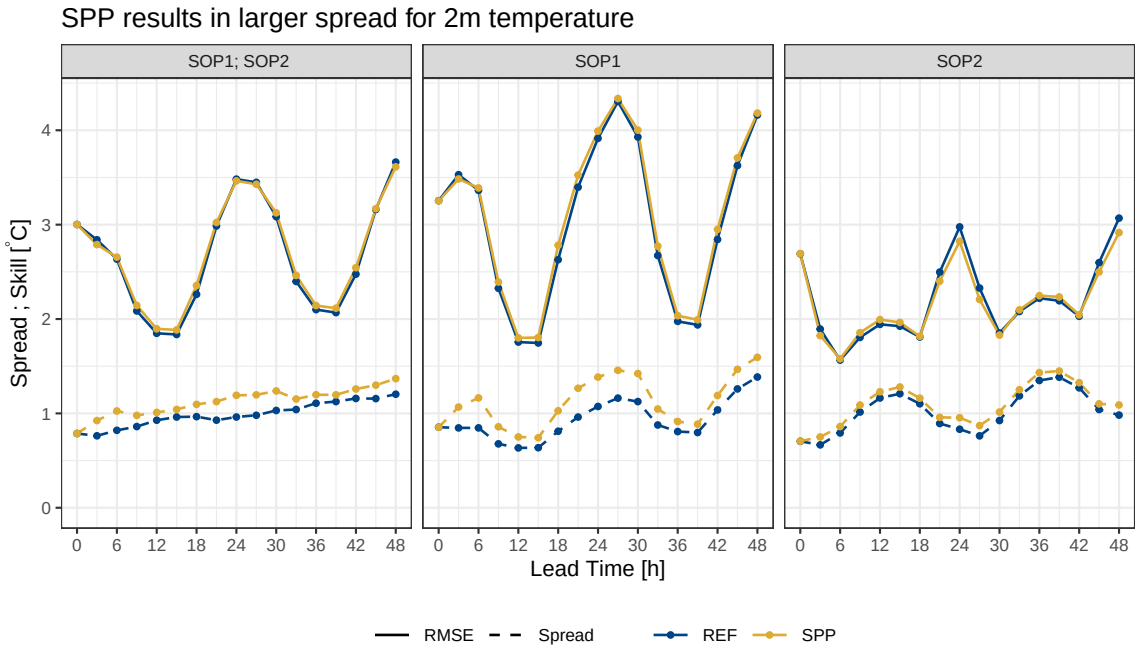

Figure 2: RMSE (solid line) and ensemble spread (dashed line) for 2m temperature for REF (blue) and SPP (gold) for (from left to right) both special observing periods combined, SOP 1 (8 - 31 March 2018) and SOP 2 (10 July - 1 August 2018).

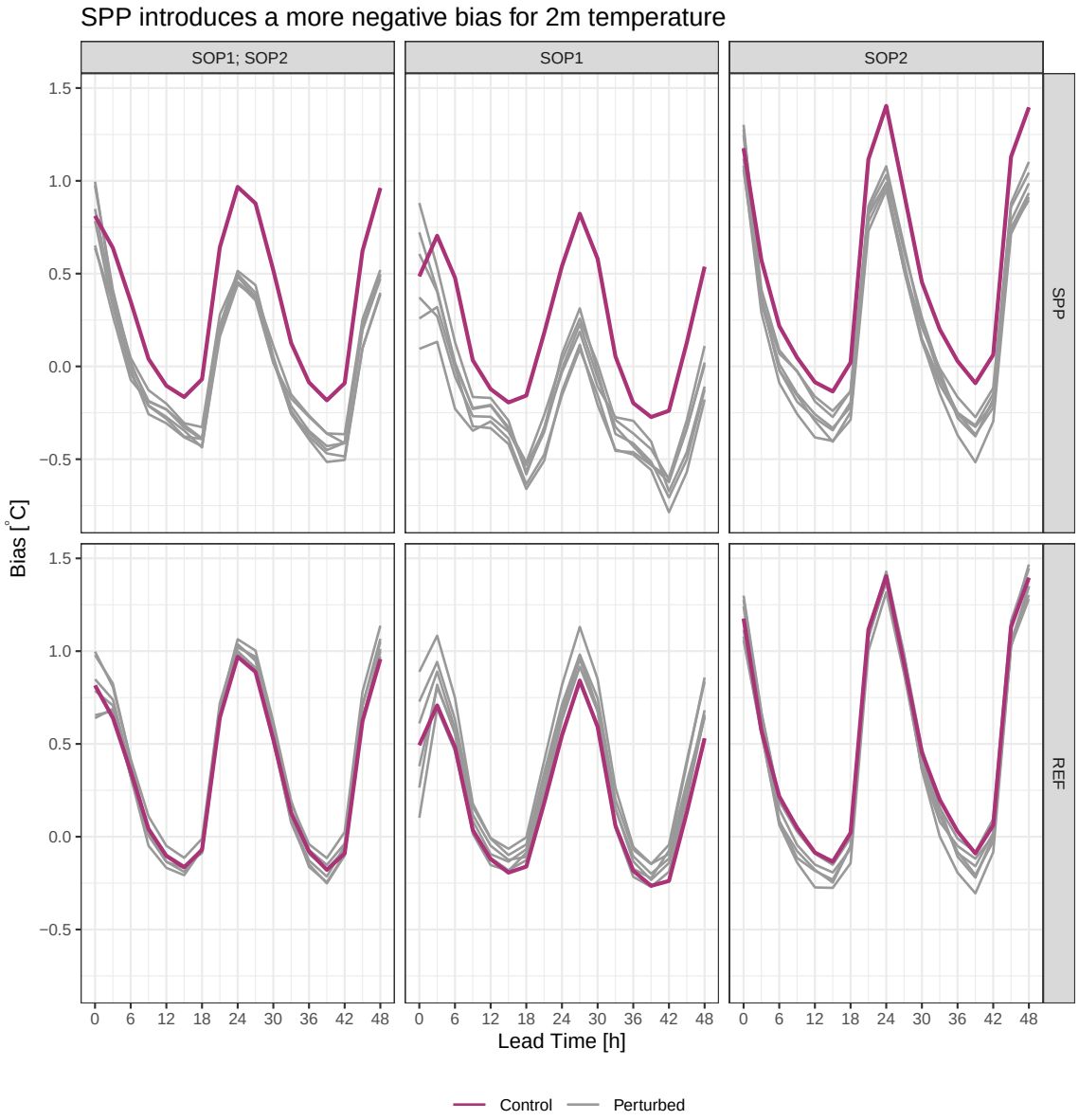

Figure 3: Bias for control (purple) and perturbed (grey) members for 2m temperature for SPP (top row) and REF (bottom row) for (from left to right) both special observing periods combined, SOP 1 (8 - 31 March 2018) and SOP 2 (10 July - 1 August 2018).

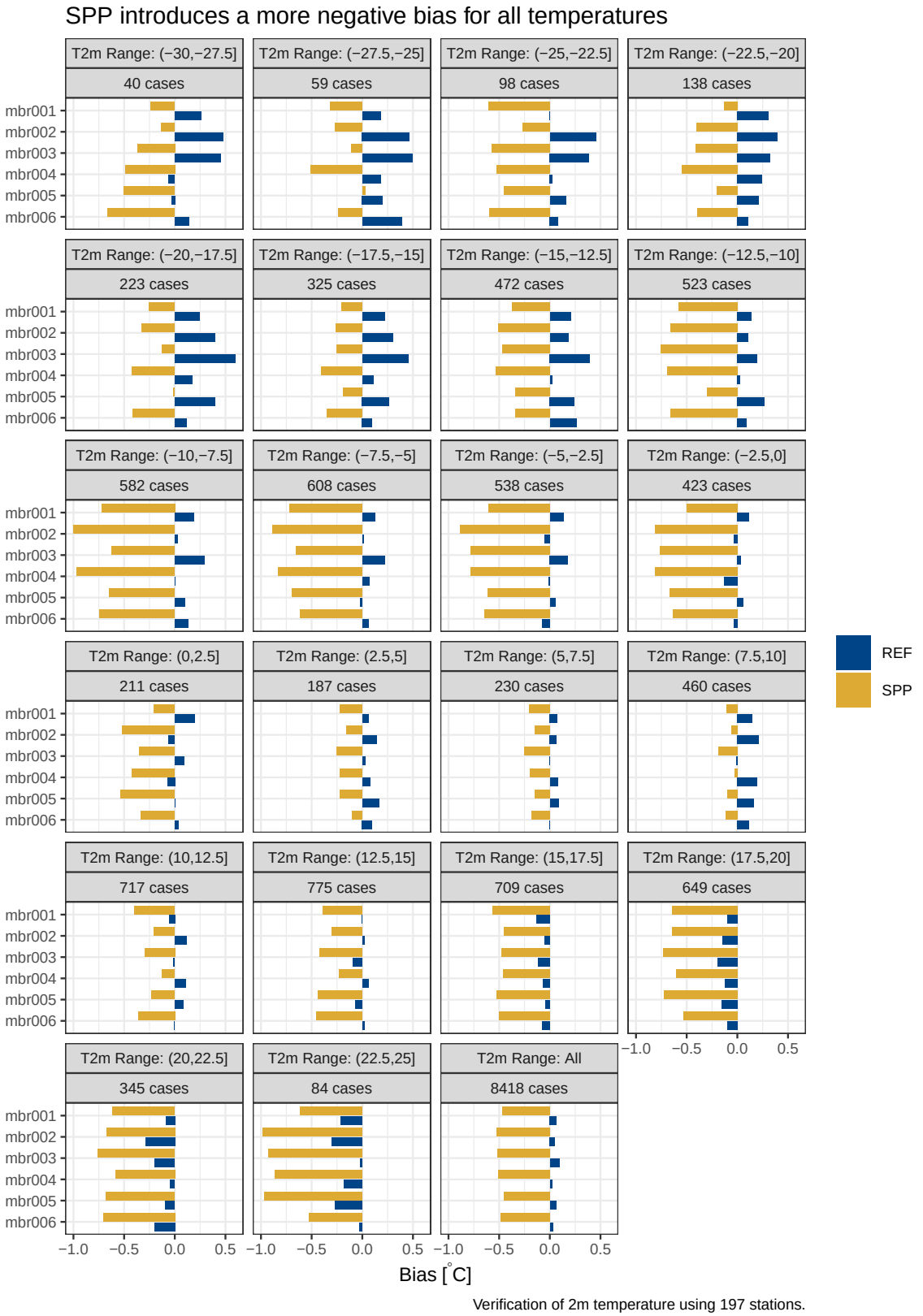

Figure 4: Bias for each ensemble member relative to the control member for 2m temperature for REF (blue) and SPP (gold) for both special observing periods combined at 24-hours lead time.

3.1.2 2m Specific Humidity

For 2m specific humidity, SPP introduces slightly more spread than REF, but for summer (SOP 2) and the the two SOPs combined, there is an increase in RMSE when the spread is at its peak values (Fig. 5). There appears to be little diurnal cycle in the spread during the winter, so the verification for the SOPs combined is dominated by the summer values.

The biases for the individual members (Fig. 6) suggest that SPP introduces a drier bias during SOP 1 and a generally moister bias during SOP 2. The biases for the individual members for REF are better balanced than those reported in Singleton and Grote (2020).

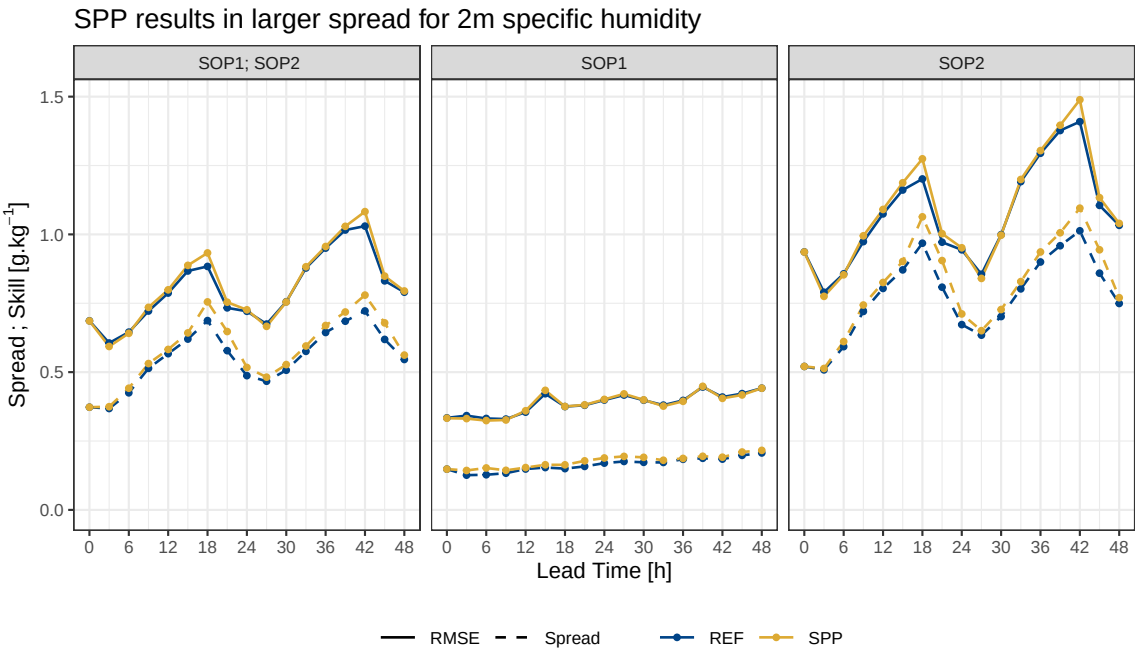

Figure 5: RMSE (solid line) and ensemble spread (dashed line) for 2m specific humidity for REF (blue) and SPP (gold) for (from left to right) both special observing periods combined, SOP 1 (8 - 31 March 2018) and SOP 2 (10 July - 1 August 2018).

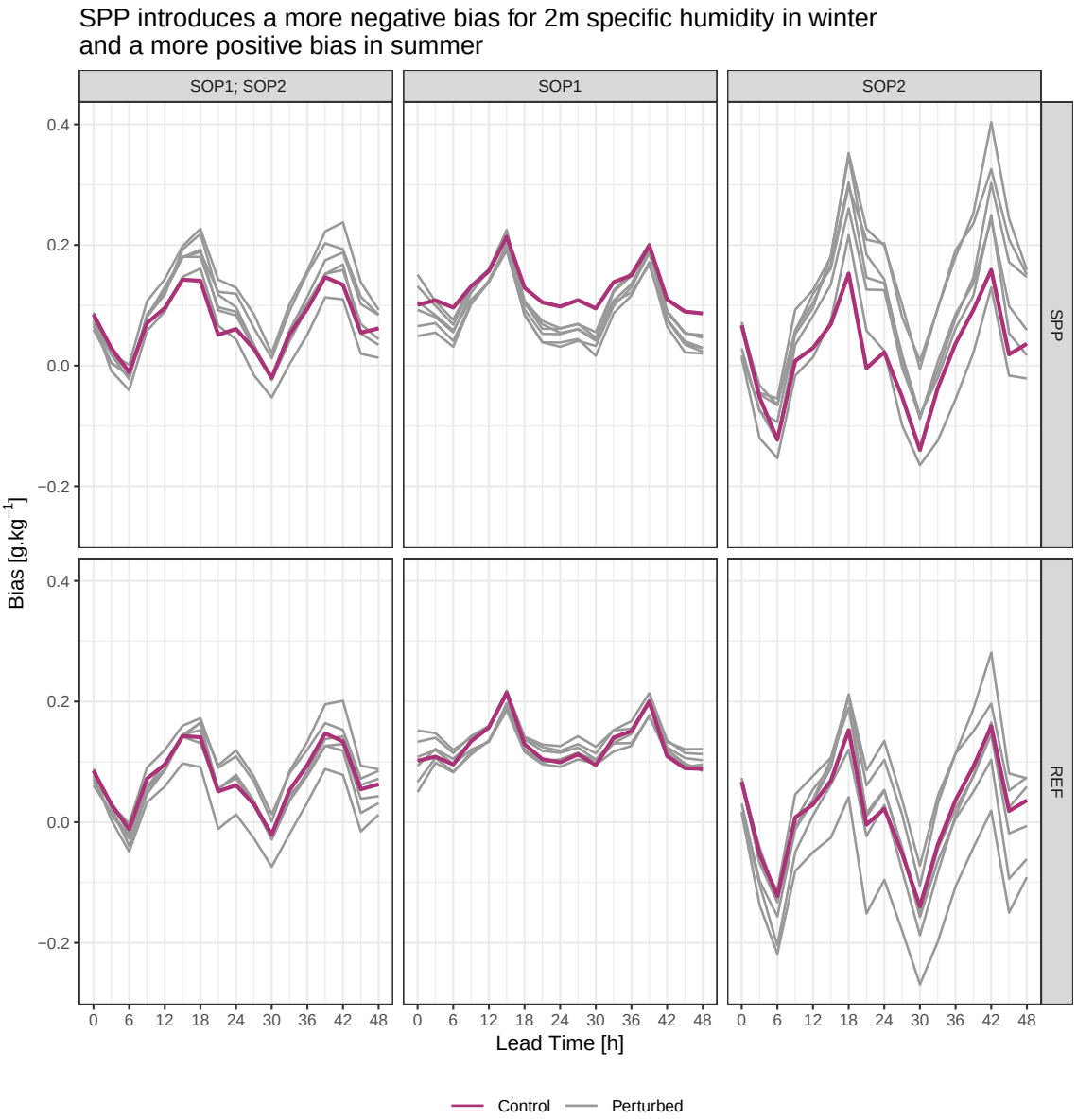

Figure 6: Bias for control (purple) and perturbed (grey) members for 2m specific humidity for SPP (top row) and REF (bottom row) for (from left to right) both special observing periods combined, SOP 1 (8 - 31 March 2018) and SOP 2 (10 July - 1 August 2018).

3.1.3 10m Wind Speed

The verification for 10m wind speed shows that the spread increases throughout the forecast at a slightly faster rate than the RMSE for both REF and SPP in both SOP 1 and SOP 2 (Fig. 7). During SOP 1, the RMSE for SPP is slightly smaller than that for REF throughout the whole forecast period, whilst during SOP 2 there is a slight diurnal cycle in both the spread and the RMSE.

Taking the biases of the individual members into account (Fig. 8), it is clear that REF introduces a more positive bias to the 10m wind speed, whereas for SPP a more negative bias is introduced, which is much stronger during SOP 1. During SOP 1, this more negative bias brings the biases closer to zero since the starting point is a positve bias, but during SOP 2 the starting point is already a negative bias so the biases get further away from zero. This could explain the reduction in RMSE that SPP resulted in for SOP 1.

Stratifying the biases by values of observed 10m wind speed suggests that the more negative biases are strongest for wind speeds below $15ms^{-1}$. For higher wind speeds there are some cases where the perturbed members have a more positive bias than the control member for SPP, but the small number of observations of such high wind speeds means that these results might not be entirely meaningful.

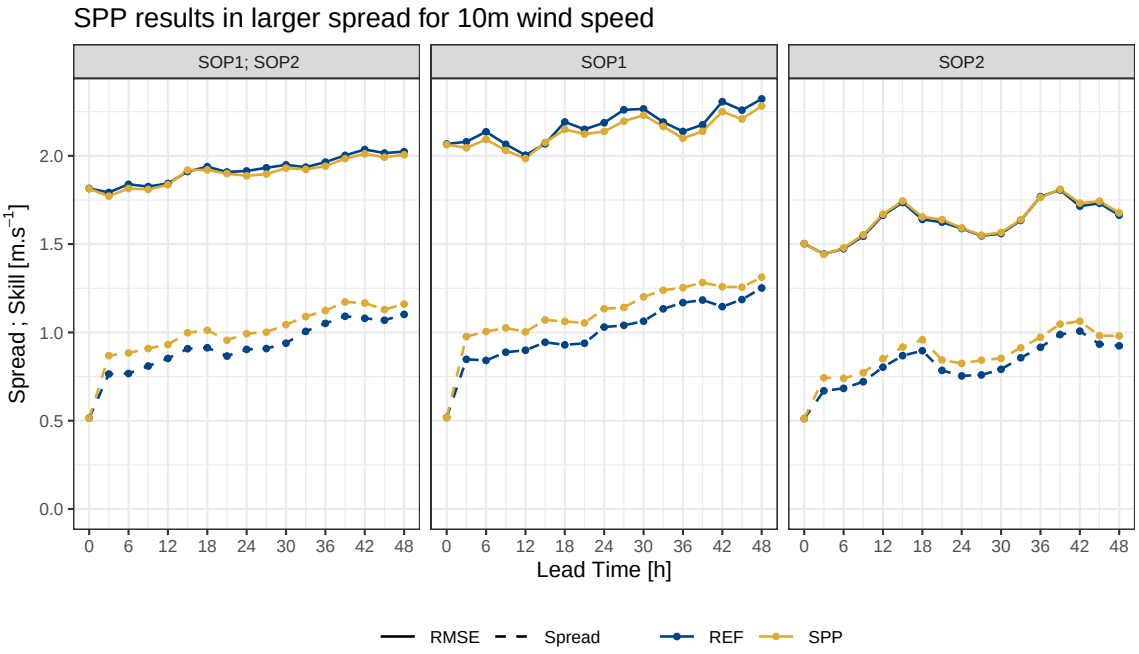

Figure 7: RMSE (solid line) and ensemble spread (dashed line) for 10m wind speed for REF (blue) and SPP (gold) for (from left to right) both special observing periods combined, SOP 1 (8 - 31 March 2018) and SOP 2 (10 July - 1 August 2018).

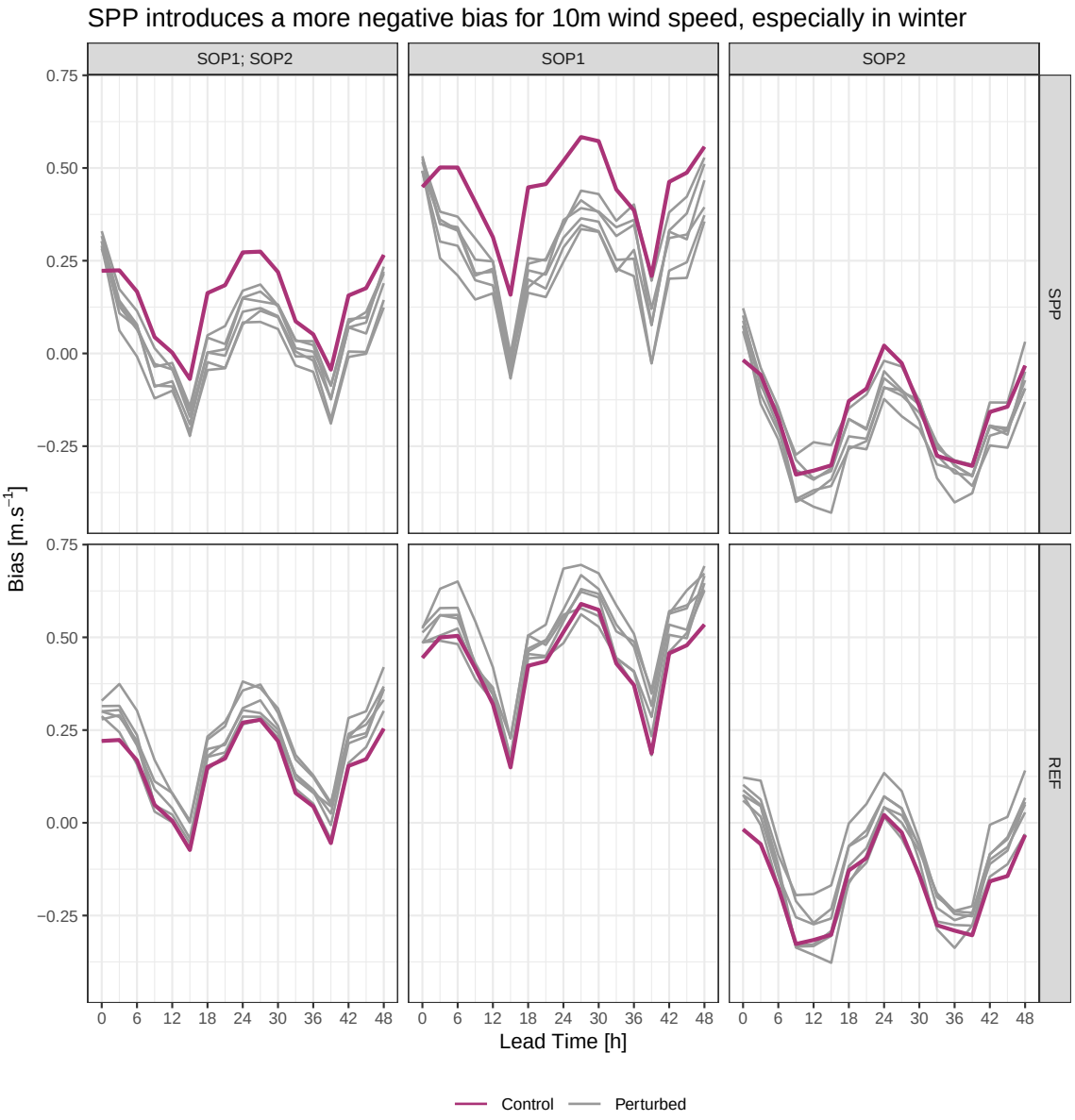

Figure 8: Bias for control (purple) and perturbed (grey) members for 10m wind speed for SPP (top row) and REF (bottom row) for (from left to right) both special observing periods combined, SOP 1 (8 - 31 March 2018) and SOP 2 (10 July - 1 August 2018).

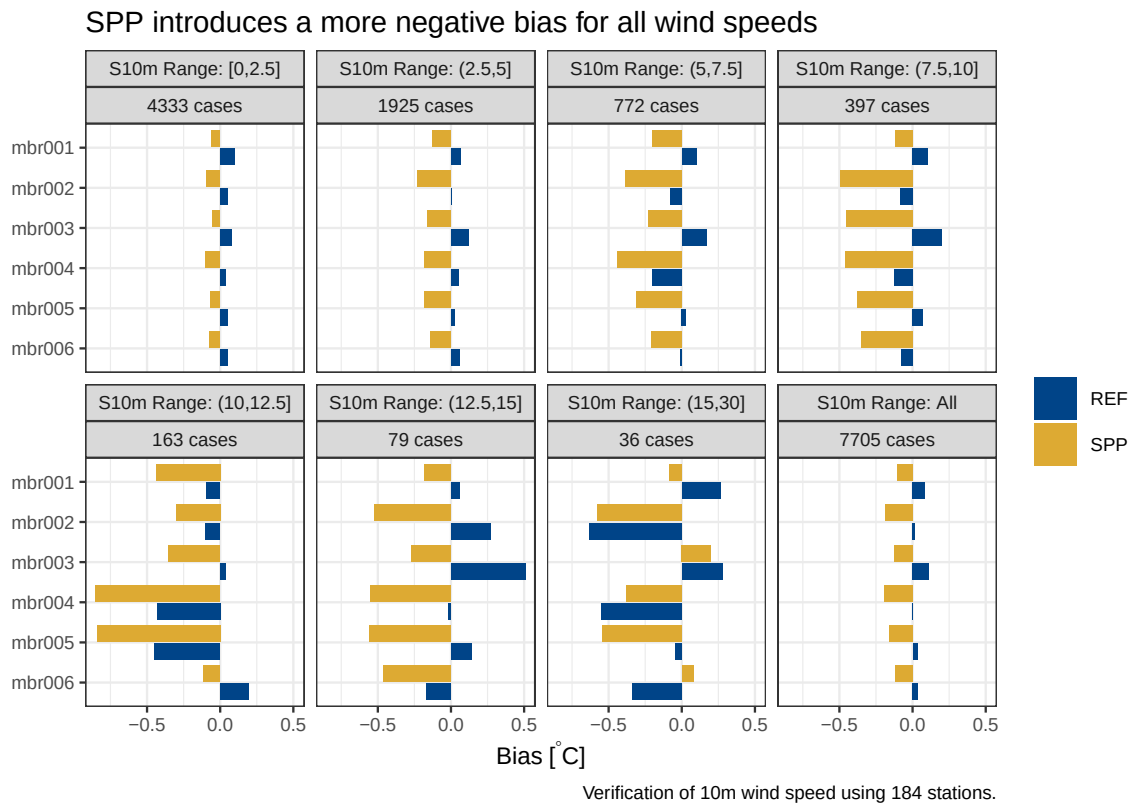

Figure 9: Bias for each ensemble member relative to the control member for 10m wind speed for REF (blue) and SPP (gold) for both special observing periods combined at 24-hours lead time.

3.1.4 Mean Sea Level Pressure

Mean sea level pressure can be thought of as summarising the the state of the atmosphere throughout the atmospheric column. Fig. 10 shows that for both SOP 1 and SOP 2 the spread for both REF and SPP increases continuously throughout the forecast. SPP generates more spread than REF in both the winter and summer, though the increase is slightly larger in the winter. However, this increase in spread is accompanied by an increase in the RMSE that is, like the spread, larger in the winter than in the summer. The balance of the biases for the individual members appears to be largely unaffected by SPP for mean sea level pressure (Fig. 11).

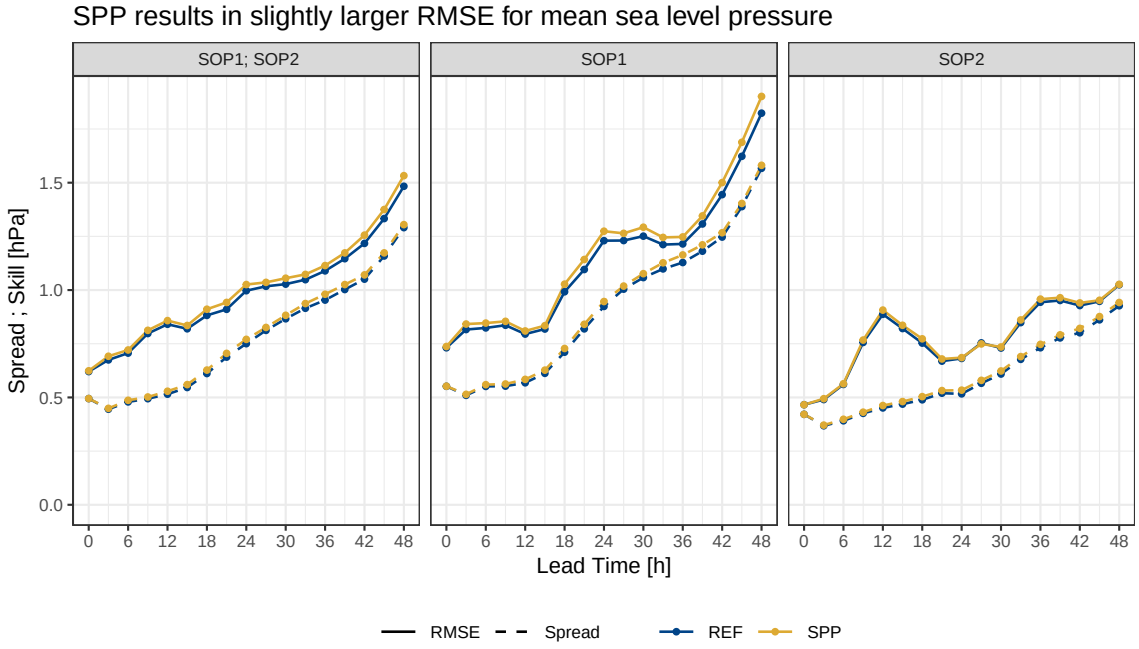

Figure 10: RMSE (solid line) and ensemble spread (dashed line) for mean sea level pressure for REF (blue) and SPP (gold) for (from left to right) both special observing periods combined, SOP 1 (8 - 31 March 2018) and SOP 2 (10 July - 1 August 2018).

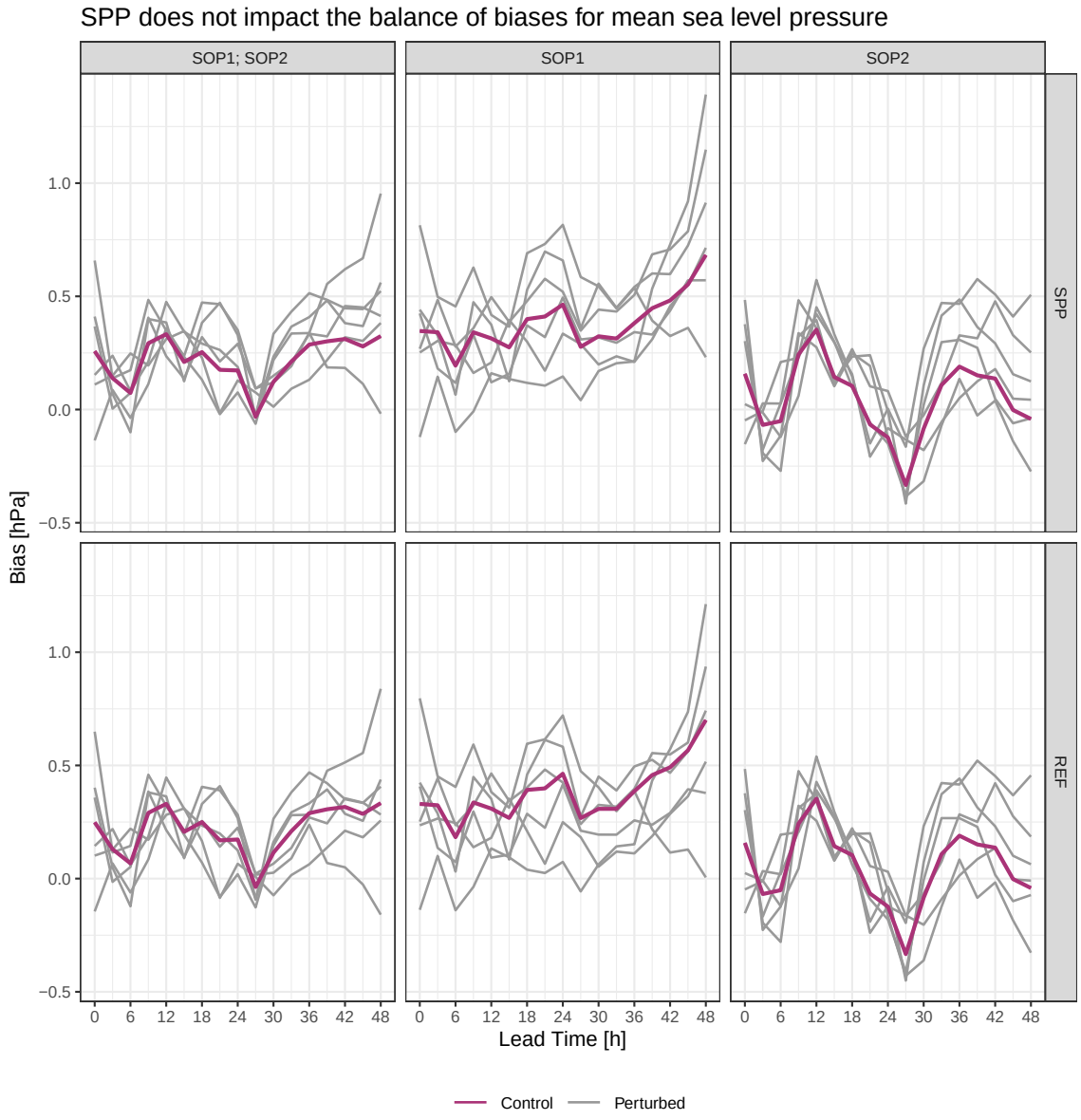

Figure 11: Bias for control (purple) and perturbed (grey) members for mean sea level pressure for SPP (top row) and REF (bottom row) for (from left to right) both special observing periods combined, SOP 1 (8 - 31 March 2018) and SOP 2 (10 July - 1 August 2018).

3.1.5 6-hour Accumulated Precipitation

The verification for 6-hour accumulated precipitation suggests that, in general, SPP results in lower spread than REF, but also lower RMSE (Fig. 12). The lower spread for SPP is mostly confined to the first 24 hours of the forecast, and is more apparent in SOP 2 than in SOP 1. In terms of RMSE, SPP is either lower than or equal to REF throughout the forecast. There is also considerable overspread for both REF and SPP during SOP 2. The biases for the individual members are generally unaffected by SPP, though it could be argued that SPP results in better balanced perturbations than REF (Fig. 13).

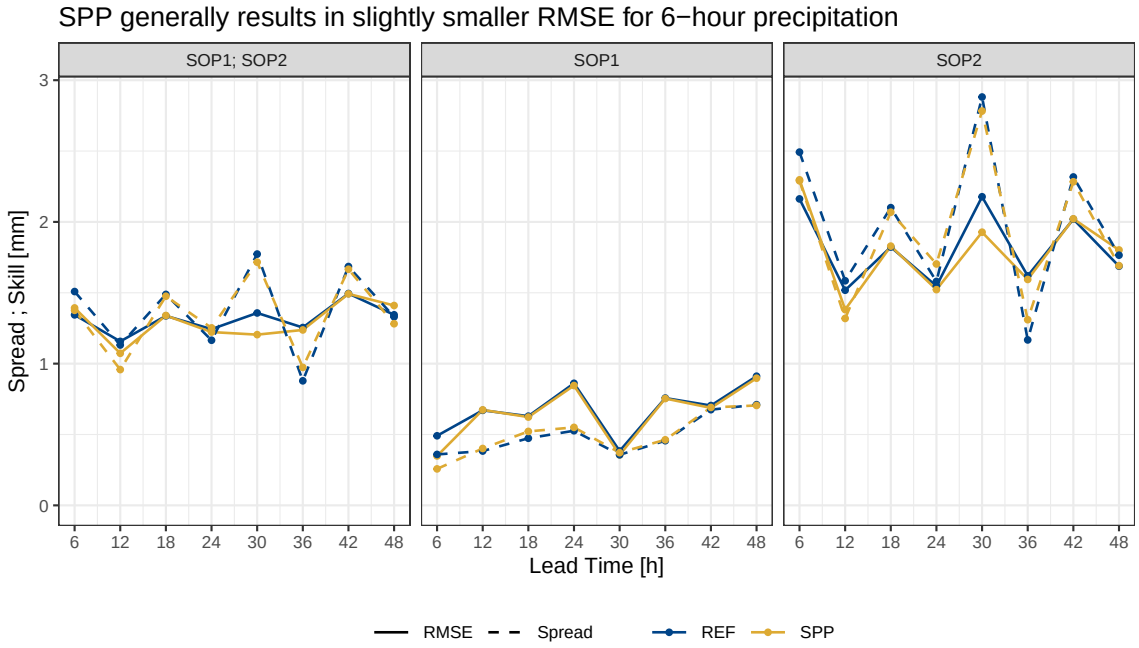

Figure 12: RMSE (solid line) and ensemble spread (dashed line) for 6-hour accumulated precipitation for REF (blue) and SPP (gold) for (from left to right) both special observing periods combined, SOP 1 (8 - 31 March 2018) and SOP 2 (10 July - 1 August 2018).

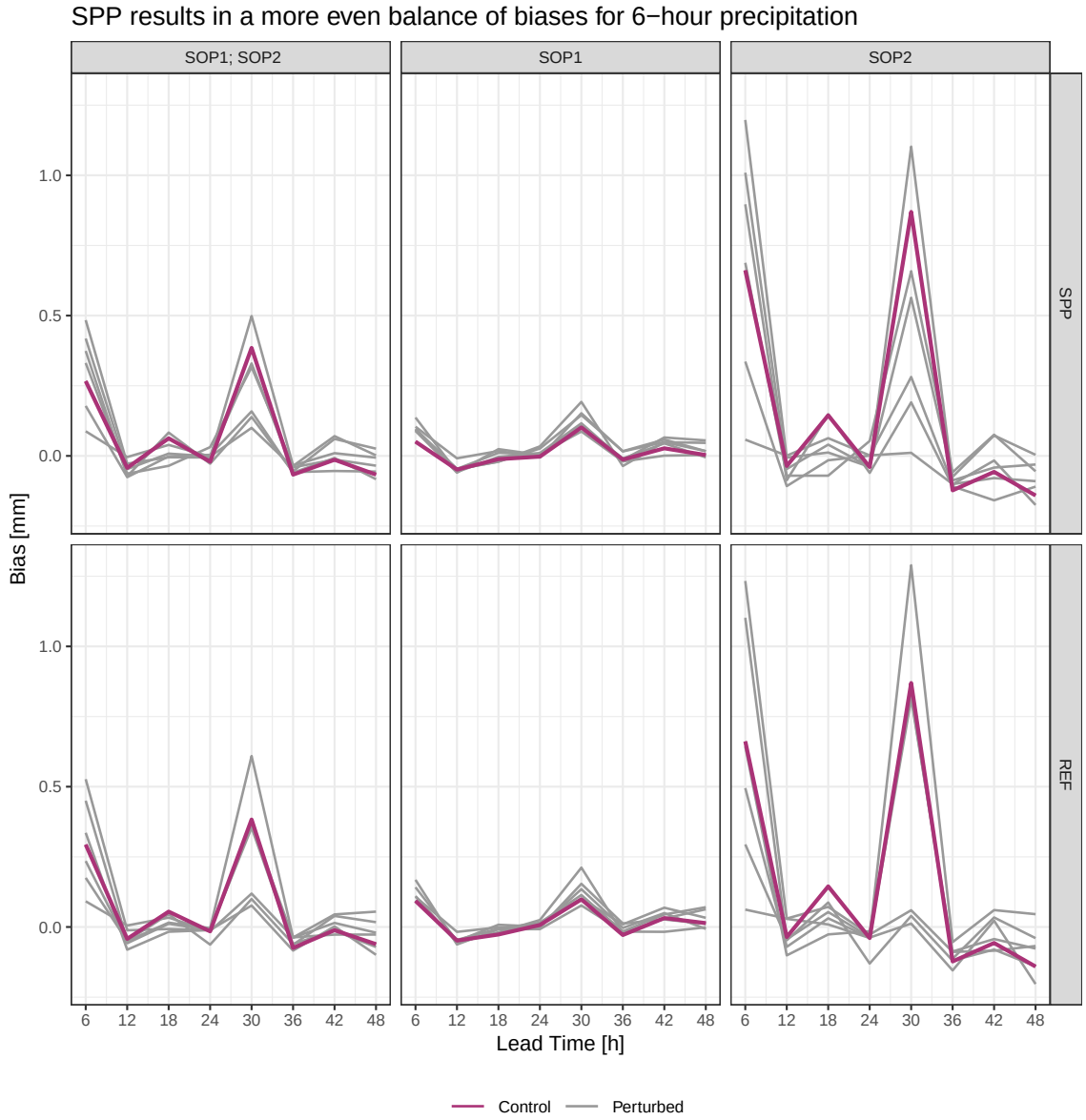

Figure 13: Bias for control (purple) and perturbed (grey) members for 6-hour accumulated precipitation for SPP (top row) and REF (bottom row) for (from left to right) both special observing periods combined, SOP 1 (8 - 31 March 2018) and SOP 2 (10 July - 1 August 2018).

3.1.6 Total Cloud Cover

SPP has a significant positive impact on both spread and RMSE compared to REF (Fig. 14). The spread for both SPP and REF grows throughout the forecast whilst the RMSE remains relatively static. The difference between the spread for SPP and REF is largest early in the forecast and gets slightly smaller as the spread for REF grows at a slightly faster rate. The difference in spread is larger in SOP 1 than SOP 2. The RMSE for SPP is smaller than that for REF throughout the forecast, but the difference is slightly larger for SOP 2.

In terms of the biases for the individual members, REF has more positive biases for the perturbed members than the control member. With SPP on the other hand, the biases of the perturbed members are more evenly spread about the control the member, particularly for SOP 2 (Fig. 15).

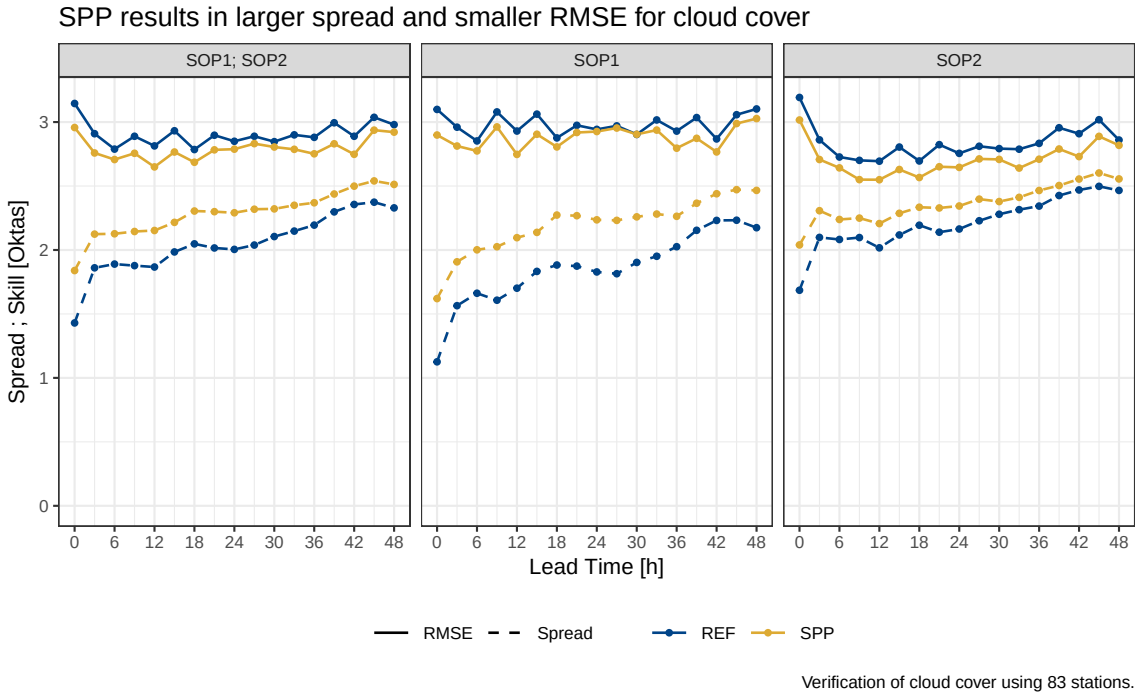

Figure 14: RMSE (solid line) and ensemble spread (dashed line) for total cloud cover for REF (blue) and SPP (gold) for (from left to right) both special observing periods combined, SOP 1 (8 - 31 March 2018) and SOP 2 (10 July - 1 August 2018).

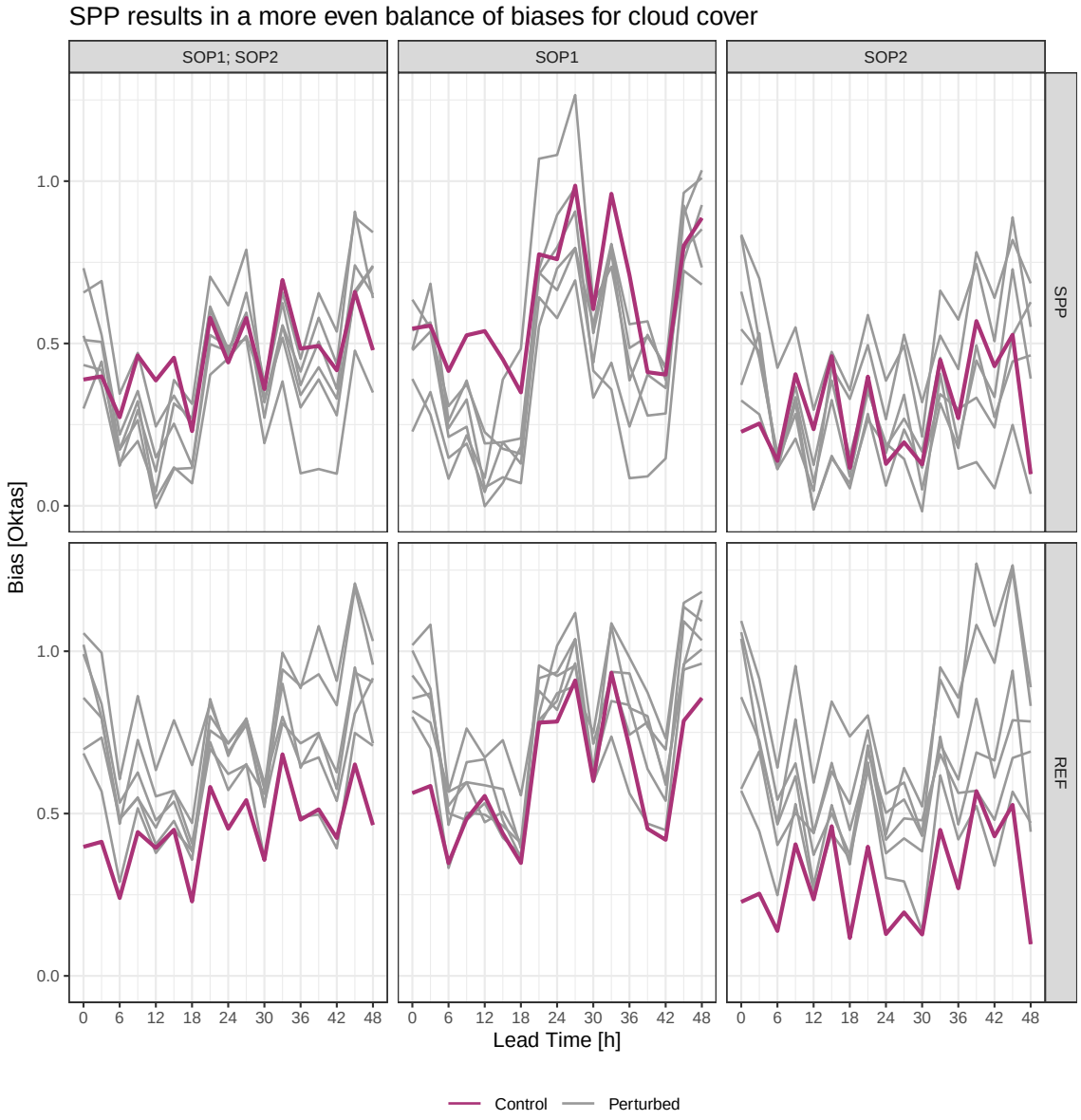

Figure 15: Bias for control (purple) and perturbed (grey) members for total cloud cover for SPP (top row) and REF (bottom row) for (from left to right) both special observing periods combined, SOP 1 (8 - 31 March 2018) and SOP 2 (10 July - 1 August 2018).

3.1.7 Upper Air Temperature

Upper air verification is limited to 8 radiosonde launch sites (Fig. 1), so the statistical robustness of the results is slightly limited, although some insights can be gained. For brevity, verification is only done at 12 hour intervals from 12- to 48-hour lead times. For upper air temperature SPP introduces more spread in the lower levels of the atmosphere for all lead times (Fig. 16). The impact is greatest below ~700hPa, and is slightly stronger at 24 hours lead time (and 36 hours durin SOP 2). SPP does result in a small increase in RMSE ar some pressure levels in the lower atmosphere between 700hPa and 900hPa during SOP 1 and at 1000hPa during SOP 2.

The cooler bias introduced by SPP for 2m temperature (Fig. 2) is limited to the atmosphere close to the surface (Fig. 17), with the perturbed members for SPP showing similar biases to those for REF for the atmospheric levels higher (in altitude) than 1000hPa. This applies to both SOP 1 and SOP 2.

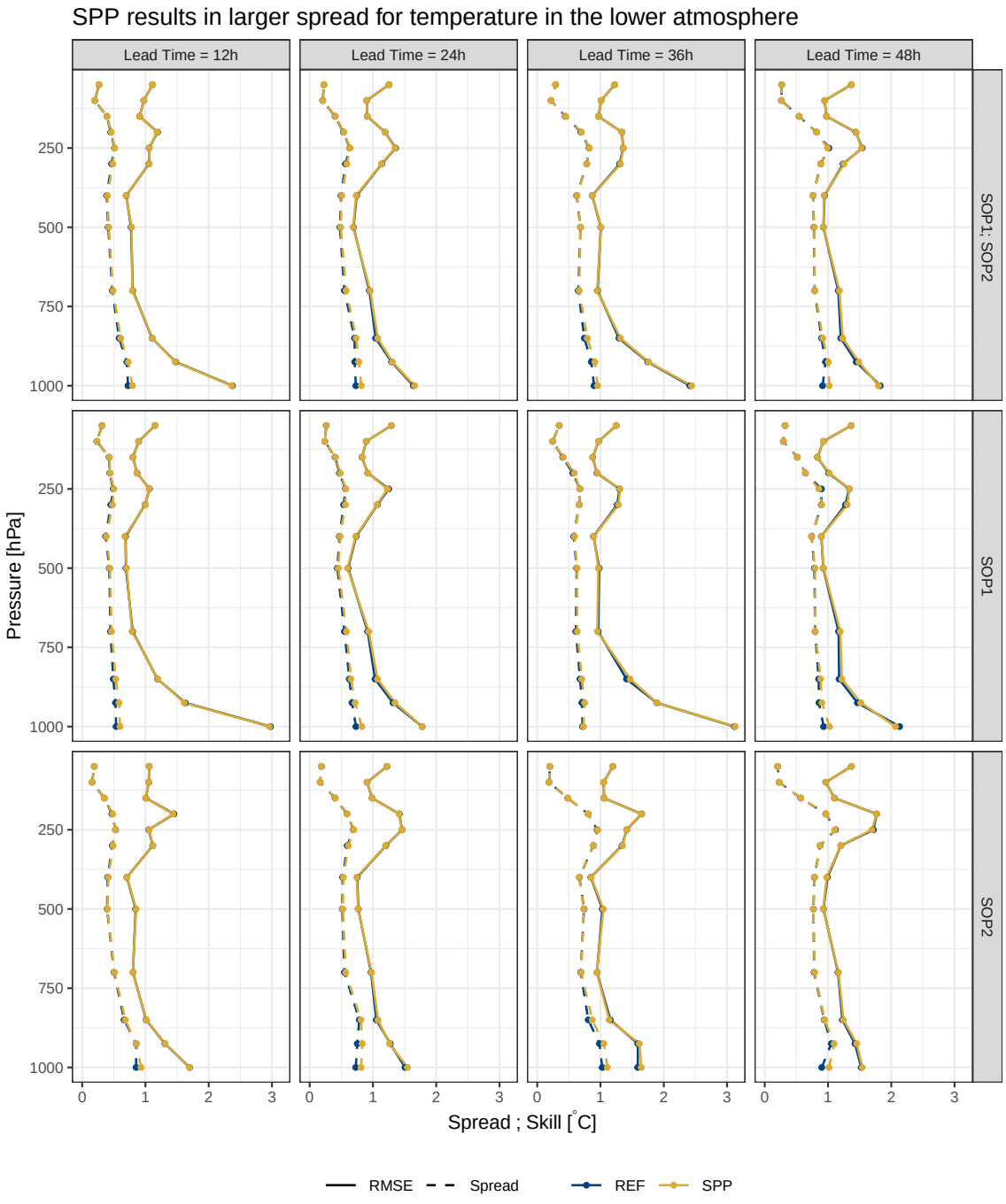

Figure 16: RMSE (solid line) and ensemble spread (dashed line) for temperature for REF (blue) and SPP (gold) for (from top to bottom) both special observing periods combined, SOP 1 (8 - 31 March 2018) and SOP 2 (10 July - 1 August 2018) for selected lead times (left to right).

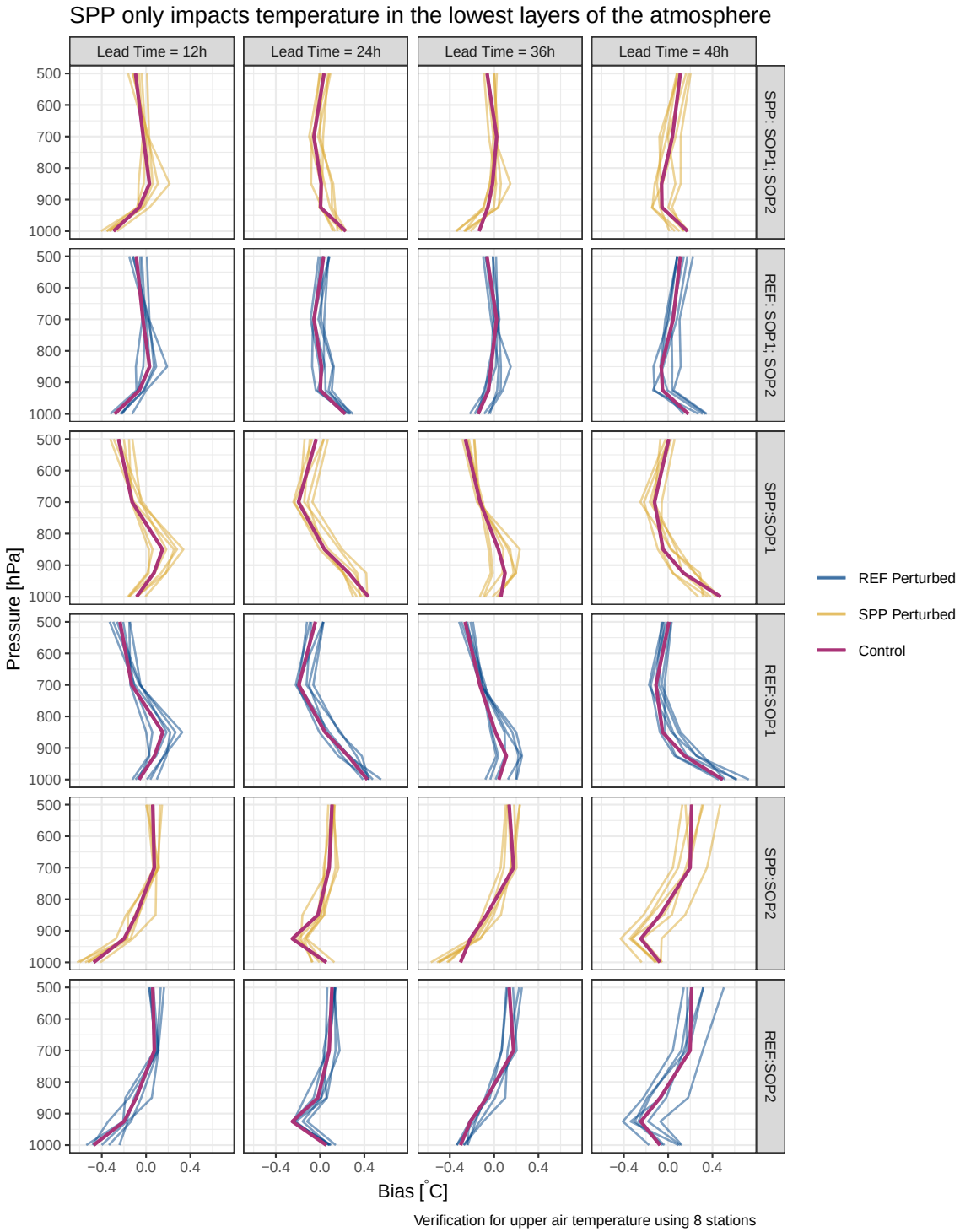

Figure 17: Bias for atmospheric temperature in the lowest 500 hPa of the atmosphere. Each pair of rows are for (from top to bottom) both special observing periods combined, SOP 1 (8 - 31 March 2018) and SOP 2 (10 July - 1 August 2018), with the first row of each pair for SPP (gold) and the second row of each pair for REF (blue). The control member (purple) is the same for both REF and SPP. The columns are for lead times from 12- to 48-hours every 12-hours.

3.1.8 Upper Air Specific Humidity

The impact of SPP on specific humidity throughout the upper atmosphere appears to be quite small and concentrated around the 850hPa level (Fig. 18). In general there is a small increase in spread that is strongest at 36 hours lead time in the summer (SOP 2). In the winter (SOP 1) an increase in spread is seen at 24 and 48 lead times (i.e. during the night). There is generally no discernible impact on the the RMSE.

The biases for individual members suggest that there is some impact on the balance of perturbations (Fig. 19), especially during the winter (SOP 1). The biases during SOP 1 suggest that the perturbed members are slightly drier for SPP than REF, though there isn't such a clear signal for SOP 2.

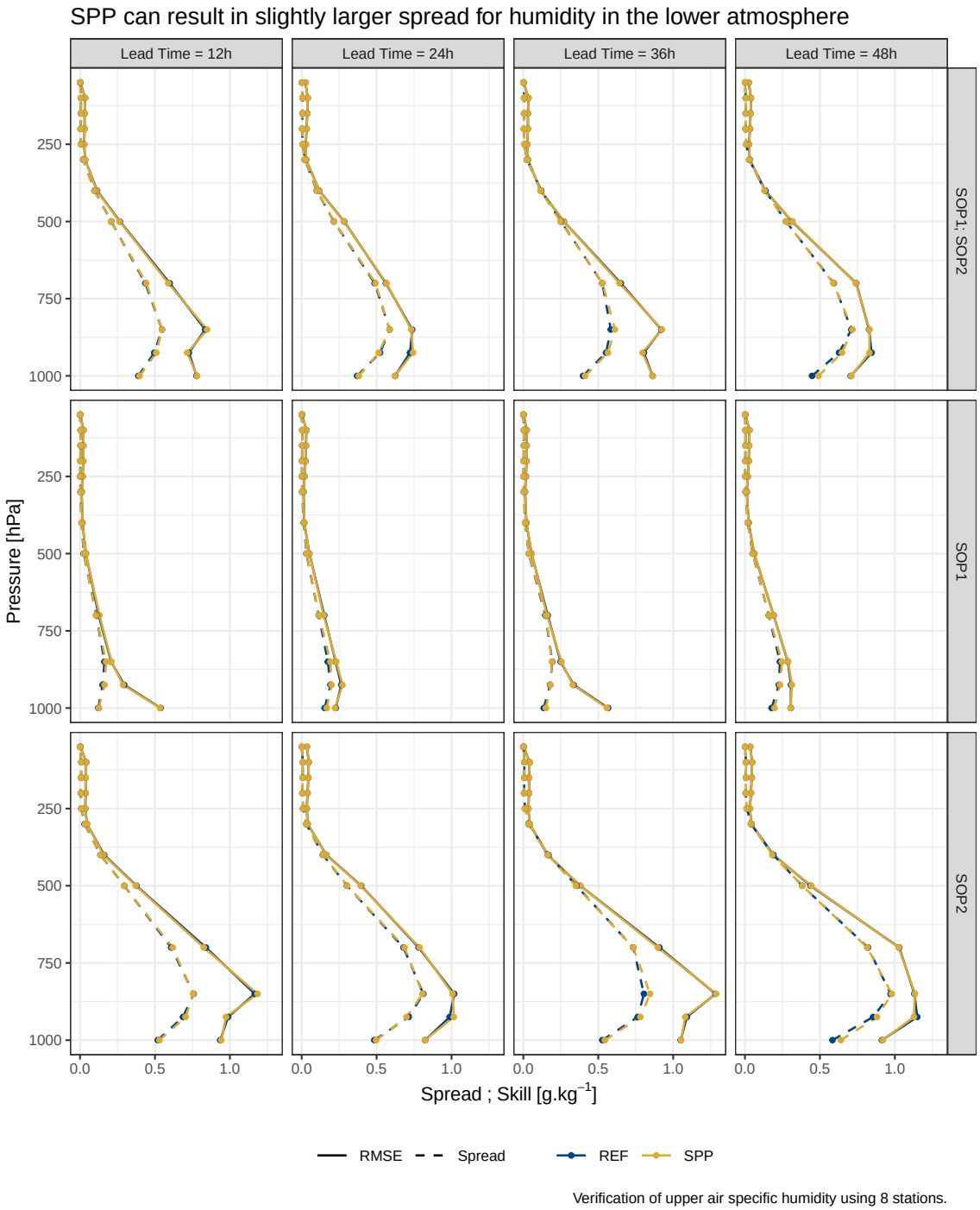

Figure 18: RMSE (solid line) and ensemble spread (dashed line) for specific humidity for REF (blue) and SPP (gold) for (from top to bottom) both special observing periods combined, SOP 1 (8 - 31 March 2018) and SOP 2 (10 July - 1 August 2018) for selected lead times (left to right).

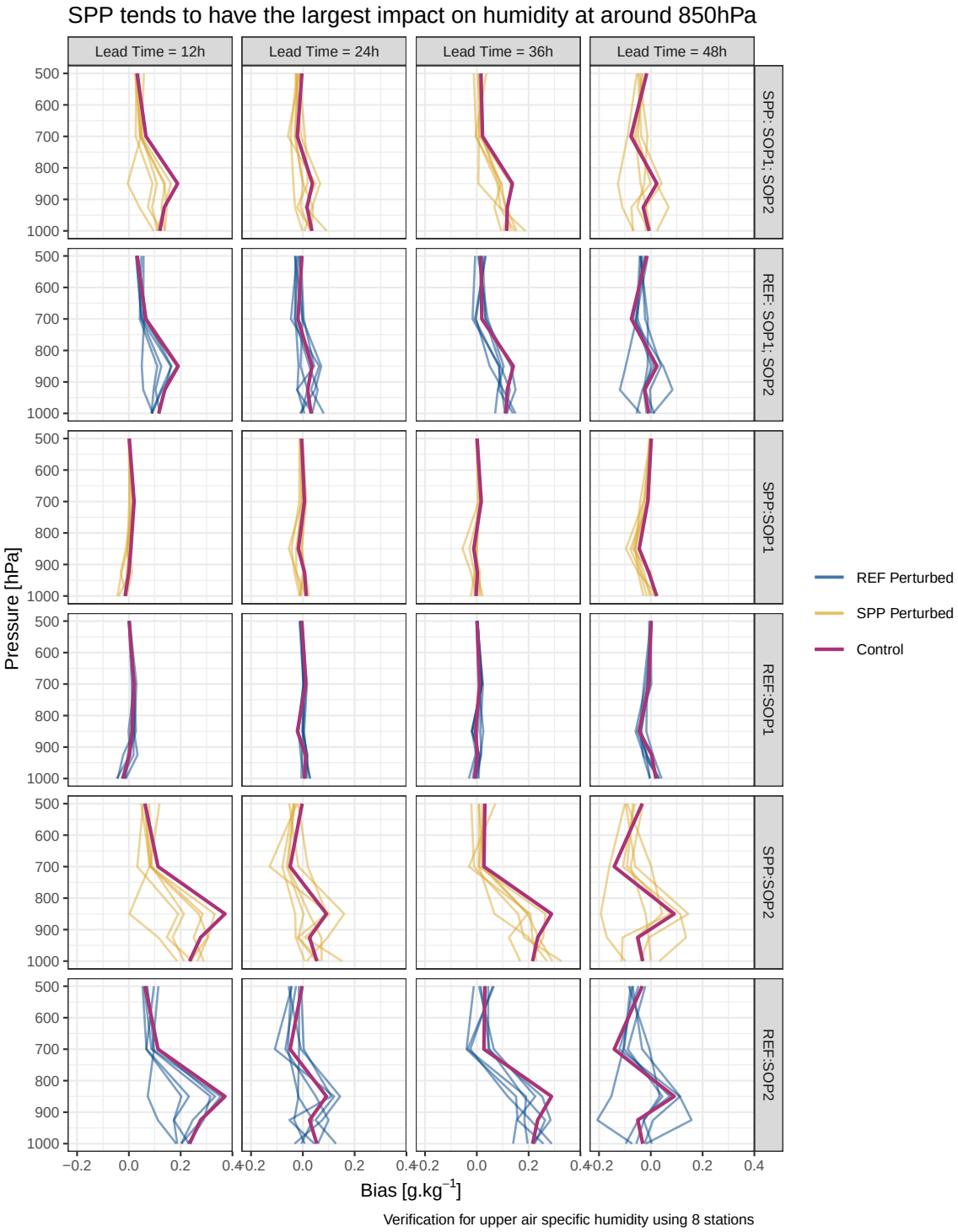

Figure 19: Bias for atmospheric specific humidity in the lowest 500 hPa of the atmosphere. Each pair of rows are for (from top to bottom) both special observing periods combined, SOP 1 (8 - 31 March 2018) and SOP 2 (10 July - 1 August 2018), with the first row of each pair for SPP (gold) and the second row of each pair for REF (blue). The control member (purple) is the same for both REF and SPP. The columns are for lead times from 12- to 48-hours every 12-hours.

3.1.9 Upper Air Wind Speed

SPP has a considerable impact on the spread of wind speed in both winter and summer that is strongest at the lower levels of the atmosphere (Fig. 20). During SOP 1, there is an increase in the RMSE for SPP compared with REF at the lowest couple of atmospheric levels that is more apparent at 14- and 48-hours lead times (i.e. during the night). The biases of the individual members suggest that there is slight move towards more positive biases for SPP compared with REF (Fig. 21).

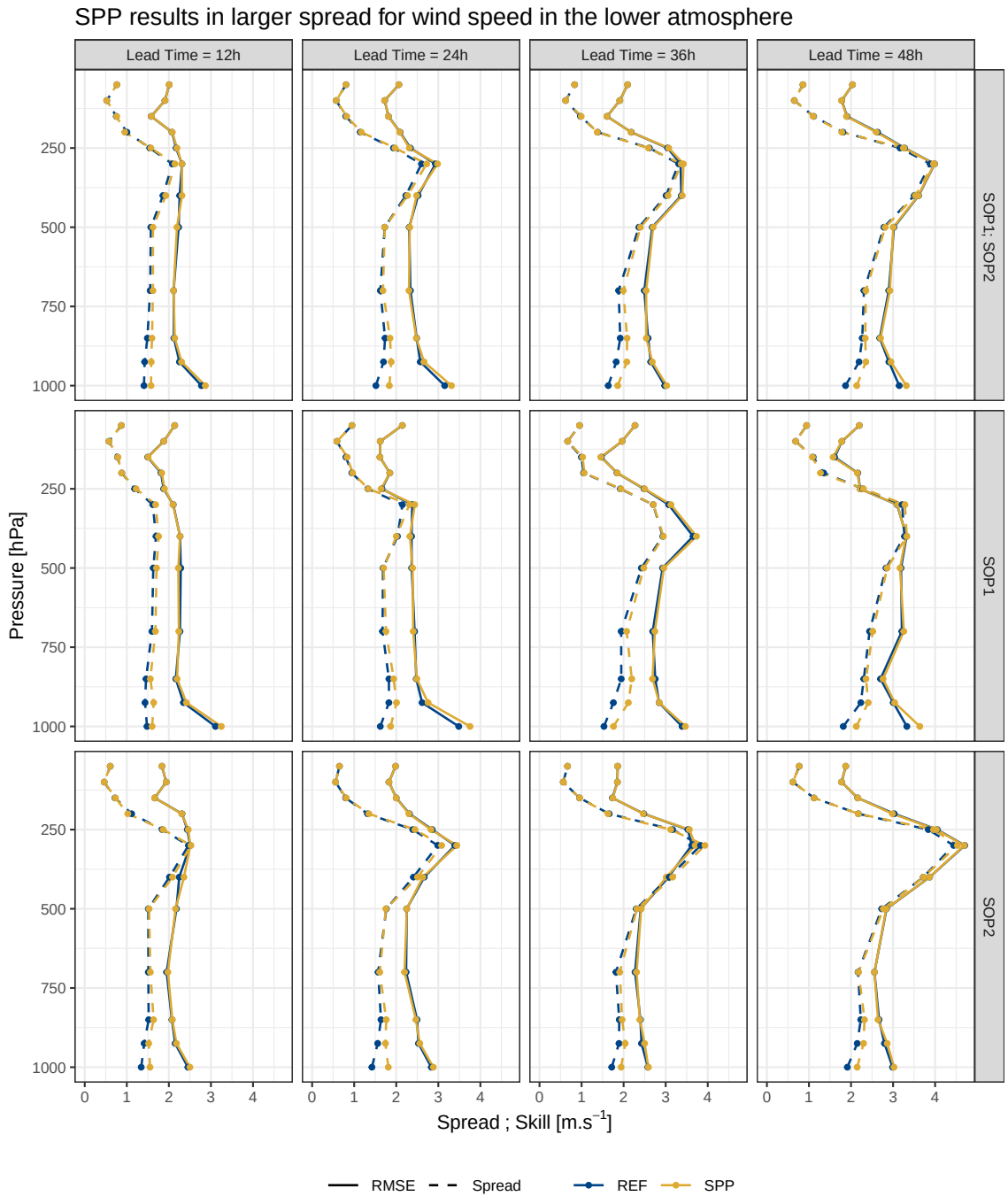

Figure 20: RMSE (solid line) and ensemble spread (dashed line) for wind speed for REF (blue) and SPP (gold) for (from top to bottom) both special observing periods combined, SOP 1 (8 - 31 March 2018) and SOP 2 (10 July - 1 August 2018) for selected lead times (left to right).

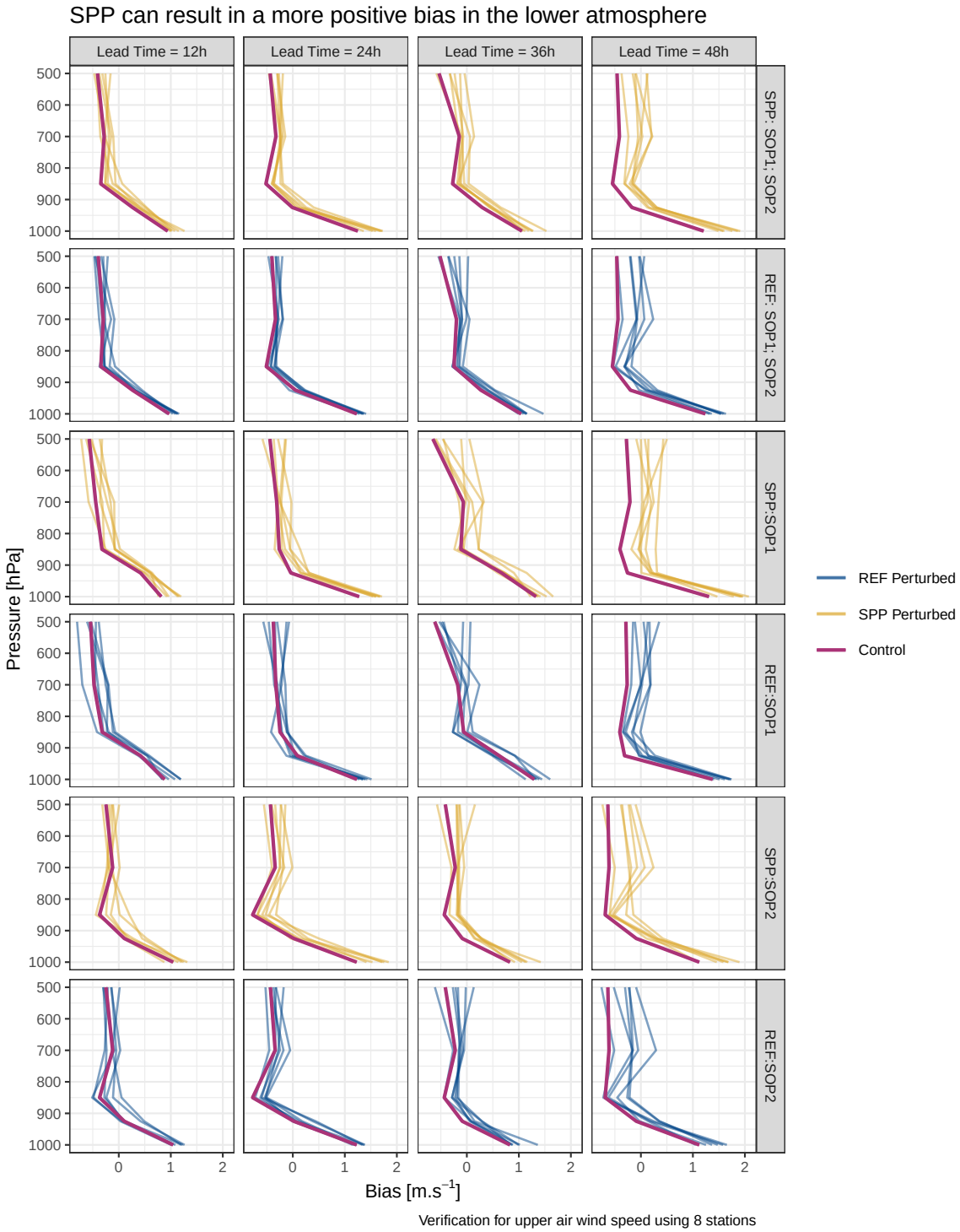

Figure 21: Bias for wind speed in the lowest 500 hPa of the atmosphere. Each pair of rows are for (from top to bottom) both special observing periods combined, SOP 1 (8 - 31 March 2018) and SOP 2 (10 July - 1 August 2018), with the first row of each pair for SPP (gold) and the second row of each pair for REF (blue). The control member (purple) is the same for both REF and SPP. The columns are for lead times from 12- to 48-hours every 12-hours.

3.1.10 Geopotential Height

The geopotential height can be used a proxy for the overall impact of SPP on the large scale atmospheric flow. Fig. 22 and Fig. 23 suggest that the impact of SPP comapred with REF on the geopotential height at the available observation stations is negligible.

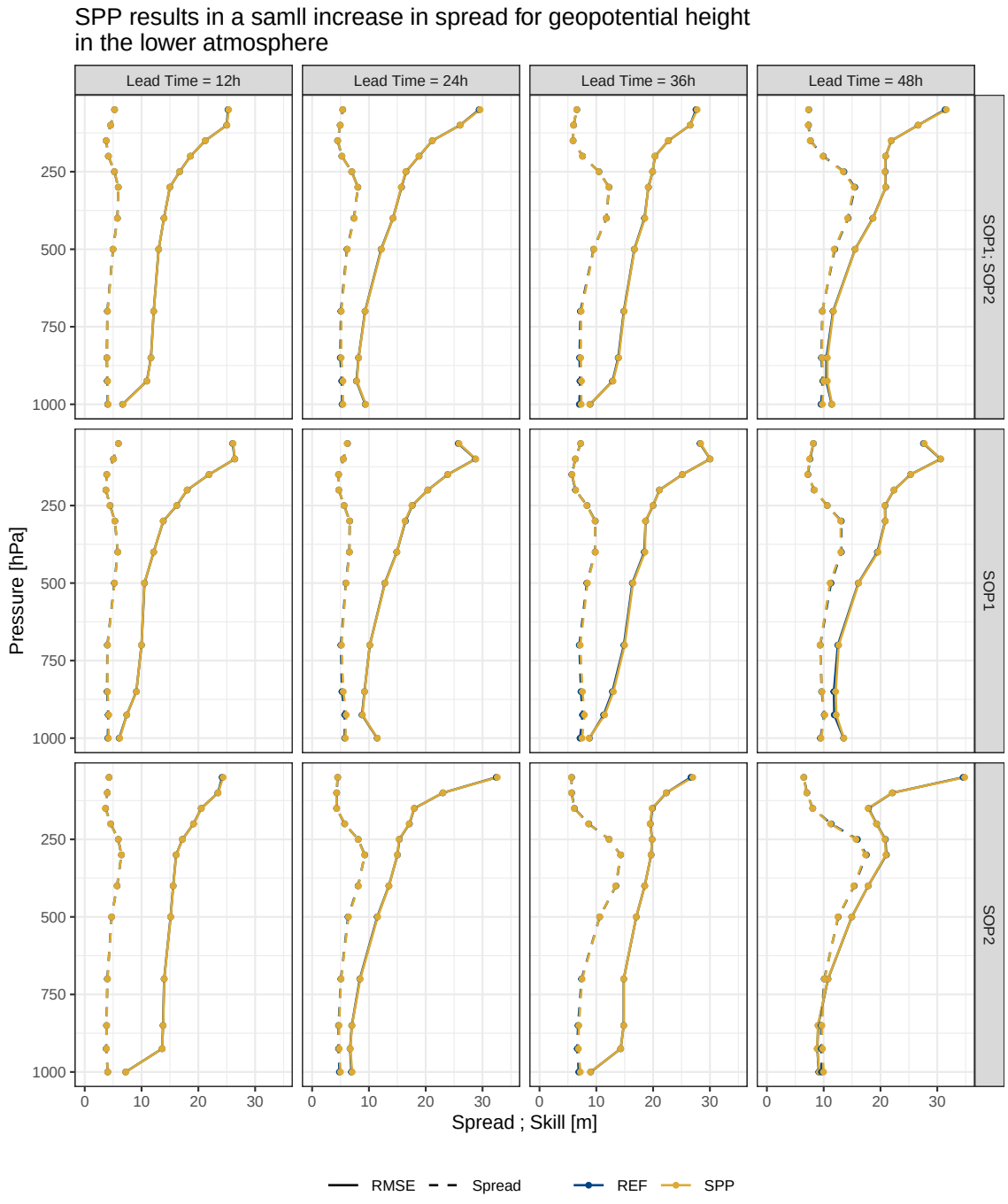

Figure 22: RMSE (solid line) and ensemble spread (dashed line) for geopotential height for REF (blue) and SPP (gold) for (from top to bottom) both special observing periods combined, SOP 1 (8 - 31 March 2018) and SOP 2 (10 July - 1 August 2018) for selected lead times (left to right).

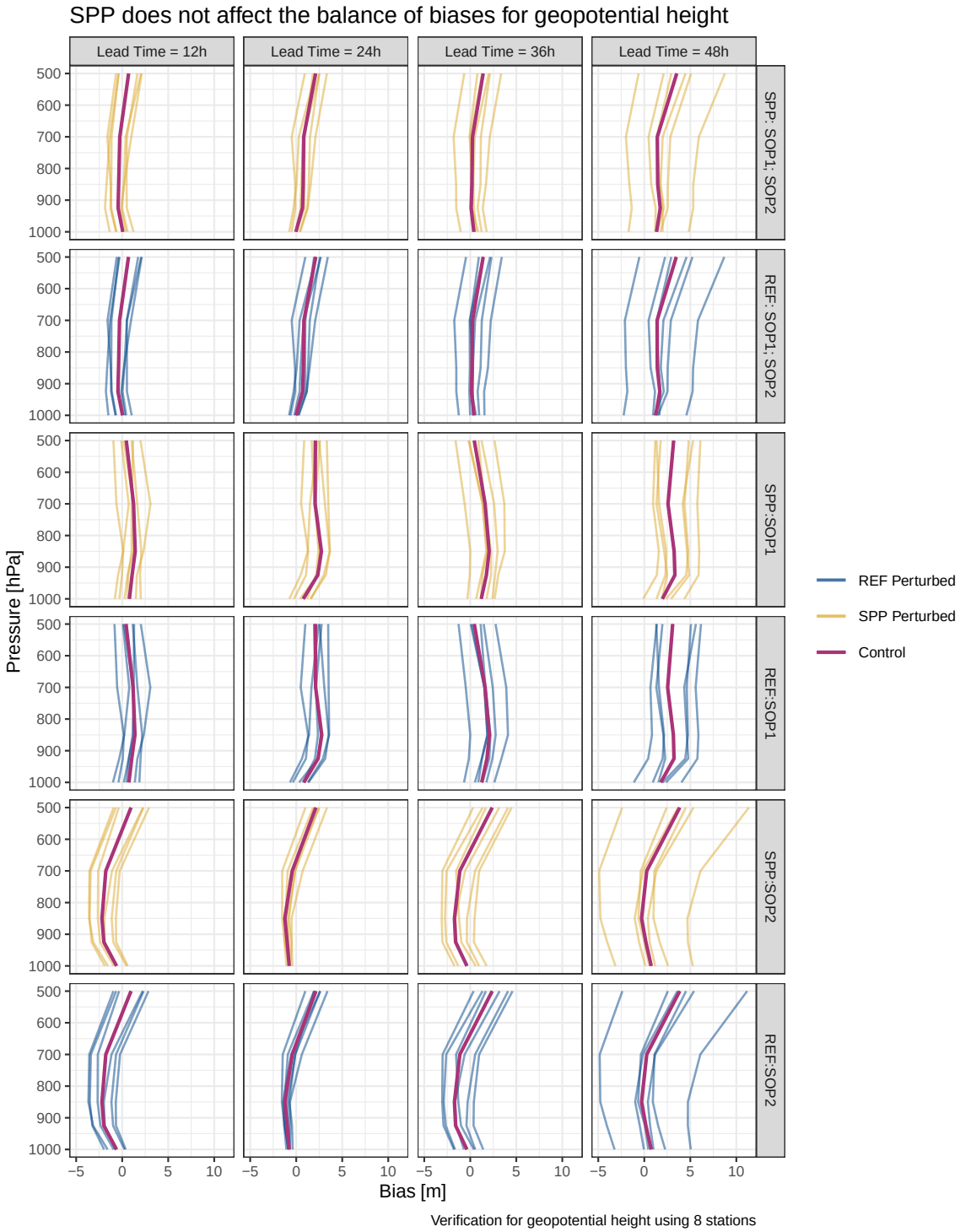

Figure 23: Bias for geopotential in the lowest 500 hPa of the atmosphere. Each pair of rows are for (from top to bottom) both special observing periods combined, SOP 1 (8 - 31 March 2018) and SOP 2 (10 July - 1 August 2018), with the first row of each pair for SPP (gold) and the second row of each pair for REF (blue). The control member (purple) is the same for both REF and SPP. The columns are for lead times from 12- to 48-hours every 12-hours.

3.2 EDA

3.2.1 Initial condition perturbations

EDA provides a method for perturbing observations in order to generate an ensemble of analyses from which to initialize each member of an ensemble forecast. The reference (REF) against which EDA is evaluated generates an ensemble of analyses using the SLAF method (see Singleton and Grote (2020); Toth and Kalnay (1993)). In order to compare the impact on the initial conditions, the mean of the standard deviation of the ensemble members at each pixel was calculated for a range of upper air parameters for each ensemble of analyses during the test period (25 October - 7 November 2019). This diagnostic is used as an indicator of total ensemble spread throughout the model domain. It is clear that EDA generates more spread in the initial conditions for all of the considered parameters at both the $850-$ and $500-hPa$ pressure levels (Fig. 24). For many of the parameters, the difference is much smaller for the first ensemble of analyses in the test period. This is because only the control member was spun up before the test period meaning that the first guess was the same for all EDA ensemble members. For all analyses forward of this point, each ensemble member was cycled individually with the first guess coming from each of the ensemble members.

For temperature, EDA adds approximately $1^{\circ}C$ on average to the spread compared with REF at both $850-$ and $500-hPa$ pressure levels. EDA adds approximately 0.05 $g.kg^{-1}$ at the 850 hPa pressure level and approximately 0.02 $g.kg^{-1}$ at the 500 hPa pressure level where the air is typically less humid anyway. For wind speed, EDA adds just under 0.5 $m.s^{-1}$ to the spread at the 850 hPa pressure level and just over 0.5 $m.s^{-1}$ at the 500 hPa pressure level. Similarly for geopotential height, approximately 0.5 m is added to the spread by EDA at the 850 hPa pressure level and approximately 1 m is added at the 500 hPa pressure level. In terms of mean sea level pressure, EDA increases the spread by approximately 0.1 hPa. For 500 hPa temperature, 500 hPa wind speed, $850-$ and $500-hPa$ geopotential height, and mean sea level pressure, there is a peak in the spread for both EDA and REF on 29 October 2019. This coincides with the beginning of a polar event and may be reflective of the increased uncertainty resulting from that weather system.

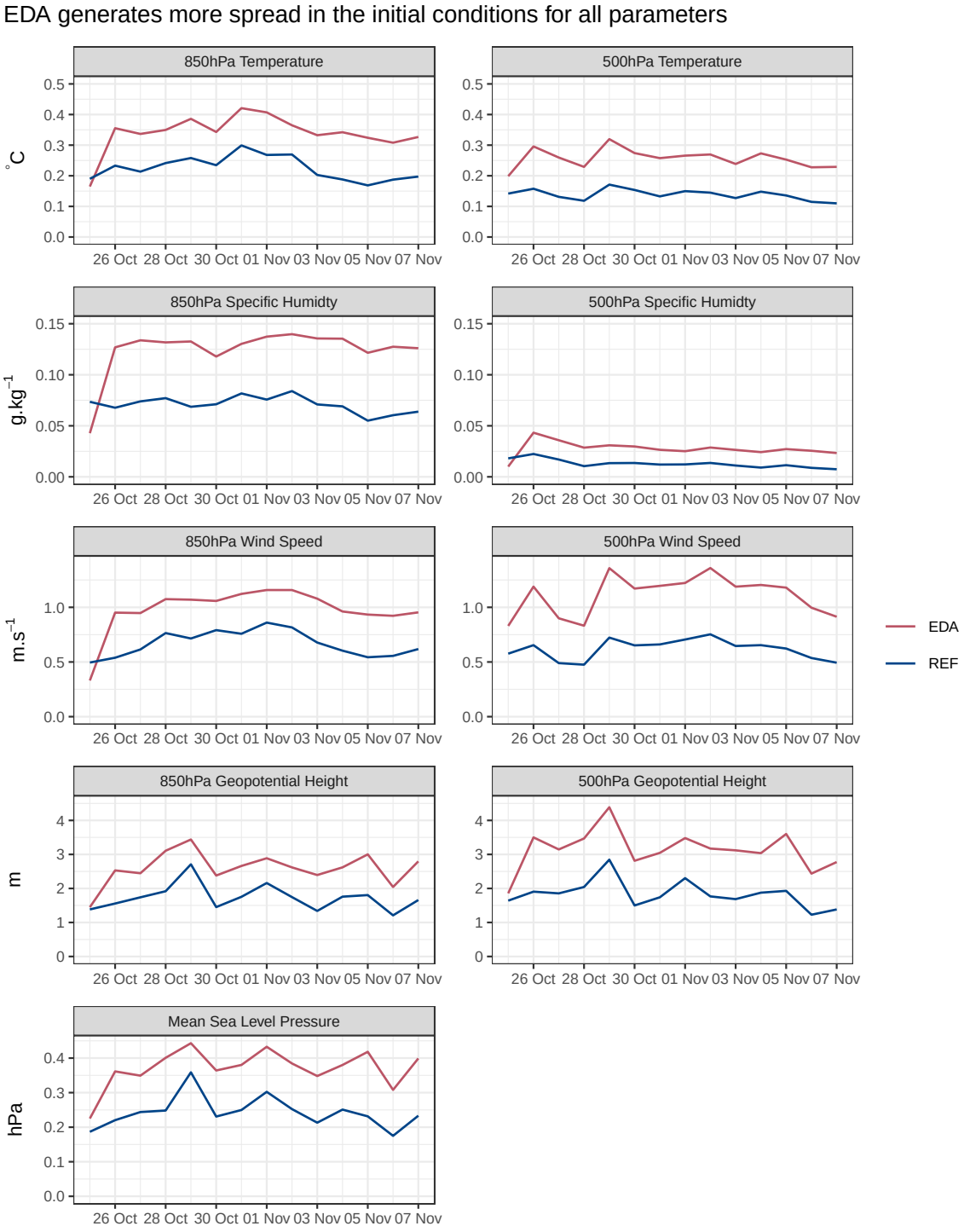

Figure 24: Mean of the standard deviation of the initial conditions for a range of parameters for REF (blue) and EDA (red) for each forecast in the period 25 Oct - 7 Nov 2019.

3.2.2 2m Temperature

For 2m temperature, EDA provides additional spread throughout the forecast (Fig. 25). The amount of additional spread is largest at the start of the forecast (approximately $0.25°C$) and becomes slightly smaller as the forecast progresses. The trend in spread is the same for both REF and EDA. Additionally, there is a very small reduction in RMSE resulting from EDA between hours 3 and 9 of the forecast. Similar to the SOP 1 and SOP 2 periods investigated for the SPP comparisons, the perturbed members of REF are warmer than the control member at the beginning of the forecast, but in this period a better balance is achieved as the forecast continues (Fig. 26) and there is no clear diurnal cycle in the biases. For EDA, the perturbed members are also warmer than the control at the beginning of the forecast, but a better balance of biases is achieved earlier in the forecast than REF with a wider spread of biases around the control member.

There is no obvious temperature range for which the warmer perturbed members at initialization time for EDA occur, although for temperatures between $0°C$ and $5°C$, some of the perturbed members are cooler than the control for EDA, whereas for REF there appears to be a switch from perturbed members warmer than control below $0°C$ and perturbed members cooler than the control for temperatures above $0°C$ (Fig. 27).

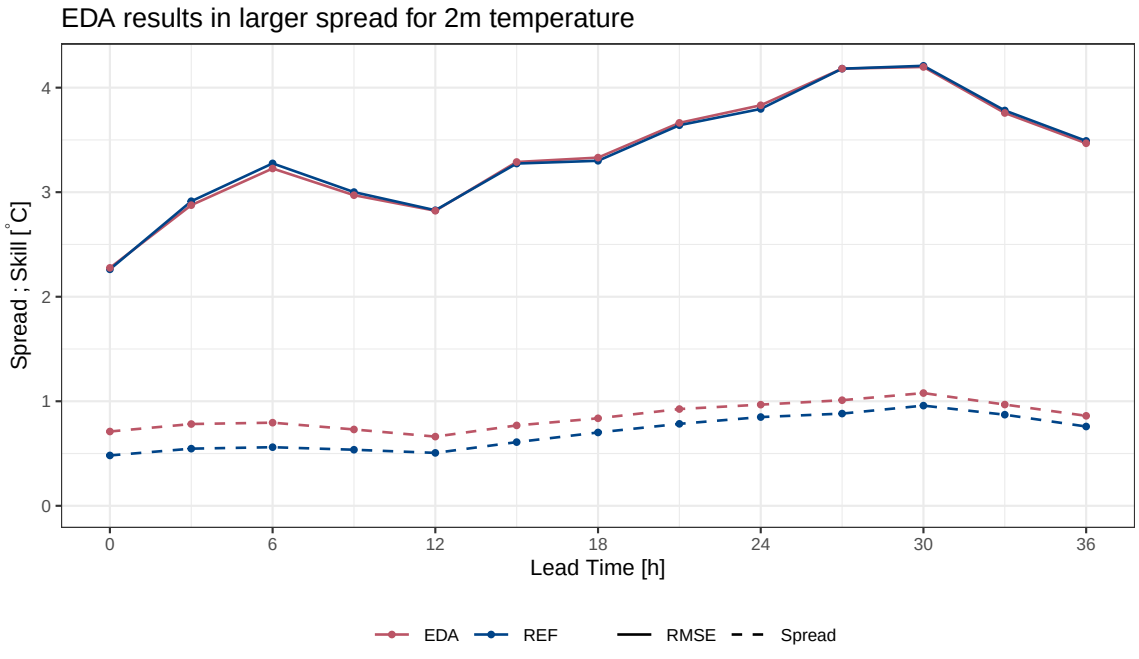

Figure 25: RMSE (solid line) and ensemble spread (dashed line) for 2m temperature for REF (blue) and EDA (red) for the period 25 Oct - 11 Nov 2019.

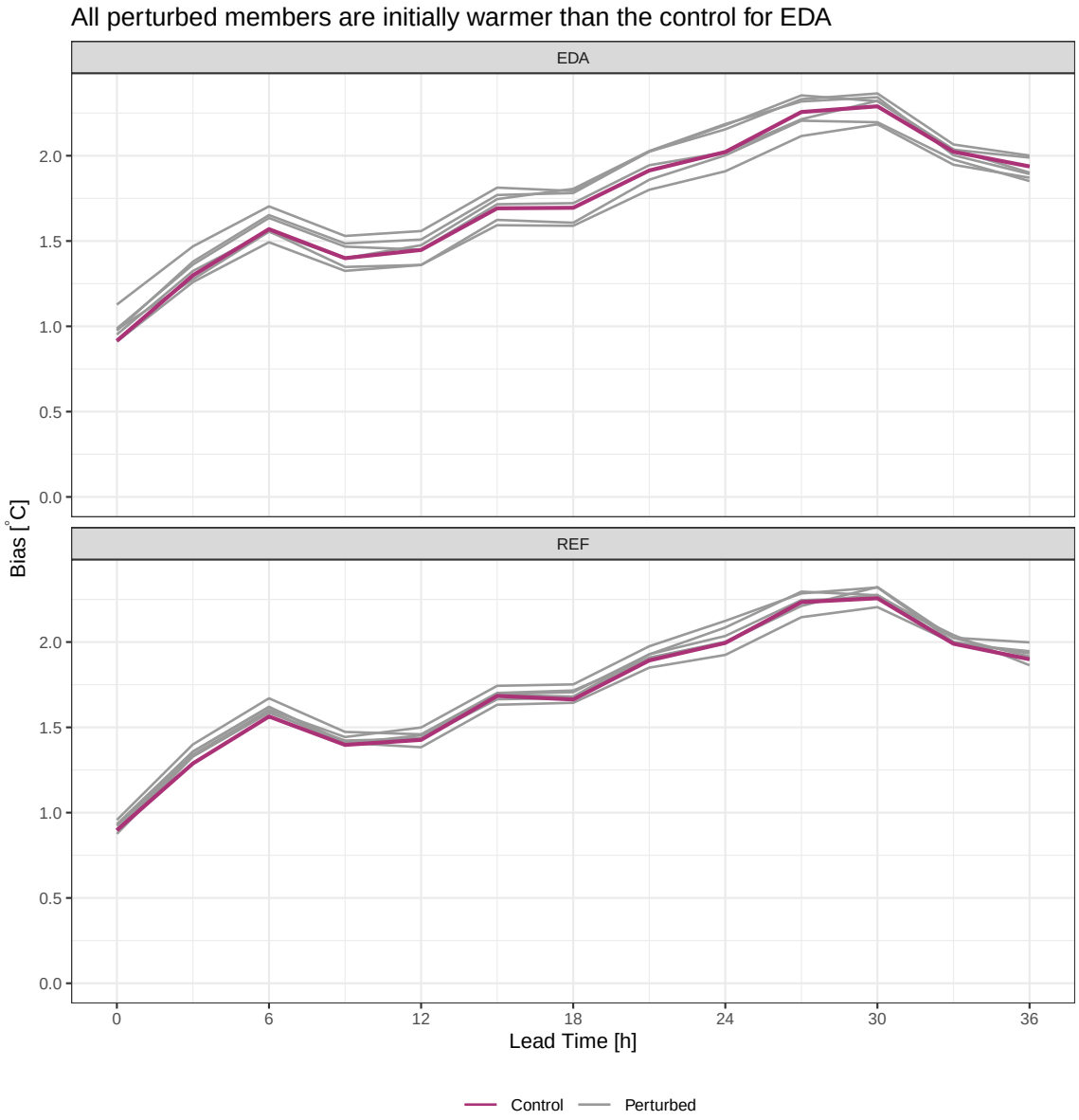

Figure 26: Bias for control (purple) and perturbed (grey) members for 2m temperature for EDA (top row) and REF (bottom row) for the period 25 Oct - 11 Nov 2019.

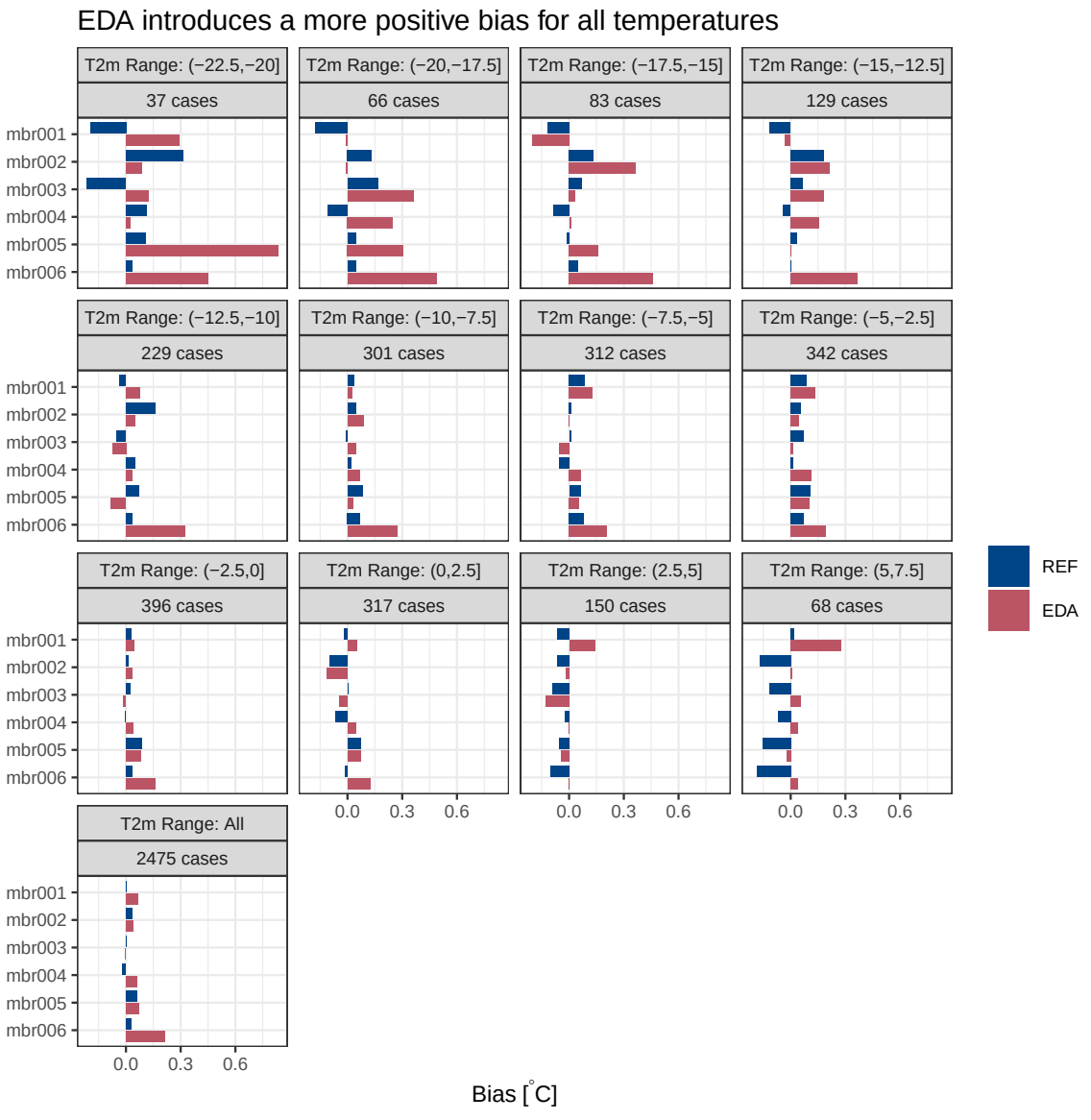

Figure 27: Bias for each ensemble member relative to the control member for 2m temperature for REF (blue) and EDA (red) for the period 25 Oct - 11 Nov 2019 at forecast initialization time.

3.2.3 2m Specific Humidity

Similar to 2m temperature, EDA adds spread for 2m specific humidity that is largest at the beginning of the forecast and becomes closer to the spread for REF as the forecast progresses (Fig. 28). However, in the case of 2m specific humidity, there is also a reduction in the RMSE resulting from EDA for the first 12 hours of the forecast. While REF has perturbed members that are all moister than the control at the beginning of the forecast the biases of the perturbed members are much better balanced for EDA (Fig. 29), and this improved balance with a wider spread of biases remains throughout the forecast, with REF only becoming well balanced from approximately 18 hours lead time.

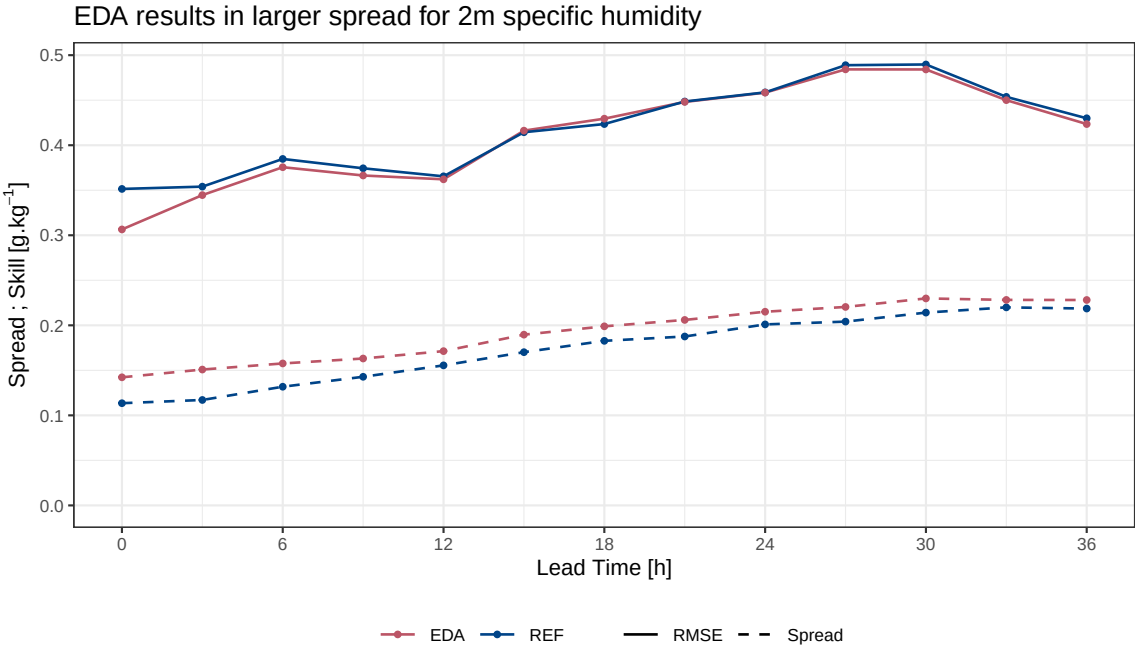

Figure 28: RMSE (solid line) and ensemble spread (dashed line) for 2m specific humidity for REF (blue) and EDA (red) for the period 25 Oct - 11 Nov 2019.

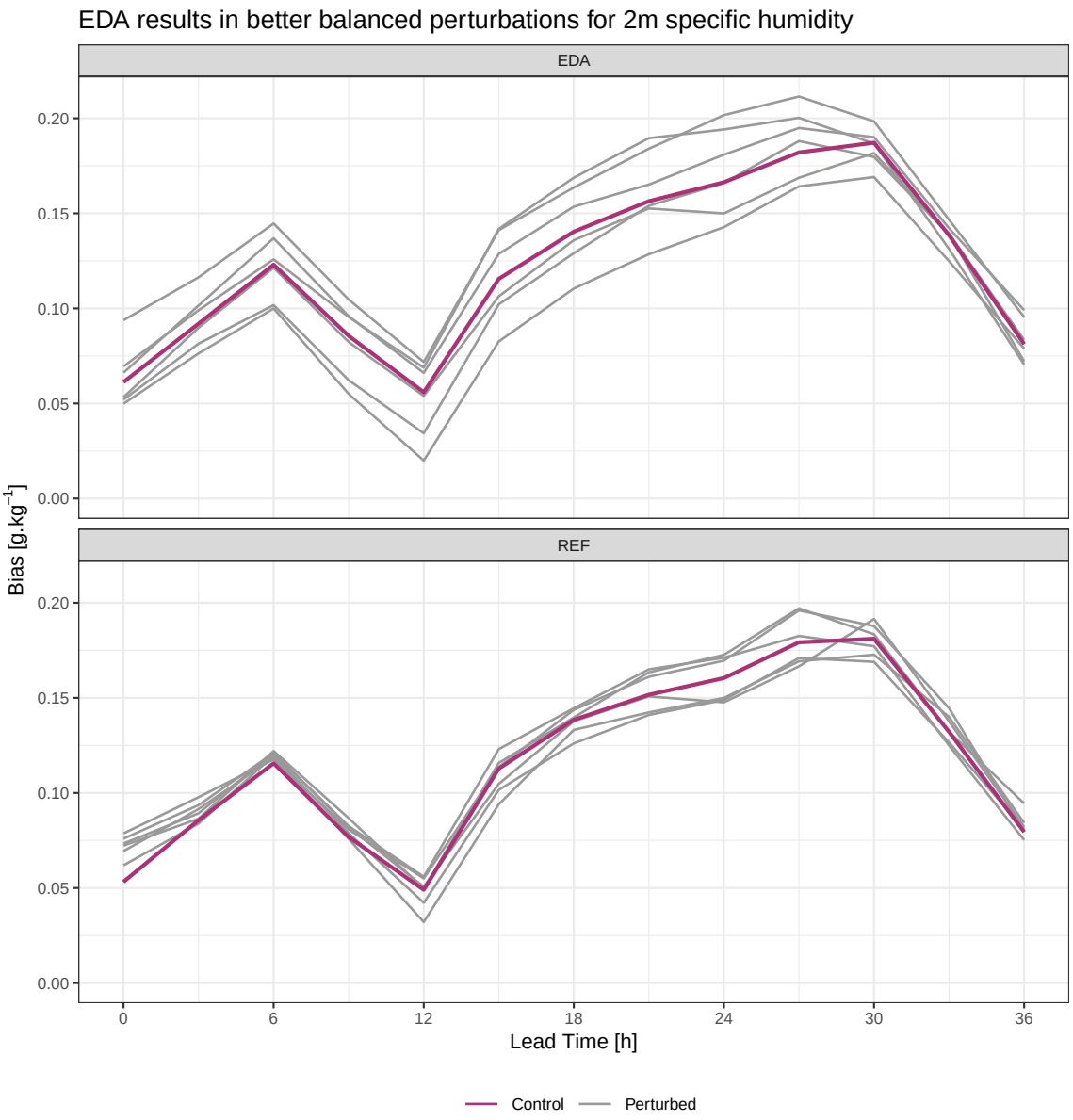

Figure 29: Bias for control (purple) and perturbed (grey) members for 2m specific humidity for EDA (top row) and REF (bottom row) for the period 25 Oct - 11 Nov 2019.

3.2.4 10m Wind Speed

Similar to both 2m temperature and 2m specific humidity, the 10m wind speed for EDA has more spread towards the beginning of the forecast than REF, with the difference most pronounced at initialization time (Fig. 30). The larger difference at initialization time is more because the spread takes a few hours to grow for REF at the beginning of the forecast, whereas for EDA the spread grows fairly consistently throughout the forecast. The RMSE for EDA is either roughly equal to, or slightly smaller than the RMSE for REF for the whole forecast.

In terms of the biases of the perturbed members, both REF and EDA follow roughly the same trend, but the biases tend to be slightly more positive for EDA than for REF (Fig. 31). The distribution of the biases at initialization time stratified by observed 10m wind speed suggest that EDA tends to produce more positive biases then REF for lower wind speeds, whereas for observed wind speeds greater than 10 $m.s^{-1}$ the biases for EDA tend to be in the same direction as those for REF but stronger (Fig. 32).

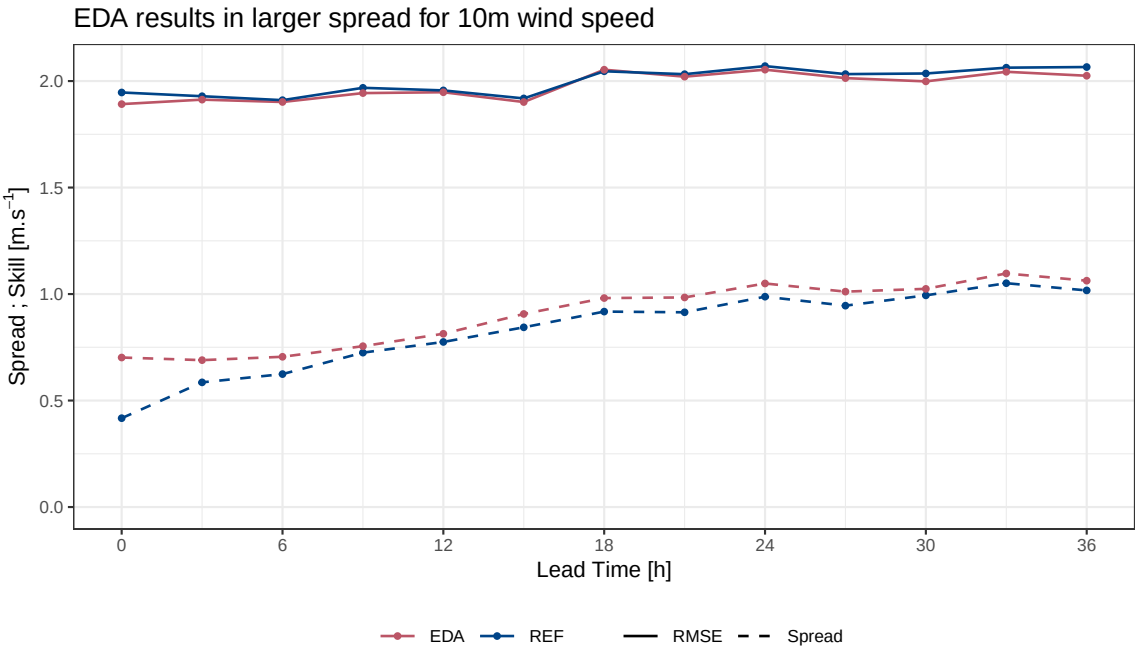

Figure 30: RMSE (solid line) and ensemble spread (dashed line) for 10m wind speed for REF (blue) and EDA (red) for the period 25 Oct - 11 Nov 2019.

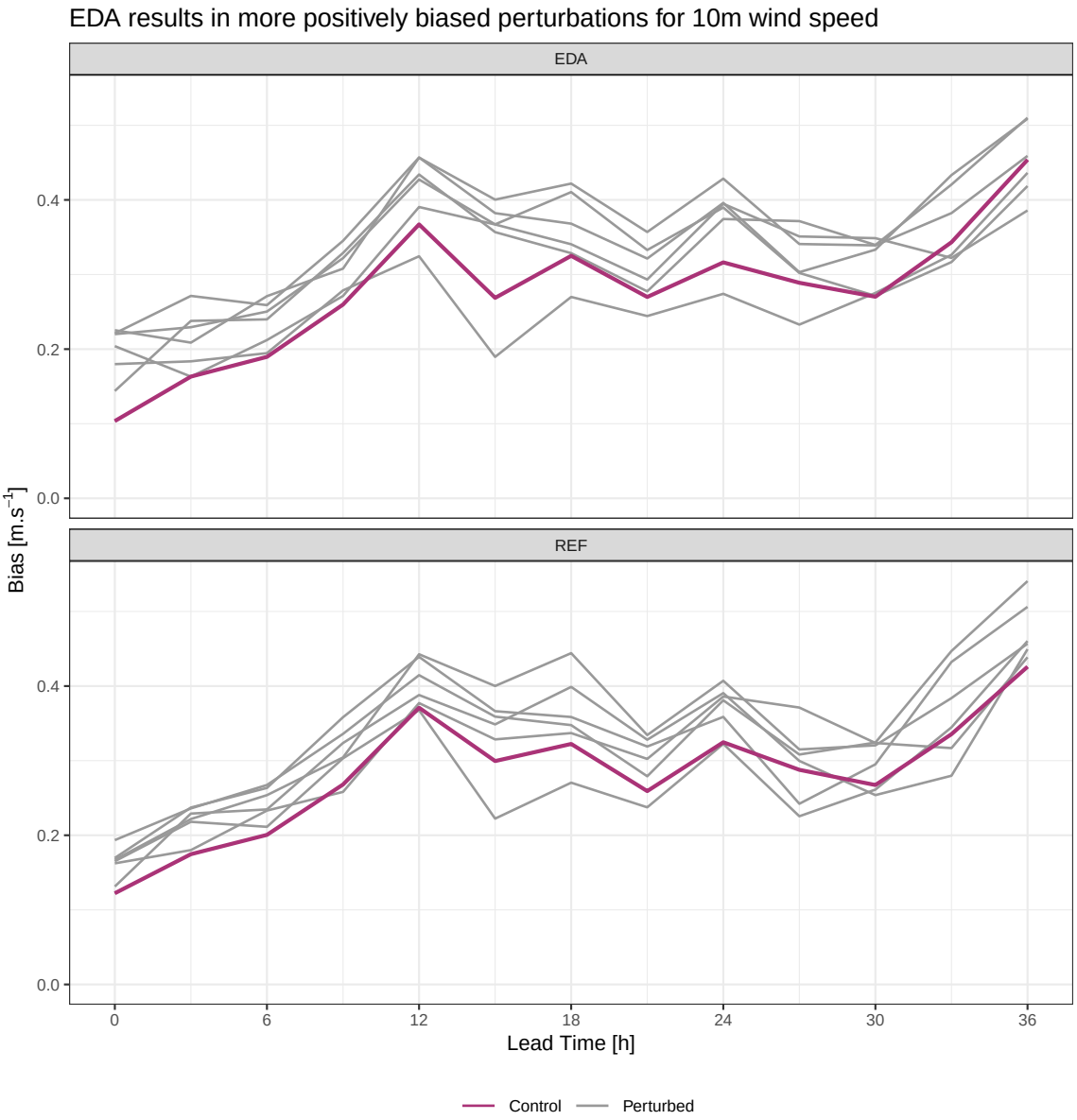

Figure 31: Bias for control (purple) and perturbed (grey) members for 10m wind speed for EDA (top row) and REF (bottom row) for the period 25 Oct - 11 Nov 2019.

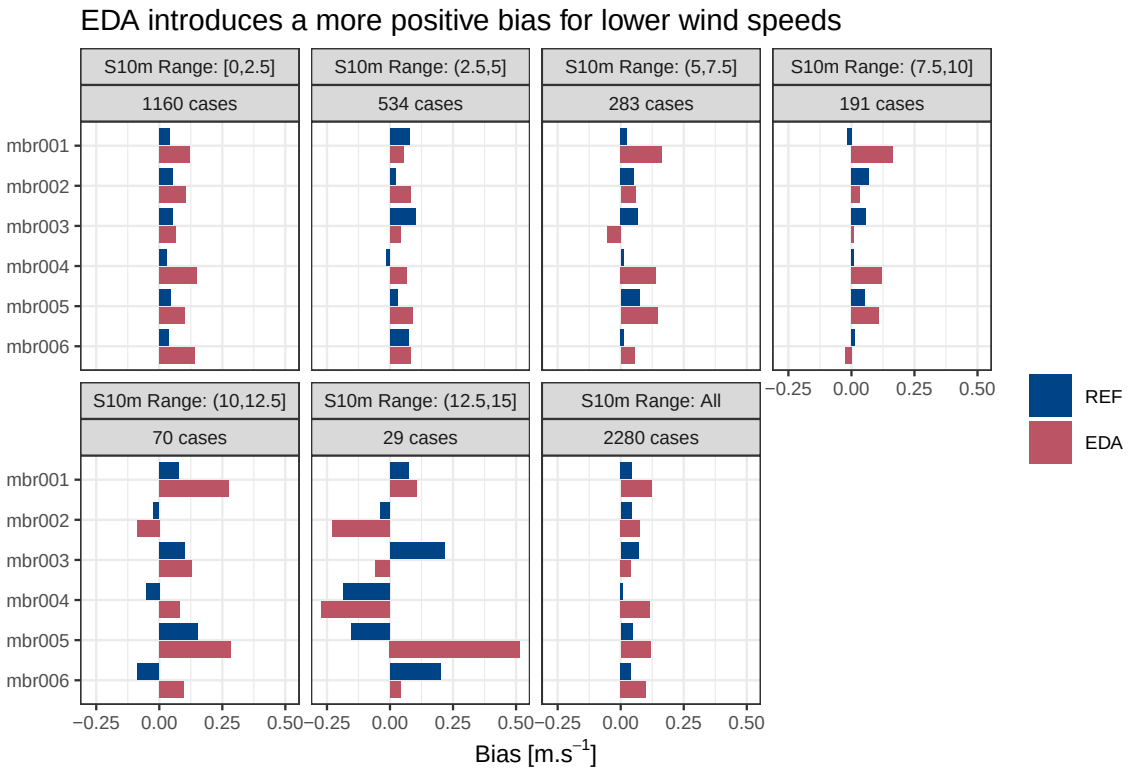

Figure 32: Bias for each ensemble member relative to the control member for 10m wind speed for REF (blue) and EDA (red) for the period 25 Oct - 11 Nov 2019 at forecast initialization time.

3.2.5 Mean Sea Level Pressure

The impact of EDA on mean sea level pressure is relatively small (Fig. 33). There is a small increase in spread at initialization time, but then it varies as to whether EDA or REF has more spread. In general, the RMSE for EDA is slightly smaller than or roughly equal to that for REF. When it comes to the biases of the perturbed members, there appears to be a more even spread between the members for EDA than for REF at initialization time as the fact that REF uses pairs of perturbations that are the reverse of each other are reflected in the biases (Fig. 34).

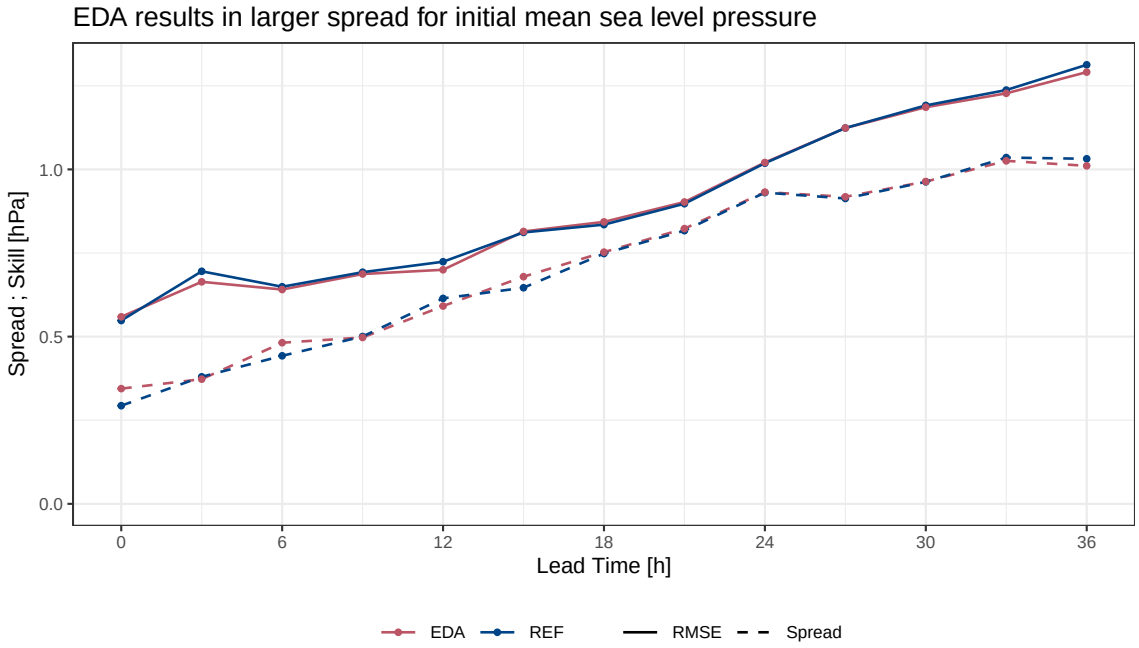

Figure 33: RMSE (solid line) and ensemble spread (dashed line) for mean sea level pressure for REF (blue) and EDA (red) for the period 25 Oct - 11 Nov 2019.

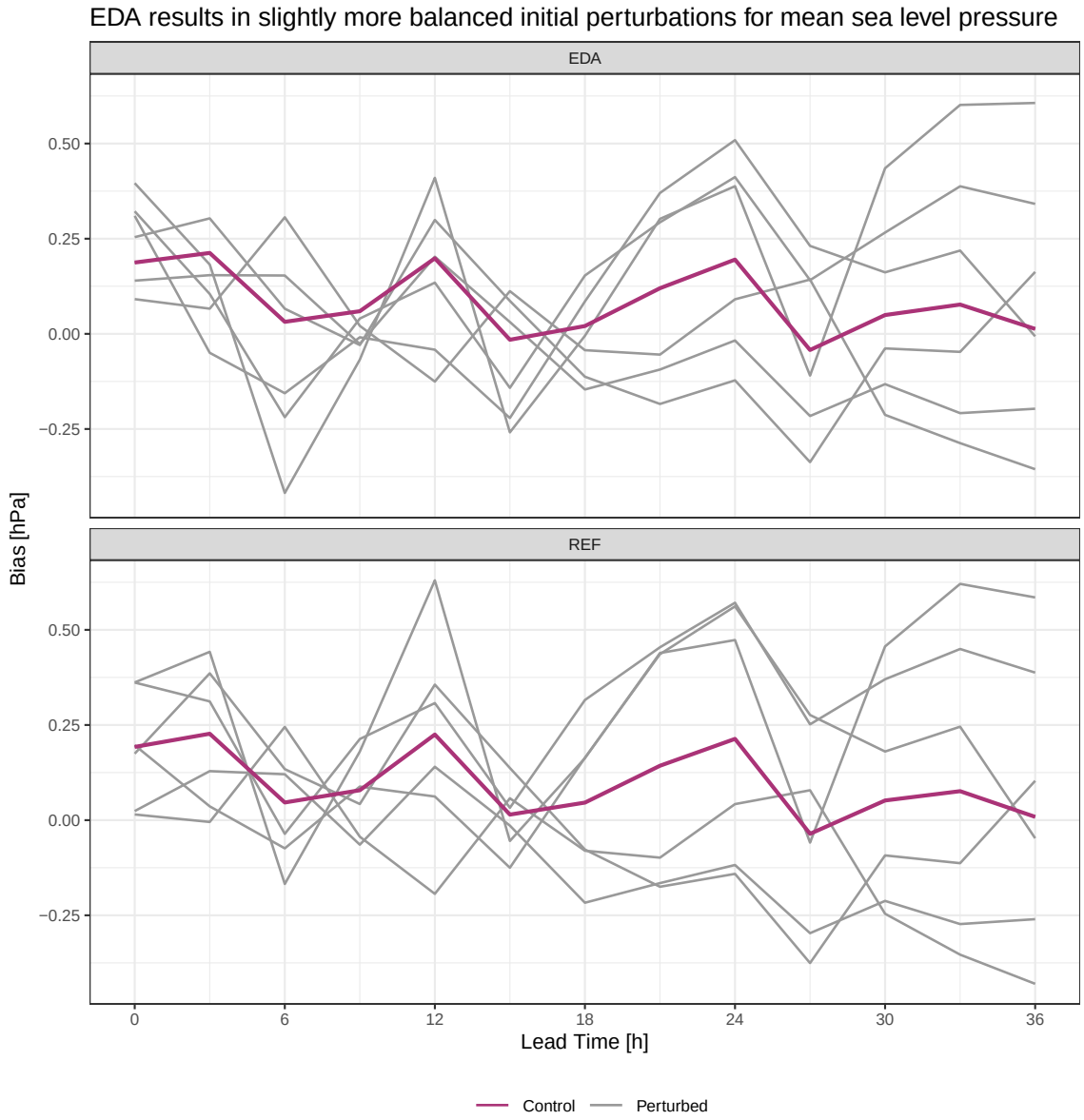

Figure 34: Bias for control (purple) and perturbed (grey) members for mean sea level pressure for EDA (top row) and REF (bottom row) for the period 25 Oct - 11 Nov 2019.

3.2.6 6-hour accumulated precipitation

Unlike the other near surface parameters discussed so far, EDA has little impact on the spread of the forecast for 6-hour accumulated precipitation (Fig. 35). It should be noted that very few observations of 6-hour precipitation were available at 6am UTC so these were not included in the verification, meaning that the first verification time is at 12 hours lead time. However, EDA appears to have a larger impact on the RMSE, having a smaller RMSE than REF through the first day of the forecast, but slightly larger at 36 hours lead time. The balance of the biases of the perturbed members is slightly narrower for EDA than for REF early in forecast, but becomes wider at 24 hours lead time (Fig. 36).

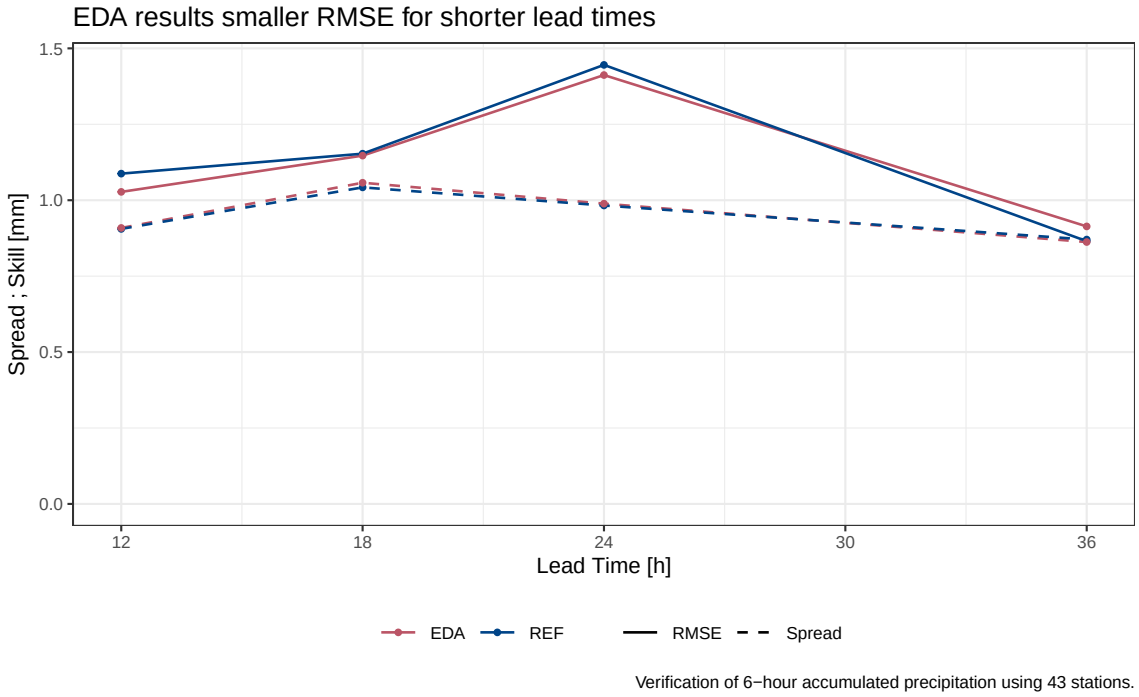

Figure 35: RMSE (solid line) and ensemble spread (dashed line) for 6-hour accumulated precipitation for REF (blue) and EDA (red) for the period 25 Oct - 11 Nov 2019.

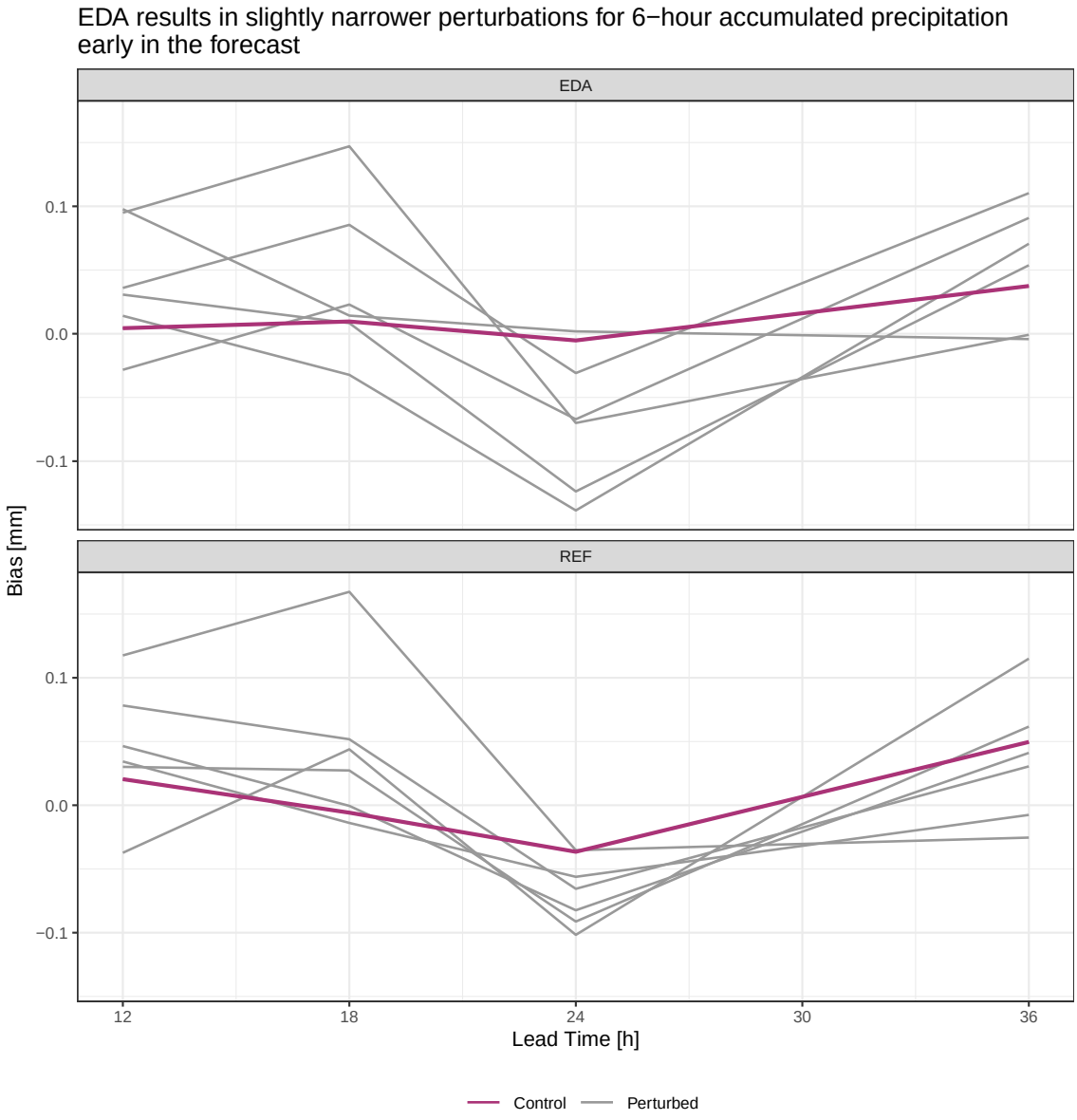

Figure 36: Bias for control (purple) and perturbed (grey) members for 6-hour accumulated precipitation for EDA (top row) and REF (bottom row) for the period 25 Oct - 11 Nov 2019.

3.2.7 Total cloud cover

EDA has an impact on both the spread and RMSE of the forecast for total cloud cover (Fig. 37). The spread for EDA is almost 1 okta wider for EDA than for REF at the forecast initialization time, while the RMSE is slightly lower. A large proportion of the difference in spread at initialization time would appear to be due to REF taking until 3 hours into the forecast for the spread in total cloud cover to properly spin up. From about 12 hours lead time, the spread for both EDA and REF grows at about the same rate and the RMSEs for each set up become approximately equal.

The biases of the perturbed members suggest that there actually may not be much spread at the forecast initialization time in EDA when compared with REF. For EDA, the biases of the perturbed members cluster together and are less negative (closer to zero) than the control member, whereas the biases of the perturbed members for REF are more evenly spread out (Fig. 38).

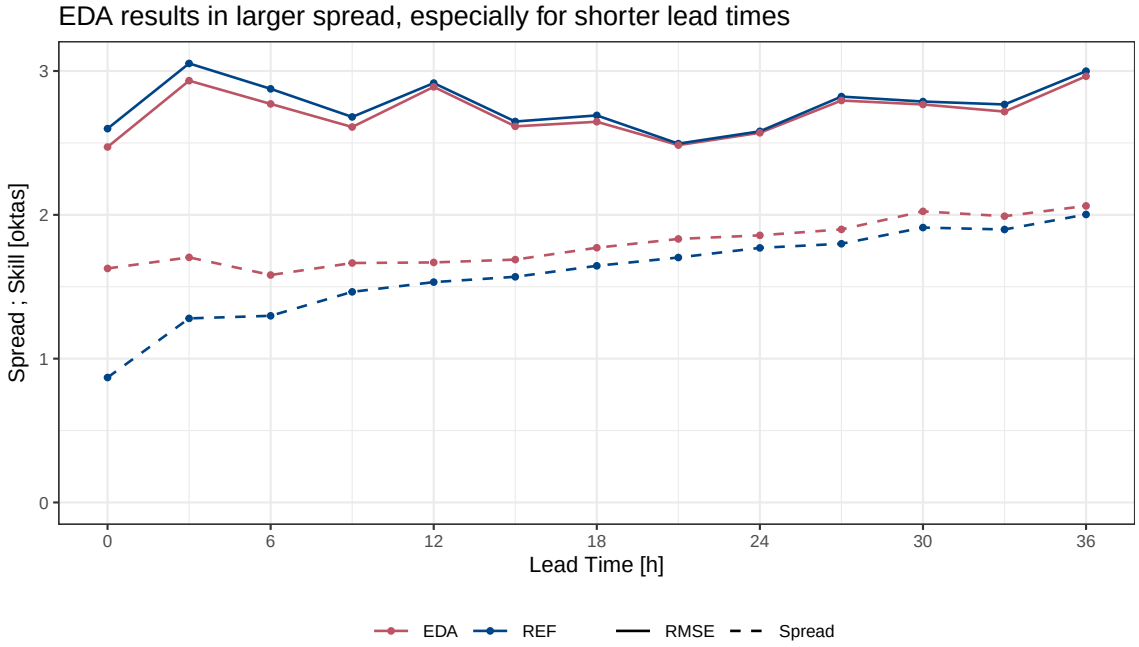

Figure 37: RMSE (solid line) and ensemble spread (dashed line) for total cloud cover for REF (blue) and EDA (red) for the period 25 Oct - 11 Nov 2019.

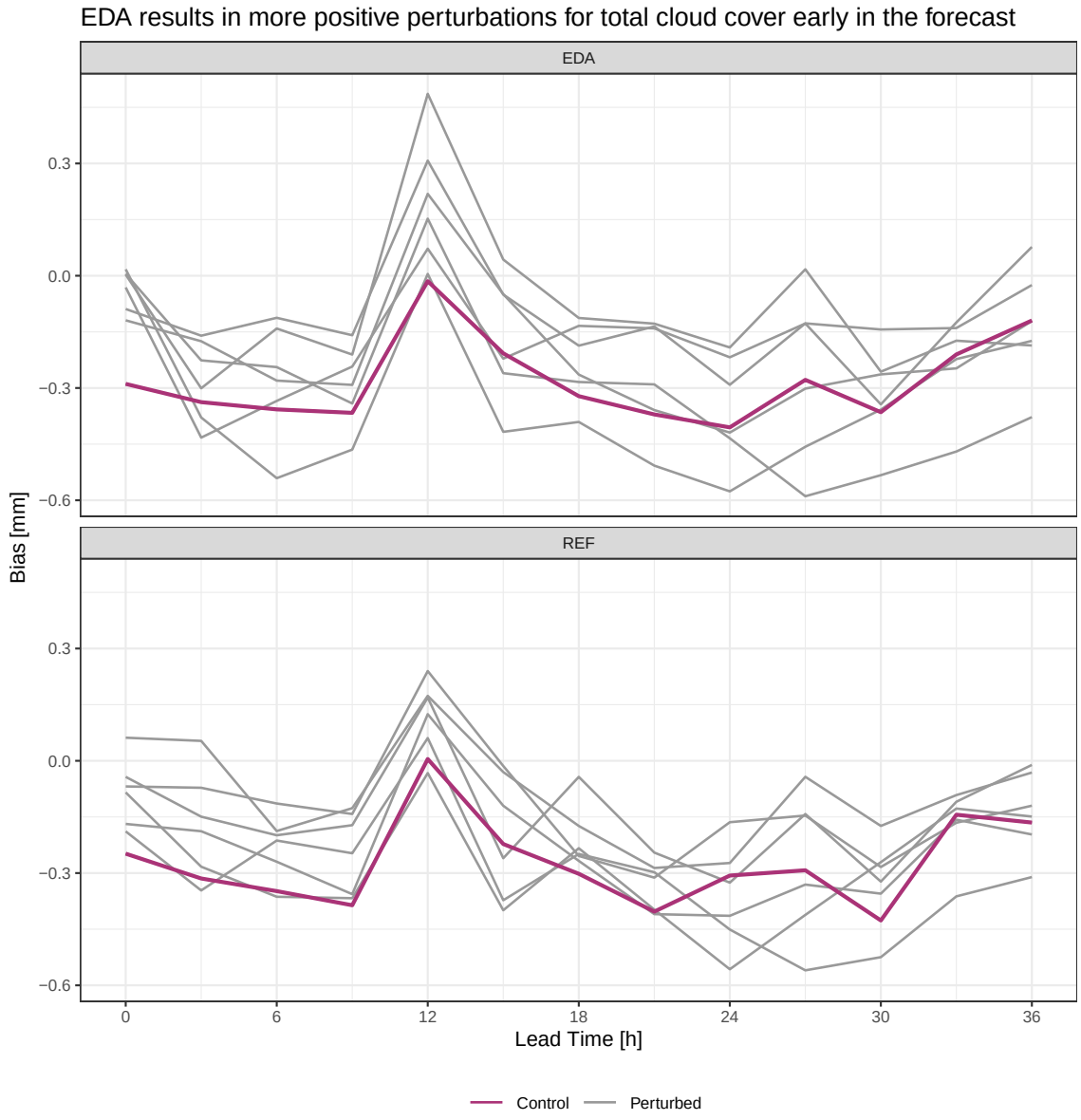

Figure 38: Bias for control (purple) and perturbed (grey) members for total cloud cover for EDA (top row) and REF (bottom row) for the period 25 Oct - 11 Nov 2019.

3.2.8 Upper Air Temperature

EDA has a clear impact on the spread of upper air temperature at the forecast initialization time, with EDA having larger spread than REF throughout the whole atmosphere (Fig. 39). Between 900- and 400-hPa, the spread for EDA is approximately equal to the RMSE. The spread difference between EDA and REF gets smaller as the forecast continues, reducing to close to zero by 24 hours lead time. The RMSE for EDA and REF is generally the same except for at the 1000 *hPa* level where EDA tends to have smaller RMSE than REF. The biases of the perturbed members confirms that the perturbations are well balanced around the control for both EDA and REF (Fig. 40).

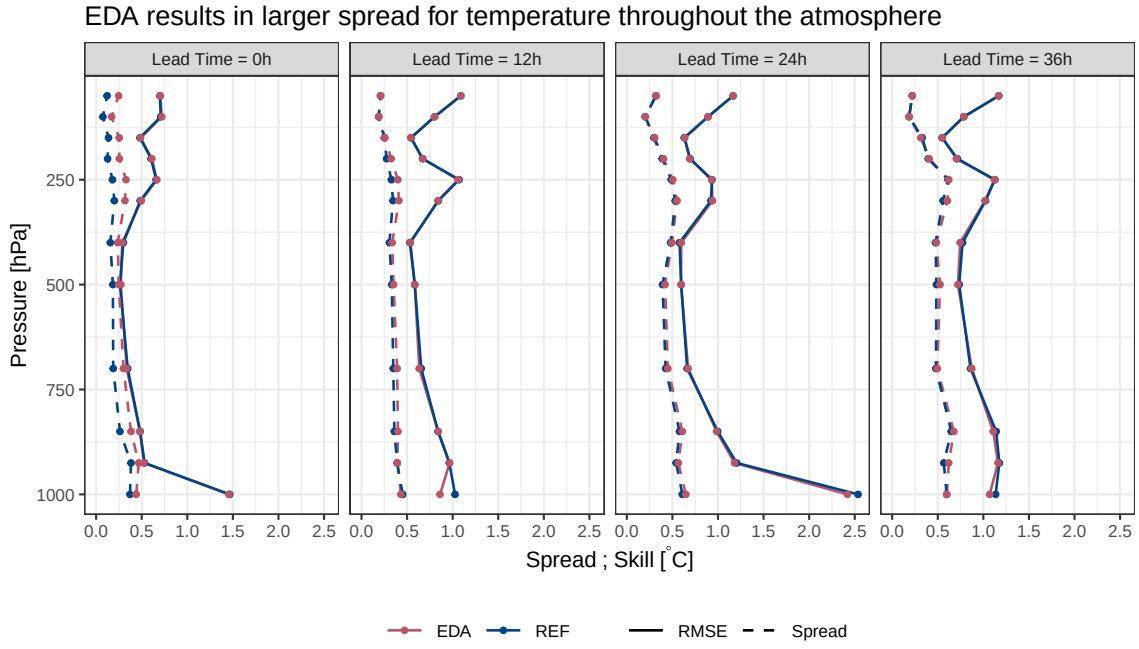

Figure 39: RMSE (solid line) and ensemble spread (dashed line) for upper air temperature for REF (blue) and EDA (red) for the period 25 Oct - 11 Nov 2019.

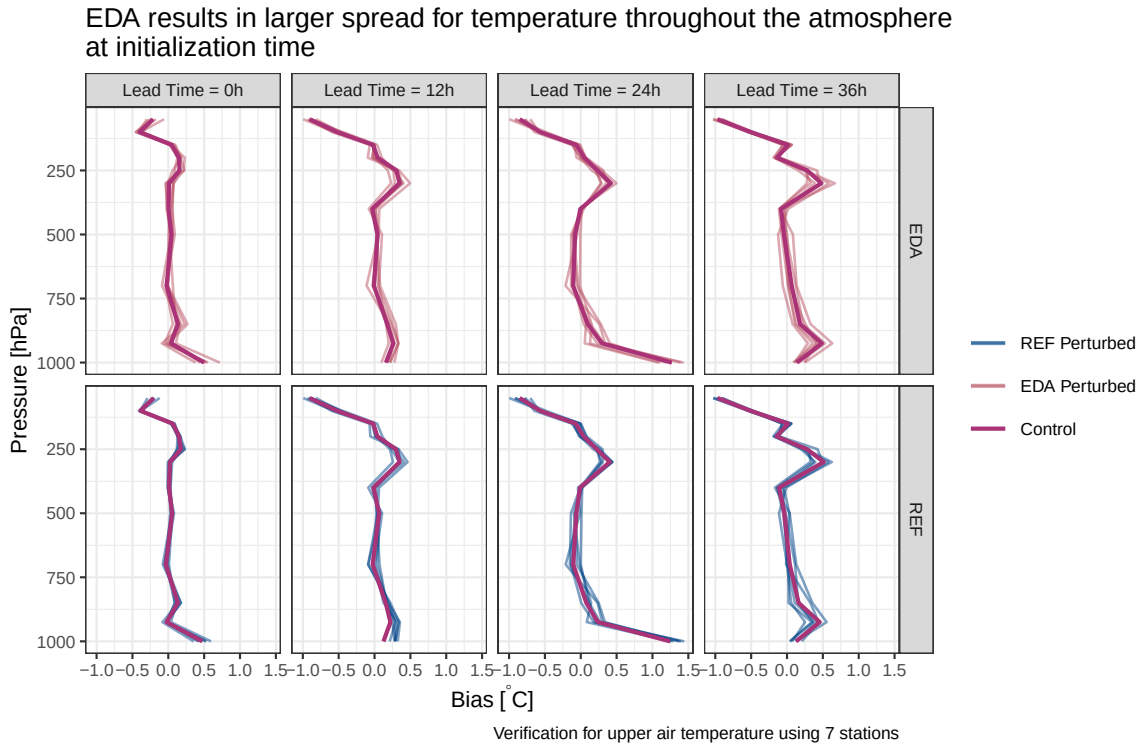

Figure 40: Bias for atmospheric temperature throughout the atmosphere for the period 25 Oct - 7 Nov 2019 for EDA (top row, with pertubed members red) and REF (bottom row, with perturbed members blue). The control member (purple) is the same for both REF and EDA. The columns are for lead times from 0- to 48-hours every 12-hours.

3.2.9 Upper Air Specific Humidity

EDA results in both larger spread and smaller RMSE than REF at initialization time below approximately 600 *hPa* in the atmosphere (Fig. 41). In fact, the spread is slightly larger than the RMSE for EDA between the 850− and 700 − *hPa* levels. For longer lead times, the larger spread and smaller RMSE for EDA persists below the 850 *hPa* level. The biases of the perturbed members suggest that the spread resulting from EDA penetrates higher into the atmosphere than for REF and that the perturbations are generally well balanced in both cases (Fig. 42).

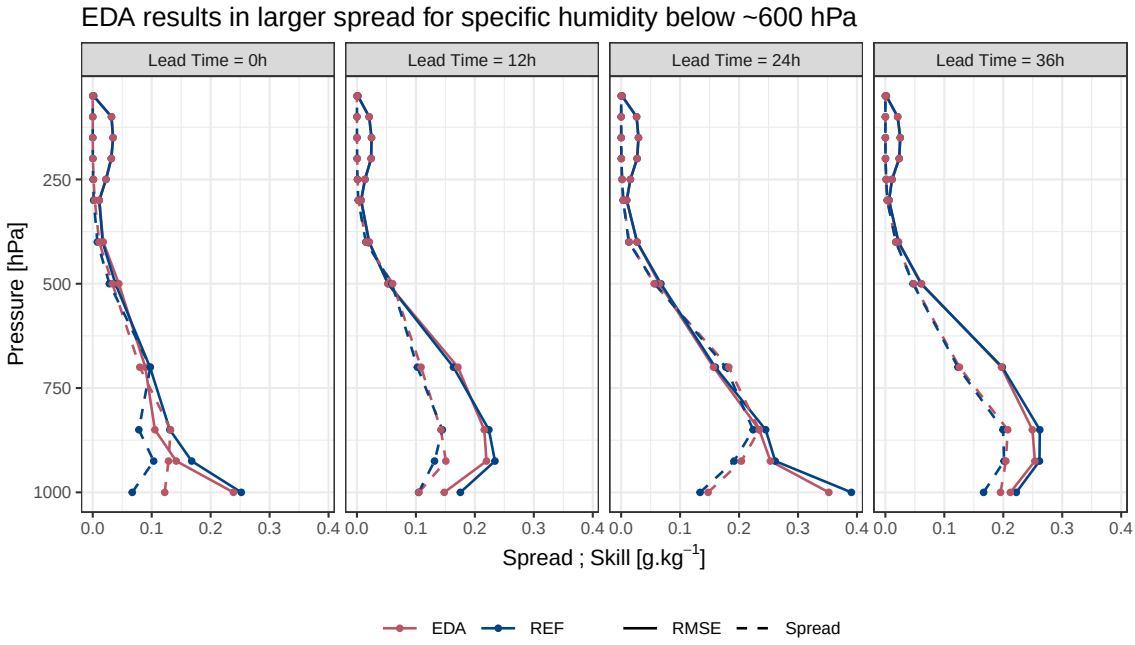

Figure 41: RMSE (solid line) and ensemble spread (dashed line) for upper air specific humidity for REF (blue) and EDA (red) for the period 25 Oct - 11 Nov 2019.

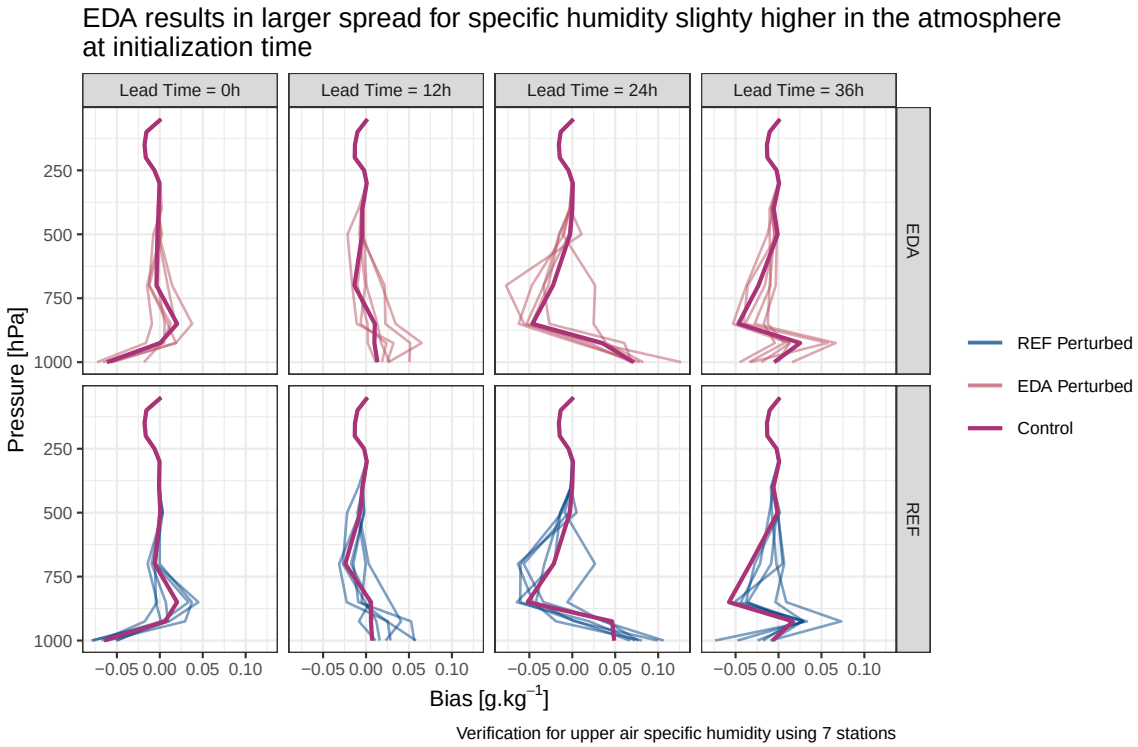

Figure 42: Bias for specific humidity throughout the atmosphere for the period 25 Oct - 7 Nov 2019 for EDA (top row, with pertubed members red) and REF (bottom row, with perturbed members blue). The control member (purple) is the same for both REF and EDA. The columns are for lead times from 0- to 48-hours every 12-hours.

3.2.10 Upper Air Wind Speed

EDA results in an increase in the spread of wind speed compared with REF at initialization time, although between the $700-$ and $500-hPa$ levels the impact is close to zero (Fig. 43). As the lead time increase, some extra spread develops in EDA between these lower / mid levels of the atmosphere, but diminishes high up in the atmosphere in particular. By 36 hours lead time, there is almost no impact on spread due to EDA except at the very lowest levels of the atmosphere. The impact of EDA on RMSE is more mixed, with a reduction in RMSE due to EDA low in the atmosphere at initialization time and a small increase at 12 hours lead time. The biases of the perturbed members suggest that the balance of the perturbations is similar for both EDA and REF (Fig. 44).

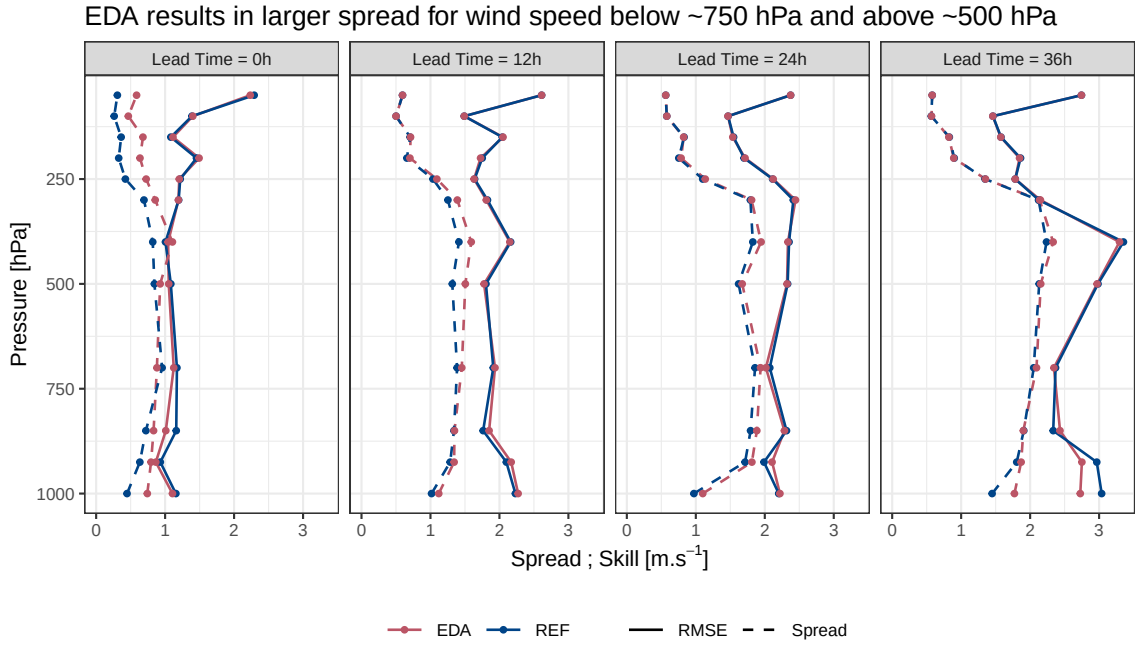

Figure 43: RMSE (solid line) and ensemble spread (dashed line) for upper air wind speed for REF (blue) and EDA (red) for the period 25 Oct - 11 Nov 2019.

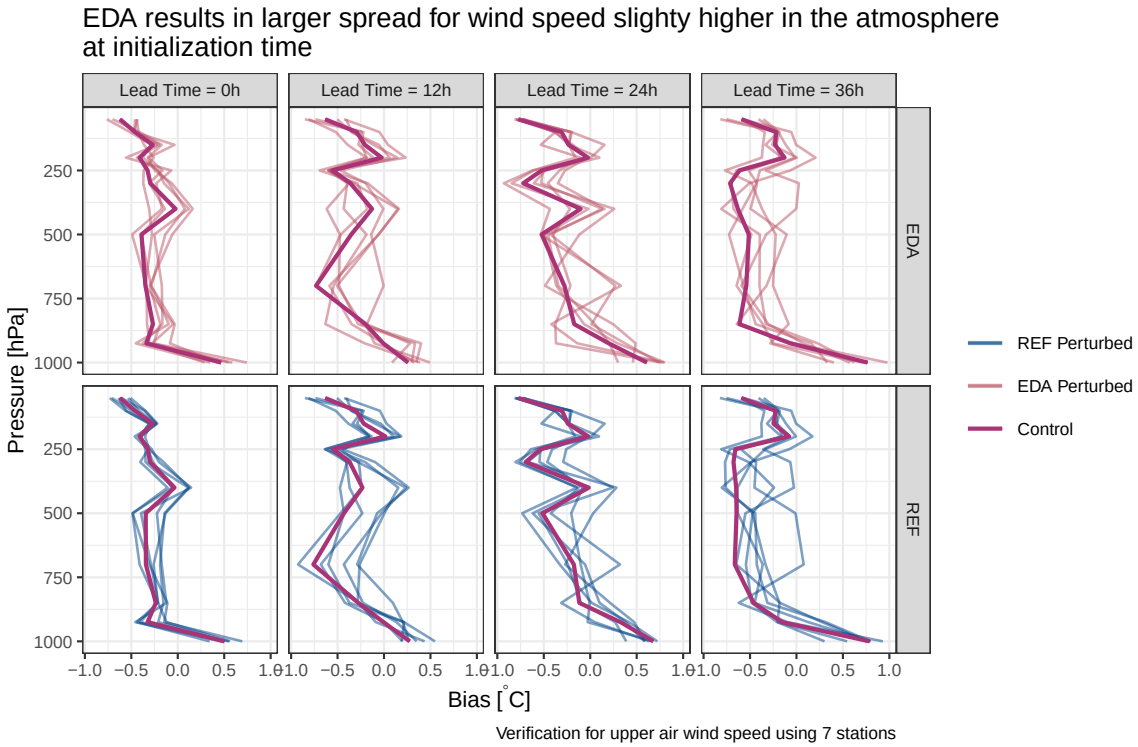

Figure 44: Bias for wind speed throughout the atmosphere for the period 25 Oct - 7 Nov 2019 for EDA (top row, with pertubed members red) and REF (bottom row, with perturbed members blue). The control member (purple) is the same for both REF and EDA. The columns are for lead times from 0- to 48-hours every 12-hours.

3.2.11 Geopotential Height

EDA results in a clear and consistent increase in spread over REF throughout the atmosphere at initialization time together with a very small increase in RMSE that gets smaller with height (Fig. 45). At 12 and 24 hours lead time, EDA has almost identical spread and RMSE as REF, and at 36 hours lead time EDA has slightly more spread than REF. The biases of the perturbed members suggest that EDA introduces a small more positive bias to the geopotential height at initialization that dissipates by 12 hours lead time (Fig. 46).

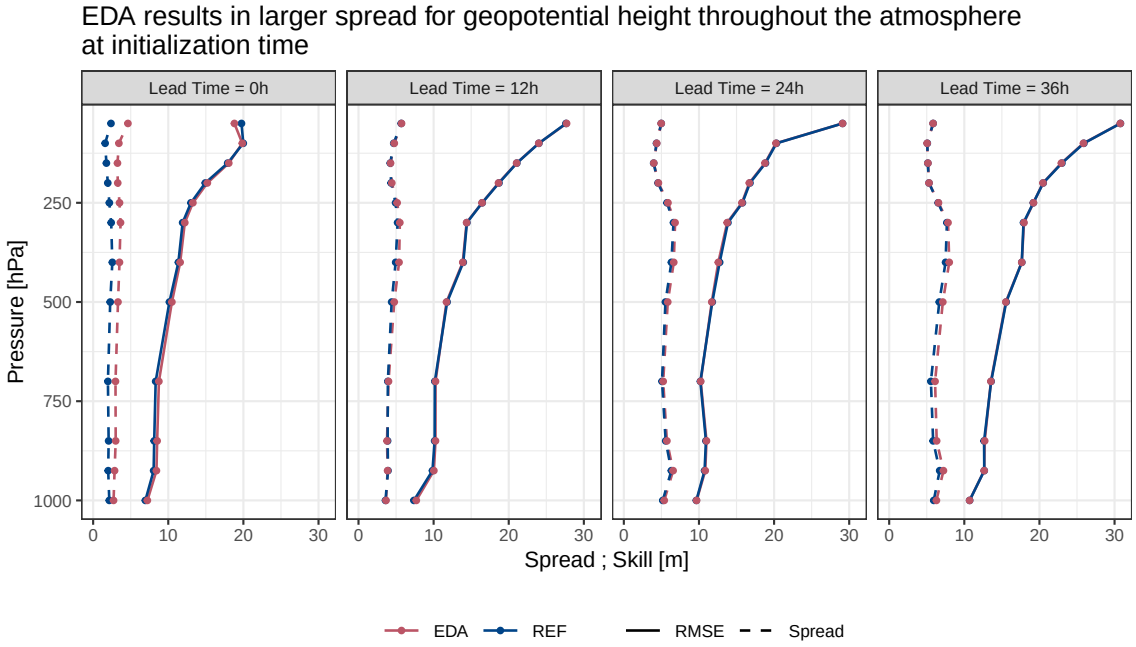

Figure 45: RMSE (solid line) and ensemble spread (dashed line) for geopotential height for REF (blue) and EDA (red) for the period 25 Oct - 11 Nov 2019.

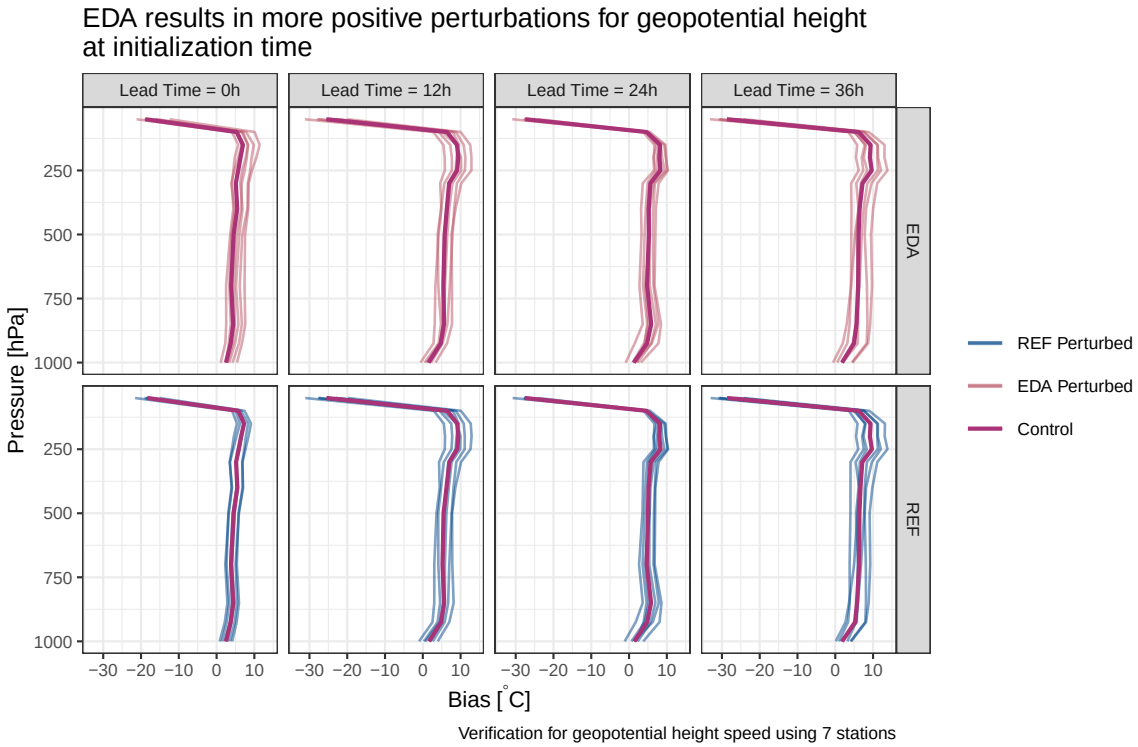

Figure 46: Bias for geopotential height for the period 25 Oct - 7 Nov 2019 for EDA (top row, with pertubed members red) and REF (bottom row, with perturbed members blue). The control member (purple) is the same for both REF and EDA. The columns are for lead times from 0- to 48-hours every 12-hours.

4 Discussion

The purpose of this work is to explore some methods for modelling uncertainty in weather forecasts for the Norwegian Arctic area. The broad sources of uncertainty addressed are in the physics parameterizations used by the model and in the initial conditions from which each forecast is initialized. The methods assessed are the use of stochastically perturbed parameters (SPP) in selected physics parameterization schemes for the model uncertainty; and ensembles of data assimilations (EDA), whereby observations are perturbed in a way such that their uncertainties are taken into account in the model initial conditions. The efficacy of these methods is estimated using a comparison of the spread and skill of the ensemble forecasts using these methods with a reference ensemble forecast. Due to the availability of observational data and different development priorities it was unfortunately not possible to test both SPP and EDA for the same time periods using the same cycle of the NWP model.

SPP is designed to address model uncertainty throughout the forecast and is thus related to the predictability characteristic of the model itself. Selected parameters for a range of parameterization schemes are drawn from log-normal distributions of those parameters. The pattern that is used to draw the parameters is both spatially and temporally correlated such that there are not big jumps in the parameter values between grid points and time steps. Given that SPP affects the parameterization schemes throughout the forecast, it is expected that the verification scores will be impacted for all lead times. An initial impression from the spread and skill analysis is that SPP generally results in improved ensemble forecasts with larger spread and either similar or smaller RMSE than the reference forecast for a wide range of weather parameters. However, for 2m temperature, for example, SPP introduces a cool bias that occurs in both the tested winter and summer seasons and for the whole range of observed temperatures. On the other hand, the perturbed members of the reference forecast have a warm bias during the winter and a cool bias during the summer. The cool bias due to SPP meant that during the day time, the cool bias that was present in the control member forecast was exacerbated by the members with SPP, but the warm bias that was present in the control member during the night time was reduced by the cooler SPP members. A similar situation is seen for both 2m specific humidity and 10m wind speed, whereby SPP appears to add more spread, but this is due to the introduction of biases in the perturbed members of the ensemble. A similar behaviour was noted by Frogner et al. (2022), who suggest that the parameter perturbations being sampled from a log-normal distribution is the cause. By sampling from a log-normal distribution, the perturbations are not equally distributed. In many cases this may be representative of the error distributions of the parameters being perturbed, but the result is that the ensemble members with those perturbations tend to be skewed in one direction. Work is ongoing within the ACCORD consortium to test the use of normal distributions as PDFs from which to sample the perturbed parameters.

Using mean sea level pressure as a proxy for the general circulation of the atmosphere in the model domain suggests that SPP has an impact throughout the atmosphere. However, it seems as though any impact of SPP on the spread of the ensemble is matched by an increase in the errors. This is not the desired behaviour of a well calibrated ensemble. It is not clear whether this is due to changes in the biases of the perturbed members of the ensemble, but some of the near surface biases in temperature, humidity and wind speed

may be contributing. Furthermore, SPP does appear to add a small positive bias to the wind speed for the perturbed ensemble members at the lower levels of the atmosphere, though for other upper air parameters the impact on biases is more balanced and increases in spread are achieved without increasing the RMSE as well.

For both cloud cover and precipitation, the physics schemes, particularly the cloud microphysics parameterizations, acted on by SPP would be expected to have an impact on the forecast. Indeed, large increases in spread together with a reduction in the RMSE were achieved for total cloud cover when SPP was used. Furthermore, a slightly better balanced spread of biases was also achieved. However, for precipitation the results were more mixed with SPP resulting in slightly more spread in the winter and at times slightly less spread in the summer. It should be noted that the majority of cases in both SOP 1 and SOP 2 are dry. Of 10353 observations of 6 hour accumulation during SOP 1, only 3435 (33%) included any precipitation, and of 9645 cases in SOP 2, only 1756 (18%) included any precipitation (Fig. 47). This means that the sample size is relatively small, especially for SOP 2, and thus the results may not be that robust.

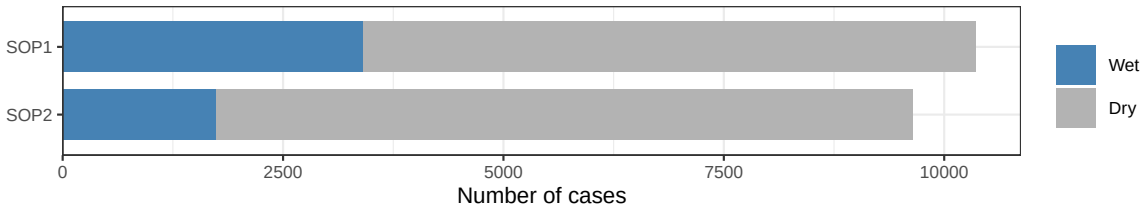

Figure 47: Number of wet and dry cases during SOP 1 and SOP 2. A wet case is defined as observed precipitation during the 6 hour observation period.

While SPP attempts to estimate the impacts of uncertainties arising from the model itself, EDA attempts to address the atmospheric predictability when the forecast is initialized. In less predictable situations small changes in the initial conditions will have a sizeable impact on the forecast, whereas for more predictable situations the impact of small changes will be smaller. A number of methods have been developed to perturb the initial conditions, the simplest of which for limited area models is to use an ensemble of initial conditions from the (global) model which provides it the boundary conditions. For the reference forecast used in this study (REF), this ensemble of initial conditions was derived from time lagged forecasts from the deterministic global model (IFSHIRES) using a process known as SLAF (Scaled Lagged Average Forecasts; Toth and Kalnay (1993)). SLAF only provides perturbations to the initial conditions at the spatial scale of IFSHIRES (~9km) and does not properly address where the uncertainties in the initial conditions arise since it relies on the uncertainties in forecasts between consecutive model runs. EDA on the other hand perturbs the observations used to derive the forecast initial conditions such that these uncertainties drive the uncertainties in the initial conditions. EDA was shown to result in more variability between the ensemble members than REF for the initial conditions for a range prognostic variables. Furthermore, for an event that might be considered less predictable (a polar low) both REF and EDA had more variability for wind speed than for other dates, with EDA resulting in more variability than REF.

Since EDA acts on the initial conditions of the forecast, it would be expected to have the largest impact early in the forecast before perturbations from the lateral boundaries begin to dominate. This is indeed the case for all of the parameters tested, with forecasts showing increases in spread due to EDA over REF becoming

smaller as the forecast progresses. In most cases, except for mean sea level pressure and 6-hour accumulated precipitation, the spread for EDA is larger than that for REF until the end of the forecast period (36 hours in the case of EDA). For upper air parameters, the impact is generally confined to the first 12 hours of the forecast. If there is any impact of EDA on the RMSE, it is to decrease it over the first few hours of the forecast. In only very few cases is the RMSE for EDA larger than that for REF.

Although the fact the EDA results in larger spread than REF without having a detrimental impact on the RMSE is promising a deeper inspection revealed some potential issues. While the perturbed members for REF have a positive bias for 2m temperature, 10m wind speed and total cloud cover, those members in EDA have an even more positive bias. Conversely, for 2m specific humidity the positive bias of the perturbed members in REF become more evenly spread around the control member when EDA is used. In general these biases are not seen in the upper air, except for geopotential height where the well balanced perturbed members for REF become positively biased when EDA is used. Stratifying the results suggested that the biases were present across the range of observed values for all of the parameters, suggesting that further tuning of the observation perturbations used in EDA might result in a better balance.

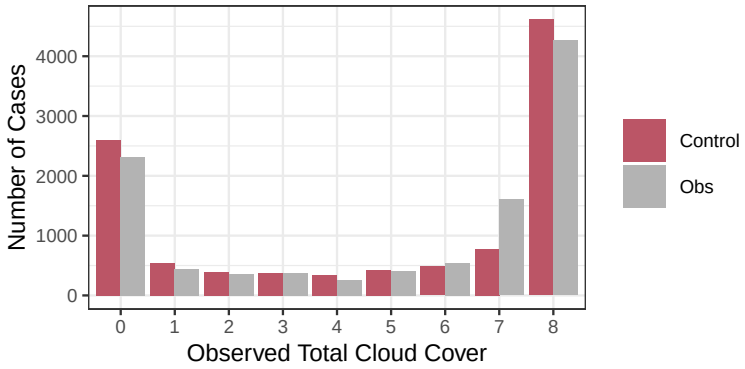

Figure 48: Number of observations (grey) and control member forecasts (red) for each level of cloud cover in oktas at forecast initialization time from 25 October to 7 November 2019.

Taking total cloud cover separately, since the values are bound between zero and full cloud cover, it may be reasonable to assume that the data set was dominated by clear days since the perturbations are predominantly positive for both REF and EDA. From Fig. 48, which shows the observed total cloud cover and the total cloud cover forecasted by the control member for all cases used in the verification, it is clear that it is not the case that the dataset is dominated by clear days. Indeed there are more cases with full cloud cover than completely clear. A closer look at the biases of the perturbed members for each observed amount of cloud cover reveals that EDA has more positive biases than REF for observations with less than half cloud cover (Fig, 49). For cloudier cases, EDA has a less positive bias than REF when 5 oktas are observed, a less negative bias than REF when 6 oktas are observed, and a more negative bias than REF when 7 or 8 oktas are observed. This suggests that in general, EDA has the same sort of impact on cloud cover as REF, but acts more strongly.

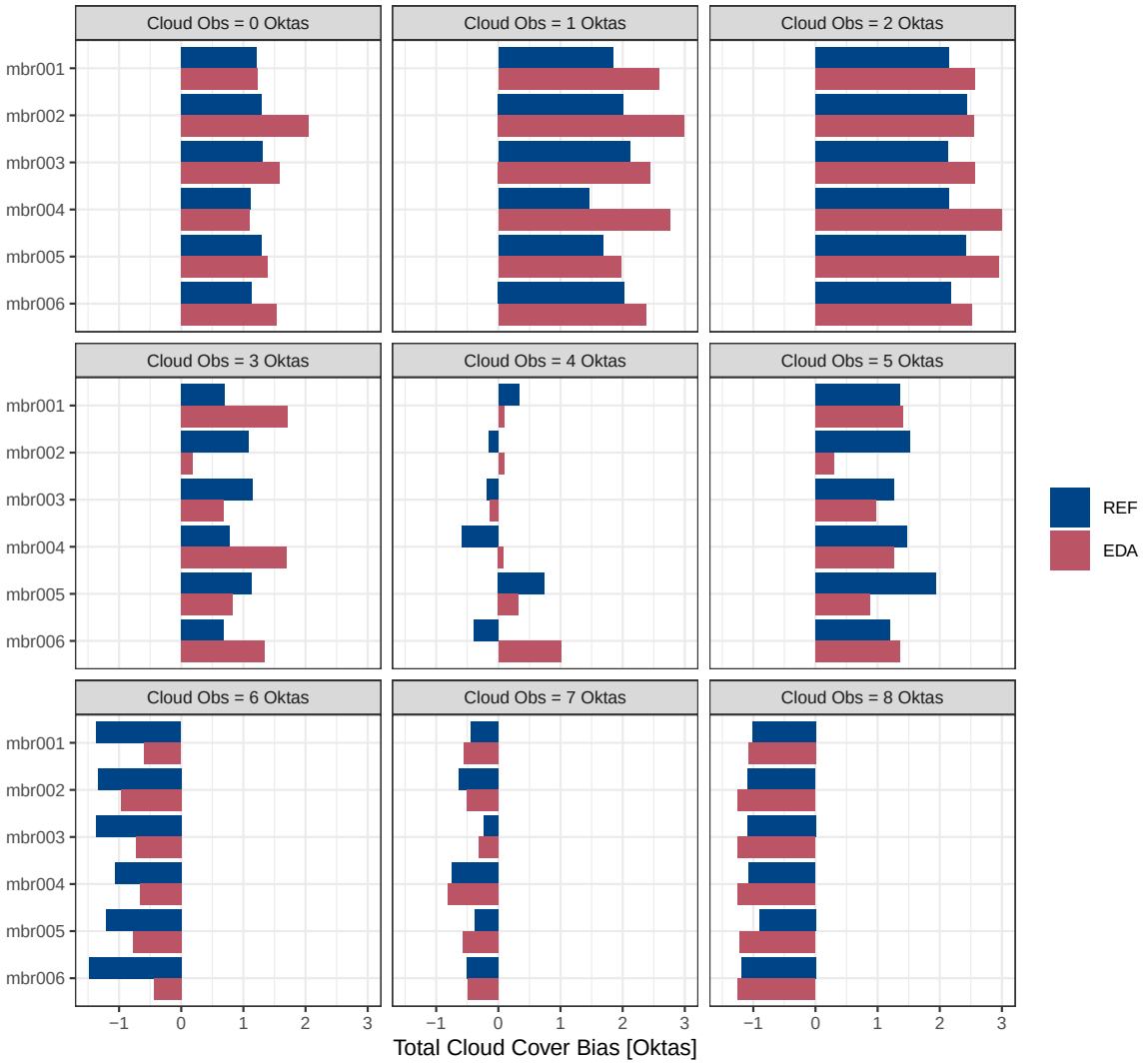

Figure 49: Biases for the perturbed members for REF(blue) and EDA (red). Each panel is for a different level of observed total cloud cover in oktas at forecast initialization time from 25 October to 7 November 2019.

5 Conclusions and Recommendations

Predictability is the main driver of uncertainty in weather forecasts. Weather forecasts should be accompanied with an estimate of this uncertainty in order for users to make the best informed decisions that they can. From the perspective of a weather forecast predictability is governed by the inherent predictability of the atmosphere on a given day - the intrinsic predictability, and the predictability that a numerical model of the atmosphere can aim to achieve - the practical predictability. This study assesses methods that attempt to estimate uncertainties resulting from both of these drivers of predictability in the European Arctic - namely EDA for intrinsic predictability and SPP for practical predictability. These methods were compared with a reference a forecast and the spread-skill verification metric was used to quantify the results.

In general both EDA and and SPP led to better quantification of the uncertainty in the forecasts according to the spread-skill metric. However, for some parameters extra biases were also introduced which is undesirable. In particular SPP introduced a cooler bias with weaker winds, while EDA introduced a slightly warm bias, with wind speeds that were slightly too strong and too much cloud cover. With regard to SPP, it is suspected that the biases come from using a lognormal distribution to sample the perturbations and work is ongoing in the ACCORD consortium to test the use of normal distributions instead. It is also suspected that further tuning of the perturbations to the observations in EDA would improve the balance of the perturbations in the initial conditions. It should also be noted that the vast majority of the observations used in the verification were over land and the AROME-Arctic domain is mostly ocean, so the results reported here are only valid over land and the impact of both SPP and EDA over the ocean in terms of verification scores remains an open question.

Following the work done here, the following are recommendations to be taken into account when preparing the AROME-Arctic EPS for production.

- SPP should not be taken into use while it adds biases to certain parameters.
- Developments with using a normal distribution for sampling perturbations for SPP within the ACCORD consortium should be followed closely.
- Once SPP results in balanced perturbations, more work should be done to tune the magnitude of the perturbations to give the best results for the Arctic region.
- EDA can be taken into use as it currently stands.
- More work should be done to tune the perturbations for EDA to be better balanced and to give the best estimates of initial conditions uncertainty for the AROME-Arctic EPS.

6 Acknowledgements

This study was supported by the Norwegian Research Council Project 280573 ‘Advanced models and weather prediction in the Arctic: enhanced capacity from observations and polar process representations (ALERTNESS).’ This work is a contribution to the Year of Polar Prediction (YOPP), a flagship activity of the Polar Prediction Project (PPP), initiated by the World Weather Research Programme (WWRP) of the

World Meteorological Organization (WMO).

In addition, thanks are due to the following people: Roger Radriamampiannina (MET) implemented the code for running EDA in the AROME-Arctic ensemble. Yurii Batakk ran the EDA experiments. Rafael Gröte (MET) set up the modelling system for AROME-Arctic EPS. Inger-Lise Frogner (MET) and Ulf Andrae (SMHI) provided code and assistance for running the SPP experiments.

7 References

Bouttier, F., L. Raynaud, O. Nuissier, and B. Ménétrier, 2016: Sensitivity of the AROME ensemble to initial and surface perturbations during HyMeX. *Quarterly Journal of the Royal Meteorological Society*, **142**, 390–403, https://doi.org/10.1002/qj.2622.

Frogner, I.-L., and Coauthors, 2019a: HarmonEPS—the HARMONIE ensemble prediction system. *Weather and Forecasting*, **34**, 1909–1937, https://doi.org/10.1175/WAF-D-19-0030.1.

——, A. T. Singleton, M. O. Koltzow, and U. Andrae, 2019b: Convection-permitting ensembles : Challenges related to their design and use. *Quarterly Journal of the Royal Meteorological Society*, **145**, 90–106, https://doi.org/10.1002/qj.3525.

——, ——, P. Ollinaho, A. Hally, K. Hämäläinen, J. Kauhanen, K.-I. Ivarsson, and D. Yazgi, 2022: Model uncertainty representation in a convection-permitting ensemble—SPP and SPPT in HarmonEPS. *Monthly Weather Review*, **150**, 775–795, https://doi.org/10.1175/MWR-D-21-0099.1.

Isaksen, L., M. Fisher, and J.Berner, 2007: Use of analysis ensembles in estimating flow-dependent background error variance. *Proc. ECMWF workshop on flow-dependent aspects of data assimilation*, Reading, United Kingdom, ECMWF, 65–86 https://www.ecmwf.int/sites/default/files/elibrary/2007/10127-use-analysis-ensembles-estimating-flow-dependent-background-error-variance.pdf.

——, M. Bonavita, R. Buizza, M. Fisher, J. Haseler, M. Leutbecher, and L. Raynaud, 2010: Ensemble of data assimilations at ECMWF. *Research Department Tech. Memo*, **636**, 45 pp, https://doi.org/10.21957/obke4k60.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, **20**, 130–141, https://doi.org/10.1175/1520-0469(1963)020%3C0130:DNF%3E2.0.CO;2.

——, 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289–307, https://doi.org/10.3402/tellusa.v21i3.10086.

Müller, M., Y. Batrak, J. Kristiansen, M. A. Ø. Køltzow, G. Noer, and A. Korosov, 2017: Characteristics of a convective-scale weather forecasting system for the european arctic. *Monthly Weather Review*, **145**, 4771–4787, https://doi.org/10.1175/MWR-D-17-0194.1.

Ollinaho, P., and Coauthors, 2017: Towards process-level representation of model uncertainties: Stochastically perturbed parametrizations in the ECMWF ensemble. *Quarterly Journal of the Royal Meteorological Society*, **143**, 408–422, https://doi.org/10.1002/qj.2931.

Singleton, A. T., and R. Grote, 2020: *Verification of EPS forecasts using AROME-arctic*. Norwegian Meteorological Institute MET Report 06-2020, https://www.met.no/publikasjoner/met-report/met-report-2020/_/attachment/download/76218c4b-c061-41b7-aa6c-b3c4feaf5af2:bb7271f45a96a5210e957e6c4e41e41cbfa32ba7/MET-report-06-2020.pdf.

Surcel, M., I. Zawadzki, and M. K. Yau, 2015: A study on the scale dependence of the predictability of precipitation patterns. *Journal of the Atmospheric Sciences*, **72**, 216–235, https://doi.org/10.1175/JAS-D-14-0071.1.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bulletin of the American Meteorological Society*, **74**, 2317–2330, https://doi.org/10.1175/1520-0477(1993)074%3C2317:EFANTG%3E2.0.CO;2.

Tsyrulnikov, M., and D. Gayfulin, 2017: A limited-area spatio-temporal stochastic pattern generator for simulation of uncertainties in ensemble applications. *Meteor. Z.*, **26**, 549–566.