Authors Response to comments on "Adding value to Extended-range Forecasts in Northern Europe by Statistical Post-processing Using Stratospheric Observations" by Natalia Korhonen et al.

The comments are in Black and the responses in blue.

We thank the editor and the reviewer for their comments on the revised manuscript.

We have now done further changes to the manuscript. The *SWI* categories have now been defined using the ZMZW only and the forecast skill scores have been recalculated according to this. In addition to these, we have done further editing to the manuscript, to clarify it, according to the comments. Below we respond to the reviewer point-by-point.

Best regards

Natalia Korhonen

Submitted on 16 Apr 2020
Anonymous Referee #3

Having read the latest revision of the paper, I still don't feel that my main reservation about the analysis has been adequately addressed, and I therefore do not feel the paper is suitable for publication in its current form. In my previous reviews, I expressed scepticism about the use of the QBO as a predictor of the Arctic Oscillation (AO). It seemed more plausible to me that the polar vortex strength was a stronger predictor of the AO, and that the authors should more fully analyse this before focusing on the QBO.

In the latest draft the authors have revised their methods, but I now find the latest analysis rather confusing and hard to follow. If I understand correctly, the authors now use the polar vortex strength (ZMZW) as a predictor of the AO if the vortex is weak (ZMZW below its 10th percentile) or strong (ZMZW above its 80th percentile) and demonstrate that the AO and temperature are influenced by the polar vortex strength. For post-processing of temperature forecasts, however, they only consider cases where ZMZW is between the 10th and 80th percentiles, and use the QBO to categorise different forecast ensemble members, with a selection of different QBO wind thresholds tried. This choice of method appears arbitrary and is not adequately justified in the paper. It is hard to see a chain of reasoning that led to the final method, which seems rather complicated. Why restrict the post-processing to those cases where the ZMZW is between the 10th and 80th percentiles, when you've just shown that a weak or strong polar vortex has an effect on the AO and temperature?

The authors argue that the QBO is important, since they have left out the weak and strong ZMZW cases from the post-processing, and still find improved forecast skill when using the

QBO only to define their SWI_neg and SWI_plain categories. But this misses the point, in my opinion - if the QBO is influencing the AO via the polar vortex, then ZMZW is the only predictor needed. The SWI categories should therefore first be defined using ZMZW only, and the forecast skill calculated. Including the QBO in the SWI categories' definition is only then justified if it leads to a significant improvement in skill compared to using ZMZW only.

Response: As suggested in the third paragraph of the comment above, the SWI categories have now been defined using the ZMZW only and the forecast skill calculated. As the forecast skill was similar (i.e., not significant difference) to the SWI including QBO, the QBO was no longer included in the SWI. We have updated Figures 2-5 and edited the text according to this modification, i.e., QBO is left out elsewhere but the Discussion part.

Minor comments

p6 line 15: Where did the choice of QBO wind threshold values come from? (7, 10, 13 ms-1)

Response: this part was now left out.

Figure 2 : the caption states that the p-values relate to pairs of distributions, but (unlike figure 3) it's not clear what these pairs are. Also, why are there only 57 members above the 80th percentile, but 149 members below the 20th percentile?

Response: The Fig. 2 has now been replotted, and the pairs of distributions are now visible.

The quantiles in the previous version were calculated directly from the November-March daily data. As the members represented 10 days minima, there were more members below the 20^{th} percentile than above the 80^{th} percentile. In the current revised manuscript the quantiles are now defined directly from the 10 days minima of the November-February ZMZW at 60N 10 hPa.

Figure 5,6 : The numerical values in the colour bar don't correspond to the boundaries between colours.

Response: the figures have now been replotted and now the numerical values in the color bar do correspond to boundaries between the colors.

Adding value to Extended-range Forecasts in Northern Europe by Statistical Post-processing Using Stratospheric Observations

Natalia Korhonen[1,2], Otto Hyvärinen[1], Matti Kämäräinen[1], David S. Richardson[3], Heikki Järvinen[4], Hilppa Gregow[1]

[1]Weather and Climate Change Impact Research, Finnish Meteorological Institute, Helsinki, 00101, Finland
[2]Doctoral Programme in Atmospheric Sciences, University of Helsinki, Finland
[3]ECMWF, Reading, UK
[4]Institute for Atmospheric and Earth System Research/Physics, University of Helsinki, 00014, Finland

Correspondence to: Natalia Korhonen (Natalia.Korhonen@fmi.fi)

Abstract. The strength of the stratospheric polar vortex influences the surface weather in the Northern Hemisphere in winter; a weaker (stronger) than average stratospheric polar vortex is connected to negative (positive) Arctic Oscillation (AO) and colder (warmer) than average surface temperatures in Northern Europe within weeks or months. This holds a potential for forecasting in that time-scale. We investigate here if the strength of the stratospheric polar vortex at the start of the forecast could be used in improving the Extended-Range temperature Forecasts of the European Centre for Medium-Range Weather Forecasts (ECMWF)~~The skill scores of the Extended-Range Forecasts (ERF) of the European Centre for Medium-Range Weather Forecasts (ECMWF) are still quite modest for the forecast weeks 3–6 in Northern Europe. As there are known stratospheric precursors impacting the surface weather with~~ and in finding periods with higher prediction skill scores. ~~potential to improve ERFs, we aim to quantify the effect of these predictors and post-process the ERFs with them.~~

~~During boreal winter the quasi-biennial oscillation (QBO) affects the stratospheric polar vortex; the easterly (westerly) QBO often coincides with weaker (stronger) than average polar vortex. Consequently, the weaker (stronger) than average stratospheric polar vortex is connected to negative (positive) Arctic Oscillation (AO) and colder (warmer) than average surface temperatures in Northern Europe.~~ For this, w~~W~~e developed a stratospheric wind indicator, *SWI*, based on the ~~previous weeks' stratospheric wind conditions~~strength of the stratospheric polar vortex and the phase of the AO during the following weeks. We demonstrate that there was a statistically significant difference in the observed surface temperature in Northern Europe within the 3–6 weeks depending on the *SWI* at the start of the forecast. ~~These temperature anomalies were underestimated by the ECMWF's reforecasts.~~

When our new *SWI* was applied in post-processing the ECMWF's two-week mean temperature reforecasts for weeks 3–4 and weeks 5–6 in Northern Europe during boreal winter, the skill scores of those weeks were slightly improved. This indicates there is some room to improve the Extended Range Forecasts~~ERFs~~, if the stratosphere-troposphere links were better captured in the modelling. In addition to this, we found that during the boreal winter, in cases the polar vortex was weak at the start of

Formatted: English (United States)

the forecast, the mean skill scores of the 3–6 weeks surface temperature forecasts were higher than average ~~in cases the polar vortex was weak at the start of the forecast~~.

1 Introduction

Extended-range forecasts (ERF; lead time up to 46 days) by dynamical models have been developed since the 1990s with the aim to fill the gap between the medium-range weather forecasts and the seasonal forecasts. It is known that ERF skills are still rather modest in forecast weeks 3–6 especially in the Northern latitudes. If the skill of the forecasts improves, ERFs have the potential to become an essential element in climate services e.g., in the form of early warnings of climatic extremes. In an academic project CLIPS (CLImate services supporting Public mobility and Safety), climatic impact outlooks and early warnings of extremes (CLIPS forecasts) were developed by employing the ERF datasets (Ervasti et al. 2018). The CLIPS forecasts were co-designed with the general public in Finland and experimented by a one year piloting phase. As many industries, e.g., energy and food production, as well as users from the general public considered they could use and would benefit from reliable ERFs (Ervasti et al. 2018), development of more skillful ERFs is clearly needed.

The European Centre for Medium-Range Weather Forecasts (ECMWF) has produced ERFs routinely since March 2002 (Vitart 2014). The verification results of the ECMWF model's ERF (Buizza and Leutbecher 2015; Vitart 2014) on a sub-continental and a regional scale (e.g., Monhart et al. 2018) demonstrated predictive skill beyond 2 weeks for temperature reforecasts over Northern Europe. ECMWF uses bias-correction of the mean in their automatic products, removing the mean bias computed from the reforecasts, depending on the time of the year (Buizza and Leutbecher 2015). We consider the bias over Northern Europe not to be dependent only on the time of the year but also on the prevailing weather pattern, and therefore, we aim to explore whether known teleconnections such as the strength of the stratospheric polar vortex and~~,~~ the phase of the Arctic Oscillation (AO)~~, and the phase of the Quasi-Biennial Oscillation (QBO)~~ could be used in improving the forecasts.

The stratospheric polar vortex is an upper-level low-pressure area that forms over both the northern and southern poles during winter due to the growing temperature gradient between the pole and the tropics. Strong westerly winds circulate the polar vortex, isolating the gradually cooling polar cap air. The strength of the northern polar vortex varies from year to year and can be indicated by, e.g., the zonal mean zonal wind (ZMZW) at 60°N and 10 hPa or polar cap temperatures. The stronger the circumpolar winds and the colder the polar cap temperatures are, the stronger is the polar vortex. Planetary waves from the troposphere disturb the northern stratospheric polar vortex, leading to meandering and weakening of the westerlies and occasionally to reverse, i.e., easterly flow (Schoeberl, 1978). This weakening of the stratospheric polar vortex also leads to warming of the polar cap temperatures, sometimes even > 30–40 K within several days. A warming of this magnitude together with a reversal of the ZMZW at 60°N and 10 hPa is commonly defined as a major sudden stratospheric warming (SSW), albeit other definitions have also been used (Butler et al. 2015).

During boreal winter the strength of the polar vortex affects the phase of the AO, which characterizes air mass flow between the Arctic and the mid-latitudes. At the surface, the AO index is affected by the strength of the polar vortex with a time lag of about two to three weeks (Baldwin and Dunkerton 1999). A strong polar vortex is characterized by lower than average surface pressure in the Arctic, positive AO index, and strong westerly winds keeping the cold Arctic air locked in the polar region and bringing milder and wetter than average weather to Northern Europe (Limpasuvan et al. 2005). In contrast, a weak polar vortex is characterized by higher than average surface pressure in the Arctic, negative AO index, and the meandering and/or weakening of the polar jet stream and tropospheric jet stream enabling cold arctic/polar air outbreaks to Northern Europe (Thompson et al. 2002, Tomassini et al. 2012).

During boreal winters, the strength of the stratospheric polar vortex influences the surface weather in the Northern Hemisphere within weeks or months (Baldwin and Dunkerton 2001, Kidston et al. 2015), hence holding a potential for forecasting in that time scale. ~~When making forecasts based on the strength of the polar vortex, a noteworthy phenomenon is also the QBO, a quasiperiodic oscillation of the equatorial zonal wind between downwards propagating easterlies and westerlies in the tropical stratosphere with a mean period of 28 to 29 months (Baldwin et al. 2001). Holton and Tan (1980) found that during the easterly QBO at level 50 hPa the polar vortex was statistically significantly weaker than during westerly QBO at the same level. Further, Scaife et al. (2014) demonstrated indicators of a more negative AO in the easterly QBO at level 30 hPa than in the westerly QBO phase at this level. There is no precise consensus of the mechanisms of this tropical-extratropical connection (Garfinkel et al. 2012), but the most common explanation is that the QBO affects the polar vortex via the Holton Tan effect: During easterly QBO, small amplitude planetary waves are reflected back towards the North Pole weakening the polar vortex (Holton and Tan 1980, 1982; Watson and Gray 2014, Gray et al. 2018).~~

However, ~~C~~challenges related to the realistic modelling of the dynamical stratosphere-troposphere coupling have been adduced, e.g., by Shepherd et al. (2018) and Polichtchouk et al. (2018). Therefore, we investigate if the known stratospheric-tropospheric connection could be used to improve the ERFs by statistical post-processing. ~~In the ECMWF model the skill of the QBO forecasts decreases substantially after the first 2 months of the forecast. This is shown to be sensitive to the parametrization of the tropical non-orographic gravity wave drag in the model (Polichtchouk et al. (2018), Polichtchouk et al. (2017)). The amplitude of the QBO tends to weaken (in the ECMWF model) through the forecast. Even though some S2S models, including the ECMWF's Integrated Forecasting System (IFS, Vitart, 2014), are already able to reproduce the QBO's effect on the polar vortex, they are still underestimating the effect on surface weather (Garfinkel et al. 2018). Furthermore, the anomalous QBO disruption in winter 2015/2016 was not forecasted by the models (Newman et al. 2016).~~

In this paper, we first verify the raw and the mean bias-corrected surface temperature reforecasts of the ECMWF's ERFs for forecast weeks 1 to 6 over Northern Europe against ERA-Interim surface temperature re-analysis (Dee et al. 2011). After that,

our aim is to find out which thresholds of the ZMZW at 60°N and 10 hPa~~stratospheric observations available at the start of the forecast~~ are followed by a statistically significantly weaker AO index. For this, we explore the observed daily AO index during boreal winters 1981–2016, 1–2 weeks, 3–4 weeks, and 5–6 weeks after different ~~phases of QBO and~~ strengths of the observed the ZMZW at 60°N and 10 hPa. ~~stratospheric zonal mean winds.~~ According to the observed daily AO index, after different thresholds of the ZMZW at 60°N and 10 hPa~~phases of QBO, and strengths of the observed stratospheric zonal mean winds~~, we define a novel Stratospheric Wind Indicator (*SWI*)~~, which is a novel indicator for the strength of the AO index in the following 1 to 6 weeks~~. For a statistically significantly weaker mean AO index, the *SWI* is defined as SWI_{neg}; otherwise, *SWI* is defined as SWI_{plain}. Further, we study the mean surface temperature anomalies observed in Northern Europe 1–2 weeks, 3–4 weeks, and 5–6 weeks after ~~strong or weak ZMZW at 60°N and 10hPa and after~~ SWI_{neg} in comparison to~~versus~~ SWI_{plain}~~.~~ ~~W~~ and we utilize these anomalies ~~SWI_{neg} and SWI_{plain}~~ in post-processing the ~~mean of the~~ temperature forecasts of the ECMWF reforecasts. Finally, we compare the *SWI* based post-processed ECMWF reforecasts with the mean bias-corrected ECMWF reforecasts. Our paper is constructed as follows: First, we present the datasets and methods. Then, we present results about the selection of the SWI_{neg} and SWI_{plain} and the skill scores of the forecasts without post-processing and with post-processing. In the Discussions and Conclusions section, we present our view on our findings and the possible next steps.

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

2 Datasets and Methods

We verified and post-processed ERFs of the ECMWF's IFS Cycle 43r1 (Vitart, 2014), which belongs to the models of the Sub-seasonal to seasonal (S2S) prediction project of the World Weather Research Program/World Climate Research Program (Vitart et al. 2017). These forecasts are run twice a week, on Mondays and Thursdays, in a horizontal resolution of 0.4 degrees. We first studied the weekly mean temperatures of the Monday runs over Northern Europe (52° N to 71.2° N and 10° E to 33.2° E) with lead times of 1 to 6 weeks, here called forecast weeks 1 to 6. We verified the 20 years × 52 weeks = 1040 reforecasts (11 members ensemble) for 1997–2016 run for the same dates as the operational forecasts, i.e., as Mondays in 2017. The weekly averages of the raw, mean bias-corrected (Section 2.2), and post-processed (Sections 2.3 and 2.4) surface temperature forecasts over Northern Europe were verified against ERA-Interim 1981–2016 temperature re-analyses (Dee et al. 2011). Years 1981–2010 of the ERA-Interim data were used as the climatological reference period and as the statistical/climatological forecast.

2.1 Skill scores of the forecasts

A commonly used measure for the probabilistic forecasts is the continuous ranked probability score (CRPS, Hersbach 2000) calculated by the following Eq. (1):

$$CRPS = \int |F(y) - F_o(y)|^2 dx, \quad (1)$$

where $F(y)$ and $F_o(y)$ are the cumulative distribution functions of the forecast and the observation, respectively.

The CRPSs were calculated by the R package 'ScoringRules' (Jordan et al. 2018) for the ECMWF's reforecast ($CRPS_{rf}$) and the climatological forecasts (ERA-Interim weekly mean temperatures in 1981–2010), which were used as the reference ($CRPS_{clim}$). As the ensemble size of the reforecasts, *m*, was only 11, and the ensemble size of the operational forecasts of the ECMWF's IFS, *M*, was 51, the expected CRPS, the $CRPS_{RF}$ of the ECMWF's reforecast was calculated for 51 members using equation 26 in Ferro et al. (2008):

$$CRPS_{RF} = \frac{m(M+1)}{M(m+1)} CRPS_{rf} \quad (2).$$

We calculated the annual means of the expected $CRPS_{RF}$ across all weeks (1 to 52) of the years 1997–2016 reforecasts. These annual means were computed separately for lead times of 1 week, 2 weeks, 3 weeks, 4 weeks, 5 weeks, and 6 weeks, here called forecast week 1, forecast week 2, forecast week 3, forecast week 4, forecast week 5, and forecast week 6, respectively. Further, the skill scores of the annual mean CPRSs, the annual mean CRPSSs, for each lead time were calculated as follows:

$$CRPSS = 1 - \frac{CRPS_{RF}}{CRPS_{clim}} \quad (3).$$

The statistical significances of each forecast week's annual mean CRPSS was determined for each grid point. The p-value with the null hypothesis that the CRPSS is zero was calculated by bootstrap resampling procedure with replacement and a sample size of 5000 for significance level 0.05.

2.2 Bias-correction of the ensemble mean

The mean bias-correction (as in Buizza and Leutbecher 2015, eq. 7a) removed the mean bias computed from the ensemble reforecasts for the 20 years (1997–2016) depending on the forecast week date. For the 1997–2016 reforecasts, the average bias was calculated considering 19 × 11 × 5 = 1045 ensemble reforecast members: 11 members' reforecast with initial dates defined by five weeks centred on the forecast week date for the 19 years reforecasts (1997–2016 excluding the reforecast year). The mean bias-corrected weekly mean temperatures were verified against the ERA-Interim data by calculating the annual mean CRPS separately for each lead time, i.e., forecast weeks 1 to 6. The skill scores of the mean bias-corrected forecasts and their statistical significance were calculated as explained in Section 2.1.

2.3 Definition of the stratospheric wind indicator (*SWI*)

As numerous observational and modelling studies have shown, the stratospheric polar vortex influences the weather in the Northern Hemisphere during boreal winter; strong polar vortex coincides more often with positive AO index and mild surface weather in Northern Europe, whereas weak polar vortex is more often followed by negative AO index and cold air outbreaks (Thompson and Wallace 1998, 2001, Kidston et al. 2015 and references therein). We aimed to find a stratospheric precursor~~s~~ for a statistically significantly weaker AO index available at the start of the forecast. The daily surface AO index were downloaded from the National Centers for Environmental Prediction (NCEP), Climate Prediction Center (CPC). This daily

AO index from the NCEP CPC is produced by projecting the daily 1000 hPa geopotential height anomalies north of 20° N onto the loading pattern of AO, which is defined as the first leading mode from the Empirical Orthogonal Function (EOF) analysis of monthly mean 1000 hPa height anomalies poleward of 20° N during 1979–2000. As a precursor~~s~~ for the AO index we used ~~two stratospheric wind data sets. The first data were~~ the daily ZMZW at 60°N and 10 hPa during 1981–2016 of the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2; Rienecker et al. 2011) reanalysis data provided by the National Aeronautics and Space Administration (NASA). ~~The second data were the monthly mean zonal wind components at levels 70 hPa, 50 hPa, 40 hPa, 30 hPa, 20 hPa, 15 hPa, and 10 hPa from the Singapore radio soundings, during 1981–2016, provided by the Free University of Berlin, representing the equatorial stratospheric monthly mean zonal wind components, the QBO (Naujokat 1986).~~

We explored the mean AO index 1 to 6 weeks after the beginning of each week in November–February (1981–2016) and~~. First, we investigated the mean AO index 1–2, 3–4, and 5–6 weeks after~~ the minimum daily ZMZW at 60°N and 10 hPa during the preceding 10 days ~~had been:~~to find a threshold for the ZMZW at 60°N and 10 hPa to be followed by statistically significantly weaker AO index 1–2, 3–4, and 5–6 weeks later.

- ~~below its overall wintertime (November–March 1981–2016) 10th percentile, corresponding a value of 3.8 ms^{-1}, indicating a weak polar vortex,~~
- ~~between its overall wintertime (November–March 1981–2016) 10th and 20th percentile corresponding to values greater than 3.8 ms^{-1} and lower than 12 ms^{-1},~~
- ~~between its overall wintertime (November–March 1981–2016) 20th and 50th percentile corresponding to values greater than 12 ms^{-1} and lower than 27 ms^{-1};~~
- ~~between its overall wintertime (November–March 1981–2016) 50th and 80th percentile corresponding to values greater than 27 ms^{-1} and lower than 41 ms^{-1}; and~~
- ~~above its overall wintertime (November–March 1981–2016) 80th percentile corresponding to values greater than 41 ms^{-1}, indicating a strong polar vortex.~~

~~In the cases the ZMZW at 60°N and 10 hPa was between 3.8 ms^{-1} and 41 ms^{-1} during the preceding 10 days, we explored the mean AO index 1–2, 3–4, and 5–6 weeks after following predictors:~~

- ~~westerly QBO at 30 hPa, the *WQBO*,~~
- ~~easterly QBO at 30 hPa, the *EQBO*,~~
- ~~*EQBO* with the maximum of the monthly mean zonal wind components of the QBO between 70 hPa and 10hPa restricted to 7ms^{-1}, 10ms^{-1}, and 13ms^{-1};~~

The statistical significance of the difference between the AO index following the different thresholds of the ZMZW at 60°N and 10 hPa ~~two different stratospheric situations, e.g., the *EQBO* and the *WQBO*~~, was determined using a two-sided Student's

Formatted: Normal, No bullets or numbering

t-test with the null hypothesis that there is no difference. The threshold of the ZMZW at 60°N and 10 hPa for statistically significantly ~~predictors for~~ weaker (at the 99% confidence level) AO index observed 1–2 weeks, 3–4 weeks, and 5–6 weeks later~~after these stratospheric situations~~, was~~were~~ used to define the~~a~~ *SWI* as follows:~~-~~ below the threshold the *SWI* was defined negative,~~to be~~ SWI_{neg}~~; otherwise~~, and above the threshold the *SWI* ~~it~~ was defined as plain, SWI_{plain} ~~for the beginning of each winter week (November–February) in 1981–2016~~.

Formatted: Font: Italic

Formatted: Font: Italic

2.4 Utilizing the stratospheric winds indicator (*SWI*) in forecasting

In this section, we investigated the observed and reforecasted surface temperature anomalies 1–2 weeks, 3–4 weeks, and 5–6 weeks after ~~weak/strong ZMZW at 60°N and 10 hPa and~~ SWI_{neg}/SWI_{plain} defined in Section 2.3. First, we calculated the two-week mean temperature anomalies of the ERA-Interim reanalyses (Dee et al. 2011) of the 1–2 weeks, the 3–4 weeks, and the 5–6 weeks from the beginning of each week in January, February, November, and December in 1981–2016 in Northern Europe. Subsequently, we divided the observed two-week mean temperature anomalies to sets of anomalies, representing ~~weak ZMZW at 60°N and 10 hPa (below 3.8 ms^{-1}), strong ZMZW at 60°N and 10 hPa (above 41 ms^{-1}),~~ SWI_{neg}, and SWI_{plain} according to the ~~previous weeks' stratospheric wind conditions~~minimum ZMZW at 60°N and 10 hPa during the preceding 10 days. Thereafter, we determined the statistical significance of the difference between the surface temperatures after SWI_{neg} and SWI_{plain} ~~(and after weak/strong ZMZW at 60°N and 10 hPa in comparison to the rest of the cases)~~ using a two-sided Student's t-test with the null hypothesis that there is no difference between SWI_{neg} and SWI_{plain} ~~(and no difference between weak/strong ZMZW at 60°N and 10 hPa and the rest of the cases)~~. This same procedure to define the difference between the surface temperatures after SWI_{neg} and SWI_{plain} ~~(and weak and strong ZMZW at 60°N and 10 hPa)~~ was used for the ERA-Interim reanalyses for the period 1997–2016 to see how the selection of a shorter period affects the temperature anomalies. Further, the mean surface temperature anomalies 1–2 weeks, 3–4 weeks, and 5–6 weeks after SWI_{neg} and SWI_{plain} ~~(and after weak and strong ZMZW at 60°N and 10 hPa)~~ in the ECMWF reforecasts run in the beginning of each week in November–February 1997–2016 were calculated to examine how the model reproduced the anomalies.

For post-processing the ECMWF reforecasts ~~in cases the ZMZW at 60°N and 10 hPa was between 3.8 ms^{-1} and 41ms^{-1} at the start of the forecast,~~ we calculated TA_{SWIneg} and $TA_{SWIplain}$ representing mean temperature anomalies in November–February 1981–2016 after SWI_{neg} and SWI_{plain}, respectively. The ~~means,~~ TA_{SWIneg} and the $TA_{SWIplain}$ ~~representing SWI_{neg} and SWI_{plain}~~ were calculated separately for each 0.4° × 0.4° grid point over Northern Europe.

For the post-processing of the ECMWF reforecasts, we first defined the *SWI* either SWI_{neg} or SWI_{plain} at the start of the forecast according to the minimum ZMZW at 60°N and 10 hPa during the preceding 10 days~~previous weeks' stratospheric wind conditions~~. According to the *SWI*, we added either TA_{SWIneg} or $TA_{SWIplain}$ to the ERA-Interim mean temperature during 1981–2016, corresponding to forecast weeks 1–2, 3–4, and 5–6 to get a SWI_{neg} and SWI_{plain} based mean temperatures, T_{SWIneg} and

$T_{SWIplain}$, for weeks 1–2, 3–4, and 5–6, respectively. The T_{SWIneg} and $T_{SWIplain}$ were used in post-processing the ECMWF reforecasts' mean bias-corrected ensemble members, T_{BC}, by calculating a weighted average, T_{SWI_BC}, for SWI_{neg} as follows:

$$T_{SWI\_BC} = (1 - k_{SWI}) * T_{BC} + k_{SWI} * T_{SWIneg} \quad (4)$$

And for SWI_{plain},

$$T_{SWI\_BC} = (1 - k_{SWI}) * T_{BC} + k_{SWI} * T_{SWIplain} \quad (5)$$

where T_{SWI_BC} was a post-processed ensemble member. k_{SWI} was the weight of the T_{SWIneg} or $T_{SWIplain}$, which was tested between 0–1 and defined according to the best improvement in the skill scores of the post-processed forecast. By Eq. (4) and Eq. (5), we adjusted each ensemble member with the same weight, and hence, the original spread of the ECMWF reforecasts remained unchanged. The skill scores of the *SWI* based post-processed forecasts, and their statistical significance, were calculated as explained in Section 2.1.

3 Results

3.1 Skill scores of the forecasts

The annual mean of the expected CRPSS and its 95% level of confidence of the raw and the mean bias-corrected (Section 2.2) weekly mean temperature of the ECMWF reforecasts for 1997–2016 are displayed in Figure 1. In grid points where the CRPSS was higher than zero and the confidence level was higher than 95% (dotted areas), the reforecasts were statistically significantly better than just the statistical forecast based on 1981–2010 climatology. Figure 1 illustrates that for forecast weeks 1–6 the mean bias-corrected ERF reforecasts were on average significantly better than climatology. The annual mean CRPSS values show that in forecast weeks 1–3 the CRPSSs are for the most part above 0.1, whereas on in forecast weeks 4–6 they are mostly lower, between 0 and 0.1.

3.2 The stratospheric observations and the thereafter observed AO index and surface temperature

Figure 2 ~~and Fig. 3~~ shows boxplots of the observed mean of the daily AO index 1–2 weeks, 3–4 weeks, and 5–6 weeks after different strengths of the ZMZW at 60°N and 10 hPa ~~(Fig. 2) and after different phases of QBO and restrictions in the strength of the stratospheric winds in 1981–2016 (Fig. 3)~~. In Fig. 2 the first box (brown) represents the mean AO index after all the cases in 1981–2016 November–February, i.e., 36 years * 17 weeks=612 cases. The blue, yellow, and red boxes ~~2nd box (blue)~~ in Fig. 2 show~~s~~ the mean AO index after cases in which~~where~~ the daily ZMZW at 60°N and 10 hPa was during the preceding 10 days below its 10th (2.5ms^{-1}), 15th (6.7ms^{-1}), and 20th (10ms^{-1}) percentile, respectively ~~(3.8ms^{-1}) during the preceding 10 days, corresponding to cases with a weak polar vortex~~. The observed mean AO index was statistically significantly weaker at the 99% confidence level 1–2 weeks, 3–4 weeks, and 5–6 weeks after the daily ZMZW at 60°N and 10 hPa had been below its overall wintertime 150th percentile, 6.7ms^{-1} ~~(indicating a weak polar vortex)~~. ~~The third, fourth, fifth, and sixth (green,~~

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

yellow and purple, red) boxes in Fig. 2 represent the mean AO index after cases the daily ZMZW at 60°N and 10 hPa was between its 10th and 20th percentile (between 4 ms^{-1} and 12 ms^{-1}), 20th and 50th percentile (between 12 and 27 ms^{-1}), 50th and 80th percentile (between 27 ms^{-1} and 41 ms^{-1}), and above 41 ms^{-1}, respectively. In these cases the AO index seems to be the statistically significantly higher 1–2 weeks (Fig. 2a) after the ZMZW at 60°N and 10 hPa was between 27 ms^{-1} and 41 ms^{-1} and above 41 ms^{-1}.

In Fig. 3 the first box (light brown) represents the mean AO index after all the cases in 1981–2016 November–February, with ZMZW at 60°N and 10 hPa between 3.8 ms^{-1} and 41 ms^{-1}. The second and third boxes show the mean AO index after easterly (*EQBO*, blue) and westerly (*WQBO*, pink) QBO at the 30 hPa level, respectively, in cases the ZMZW at 60°N and 10 hPa was between 3.8 ms^{-1} and 41 ms^{-1}. The p-value written below each boxplot pair indicates the likelihood of such a pair of distributions arising from a random sampling of a single distribution as given by a Student's t-test, i.e., p-values less than 0.05 indicate that the means of the data sets differ significantly at the 95% level of confidence. The median and the mean of the mean AO index 5–6 weeks after *EQBO* were statistically significantly lower than after *WQBO*. The *EQBO* (blue) box shows all the cases of *EQBO* with no restriction in the QBO's monthly mean zonal wind components, whereas the fourth, the sixth, and the eighth blueish boxes show the mean AO index after *EQBO* with all the QBO's monthly mean zonal wind components between levels 70…10hPa being below 13 ms^{-1}, 10 ms^{-1}, and 7 ms^{-1}, respectively. Restricting the *EQBO* cases by a maximum of the QBO's monthly mean zonal wind components in levels 70…10hPa being below 13 ms^{-1} or 10 ms^{-1} led to a statistically significantly lower mean AO during the following 1–2, 3–4, and 5–6 weeks in comparison to the rest of the cases.

Aiming to select stratospheric precursors indicating weak AO index, Based on this, we defined the *SWI* to be negative (positive) and to indicate statistically significantly lower (higher) AO index in cases the minimum ZWZW at 60°N and 10 hPa was below (above) its 15th percentile, 6.7ms^{-1}, during the preceding 10 days. when the QBO was easterly at 30 hPa and the QBO's monthly mean zonal wind components in levels 70…10hPa were weaker than 10ms^{-1} and the minimum of the daily ZMZW at 60°N and 10 hPa during the 10 last days was between its overall wintertime 10th and 80th percentile, i.e., above 3.8 ms^{-1} and below 41 ms^{-1}. In other cases, when the minimum of the daily ZMZW at 60°N and 10 hPa during the 10 last days month was above 3.8 ms^{-1} and below 41 ms^{-1}, the *SWI* was defined as plain. This decision tree for the *SWI* is depicted in Fig. 4.

Formatted: Superscript

Formatted: Superscript

Figure 35 shows the ERA-Interim (1981–2016 and 1997–2016) and model forecasted mean temperature anomalies 1–6 weeks after SWI_{neg} and SWI_{plain} weak (below 3.8 ms^{-1}) and strong (higher than 41 ms^{-1}) ZMZW at 60°N and 10 hPa. Cases with ZMZW at 60°N and 10 hPa weaker than 6.7ms^{-1} 3.8 ms^{-1}, i.e., SWI_{neg}, in Fig. 35a–3c and Fig 53g–3i (stronger than 6.7ms^{-1}, i.e., SWI_{plain}, 41 ms^{-1} in Fig. 35d–3f and Fig. 35j–l) were on average followed by colder (warmer) than average mean temperature. The ECMWF reforecasts (Fig. 53-m–r) capture these mean anomalies clearly, in some areas even too strong in comparison to the ERA-Interim 1997–2016 (Fig. 3g–l).

~~Figure 6 shows the ERA-Interim (periods 1981–2016 and 1997–2016) and model forecasted (the period 1997–2016) mean temperature anomalies of the weeks' 1–2, 3–4, and 5–6 in November–February after SWI_{neg} and SWI_{plain}. The ERA-Interim showed on average lower mean temperatures for the weeks' 1–2, 3–4 and 5–6 after SWI_{neg} (Fig. 6a–c and 6g–i). The ECMWF reforecasts also showed cold anomalies after SWI_{neg} for forecast weeks 1–2 (Fig. 6m) but weaker than the observed ones. For forecast weeks 3–6 (Fig. 6n–o) there was no sign of cold anomalies in the mean of the reforecast. Further, the ERA-Interim reanalyses showed on average higher mean temperatures for weeks 1–2, 3–4, and 5–6 after SWI_{plain} (Fig. 6d–f and 6j–l). This warm anomaly was well forecasted in the forecasts weeks 1–2 (Fig. 6p), and it was partly but weaker forecasted in forecast weeks 3–4 (Fig. 6q) and 5–6 (Fig. 6r). The mean temperature anomalies 1–6 weeks after SWI_{neg} (Fig. 6a–c) and SWI_{plain} (6d–f) during 1981–2016 were statistically significantly different using a Student's t-test, with anomalously cold surface temperatures more common 1-6 weeks after SWI_{neg}. When examining the years 1997–2016 (Fig. 6g–i and Fig. 6j–l), which was the reforecast period, the temperature anomalies were mostly of the same sign than during the longer 1981–2016 period (Fig. 6a–c and Fig. 6d–f).~~

3.3 The *SWI* and the forecasted mean temperatures

The mean temperature anomalies in Fig. 3~~6~~(a–f) for Northern Europe were used for the *SWI* based post-processing as described in Section 2.4. ~~The CRPSS of the mean temperature of the forecast weeks 1–2 were not improved by the *SWI* (no figure), whereas t~~Figure 4 shows how the post-processing based on the *SWI* affected the forecasting skill scores, in cases of SWI_{neg} and SWI_{plain}. The CRPSSs of the mean temperatures of the forecast weeks 3–4 and 5–6 were improved by the *SWI* based post-processing ~~(Fig. 7a and 7b). T~~and the best median CRPSS was achieved by k_{SWI}=0.34, for forecast weeks 3–4 in cases of SWIplain, and by k_{SWI}=0.~~67~~ for all the other cases ~~forecast weeks 5–6~~.

Figure ~~8~~5a–f shows the forecast skill of the mean bias-corrected mean temperature reforecasts of forecast weeks 3–4 and 5–6 in all cases (Fig. 5a and 5c), in cases the ZMZW at 60°N and 10 hPa was ~~weak (~~below 6.7~~3.8~~ ms^{-1} at the start of the forecast~~,~~ (~~SWI_{neg}~~, Fig. ~~8~~5e and g~~a–b~~), ~~strong (above 41 ms^{-1}, Fig. 8c–d), and between 3.8 ms^{-1} and 41ms^{-1} (Fig. 8e–f)~~ and in cases the ZMZW at 60°N and 10 hPa was above 6.7 ms^{-1} at the start of the forecast (~~SWI_{plain}~~, Fig. 5i and 5k). In cases of ~~the~~ weak ZMZW at 60°N and 10 hPa at the start of the forecast (Fig. 5e and g~~8a–b~~) the CRPSSs of forecast weeks 3–4 and 5–6 reached even higher than 0.4 values in some areas, indicating there higher predictability in comparison to~~than~~ cases in which the~~with~~ ZMZW at 60°N and 10 hPa was stronger than 6.7~~3.8~~ ms^{-1} at the start of the forecast (Fig. 5i and 5k~~8c–f~~). ~~Also in cases the ZMZW at 60°N and 10 hPa was strong (higher than 41 ms^{-1}) at the start of the forecast, in forecast weeks 5–6 the mean CRPSS reached in some areas 0.2 (Fig. 8d), showing higher predictability than in cases the ZMZW at 60°N and 10 hPa was between 3.8 ms^{-1} and 41 ms^{-1} (see Fig. 8f).~~

Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic, Subscript
Formatted: Font: Italic
Formatted: Font: Italic, Subscript
Formatted: Font: Italic
Formatted: Font: Italic, Subscript

Figures 8e–f show the forecasts skill of the mean temperature of the forecast weeks 3–4 and weeks 5–6 forecasted by the mean bias-corrected reforecasts alone in cases the ZMZW at 60ºN and 10 hPa was between 3.8 ms^{-1} and 41 ms^{-1} at the start of the forecast. Figure 5 8g and Fig. 8h depicts also the mean CRPSS of the mean bias-corrected and *SWI* post-processed reforecasts in all cases (Fig. 5b and 5d), in cases the ZMZW at 60°N and 10 hPa was below 6.7 ms^{-1} between 3.8 ms^{-1} and 41 ms^{-1} (SWI_{neg}) at the start of the forecast (k_{SWI} =0.64 and k_{SWI} =0.67 in Fig. 5e and Fig. 5g Fig. 8g and Fig 8h, respectively), and in cases the ZMZW at 60°N and 10 hPa was above 6.7 ms^{-1}, (SWI_{plain}) at the start of the forecast (k_{SWI} =0.3 and k_{SWI} =0.6 in Fig. 5j and Fig. 5l, respectively). In comparison to the only mean bias corrected ECMWF reforecasts (see Fig. 5a, 5c, 5e, 5g, 5i, 5k), adding By using the *SWI* based post-processing to the ECMWF reforecasts (see Fig. 5b, 5d, 5f, 5h, 5j), the CRPSSs for weeks 3–4 and weeks 5–6 were slightly improved and the area of these forecasts being significantly better than just the climatological forecast was expanded.

4 Discussion and Conclusions

Based on ECMWF's extended-range reforecasts for the period 1997–2016, we found that the weekly mean surface temperature forecasts over Northern Europe were on average significantly better than just the climatological forecast in weeks 1–6, however, in weeks 4–6, the CRPSSs were quite low, mostly between 0 and 0.1.

We studied the mean AO index after different thresholds of ZMZW at 60°N and 10 hPa. We found that the mean AO index was statistically significantly weaker 1–2 weeks, 3–4 weeks, and 5–6 weeks after the daily ZMZW at 60°N and 10 hPa had been below its November-February their overall wintertime 10th 15th percentile, 6.7 3.8 ms^{-1} (indicating a weak polar vortex).

We separated the weak (below 3.8 ms^{-1}) and strong (above 41 ms^{-1}) ZMZW at 60°N and 10 hPa cases and investigated the rest of the data based on the QBO. We showed that in addition to the previously demonstrated more negative AO during easterly QBO in comparison to westerly QBO at 30 hPa (Scaife et al. 2014), the mean AO index was sensitive to the maximum strength of the QBO's monthly mean zonal wind components in levels 70…10hPa during the easterly QBO at 30 hPa. Based on observations, we found that the mean AO index was statistically significantly weaker 1–2 weeks, 3–4 weeks, and 5–6 weeks after the monthly mean QBO was easterly at 30 hPa, and all the QBO's monthly mean zonal wind components in levels 70…10hPa were less than 10 ms^{-1}. Selecting the SWI_{neg} to include cases where the QBO was easterly at 30 hPa and all the QBO's monthly mean zonal wind components in levels 70…10hPa were less than 10 ms^{-1} resulted in a statistically significantly weaker AO index within the following 1–6 weeks in comparison to the rest of the data, defined as SWI_{plain}. Cases preceded by weaker (stronger) than 6.7 ms^{-1} ZMZW at 60°N and 10 hPa were defined as SWI_{neg} (SWI_{plain}). As negative AO index enables cold air outbreaks to Northern Europe (Thompson et al. 2002, Tomassini et al. 2012) and positive AO index tends to bring milder and wetter than average weather to Northern Europe (Limpasuvan et al. 2005), we investigated how the mean surface temperatures were in November-February (1981–2016) in Northern Europe 1–6 weeks after weak/strong ZMZW

~~at 60°N and 10 hPa and~~ after SWI_{neg}/SWI_{plain}. We found that the mean surface temperature anomalies in Northern Europe in November–February in 1981–2016 after ~~weak and strong ZMZW at 60°N and 10 hPa (~~SWI_{neg} and SWI_{plain}~~)~~ were in many places~~mostly~~ statistically significantly different, with anomalously cold surface temperatures more common 1–6 weeks after ~~weak ZMZW at 60°N and 10 hPa (~~SWI_{neg}~~)~~. The mean temperature anomalies corresponding to SWI_{neg}/SWI_{plain} were used in post processing the ECMWF's mean temperature reforecast for weeks 3–4 and 5–6 in Northern Europe during boreal winter, and thereby, those weeks' forecast skills were slightly improved.

We also investigated the forecast skill in cases of ~~strong or weak~~ ZMZW at 60°N and 10 hPa below or above the threshold of 6.7 ms^{-1}. We found that cases of weaker than 6.7 ms^{-1} ZMZW at 60°N and 10 hPa at the start of the forecast ~~led to~~were followed by higher than average forecasting skill scores~~predictability~~ of mean surface temperature ~~anomalies~~ for forecast weeks 3–4 and 5–6. Also earlier studies have reported enhanced forecast skill during periods of negative AO, e.g., in 500 hPa geopotential height forecasts in the northern mid-latitudes in both medium-range (Langland and Maue 2012) extended range (Minami and Takaya 2020).~~This could be valuable information for both forecasters and customers and should be further researched.~~

~~This study demonstrates that the QBO-polar vortex connection should be better integrated into the extended-range surface temperature forecasts over Northern Europe. T~~In future the *SWI* based post-processing method introduced in this paper could also be tested for other northern areas affected by the polar vortex and to precipitation and windiness forecasts, and it could be further developed by, e.g., the Madden-Julian-Oscillation (Madden and Julian 1994; Zhang 2005; Jiang et al. 2017; Vitart 2017; Vitart and Molteni 2010; Robertson et al. 2018, Cassou 2008) and the quasi-biennial oscillation (Watson and Gray 2014; Scaife et al. 2014; Garfinkel et al. 2018; Gray et al. 2018). In this study, the effect of global warming was not filtered from the temperature anomalies used for statistical post-processing. In future work, the impact of filtering the effect of global warming could be tested. Moreover, the next step would be looking for the stratospheric signal from the forecast model.

Data availability. ERA-Interim data available at https://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/ (last accessed 24 June 2019). ECMWF reforecasts data available at https://apps.ecmwf.int/mars-catalogue/ (last accessed 28 June 2019). AO index data available at https://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_ao_index/ao.shtml (last accessed 24 June 2019). The daily ZMZW at 60°N and 10 hPa data available at https://acd-ext.gsfc.nasa.gov/Data_services/met/ann_data.html (last accessed 24 June 2019). ~~The QBO data data available at https://www.geo.fu-berlin.de/met/ag/strat/produkte/qbo/qbo.dat (last accessed 24 June 2019).~~ The data of Figures 1~~–3 and~~ 5~~–8~~ available at https://etsin.fairdata.fi/dataset/34d0f8b3-a593-46aa-8fcf-358d72f6cac1.
~~https://etsin.fairdata.fi/dataset/9f35af27-c3bb-4115-9e31-a9ef339b10ed.~~

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

Author contributions. NK designed the study, analysed the results and prepared the manuscript with contributions from all co-authors. OH participated in the study design and analysing the results. MK contributed to the discussions and fine-tuned the experiments. DSR contributed to the discussions and to the interpretation of the results. HJ provided supervision during the experiments and writing. HG contributed to the study design and was in charge of the management and the acquisition of the financial support for the CLIPS-project leading to this publication.

Acknowledgements. We wish to thank Academy of Finland for funding the project (number 303951 SA CLIPS). We also acknowledge the ECMWF for monthly forecast data and ERA-Interim data, NOAA/CPC for providing the AO index data, and NASA for providing 10hPa wind data~~, and Free University of Berlin for providing the QBO data~~. We thank the CLIPS team and developers of the R cran calculation package 'ScoringRules'. We thank the three anonymous reviewers for their good and constructive comments.

References

Baldwin, M. P., and Dunkerton, T.J.: Propagation of the Arctic Oscillation from the stratosphere to the troposphere, J. Geophys. Res, 104, D24, 30937-30946, https://doi.org/10.1029/1999JD900445, 1999.

Baldwin, M. P., and Dunkerton, T. J.: Stratospheric harbingers of anomalous weather regimes. Science, 294, 581–584, doi:10.1126/science.1063315, 2001.

~~Baldwin, M. P., Gray, L. J., Dunkerton, T. J., Hamilton, K., Haynes, P. H., Randel, W. J., et al.: The quasi-biennial oscillation, Rev. Geophys., 39(2), 179–229, 2001.~~

Buizza, R. and Leutbecher, M.: The forecast skill horizon, Q. J. R. Meteorol. Soc., 141, 3366–3382, doi:10.1002/qj.2619, 2015.

Butler, A. H., Seidel, D. J., Hardiman, S. C., Butchart, N., Birner, T., and Match, A.: Defining sudden stratospheric warmings, Bull. American Meteor. Soc., 96, 1913-1928, doi: http://dx.doi.org/10.1175/BAMS-D-13-00173.1, 2015.

Cassou C.: Intraseasonal interaction between the Madden–Julian Oscillation and the North Atlantic Oscillation. Nature, 455, 523–527, 2008.

Chambers, J. M., Cleveland, W. S., Kleiner, B, and Tukey, P.A.: Graphical Methods for Data Analysis, The Wadsworth statistics/probability series. Wadsworth and Brooks/Cole, Pacific Grove, CA, 1983.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. Roy. Meteorol. Soc., 137, 553–597, 2011.

Minami, A., Takaya, Y.; (2020) Enhanced Northern Hemisphere Correlation Skill of Subseasonal Predictions in the Strong Negative Phase of the Arctic Oscillation. Journal of Geophysical Research: Atmospheres 125:10.

Ervasti, T., Gregow, H., Vajda, A., Laurila, T. K., and Mäkelä, A.: Mapping users' expectations regarding extended-range forecasts. Adv. Sci. Res., 15, 99–106, doi: 10.5194/asr-15-99-2018, 2018.

Ferro C. A. T., Richardson, D. S., and Weigel, A. P.: On the effect of ensemble size on the discrete and continuous ranked probability scores. Meteorol. Appl. 15: 19–24, doi: 10.1002/met.45, 2008.

Garfinkel, C. I., Shaw, T. A., Hartmann, D. L., and Waugh, D. W.: Does the Holton-Tan mechanism explain how the quasi-biennial oscillation modulates the arctic polar vortex? Journal of the Atmospheric Sciences, 69(5), 1713-1733. https://doi.org/10.1175/JAS-D-11-0209.1, 2012.

Garfinkel, C. I., Schwartz, C., Domeisen, D. I. P., Son, S-W., Butler, A. H., and White, I. P.: Extratropical stratospheric predictability from the Quasi-Biennial Oscillation in Subseasonal forecast models, J. Geophys. Res: Atmospheres, doi: 10.1029/2018JD028724, 2018.

Hersbach, H.: Decomposition of the continuous ranked probability score for ensemble prediction systems. Wea. Forecasting, 15, 559–570, doi:10.1175/1520-0434(2000)015,0559: DOTCRP.2.0.CO;2, 2000.

Gray, L. J., Anstey, J. A., Kawatani, Y., Lu, H., Osprey, S., and Schenzinger, V.: Surface impacts of the Quasi Biennial Oscillation, Atmos. Chem. Phys., 18, 8227–8247, https://doi.org/10.5194/acp-18-8227-2018, 2018.

Holton, J. R. and Tan, H. C.: The influence of the equatorial quasi-biennial oscillation on the global circulation at 50mb, J. Atmos. Sci., 37, 2200–2208, 1980.

Holton, J. R. and Tan, H. C.: The quasi-biennial oscillation in the Northern Hemisphere lower stratosphere, J. Meteor. Soc. Japan, 60, 140–148, 1982.

Kidston, J.; Scaife, A. A.; Hardiman, S. C.; Mitchell, D. M.; Butchart, N.; Baldwin, M. P., and Gray, L. J.: Stratospheric influence on tropospheric jet streams, storm tracks and surface weather. Nature Geoscience, 8(6), 433–440, 2015.

Langland, R. H. & Maue, R. N.: (2012) Recent Northern Hemisphere mid-latitude medium-range deterministic forecast skill, Tellus A: Dynamic Meteorology and Oceanography, 64:1, DOI: 10.3402/tellusa.v64i0.17531

Limpasuvan, V., Hartmann, D. L., Thompson, D. W. J., Jeev, K., and Yung, Y. L.: Stratosphere-troposphere evolution during polar vortex intensification. J. Geophys. Res., 110, D24101, doi: 10.1029/2005JD006302, 2005.

Madden, R. A., and Julian, P. R.: Observations of the 40–50-day tropical oscillation—A review. Mon. Wea. Rev., 122, 814–837, 1994.

Jiang, Z., Feldstein, S. B., and Lee S.: The relationship between the Madden–Julian Oscillation and the North Atlantic Oscillation. Q. J. R. Meteorol.Soc. 143: 240–250, January 2017 A DOI:10.1002/qj.2917, 2017.

Monhart, S., Spirig, C., Bhend, J., Bogner, K., Schär, C., and Liniger, M. A.: Skill of subseasonal forecasts in Europe: Effect of bias correction and downscaling using surface observations. J. Geophys. Res: Atmospheres, 123, 7999–8016. 2018.

Jordan A., Krueger F., and Lerch, S.: Evaluating Probabilistic Forecasts with scoringRules. Journal of Statistical Software, forthcoming, 2018.

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Naujokat, B.: An update of the observed quasi-biennial oscillation of the stratospheric winds over the tropics. J. Atmos. Sci., 43, 1873–1877. 1986.

Newman, P. A., L. Coy, S. Pawson, and L. R. Lait: The anomalous change in the QBO in 2015–2016, Geophys. Res. Lett., 43, 8791–8797, 2016.

Polichtchouk, I., et al., 2017: What influences the middle atmosphere circulation in the IFS? ECMWF Technical Memorandum No. 809.

Polichtchouk, I., Shepherd, T. G., Byrne, N. J.: Impact of Parametrized Nonorographic Gravity Wave Drag on Stratosphere-Troposphere Coupling in the Northern and Southern Hemispheres, Geophys. Res. Lett., 45, 8612-8618, doi: 10.1029/2018gl078981, 2018.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Emily Liu, J. B., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G. K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., Woollen, J.: MERRA: NASA's modern-era retrospective analysis for research and applications. J. Clim. 24: 3624–3648. https://doi.org/10.1175/JCLI-D-11-00015.1, 2011.

Robertson, A. W., Camargo, S. J., Sobel, A., Vitart, F., and Wang, S.: Summary of workshop on sub-seasonal to seasonal predictability of extreme weather and climate. npj Climate and Atmospheric Science, 1, 8, doi: 10.1038/s41612-017-0009-1, 2018.

Scaife, A. A., Athanassiadou, M., Andrews, M., Arribas, A., Baldwin, M., Dunstone, N., Knight, J., MacLachlan, C., Manzini, E., Müller, W. A., Holger Pohlmann, H., Smith, D., Stockdale, T., Williams, A.:et al.: Predictability of the quasi-biennial oscillation and its northern winter teleconnection on seasonal to decadal timescales, Geophys. Res. Lett., 41, 1752–1758, doi:10.1002/2013GL059160, 2014.

Shepherd T. G., Polichtchouk, I., Hogan, R., Simmons, A. J.: Report on Stratosphere Task Force, ECMWF Technical Memorandum n. 824, doi: 10.21957/0vkp0t1xx, 2018.

Thompson, D. W. J. and Wallace, J. M.: The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. Geophys. Res. Lett., 25, 1297– 1301, 1998.

Schoeberl, M. R.: Stratospheric warmings: Observations and theory. Rev. Geophys., 16, 521–538, 1978.

Thompson, D. W. J., Baldwin, M. P. and Wallace J. M.: Stratospheric connection to Northern Hemisphere wintertime weather: implications for prediction. J. Clim. 15, 1421–1428, 2002.

Thompson, D. W. J., and Wallace, J. M.: Regional Climate Impacts of the Northern Hemisphere Annular Mode. Science, 293, 85–89, 2001.

Tomassini, L., Gerber, E. P., Baldwin, M. P., Bunzel, F. and Giorgetta, M.: The role of stratosphere troposphere coupling in the occurrence of extreme winter cold spells over northern Europe. J. Adv. Model. Earth Syst., 4, M00A03, 2012.

Vitart F., and Molteni F.: Simulation of the MJO and its teleconnections in the ECMWF forecast system. Q. J. R. Meteorol. Soc., 136, 842–855, 2010.

Vitart F.: Evolution of ECMWF sub-seasonal forecast skill scores. Q. J. R. Meteorol. Soc., 140, 1889–1899, doi: 10.1002/qj.2256, 2014.

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., Déqué, M., Ferranti, L., Fucile, E., Fuentes, M., Hendon, H., Hodgson, J., Kang, H., Kumar, A., Lin, H., Liu, G., Liu, X., Malguzzi, P., Mallas, I., Manoussakis, M., Mastrangelo, D., MacLachlan, C., McLean, P., Minami, A., Mladek, R., Nakazawa, T., Najm, S., Nie, Y., Rixen, M., Robertson, A. W., Ruti, P., Sun, C., Takaya, Y., Tolstykh, M., Venuti, F., Waliser, D., Woolnough, S., Wu, T., Won, D., Xiao, H., Zaripov, R., and Zhang L.: The Subseasonal to Seasonal (S2S) Prediction Project Database. Bull. Amer. Meteor. Soc., 98, 163–173, https://doi.org/10.1175/BAMS-D-16-0017.1, 2017.

Vitart, F.: Madden-Julian Oscillation prediction and teleconnections in the S2S database: MJO prediction and teleconnections in the S2S database. Q. J. R. Meteor. Soc., 143, 2210–2220, 2017.

Watson, P. A., and L. J. Gray: How Does the Quasi-Biennial Oscillation Affect the Stratospheric Polar Vortex? J. Atmos. Sci.,71, 391–409, doi: 10.1175/JAS-D-13-096.1, 2014.

Zhang, C.: Madden-Julian Oscillation. Rev. Geophys., 43, RG2003, doi:10.1029/2004RG000158, 2005.

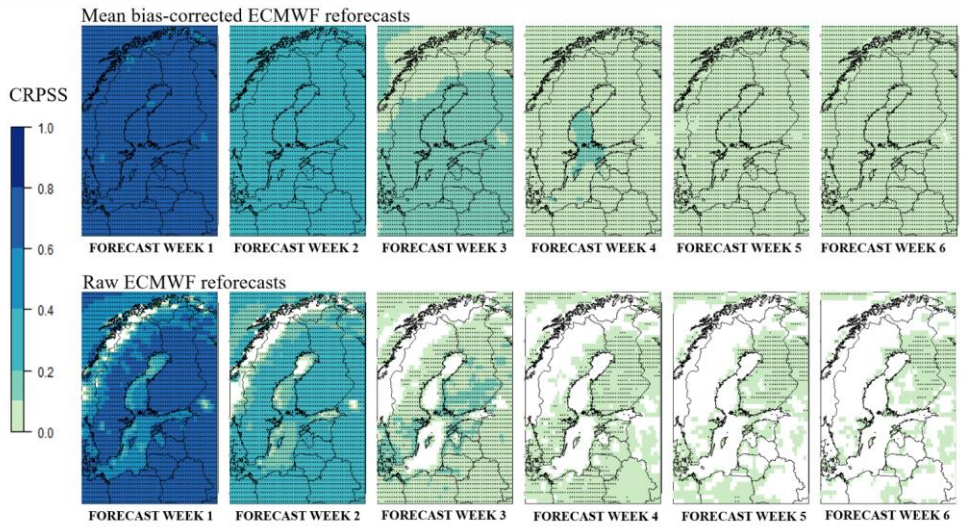


Figure 1: Annual mean of the expected CRPSS of the weekly mean temperature of the mean bias-corrected (upper row) and raw (lower row) ECMWF reforecasts for years 1997–2016 using ERA-Interim climatology of 1981–2010 as the reference. The dotted areas represent the 95% level of confidence that the CRPSS is above zero.

Formatted: Normal, Line spacing: single

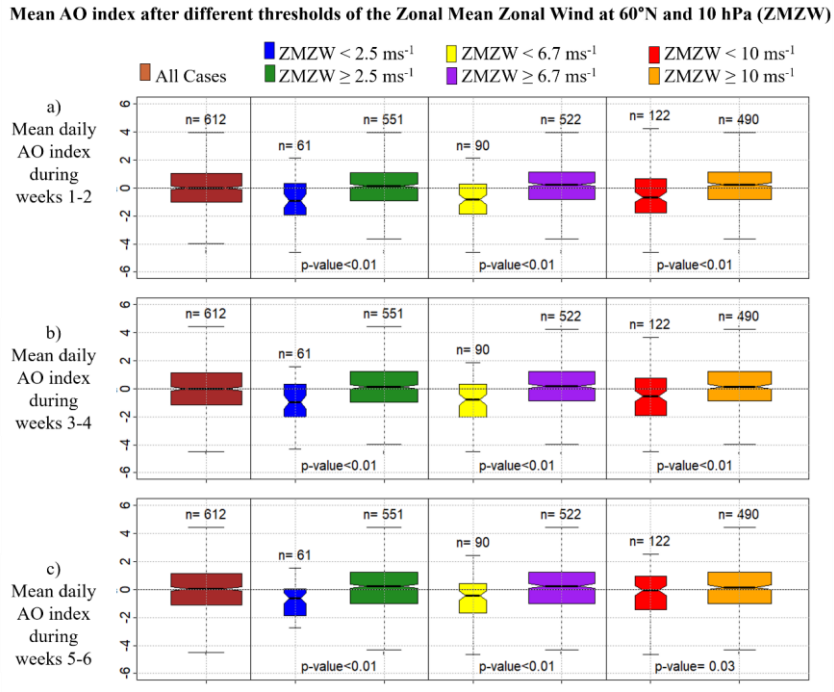


Figure 2: Observed mean AO index in November-March (1981-2016) a) 1–2, b) 3–4 and c) 5–6 weeks after different thresholds of the zonal mean zonal wind at 60°N and 10 hPa (ZMZW). The horizontal line dividing each box into two parts shows the median of the data, the ends of the box show the lower and upper quartiles, and the whiskers represent the highest and the lowest values excluding outliers. The n written above each box indicates the number of observations in each group. The p-value written below each boxplot pair indicates the likelihood of such a pair of distributions arising from a random sampling of a single distribution as given by a Student's t-test, i.e., p-values less than 0.01 indicate that the means of the data sets differ significantly at the 99% level of confidence. The notches of each side of the boxes were calculated by R boxplot.stats. If the notches of two plots do not overlap, this is 'strong evidence' that the two medians differ (Chambers et al., 1983, p. 62).

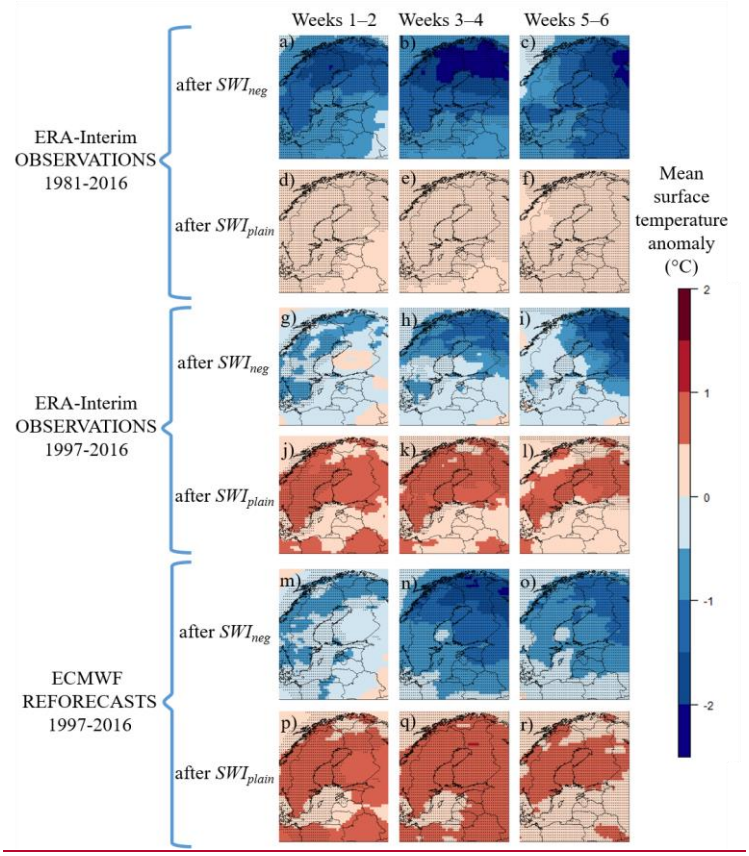


Figure 3. ERA-Interim observed (a-l) and ECMWF reforecasted (m-r) mean temperature anomalies in comparison to the 1981-2016 mean during boreal winters (November-February) in cases the previous week's *SWI* was negative (SWI_{neg}, covering about 17% of the winter weeks) or plain (SWI_{plain}, covering about 83% of the winter weeks). The dotted areas represent the 95% level of confidence where the means of surface temperature anomalies after SWI_{neg} and SWI_{plain} differ significantly.

Formatted: Normal, Line spacing: single

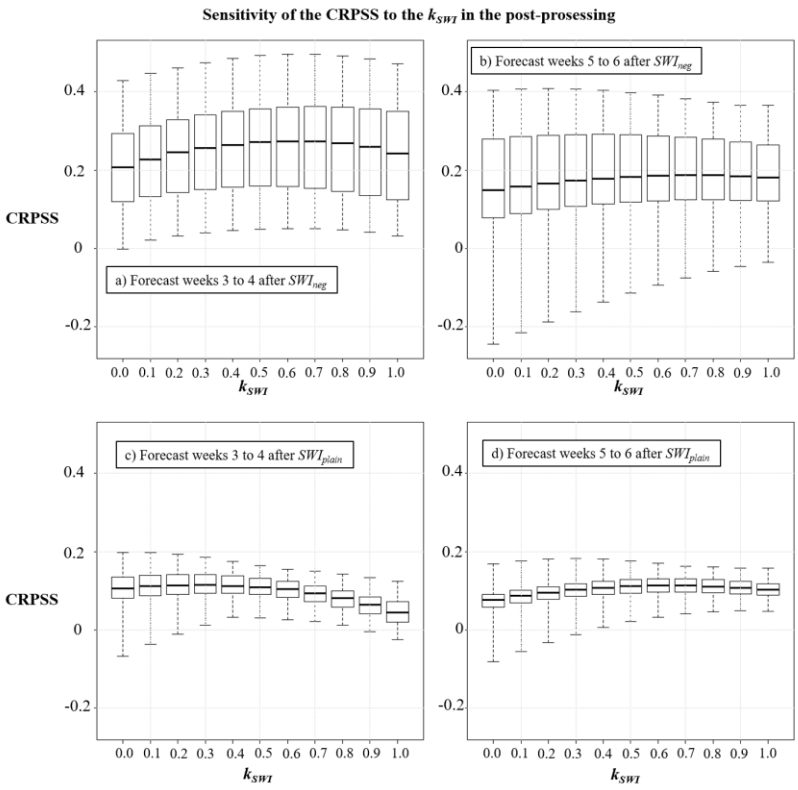


Figure 4. Sensitivity of the expected CRPSS of the post-processed ECMWF surface temperature reforecasts to the k_{SWI} ranging from 0.0 to 1.0 in forecast weeks 3–4 (a and c) and 5–6 (b and d) in cases of SWI_{neg} (upper row) and SWI_{plain} (bottom row). The black boxes show the lower and upper quartiles, and the whiskers illustrate the extremes of the November-February mean CRPSSs of all the grid points in Northern Europe.

Formatted: Normal, Line spacing: single

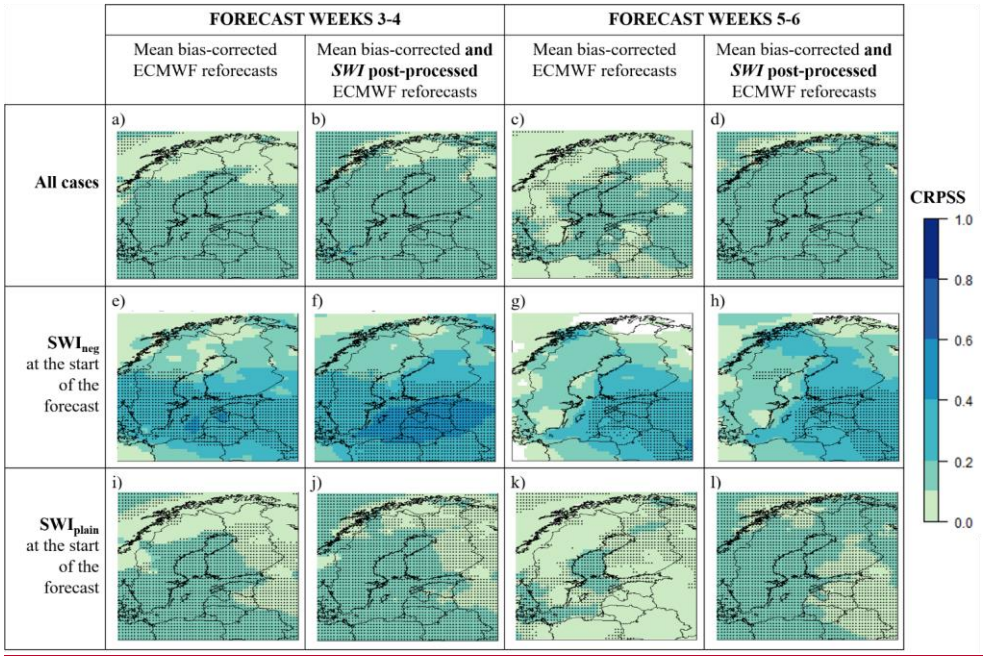
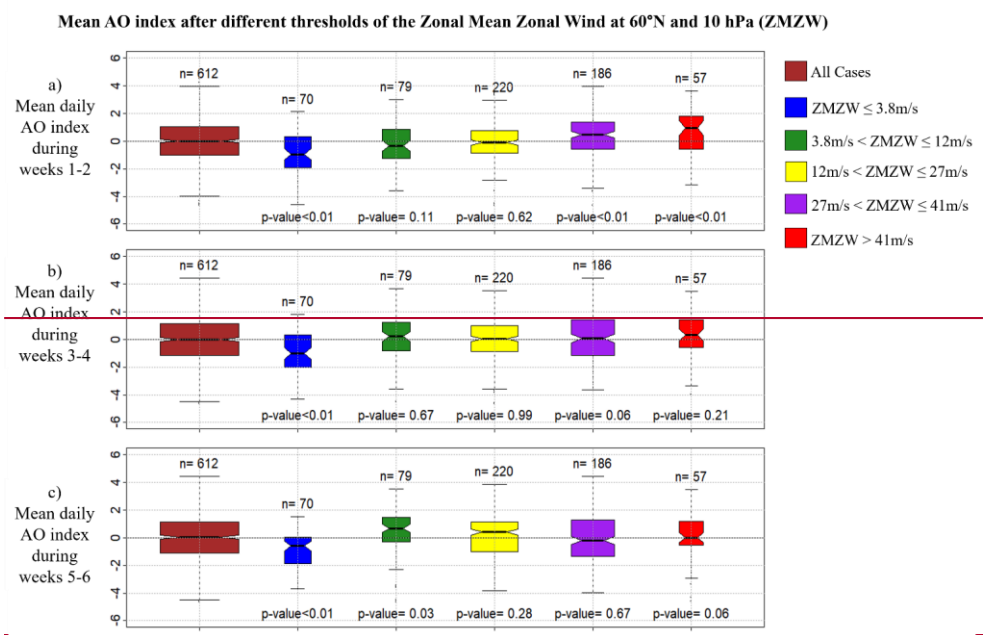


Figure 5. Expected CRPSS of forecast weeks 3–4 and 5–6 of the ECMWF's two-weeks mean temperature reforecasts for November–February 1997–2016 in all cases (upper row), after SWI_{neg} (middle row), and after SWI_{plain} (bottom row) with mean bias-correction only (a, c, e, g, i, and k) and with both mean bias-correction and *SWI* based post-processing (b, d, f, h, j, and l). ERA-Interim climatology of 1981–2010 was used as the reference. The dotted areas represent the 95% level of confidence that the CRPSS is above zero.

Formatted: Normal, Line spacing: single



~~**Figure 2 Observed mean AO index in November-March (1981-2016) a) 1–2, b) 3–4 and c) 5–6 weeks after different thresholds of the zonal mean zonal wind at 60°N and 10 hPa (ZMZW). The horizontal line dividing each box into two parts shows the median of the data, the ends of the box show the lower and upper quartiles, and the whiskers represent the highest and the lowest values excluding outliers. The n written above each box indicates the number of observations in each group. The widths of the boxes have been drawn proportional to the square-roots of n. The p-value written below each boxplot pair indicates the likelihood of such a pair of distributions arising from a random sampling of a single distribution as given by a Student's t-test, i.e., p-values less than 0.05 indicate that the means of the data sets differ significantly at the 95% level of confidence. The notches of each side of the boxes were calculated by R boxplot.stats. If the notches of two plots do not overlap, this is 'strong evidence' that the two medians differ (Chambers et al., 1983, p. 62).**~~

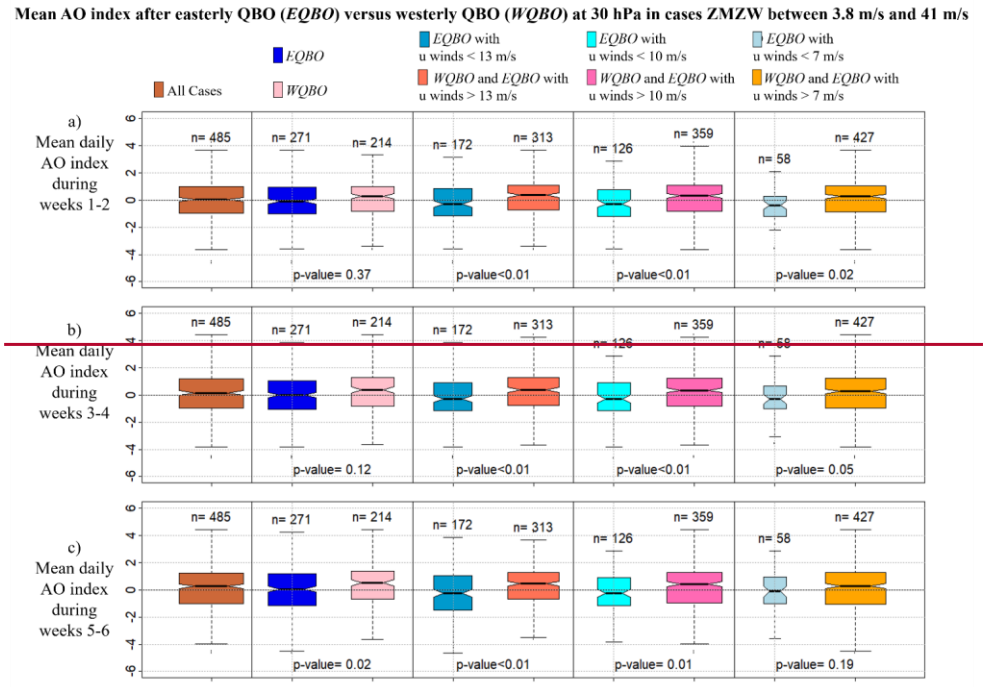
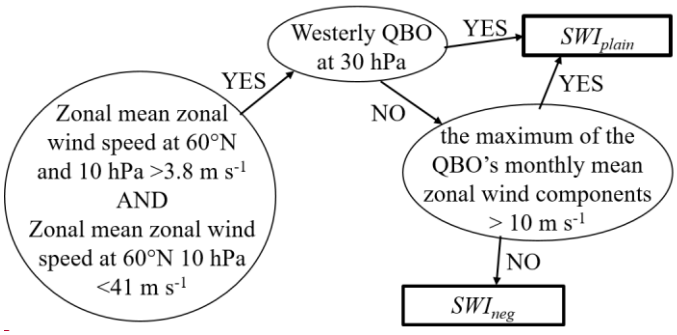
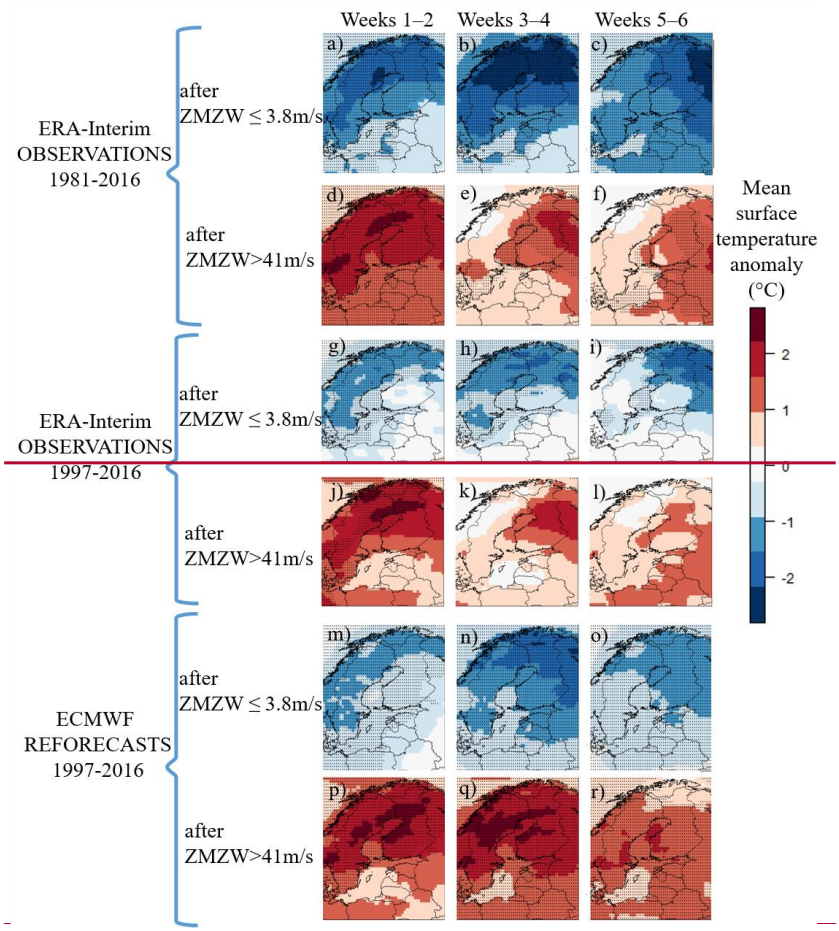


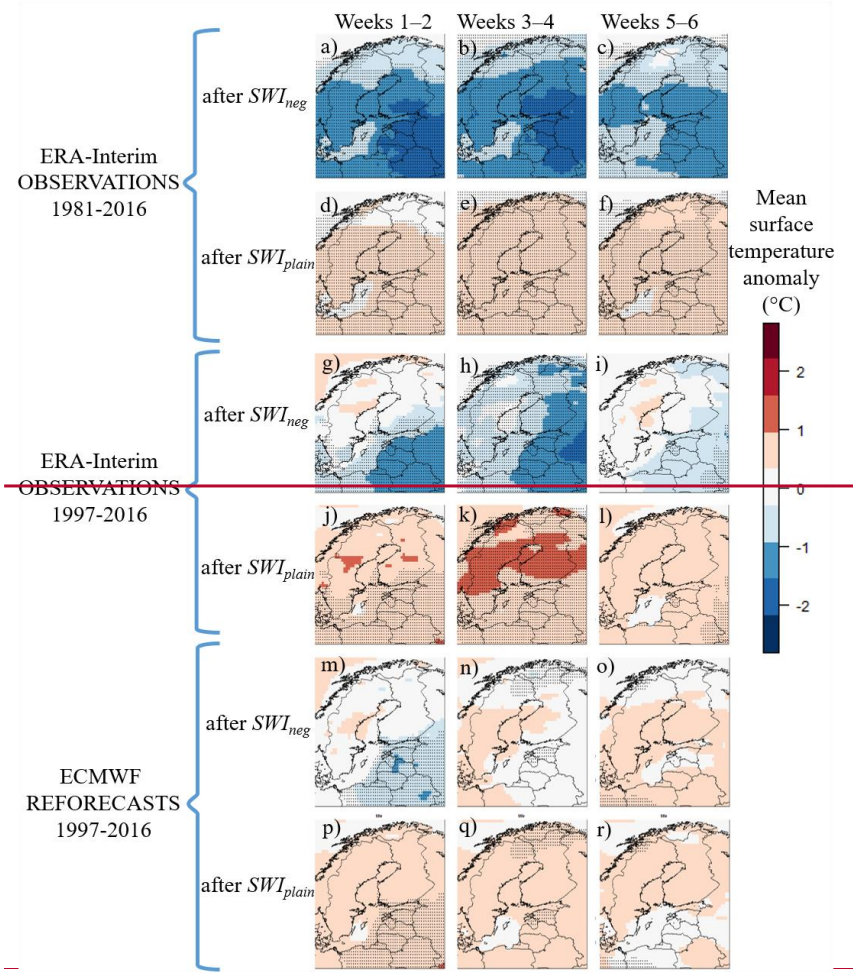
Figure 3: Observed daily AO index in November-March (1981-2016) a) 1–2, b) 3–4 and c) 5–6 weeks after different stratospheric situations. The horizontal line dividing each box into two parts shows the median of the data, the ends of the box show the lower and upper quartiles, and the whiskers represent the highest and the lowest values excluding outliers. The *n* written above each box indicates the number of observations in each group. The widths of the boxes have been drawn proportional to the square-roots of *n*. The p-value written below each boxplot pair indicates the likelihood of such a pair of distributions arising from a random sampling of a single distribution as given by a Student's t-test, i.e., p-values less than 0.05 indicate that the means of the data sets differ significantly at the 95% level of confidence. The notches of each side of the boxes were calculated by R boxplot.stats. If the notches of two plots do not overlap, this is 'strong evidence' that the two medians differ (Chambers et al., 1983, p. 62). ZMZW=zonal mean zonal wind at 60°N and 10 hPa. u winds=the QBO's monthly mean zonal wind components in levels 70 hPa…10 hPa.



~~Figure 4: Decision tree of SWI_{neg}/SWI_{plain}.~~



~~Figure 5. ERA-Interim observed (a-l) and ECMWF reforecasted (m-r) mean temperature anomalies in comparison to the 1981-2016 mean during boreal winters (November-February) in cases the minimum of the daily mean zonal mean zonal wind (ZMZW) at 60ºN and 10 hPa during the previous 10 days was below 3.8 ms-1 (covering about 11% of the winter weeks) or above 41 ms-1 (covering about 9% of the winter weeks). The dotted areas represent the 95% level of confidence where the means of surface temperature anomalies differ significantly from the rest of the cases.~~



~~**Figure 6. ERA-Interim observed (a-l) and ECMWF reforecasted (m-r) mean temperature anomalies in comparison to the 1981-2016 mean during boreal winters (November-February) in cases the previous week's *SWI* was negative (*SWI_{neg}*, covering about 21% of the winter weeks) or plain (*SWI_{plain}*, covering about 59% of the winter weeks). The dotted areas represent the 95% level of confidence where the means of surface temperature anomalies after *SWI_{neg}* and *SWI_{plain}* differ significantly.**~~

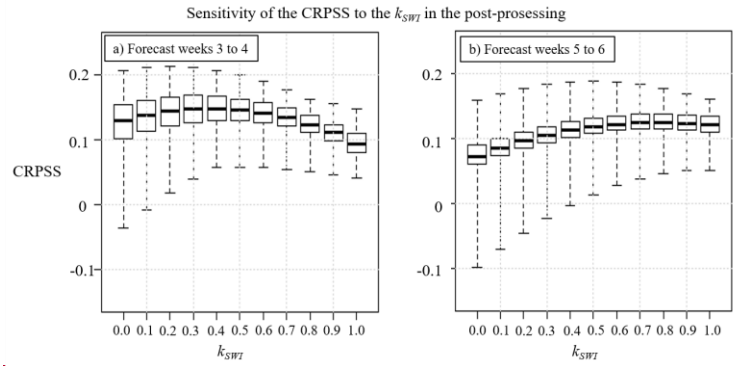
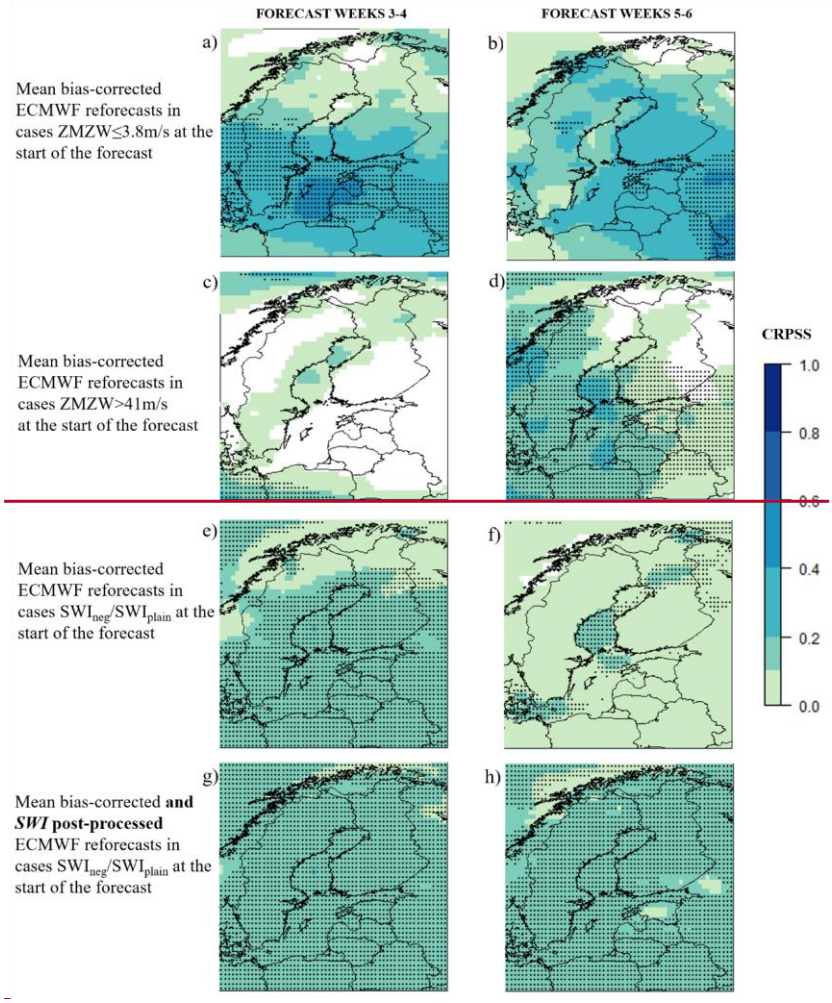


Figure 7. Sensitivity of the expected CRPSS of the post-processed ECMWF surface temperature reforecasts to the k_{SWI} ranging from 0.0 to 1.0 in forecast weeks 3–4 (a) and 5–6 (b). The black boxes show the lower and upper quartiles, and the whiskers illustrate the extremes of the November-February mean CRPSSs of all the grid points in Northern Europe.



~~**Figure 8. Expected CRPSS of forecast weeks 3–4 and 5–6 of the ECMWF's mean temperature reforecasts for November–February 1997–2016 after mean bias correction (a–f) and after both mean bias correction and the *SWI* based post-processing (g–h). ERA-Interim climatology of 1981–2010 was used as the reference. The dotted areas represent the 95% level of confidence that the CRPSS is above zero.**~~