https://doi.org/10.5194/acp-2019-679
Preprint. Discussion started: 19 August 2019

Atmospheric Chemistry and Physics Discussions

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 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 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 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. We developed a stratospheric wind indicator, *SWI*, based on the previous months' stratospheric wind observations and the phase of the AO during the following weeks. We demonstrate that there was a statistically significant difference in the observed surface temperature 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 ERFs, if the stratosphere-troposphere links were better captured in the modelling.

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

https://doi.org/10.5194/acp-2019-679
Preprint. Discussion started: 19 August 2019

Atmospheric Chemistry and Physics Discussions

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 during a one year living lab. 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, 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 10 hPa at 60 °N is commonly defined as a major sudden stratospheric warming (SSW), albeit other definitions also occur (Butler et al. 2015).

At the surface, the AO index, which is constructed from the daily 1000 hPa geopotential height field over 20°N–90°N against monthly mean values, 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).

https://doi.org/10.5194/acp-2019-679
Preprint. Discussion started: 19 August 2019

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. There is no precise consensus of the mechanisms of this tropical-extratropical connection, 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). Garfinkel et al. (2018) found by model simulation a weakened stratospheric polar vortex during the easterly QBO phase compared to the westerly phase in early winter (October–December).

Challenges related to the realistic modelling of the dynamical stratosphere-troposphere coupling have been adduced by Shepherd et al. (2018) and Polichtchouk et al. (2018). Even though many models are already able to reproduce the QBO, they are still underestimating its teleconnection to the surface weather (Scaife et al. 2014). 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 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 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 stratospheric zonal mean winds. According to the observed daily AO index, after different phases of QBO, and strengths of the observed stratospheric zonal mean winds, we define a 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 SWI_{neg} versus SWI_{plain} and utilize these in post-processing the mean of the temperature forecasts of 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.

https://doi.org/10.5194/acp-2019-679
Preprint. Discussion started: 19 August 2019

2 Datasets and Methods

We verified and post-processed ERFs of the ECMWF's Integrated Forecasting System (IFS Cycle 43r1, Vitart, 2014), which belongs to the models of the Sub-seasonal to seasonal scale (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 studied the weekly mean temperatures of the Monday runs over Northern Europe in 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):

$$\boldsymbol{CRPS = \int |F(y) - F_o(y)|^2 dx}, \qquad (1)$$

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, M, was 51, the expected CRPS, the $CRPS_{RF}$ of the $CRPS_{rf}$ was calculated for 51 members using equation 26 in Ferro et al. (2008):

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

Further, the skill scores of the CPRS, the CRPSS, were calculated as follows:

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

The annual mean CRPSSs were calculated for the ECMWF's reforecasts (1997–2016) for forecast weeks 1 to 6. 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 forecast week (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 stratospheric precursors for a statistically significantly weaker AO index available at the start of the forecast. The observed daily surface AO indexes were downloaded from the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC). We used two stratospheric wind data sets for the precursors of the AO index. The first data were the daily ZMZW at 60°N and 10 hPa during 1981–2016 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 examined the observed daily AO indexes during the 1–2 weeks, 3–4 weeks, and 5–6 weeks following different phases of QBO and strengths of the stratospheric winds. Holton and Tan (1980, 1982) demonstrated that the geopotential height at high latitudes was significantly lower during westerly QBO compared to the easterly QBO. Therefore, we explored the daily AO index 1–6 weeks after westerly QBO, the *WQBO*, and easterly QBO, the *EQBO* using the QBO winds at 30 hPa. We also examined the effect of the maximum of the monthly mean zonal wind components of the QBO between 70 hPa and 10hPa during *EQBO*. Moreover, we explored the daily AO index after the daily ZMZW at 60°N and 10 hPa during the last 10 days of the previous month falling below its overall wintertime (November–March 1981–2016) 10^{th} percentile, corresponding a value of 3.8m/s, indicating a weak polar vortex already at the start of the forecast. The statistical significance of the difference between the AO index following two different stratospheric situations, e.g., the *EQBO* and the *WQBO*, was determined using a two-sided Student's t-test with the null hypothesis that there is no difference. We used the most statistically significant predictors for weaker AO indexes observed 1–2 weeks, 3–4 weeks, and 5–6 weeks after these stratospheric situations, to define a *SWI* to be SWI_{neg}; otherwise, it was defined as SWI_{plain} for the beginning of each winter month (November–February) in 1981–2016.

https://doi.org/10.5194/acp-2019-679
Preprint. Discussion started: 19 August 2019

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 SWI_{neg} and SWI_{plain} defined in Section 2.3. First, we calculated the observed 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 the 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 SWI_{neg} and SWI_{plain} according to the previous month's stratospheric wind observations. Thereafter, we determined the statistical significance of the difference between the surface temperatures after SWI_{neg} and SWI_{plain} using a two-sided Student's t-test with the null hypothesis that there is no difference between SWI_{neg} and SWI_{plain}. This same procedure to define the difference between the surface temperatures after SWI_{neg} and SWI_{plain} 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} in the ECMWF reforecasts run of the first week of November–February 1997–2016 were defined to examine how the model reproduced the anomalies.

For post-processing the ECMWF reforecasts, we calculated TA_{SWIneg} and $TA_{SWIplain}$ representing mean temperature anomalies in November–February after SWI_{neg} and SWI_{plain}, respectively. In order to decrease the effect of the time period used for defining mean anomalies, the TA_{SWIneg} and $TA_{SWIplain}$ were calculated from the ERA-Interim 1981–2016 two-week mean temperature anomalies by dividing thirty five years' data (1981–2016 excluding the reforecast year) to 5-year-long periods. One 5-year-long period was left out, and the mean anomalies representing SWI_{neg} and SWI_{plain} were calculated of the remaining 30 years. This was repeated seven times by changing the 5-year-long period left out. The means, TA_{SWIneg} and $TA_{SWIplain}$ of the achieved seven mean temperature anomalies representing SWI_{neg} and SWI_{plain} were calculated separately for each grid 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 previous months' stratospheric wind observations. 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},

https://doi.org/10.5194/acp-2019-679
Preprint. Discussion started: 19 August 2019

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

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 shows boxplots of the observed minimum of the daily AO index 1–2 weeks, 3–4 weeks, and 5–6 weeks after different phases of QBO and restrictions in the strength of the stratospheric winds in 1981–2016. The first box (brown) represents the minimum AO indexes after all the cases in 1981–2016 November–February, i.e., 36 years * 4 months =144 cases. The second and third boxes show the minimum AO indexes after easterly (*EQBO*, blue) and westerly (*WQBO*, pink) QBO at the 30 hPa level, respectively. 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 minimum AO indexes 1–2 weeks, 3–4 weeks, and 5–6 weeks after *EQBO* were lower than after *WQBO*. However, this was statistically significant at the 95% confidence level only in weeks 1–2. 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 minimum AO indexes after *EQBO* with all the QBO's monthly mean zonal wind components between levels 70…10hPa being below 13 m/s, 10 m/s, and 7 m/s, respectively. Restricting the *EQBO* cases by a maximum of the QBO's monthly mean zonal wind components in levels 70…10hPa decreased the median of the minimum AO during the following 1–2, 3–4, and 5–6 weeks; however, only for the QBO's monthly mean zonal wind components less than 10 m/s, the minimum AO index was still at the 95% confidence level statistically significantly lower than in the rest of the data.

The 10th box (yellow) in Fig. 2 shows the minimum AO indexes after cases the daily ZMZW at 60°N and 10 hPa was below its 10th percentile (4.8m/s) during the 10 last days of the previous month, corresponding to cases with weak polar vortex already at the start of the forecast. The observed minimum AO index was statistically significantly at the 95% confidence level weaker 1–2 weeks and 3–4 weeks after the daily ZMZW at 60°N and 10 hPa, which was been below their overall wintertime 10th percentile (indicating a weak polar vortex). In weeks 5–6, the AO index was also weaker, but it was statistically insignificant at the 95% level.

Aiming to select stratospheric precursors indicating weak AO with the greatest statistical significance, we defined the *SWI* to be negative in cases when the QBO was easterly at 30 hPa and the QBO's monthly mean zonal wind components in levels 70…10hPa were weaker than 10m/s and/or if the daily ZMZW at 60°N and 10 hPa during the 10 last days of the previous month fell below its overall wintertime 10th percentile. In other cases, the *SWI* was defined as plain. This decision tree for the *SWI* is depicted in Fig. 3. The means of the minimum AO index after SWI_{neg} and SWI_{plain} (in 1981–2016) were statistically significantly different using a Student's t-test, with lower AO index more common 1–2 weeks, 3–4 weeks, and 5–6 weeks after SWI_{neg} than after SWI_{plain} (see Fig. 2 for the p-values).

Figure 4 shows the observed (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 observations showed on average lower mean temperatures for the weeks' 3–4 and 5–6 after SWI_{neg} (Fig. 4b-c and 4h-i). The reforecasts also showed cold anomalies after SWI_{neg} (Fig. 4n-o) but weaker than the observed ones. Further, the observations showed on average higher mean temperatures for weeks 1–2, 3–4, and 5–6 after SWI_{plain} (Fig. 4d-f and 4j-l). This warm anomaly was forecasted only to some degree in the forecasts weeks 1–2 (Fig. 4p), and it was totally absent in forecast weeks 3–4 (Fig. 4q) and 5–6 (Fig. 4r). The mean temperature anomalies 3-6 weeks after SWI_{neg} (Fig. 4b-c) and SWI_{plain} (4e-f) during 1981–2016 were statistically significantly different using a Student's t-test, with anomalously cold surface temperatures more common 3-6 weeks after SWI_{neg}. When examining the years 1997–2016 (Fig. 4h-i and Fig. 4k-l), which was the reforecast period, the temperature anomalies were of the same sign than during the longer 1981–2016 period (Fig. 4b-c and Fig. 4e-f) but weaker and not statistically significant all over the Northern Europe.

3.3 The *SWI* and the forecasted mean temperatures

The mean temperature anomalies in Fig. 4(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),

https://doi.org/10.5194/acp-2019-679
Preprint. Discussion started: 19 August 2019

whereas the CRPSSs of the mean temperatures of the forecast weeks 3–4 and 5–6 were improved by the *SWI* based post-processing (Fig. 5a and 5b). The best median CRPSS was achieved by k_{SWI}=0.5, for both weeks 3–4 and weeks 5–6. Figure 6 shows 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 (Fig. 6a-b) and by the mean bias-corrected reforecasts together with the *SWI* based post-processed forecast (k_{SWI} =0.5, Fig. 6c-d). By using the *SWI* based post-processing to the ECMWF forecasts, 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. The forecast skills for the weeks 3–4 and 5–6 post-processed by Eq. (4) and Eq. (5) were not sensitive to the period (within 1981–2016) used for defining SWI_{neg} and SWI_{plain} temperature anomalies. For instance, we tested periods 1981–2000 and 2001–2016 and got almost the same CRPSSs for Northern Europe (no figure).

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 showed that in addition to the previously demonstrated weaker polar vortex during easterly QBO in comparison to westerly QBO (e.g., Holton and Tan 1980, Garfinkel et al. 2018), the maximum strength of the QBO's monthly mean zonal wind components in levels 70…10hPa during the easterly QBO at 30 hPa affected the observed AO index 1–6 weeks later. Based on observations, we found that the minimum 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 m/s. We also found that the minimum AO index was statistically significantly weaker 1–2 weeks and 3–4 weeks after the daily ZMZWs at 60°N, and 10 hPa had been below their overall wintertime 10th percentile (indicating a weak polar vortex). In weeks 5–6, the AO index was weaker but statistically insignificant.

Selecting the SWI_{neg} to include the both above-mentioned situations, the level of statistical significance of a weaker AO index during the next 1–2, 3–4, and 5–6 weeks decreased in comparison to using only one of these situations. Our definition of SWI_{neg} 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}. Also, the mean surface temperature anomalies in Northern Europe in November–February in 1981–2016 after SWI_{neg} and SWI_{plain} were statistically significantly different, with anomalously cold surface temperatures more common 3–6 weeks after SWI_{neg}. The mean temperature anomalies corresponding SWI_{neg}/SWI_{plain} at the start of the forecast 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.

This study demonstrates that the QBO-polar vortex connection should be better integrated into the extended-range surface temperature forecasts over Northern Europe. 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). 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. In our studies, the use of monthly mean stratospheric observations restricted the studies to post processing only the forecasts made on the first week of each month. The next step would be looking for the stratospheric signal from the forecast model, which would also make it possible to post-process forecasts made every week.

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 indexes 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 1–2 and 4–6 available at https://github.com/fmidev/sixweeks.

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, 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’.

https://doi.org/10.5194/acp-2019-679
Preprint. Discussion started: 19 August 2019

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.

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., 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.

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.

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., 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.

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., 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.

https://doi.org/10.5194/acp-2019-679
Preprint. Discussion started: 19 August 2019

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.

Atmospheric Chemistry and Physics Discussions

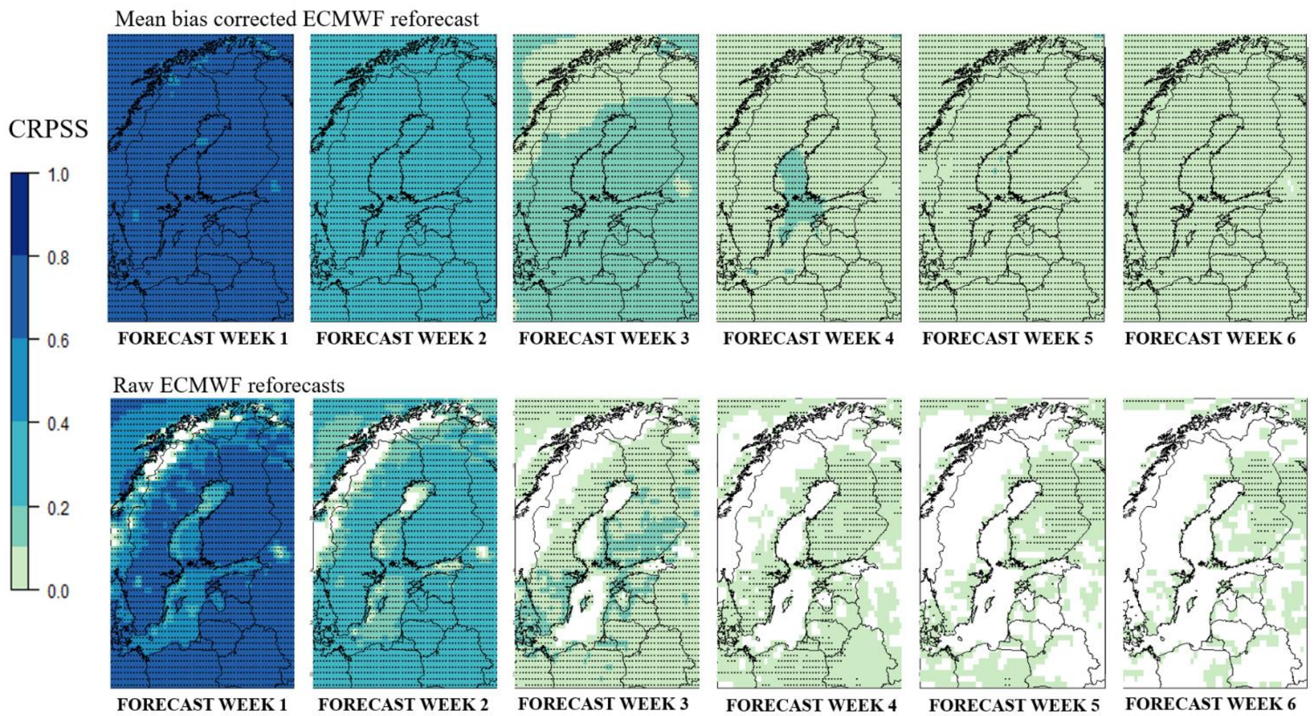

Figure 1: Expected CRPSS of the weekly mean temperature of the mean bias-corrected (upper row) and raw (lower row) ECMWF reforecasts for years 1997–2016 using ERA-Interim climatology of 1981–2010 as the reference. The dotted areas represent the 95% level of confidence that the CRPSS is above zero.

https://doi.org/10.5194/acp-2019-679
Preprint. Discussion started: 19 August 2019

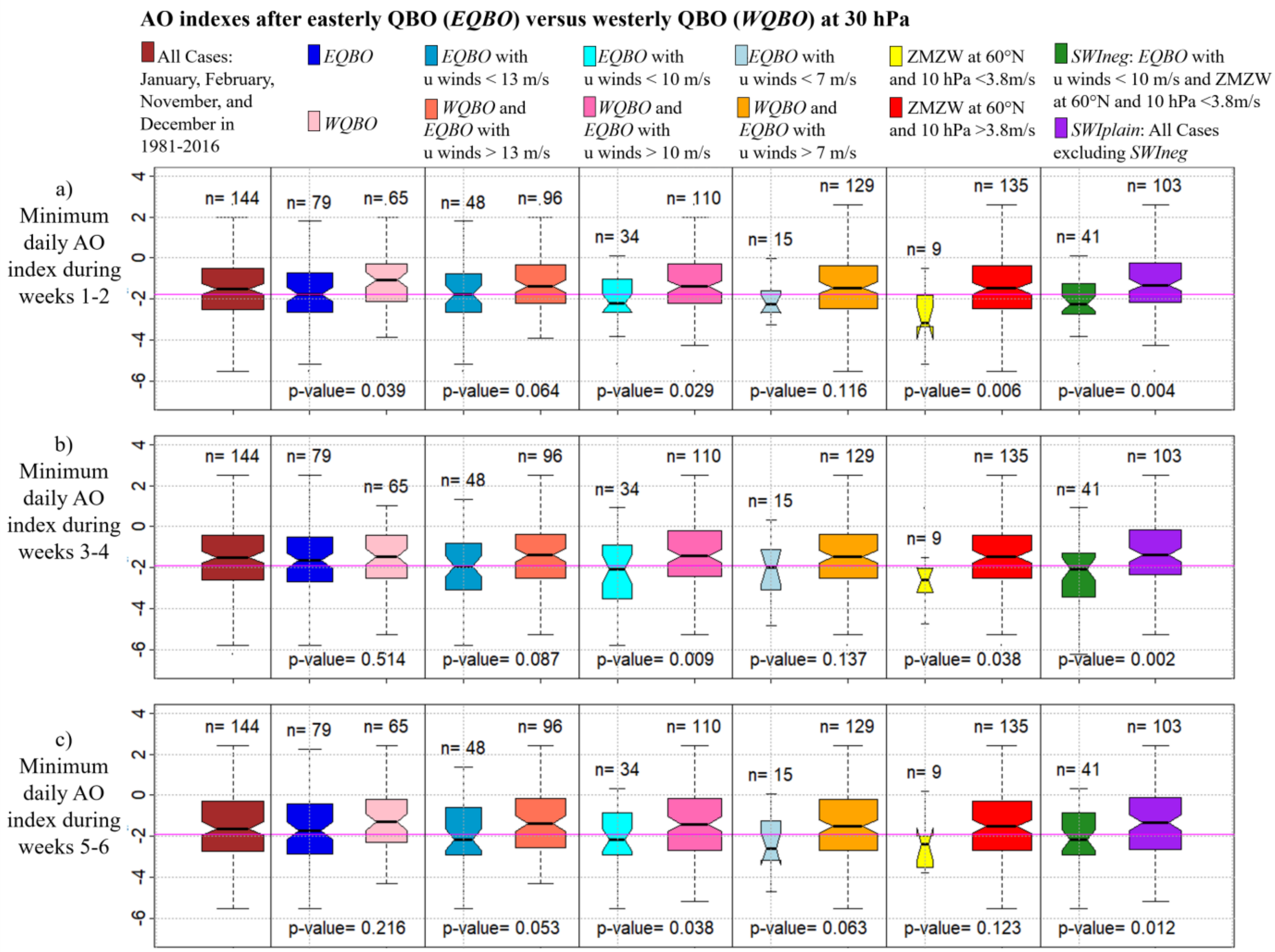

Figure 2: Observed minimum daily AO index a) 1–2, b) 3–4 and c) 5–6 weeks after different stratospheric situations. The line dividing each box into two parts shows the median of the data, the ends of the box show the lower and upper quartiles, and the whiskers represent the highest and the lowest values 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). The magenta line represents the mean minus one standard deviation of the daily AO index of all the cases. ZMZW=zonal mean zonal wind.

https://doi.org/10.5194/acp-2019-679
Preprint. Discussion started: 19 August 2019

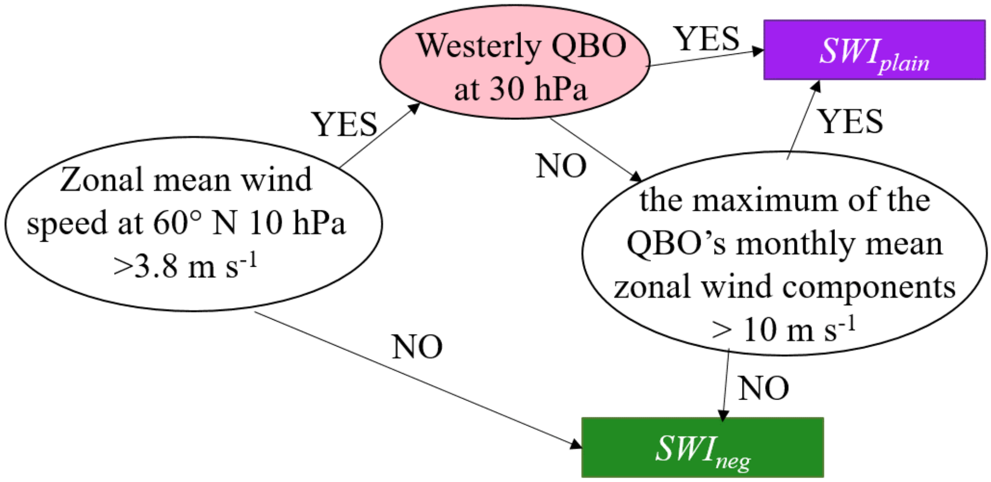

Figure 3: Decision tree of SWI_{neg}/SWI_{plain}.

https://doi.org/10.5194/acp-2019-679
Preprint. Discussion started: 19 August 2019

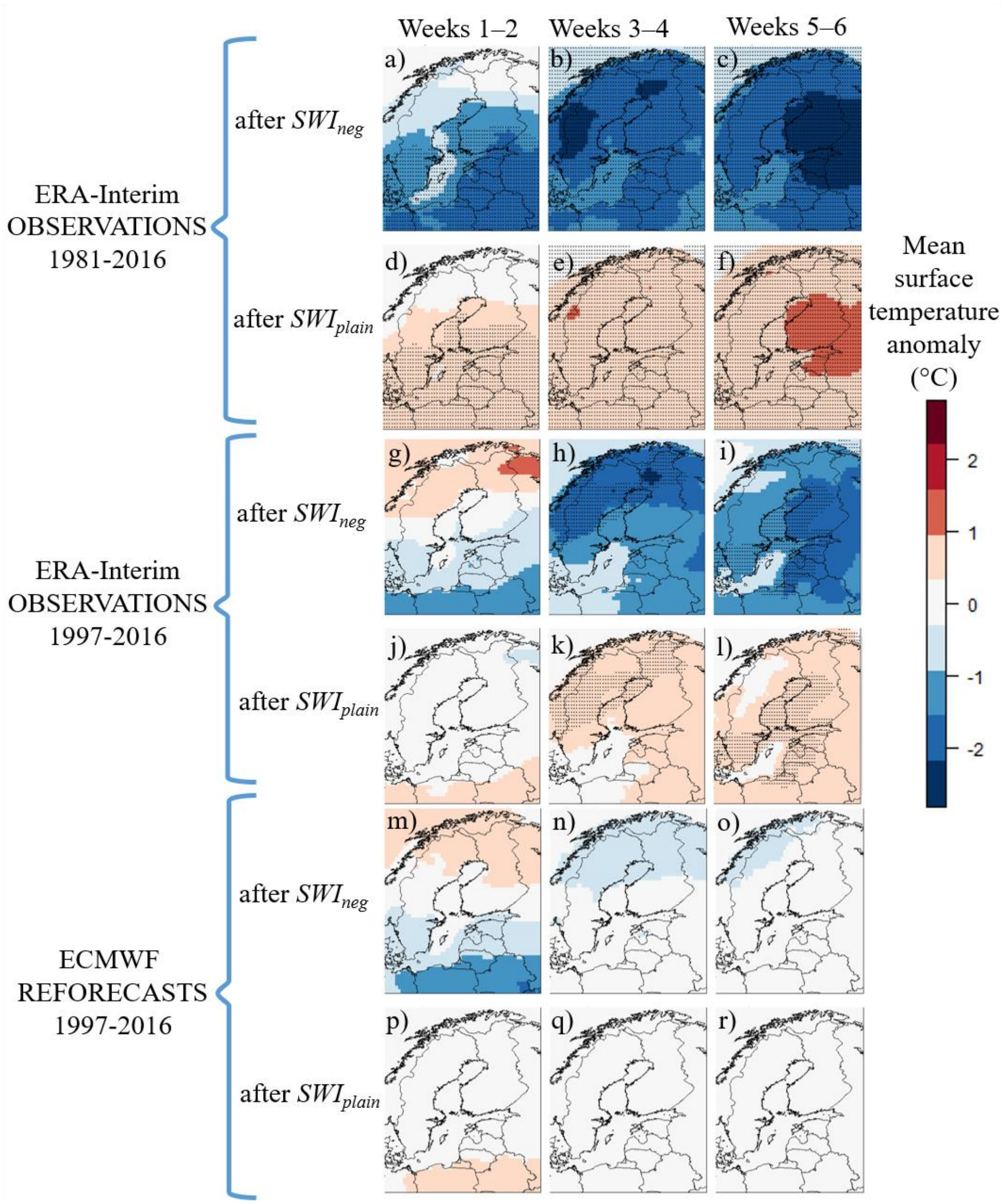

Figure 4. ERA-Interim observed (a-l) and ECMWF reforecasted (m-r) mean temperature anomalies during boreal winters (November-February) in cases the previous month's *SWI* was negative (SWI_{neg}, covering about 28% of the winter months) or plain (SWI_{plain}, covering about 72% of the winter months). The dotted areas represent the 95% level of confidence where the means of surface temperature anomalies after SWI_{neg} and SWI_{plain} differ significantly.

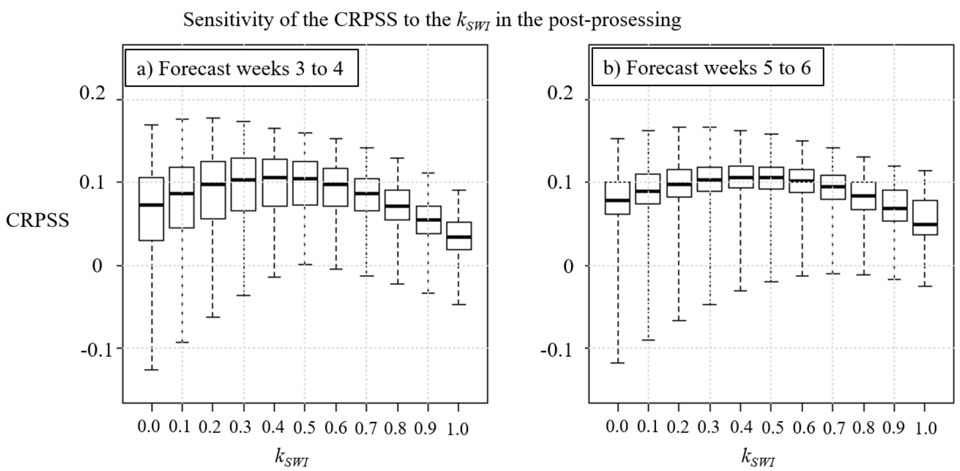

Figure 5. Sensitivity of the expected CRPSS of the ECMWF surface temperature reforecasts to the k_{SWI} ranging from 0.0 to 1.0 in forecast weeks 3–4 (a) and 5–6 (b). The black boxes show the lower and upper quartiles, and the whiskers illustrate the extremes of the November-February mean CRPSSs of all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

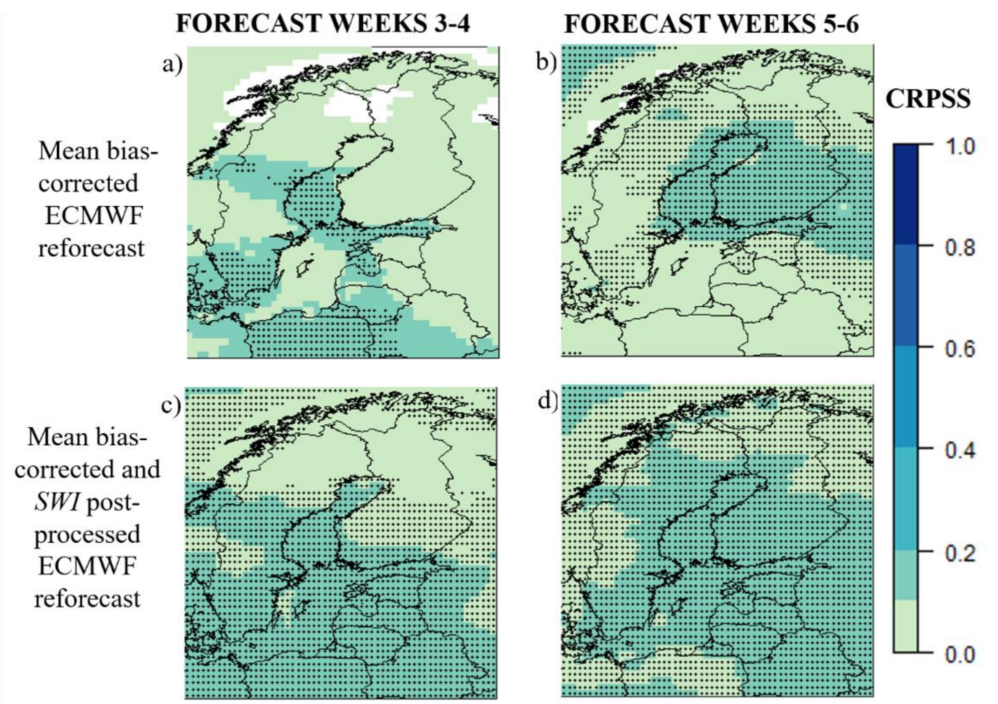

Figure 6. Expected CRPSS of forecast weeks 3–4 and 5–6 of the ECMWF's mean temperature reforecasts for November–February 1997–2016 after mean bias-correction (a-b) and after both mean bias-correction and the *SWI* based post-processing (c-d). ERA-Interim climatology of 1981–2010 was used as the reference.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