Geophysical Research Letters

RESEARCH LETTER
10.1029/2024GL109733

Flow-Dependence of Ensemble Spread of Subseasonal Forecasts Explored via North Atlantic-European Weather Regimes

J. Spaeth[1], P. Rupp[1], M. Osman[2,3,4,5], C. M. Grams[2,6], and T. Birner[1,7]

[1]Meteorologisches Insitut, Ludwig-Maximilians-Universität München, Munich, Germany, [2]Institute of Meteorology and Climate Research Troposphere Research (IMKTRO), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, [3]Facultad de Ciencias Exactas y Naturales, Departamento de Ciencias de la Atmósfera y los Océanos, Universidad de Buenos Aires, Buenos Aires, Argentina, [4]Centro de Investigaciones del Mar y la Atmósfera (CIMA), CONICET–Universidad de Buenos Aires, Buenos Aires, Argentina, [5]Instituto Franco-Argentino para el Estudio del Clima y sus Impactos (IRL 3351 IFAECI), CNRS–IRD–CONICET–UBA, Buenos Aires, Argentina, [6]Federal Office of Meteorology and Climatology, MeteoSwiss, Zurich, Switzerland, [7]Institut für Physik der Atmosphäre, Deutsches Zentrum für Luft- und Raumfahrt (DLR), Oberpfaffenhofen, Germany

Key Points:

- Ensemble spread of near-surface weather in subseasonal forecasts is sensitive to the prevailing North Atlantic-European weather regime
- Greenland blocking is linked to the smallest ensemble spread of geopotential but to the largest spread of temperature over northern Europe
- Weak polar vortex states are followed by reduced forecast uncertainty of weather regimes but increased uncertainty of European temperature

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

J. Spaeth,
jonas.spaeth@physik.uni-muenchen.de

Citation:

Spaeth, J., Rupp, P., Osman, M., Grams, C. M., & Birner, T. (2024). Flow-dependence of ensemble spread of subseasonal forecasts explored via North Atlantic-European weather regimes. *Geophysical Research Letters*, *51*, e2024GL109733. https://doi.org/10.1029/2024GL109733

Received 9 APR 2024
Accepted 12 JUL 2024

Abstract Atmospheric prediction at 2–6 weeks lead time (so-called subseasonal-to-seasonal timescales) entails large forecast uncertainty. Here we investigate the flow-dependence of this uncertainty during Boreal winter. We categorize the large-scale flow using North Atlantic-European weather regimes. First, we show that forecast uncertainty of near-surface geopotential height (Z1000) and temperature (T2m) are strongly sensitive to the prevailing regime. Specifically, forecast uncertainty of Z1000 reduces over northern Europe following Greenland Blocking (enhanced predictability) due to a southward shifting eddy-driven jet. However, due to strong temperature gradients and variable flow patterns, Greenland blocking is linked to increased forecast uncertainty of T2m over Europe (reduced predictability). Second, we show that forecast uncertainty of weather regimes is modulated via the stratospheric polar vortex. Weak polar vortex states tend to reduce regime-uncertainty, for example, due to more frequent predicted occurrence of Greenland blocking. These regime changes are associated with increased T2m uncertainty over Europe.

Plain Language Summary Weather is chaotic, and forecasts several weeks ahead are quite uncertain. Nevertheless, the degree of uncertainty varies, which can be relevant for long-term planning in various sectors, including agriculture, energy supply, and public health. Here we show that the degree of uncertainty depends on the weather at the time of forecast start, which we categorize using eight characteristic weather patterns for the North Atlantic and Europe. We analyze a large set of forecasts during winter, with lead times up to 6 weeks. For example, persistent high-pressure systems over Greenland are known to favor low temperatures over northern Europe. Our results indicate that, in addition, temperatures are highly variable in these cases, leading to unusually high forecast uncertainty. In contrast, uncertainty of near-surface air pressure tends to decrease due to less frequent storms over the North Atlantic. Furthermore, we show that forecast uncertainty of the weather patterns themselves varies, which is useful when large-scale flow conditions are more critical than local weather. We analyze forecasts under different circulation conditions in the Arctic stratosphere, as these can have long-lasting impacts on surface weather. We find forecasts of weather patterns become less uncertain when the circumpolar winds in the Arctic stratosphere are weak.

1. Introduction

Numerical weather prediction is fundamentally limited by uncertainties in initial conditions and the numerical model (e.g., Slingo & Palmer, 2011). Ensemble forecasts produce a distribution of possible future states and thereby represent a framework to quantify these uncertainties and their evolution over forecast lead time (Leutbecher & Palmer, 2008; T. Palmer, 2017). Such probabilistic forecasts offer increased economic value (T. N. Palmer, 2002; Zhu et al., 2002), making them valuable for stakeholders and decision-makers (Robbins et al., 2018). In particular, the availability of a full distribution allows for quantification of forecast uncertainty, for example, via the variance across ensemble members.

At lead times beyond about 2 weeks (the so-called extended-range) forecast uncertainty is generally large. It is therefore crucial to identify situations of reduced forecast uncertainty and thereby enhanced predictability. Such flow conditions are often referred to as "windows of opportunity" (Mariotti et al., 2020). In the northern extratropics, the tropospheric flow on extended-range time scales can be influenced by various teleconnections, including a downward impact of stratospheric circulation anomalies (Baldwin & Dunkerton, 1999, 2001; Butler et al., 2019). In particular, extreme states of the stratospheric polar vortex have been shown to project onto the North Atlantic Oscillation (NAO) (Charlton-Perez et al., 2018) and are hence associated with modulated frequencies of blocking systems (e.g., Labitzke, 1965; Vial et al., 2013) and storm series (e.g., Afargan-Gerstman & Domeisen, 2020; Rupp et al., 2022).

While the ensemble-mean response to such teleconnections has been subject of many previous studies, impact on ensemble variance, that is, forecast uncertainty, has received less attention. Spaeth et al. (2024) have recently shown that ensemble variance of 1,000 hPa geopotential height (Z1000) reduces over northern Europe and increases over southern Europe following weak polar vortex events (and vice versa for strong polar vortex events). Here, we extend this analysis by focusing on how forecast uncertainty changes following different North Atlantic-European weather regimes (Grams et al., 2017). These regimes provide a useful way to describe the large-scale flow over the Euro-Atlantic sector using a finite set of eight characteristic weather patterns. Weather regimes have successfully been employed by previous studies to analyze flow-dependence of forecast skill (Büeler et al., 2021; Ferranti et al., 2015; Matsueda & Palmer, 2018) as well as the tropospheric response to stratospheric circulation anomalies (Beerli & Grams, 2019; Charlton-Perez et al., 2018; Domeisen, Grams, & Papritz, 2020; Lee et al., 2019; Roberts et al., 2023). Moreover, we extend the analysis by studying ensemble variance of 2-m temperature (T2m) in addition to Z1000.

Specifically, we concentrate on the winter season and separate our questions about forecast uncertainty into two parts: (a) How does forecast uncertainty of near-surface weather at lead times of 2–6 weeks depend on the respective weather regime at time of initialization and (b) how does forecast uncertainty associated with these weather regimes depend on the state of the polar stratosphere? Based on a large set of extended-range ensemble forecasts, we find that even if a weather regime can be well predicted weeks ahead, this weather regime might be associated with inherently large uncertainty in some near-surface variables.

2. Methods

2.1. S2S Forecast Data

We analyze 1,197 ensemble forecasts provided by the European Center for Medium-Range Weather Forecasts (ECMWF) via the subseasonal-to-seasonal (S2S) Prediction Project Database (Vitart et al., 2017), which cover winters 1997/98 to 2020/21. These ensemble forecasts comprise 57 51-member forecasts from winters 2019/20 and 2020/21 (so-called realtime forecasts) and 1140 11-member forecasts from winters 1999/2000 to 2019/20 (so-called hindcasts or re-forecasts). Forecasts are initialized on Mondays and Thursdays and cover lead times up to +46 days. In each winter, we use forecasts initialized between 16 November and 22 February, down-sampled to a $2.5° \times 2.5°$ horizontal grid, at daily temporal resolution (geopotential: instantaneous fields at 0 UTC; temperature: daily averages). At the maximum lead time of +46 days, the forecast valid times thus range from 1 January (16 November +46 days) to 9 April (8 April in leap years; 22 February +46 days). By including the two different winters of realtime forecasts in the analyses, we combine two model versions (ECMWF IFS CY46R1 and C47R1), but this drawback is expected to be minor compared to the benefit of a larger sample size.

2.2. Ensemble Spread Anomalies

In this study, forecast uncertainty of continuous variables is quantified by means of ensemble variance. Given that ensemble variance may depend on season, lead time and the numerical model, a climatology of ensemble variance is constructed following the S2S-forecast deseasonalization procedure described in Spaeth and Birner (2022). Ensemble variance anomalies are then obtained by subtracting the corresponding climatology from the ensemble variance of each forecast. If a forecast model is reliable, such ensemble variance anomalies are related to anomalies of squared forecast error (Fortin et al., 2014). Results indicate a fair degree of reliability for Z1000 and T2m forecasts in the ECMWF model (see supplementary Text S1 in Supporting Information S1). Hence, anomalies of variance may at least qualitatively be interpreted as anomalies of error, despite deviations from the one-to-one line.

2.3. Definition of Weather Regimes

The year-round weather regimes defined by Grams et al. (2017) for the Atlantic-European region (80°W–40°E, 30–90°N) adapted to ERA5 (Hauser et al., 2023) are used in this study. This definition isolates the following distinct flow patterns: four blocked regimes (Atlantic ridge AR, European blocking EuBL, Scandinavian blocking ScBL, Greenland blocking GL) and three cyclonic regimes (Atlantic trough AT, Zonal regime ZO, Scandinavian trough ScTr). Days where the projection onto any of the above blocked or cyclonic regimes is weaker than a predefined threshold are classified as "no regime" (e.g., flow situations close to climatology, short-lived blocks or cyclones, regime transitions). The weather regime definition is based on 6-hourly 500-hPa geopotential height anomalies (Z500′) from ERA5 reanalysis (Hersbach et al., 2020) at 0.5° horizontal resolution. Z500′ are normalized with the spatially averaged 30-day running standard deviation at a given calendar time, to remove seasonal variability in Z500′ amplitude (see Osman et al., 2023). Then, a k-means clustering in the phase space spanned by the leading seven EOFs of 10-day low-pass filtered, normalized Z500′ is performed. Following Michel and Rivière (2011), we compute the normalized projection of each 6-hourly Z500′ onto the cluster mean, obtaining an index for weather regimes (IWR). We define weather regime life cycles as periods when the maximum IWR exceeds 1.0 for at least 5 days. Dates when none of the IWR values meet this definition are categorized as "no regime". Weather regimes were computed for ECMWF S2S forecasts (model cycles CY46R1 and CY47R1) following the approach outlined in Osman et al. (2023). However, these forecasts are mapped onto weather regimes identified using ERA5 reanalysis data. The computation of weather regimes for bias-corrected forecasts follows the same procedure as for the reanalysis.

2.4. Quantification of Weather Regime Uncertainty by Means of Information Entropy

To describe the distribution of weather regime likelihood and quantify the uncertainty in the occurrence of weather regimes for given boundary conditions, we use the concept of normalized entropy H. There are eight distinct weather regimes, including "no regime". The probability, p_{wr}, of a particular weather regime (denoted by subscript wr) is estimated as the ratio of the number of ensemble members, n_{wr}, classified with that regime to the total ensemble size, N, that is, $p_{wr} = \frac{n_{wr}}{N}$. The regime entropy (based on information theory, see Shannon, 1948; Cover & Thomas, 2005) is then expressed as the negative sum over all weather regimes of the product of their probability with the natural logarithm of that probability. Furthermore, the entropy is normalized by the maximum possible entropy, ln(8), which is reached when all 8 regimes are equally likely:

$$H = \frac{-\sum_{wr} p_{wr} \cdot \ln(p_{wr})}{\ln(8)}. \quad (1)$$

This formulation allows for a quantitative assessment of the uncertainty associated with the occurrence of different weather regimes within the ensemble. The applied normalization of entropy is often performed to make the entropy values comparable across different scenarios. A value of $H = 1$ corresponds to a uniform probability distribution for all weather regimes, while values $0 \leq H < 1$ describe situations where some weather regimes have increased likelihood over others. In an 11-member ensemble, the maximum possible entropy is 0.97 (3 regimes with 2 members each, 5 regimes with 1 members each). In practice, entropy is somewhat smaller on average, because neither the 11 members (hindcasts) nor the 51 members (realtime) can split exactly equally among the 8 regimes, for example, $H \approx 0.82$ given 11 members and a uniform background distribution (empirically derived using synthetic data).

3. Uncertainty of Near-Surface Weather Related to Different Weather Regimes

To investigate flow-dependent spatial patterns of ensemble variance, all ensemble forecasts are clustered by the weather regime at initial time according to ERA5. For each of the eight initial regimes the composite-mean daily anomalies of 1,000 hPa geopotential height (Z1000) ensemble variance are computed and averaged over 7–10 days lead time. By these lead times, ensemble spread has grown to a substantial level, yet the original regime is still the most likely one on average, as weather regimes typically persist for about 10 days (Büeler et al., 2021).

Figure 1 shows that Z1000 ensemble variance anomalies are largest for the two regimes that are closely linked to NAO variability (cf. Beerli & Grams, 2019; Ferranti et al., 2015; Grams et al., 2017): During the Greenland Blocking regime (associated with NAO− conditions) Z1000 ensemble spread is reduced by up to 1,800 gpm^2 near

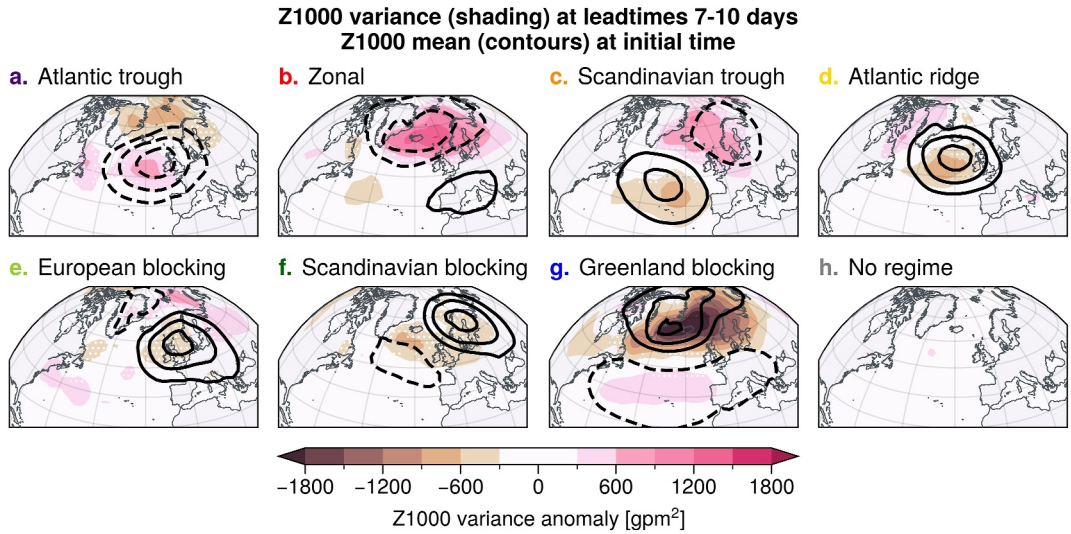

Figure 1. Composite mean of 1,000 hPa geopotential height (Z1000) ensemble variance (color-shaded), grouped by the weather regime at initialization time, averaged over lead times 7–10 days. Z1000 ensemble mean anomalies at initialization time denoted by black contours (50 gpm spacing, negative values dashed). White stippling indicates regions where color-shaded Z1000 ensemble variance anomalies are not statistically robust (based on the 95% confidence level using bootstrapping with 2,000 samples). The number of ensemble forecasts per regime are 154 (AT), 139 (ZO), 142 (ScTr), 141 (AR), 93 (EuBL), 114 (ScBL), 118 (GL) and 296 (no).

the characteristic anticyclone (about 25% relative to climatological spread; statistically significant at a 95% confidence level, derived via bootstrapping). During the Zonal regime (associated with NAO+) spread is increased near the North Atlantic jet by up to 1,000 gpm^2 (about 15%; statistically significant at a 95% confidence level). In general, regions of negative Z1000 anomalies (i.e., cyclonic patterns) tend to be associated with higher forecast uncertainty than regions of positive Z1000 anomalies (i.e., anticyclonic patterns). This is consistent with Spaeth et al. (2024) who argue that it is the occurrence of individual storms that contribute to reduced predictability of cyclonic patterns. Compared to high pressure systems, synoptic-scale storms tend to be smaller in size, are associated with a more rapid intensification and decay, propagate faster and have larger Z1000 gradients. Given the high degree of spatial and temporal de-correlation between individual ensemble members at these lead times, this can result in large ensemble variance.

While Z1000 ensemble mean and ensemble variance patterns are spatially aligned to a relatively high degree, their relation can be more complex in detail. For instance, anomalous ensemble variance can also appear in regions of near-zero ensemble mean signals (e.g., reduced variance around the North Sea during Greenland Blocking). Likewise, ensemble mean anomalies can be associated with near-zero ensemble variance anomalies (e.g., northern Europe during European Blocking). Furthermore, the relation between ensemble mean and ensemble variance can be non-linear: for example, Atlantic Trough and Atlantic Ridge are characterized by approximately opposite mean anomalies around the eastern Atlantic (up to about −180 gpm and +160 gpm, respectively), but they are both associated with reduced variance over the Greenland sea (about −800 gpm^2 and −300 gpm^2), suggesting that the synoptic-scale dynamics of the two regimes are not simply anti-symmetric.

Figure 2 reveals that 2-m-temperature (T2m) ensemble variance anomaly patterns are qualitatively different compared to Z1000, suggesting that they involve different mechanisms. In general, T2m variance is smaller over the ocean than over land (see supplementary Figure S2 in Supporting Information S1 for the spread climatology). Over land, cold anomalies tend to be associated with increased T2m variance and vice versa. For instance, Greenland blocking is followed by cold anomalies (up to −3 K) and increased T2m variance (up to +7 K^2, or +40% relative to climatological spread) over North-East Europe. However, there are many other regions where T2m ensemble mean and variance are not well-correlated. For example, North-East Canada experiences warm anomalies during Greenland Blocking and during European Blocking, but variance is decreased during the former and increased during the latter.

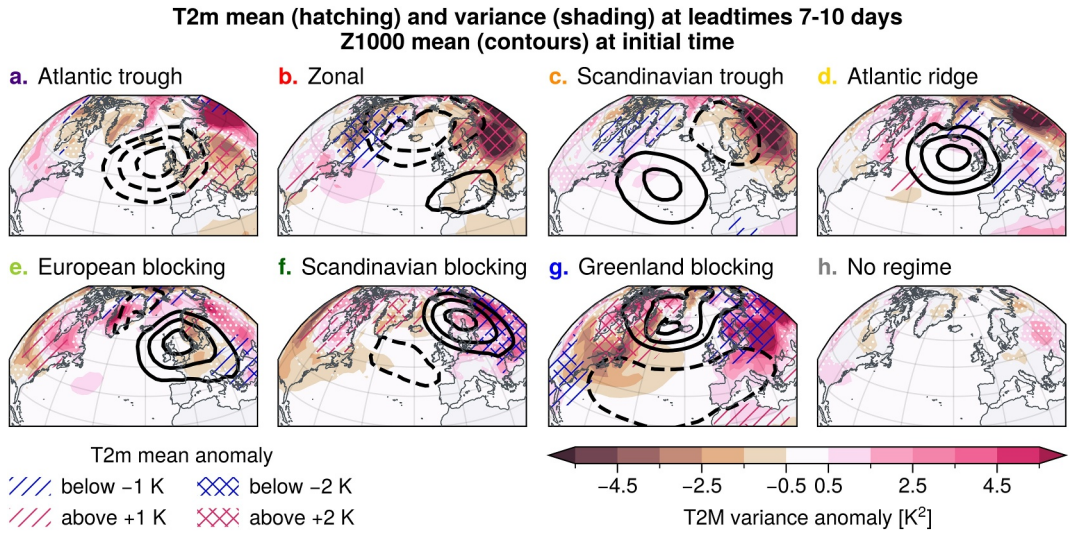

Figure 2. Same as Figure 1, but with color-shading indicating anomalies of 2-m temperature (T2m) variance. White stippling indicates regions where color-shaded T2m ensemble variance anomalies are not statistically robust (based on the 95% confidence level using bootstrapping with 2,000 samples). Anomalies of T2m ensemble mean are hatched (cold anomalies in blue, warm anomalies in red). Z1000 ensemble mean anomalies at time of initialization are marked by contours as in Figure 1.

Our results are consistent with the hypothesis that T2m ensemble variance anomalies are predominantly driven by advection, that is, sensitive to horizontal temperature gradients. Over the ocean, temperature gradients are generally smaller in magnitude. When marine air is advected to the continent, such as from the ocean to central Europe during the Zonal regime and Atlantic Trough, this leads to smaller variance in predicted T2m among ensemble members. In contrast, advection of continental air leads to larger ensemble variance. For instance, during Greenland Blocking, European weather is typically characterized by easterly winds and increased risk for cold-air outbreaks. Such cold spells are associated with strong temperature anomalies, but especially at extended-range forecast lead times, their exact location, strength and timing is differently simulated by ensemble members, leading to large ensemble spread. In addition, Greenland blocking often transitions to either Atlantic Ridge (about 1/6 of the time) or Atlantic Trough (about 1/8 of the time, see supplementary Figure S3 in Supporting Information S1 of Osman et al., 2023). Over Europe, Atlantic Ridge and Atlantic Trough are associated with opposite flow directions (easterly vs. westerly) and hence opposite temperature anomalies (cold vs. mild), which can additionally contribute to large T2m ensemble spread following Greenland blocking (see supplementary Text S3 in Supporting Information S1 for a comparison of Greenland blocking transitions to Atlantic Ridge vs. to Atlantic Trough).

This illustrates that the limited life time of each weather regime and a subsequent transition of ensemble members to different regimes can also contribute to large ensemble spread. In principle, such "inter-regime variance" could be the reason for the Z1000 and T2m ensemble variance anomalies at 7–10 days lead time that were presented in Figures 1 and 2. However, as we will show below, a substantial part is resulting from the inherent "intra-regime variance", that is, the variance across a single given weather regime.

Figure 3 presents the evolution of intra- and inter-regime ensemble variance over lead time for Z1000 over northern Europe and for T2m over central Europe. For each ensemble forecast and lead time, members are categorized based on their regime. Inter-regime variance is computed as the variance across these regime-group means, while intra-regime variance corresponds to the variance within members of a single regime-group. Note that the categorization differs from the analyses of Figures 1 and 2, because forecasts are not only grouped based on the regime at forecast initial time, but each group can include a different set of members at different lead times. Consistent with Figure 1, the Zonal and Scandinavian Trough regimes show largest intra-regime Z1000 ensemble variance over northern Europe beyond about 7 days lead time. Greenland Blocking is associated with the smallest amount of Z1000 ensemble variance. As before, this is somewhat opposite to T2m ensemble variance over Europe, which on average is twice as large during Greenland blocking than during Zonal. In fact, T2m intra-

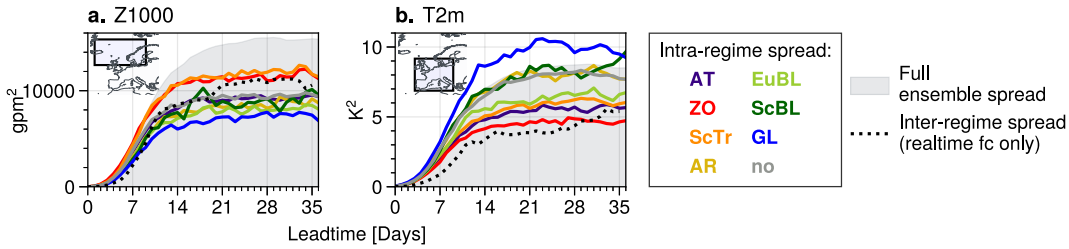

Figure 3. Evolution of intra-regime ensemble variance of (a) 1,000 hPa geopotential height over northern Europe and (b) 2-m temperature over central Europe (55–75°N/−20 to 20°E and 35–60°N/−10 to 20°E, respectively, indicated in map insets). For each lead time and ensemble forecast, the intra-regime variance is computed only across ensemble members sharing the same weather regime at that day. Ensemble variances are first computed for each grid-point and then averaged over the respective region. Gray shading denotes the ensemble variance computed across all ensemble members, that is, the full ensemble variance. Dotted black lines denote the average inter-regime spread (realtime forecasts only), that is, the variance between the regime-grouped averages within an ensemble.

regime variance during Greenland Blocking is even larger (saturating at 10 K^2) than the average total ensemble spread (saturating at 8 K^2).

This regime-dependence of forecast uncertainty is in good agreement with Figures 1 and 2, where forecasts were categorized based on the regime at initial time. Therefore, we conclude that intra-regime variance can represent a major contribution to the full ensemble variance anomaly patterns. Inter-regime variance is of similar magnitude as intra-regime spread. In general, large inter-regime variance indicates that weather regimes form distinct groups of ensemble members in terms of the chosen metric of interest. It is therefore not surprising to see Z1000 inter-regime variance being relatively larger than T2m inter-regime variance, as the regimes were constructed based on geopotential height in the first place and regimes are therefore well-separated somewhat by construction.

4. Uncertainty of Weather Regimes Under Different Stratospheric Initial Conditions

In Section 3 we have shown that different weather regimes are associated with distinctly different signatures in ensemble variance in Z1000 and T2m. The occurrence of weather regimes is known to be influenced by certain teleconnection patterns, such as the downward impact following weak or strong stratospheric polar vortex states (see introduction). In this section, we therefore consider the additional effect of the remote impact of the state of the stratosphere on forecast uncertainty of weather regimes.

Figure 4 shows climatological weather regime frequencies in forecasts and, in addition, changes in frequency at lead times of 3–4 weeks when forecasts are initialized with specific stratospheric polar vortex conditions. Here,

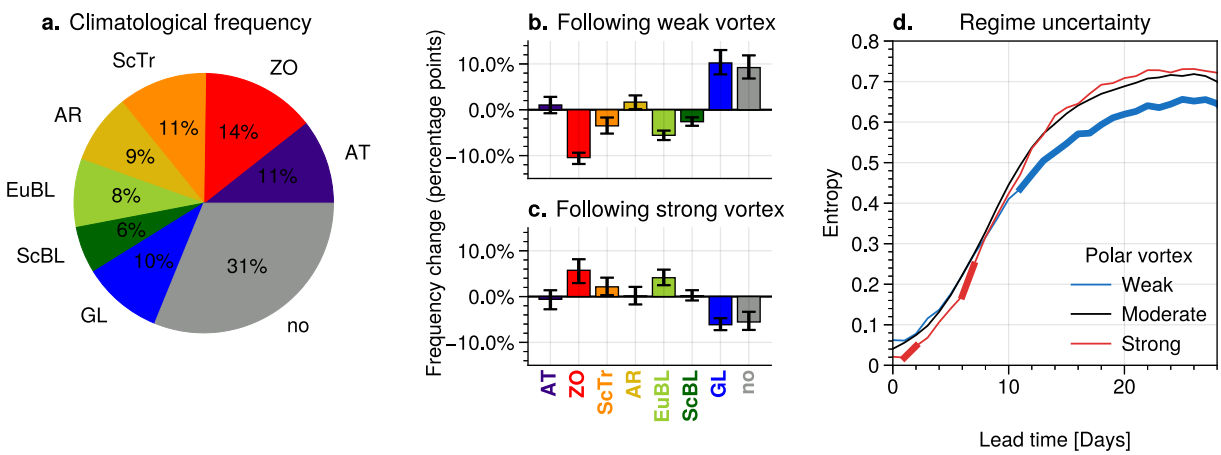

Figure 4. (a) Climatological frequency of weather regimes in ECMWF S2S forecasts averaged over lead times 14–27 days. (b), (c) Percentage point changes in predicted regime frequencies following (b) weak and (c) strong stratospheric polar vortex conditions. Error bars indicate 95% confidence intervals, obtained via bootstrapping. (d) Regime uncertainty, as quantified by the normalized entropy, for forecasts initialized under weak, moderate and strong polar vortex conditions (see text for definition). Thicker lines indicate where differences to "moderate" are statistically significant at a 95% confidence level, obtained via bootstrapping.

weak, moderate and strong polar vortex conditions are classified based on the ERA5 zonal-mean zonal wind at 10 hPa and 60°N, u_{60}^{10} (weak: $u_{60}^{10} < 0$ ms^{-1}, resulting in 86 forecasts; strong: $u_{60}^{10} > 51.82$ ms^{-1}, ensuring the same number of forecasts as for weak; moderate vortex: 1,025 forecasts). We generally find the same signals as Beerli and Grams (2019), in the sense that a strong polar vortex is followed by increased occurrence of the Zonal regime, whereas a weak polar vortex is followed by increased occurrence of Greenland blocking. It is perhaps surprising that there is a relatively large impact on the frequency of "no regime" (more often under weak vortex), though this impact seems to be larger in forecasts than ERA5 (compare to Figure 6 in Beerli & Grams, 2019) and its relative change compared to climatology is smaller (+30%) compared to Greenland blocking (+103%) and Zonal (−74%). The increased occurrence of no regime corresponds to generally weaker geopotential anomalies with associated weaker projections onto the regimes during periods of a weakened stratospheric polar vortex.

Moreover, there are about 7% more regime transitions observed under weak stratospheric polar vortex conditions compared to moderate vortex conditions (not shown), which further manifests in more days being classified as "no regime".

In addition to the mean frequency of regimes, we quantify the uncertainty of predicted regimes, in particular under different polar vortex conditions. To quantify the uncertainty associated with the classification into distinct weather regimes we use the information entropy measure introduced in Section 2.4. Figure 4d shows that after about 10 days weak polar vortex states are on average associated with reduced uncertainty in the predicted regime, as indicated by smaller entropy. This is consistent with the fact that forecasts initialized during weak polar vortex states are associated with increased occurrence of Greenland blocking and "no regime". However, entropy is not reduced under strong vortex conditions despite increased occurrence of the Zonal regime. This seems to be due to reduced occurrence of "no regime" (from 29% to 22%), which climatologically occurs most often. As a result, ensemble members are more evenly distributed among the eight regimes, which increases entropy.

5. Discussion

Our results have highlighted the added value of decomposing subseasonal forecast uncertainty over the North Atlantic-European sector during Boreal winter into uncertainty of the large-scale flow and uncertainty of near-surface weather given a particular large-scale flow pattern. Here, the large-scale flow was categorized by the eight North Atlantic-European weather regimes introduced by Grams et al. (2017). We now return to the questions formulated in the introduction.

5.1. How Does Forecast Uncertainty of Near-Surface Weather at Lead Times of 2–6 weeks Depend on Weather Regimes?

First, our analysis revealed distinct patterns of Z1000 and T2m forecast uncertainty inherently related to different weather regimes. These patterns of forecast uncertainty are found independent of whether forecasts categorized based on the initial regime (see Figures 1 and 2) or based on the regime at each day, separately (see Figure 3). The fact that the uncertainty patterns of Z1000 and T2m are quite distinct suggests that these uncertainties are driven by different underlying mechanisms. Z1000 ensemble variance seems to be mostly related to the occurrence of storms, while T2m ensemble variance seems to be predominantly driven by advection.

5.2. How Does Forecast Uncertainty of Weather Regimes Depend on the Circulation in the Polar Stratosphere?

Second, we introduced weather regime entropy as a quantitative measure for forecast uncertainty of weather regimes. The results showed that weak polar vortex states on average reduce weather regime entropy at extended-range lead times, implying a reduction of forecast uncertainty (see Figure 4). From this perspective, weak polar vortex episodes, including sudden stratospheric warmings (SSWs), form windows-of-opportunity for subseasonal prediction. However, these results require careful interpretation, as the reduction in entropy is partly due to a more frequent occurrence of "no regime", which likely is a result of weaker geopotential anomalies and more frequent regime transitions. Moreover, the reduced uncertainty only corresponds to an actual increase in predictability if the model is reliable, that is, if predicted regime likelihoods following weak polar vortex states match corresponding regime frequencies in reanalysis data. A corresponding reliability diagram did, unfortunately, not allow robust conclusions whether the model is reliable, over-confident or under-confident (not shown) due to the small hindcast ensemble size and the scarcity of weak vortex events in reanalysis data. Nevertheless, the average

increase in Greenland blocking frequency following weak polar vortex states is consistent between forecasts (from 10% climatology to 18%) and ERA5 (from 8% climatology to 18%).

While both weak and strong vortex states project onto the NAO, their impact on forecast uncertainty of near-surface weather is more nuanced. Specifically, weak vortex states such as SSWs are followed by increased Greenland Blocking frequency over the course of several weeks, which can enhance regime predictability (Büeler et al., 2021). However, Greenland Blocking is inherently associated with higher forecast uncertainty of T2m over Europe. This may create situations where the weather regime following a SSW has increased predictability at extended-range lead times, but T2m forecast skill (e.g., in terms of root-mean-squared-error) does in fact decrease. This has indeed been shown to be the case: Büeler et al. (2020) and Domeisen, Butler, et al. (2020) found decreased T2m extended-range forecast skill over Central Europe following weak polar vortex states, although there are differences between models. While the authors have discussed flow-dependent model deficiencies as a possible reason, our results suggest an additional interpretation: T2m skill is expected to decrease following SSWs even in a perfect model, due to the inherently increased T2m variance that is associated with the tropospheric weak vortex signatures. This increase in T2m variance could be driven, first, by the westward advection of continental air and, second, by co-projections of Greenland blocking onto either Atlantic Trough or Atlantic Ridge, which are related to opposite temperature anomalies over central Europe (see supplementary Text S2 in Supporting Information S1). Finally, we note that the difference between large-scale (related to regimes) and small-scale (related to Z1000 and T2m) predictability can manifest even in individual SSW cases: supplementary Text S3 in Supporting Information S1 briefly discusses how the February 2018 SSW (e.g., González-Alemán et al., 2022; Karpechko et al., 2018; Kautz et al., 2020) was associated with reduced regime uncertainty, reduced Z1000 uncertainty and increased T2m uncertainty in forecasts.

Data Availability Statement

ERA5 re-analysis (Hersbach et al., 2020, 2023a, 2023b) and the S2S forecast data (Vitart et al., 2017) used in this study can be accessed via the websites of the Copernicus Climate Data Store and ECMWF.

Acknowledgments

JS, CMG and TB have been supported by the Deutsche Forschungsgemeinschaft (DFG; grant no. SFB/TRR165,"Waves to Weather"). The contribution of MO was supported by Axpo Solutions AG. The contribution of CMG is funded by the Helmholtz Association as part of the Young Investigator Group "Sub-seasonal Predictability: Understanding the Role of Diabatic Outflow" (SPREADOUT, grant VH-NG-1243). This work is based on S2S data. S2S is a joint initiative of the World Weather Research Programme (WWRP) and the World Climate Research Programme (WCRP). We thank the two anonymous reviewers for their constructive comments. Open Access funding enabled and organized by Projekt DEAL.

References

Afargan-Gerstman, H., & Domeisen, D. I. (2020). Pacific modulation of the North Atlantic storm track response to sudden stratospheric warming events. *Geophysical Research Letters*, *47*(2), 1–10. https://doi.org/10.1029/2019GL085007

Baldwin, M. P., & Dunkerton, T. J. (1999). Propagation of the Arctic oscillation from the stratosphere to the troposphere. *Journal of Geophysical Research*, *104*(D24), 30937–30946. https://doi.org/10.1029/1999JD900445

Baldwin, M. P., & Dunkerton, T. J. (2001). Stratospheric harbingers of anomalous weather regimes. *Science*, *294*(5542), 581–584. https://doi.org/10.1126/science.1063315

Beerli, R., & Grams, C. M. (2019). Stratospheric modulation of the large-scale circulation in the Atlantic–European region and its implications for surface weather events. *Quarterly Journal of the Royal Meteorological Society*, *145*(725), 3732–3750. https://doi.org/10.1002/qj.3653

Büeler, D., Beerli, R., Wernli, H., & Grams, C. M. (2020). Stratospheric influence on ECMWF sub-seasonal forecast skill for energy-industry-relevant surface weather in European countries. *Quarterly Journal of the Royal Meteorological Society*, *146*(733), 3675–3694. https://doi.org/10.1002/qj.3866

Büeler, D., Ferranti, L., Magnusson, L., Quinting, J. F., & Grams, C. M. (2021). Year-round sub-seasonal forecast skill for Atlantic–European weather regimes. *Quarterly Journal of the Royal Meteorological Society*, *147*(741), 4283–4309. https://doi.org/10.1002/qj.4178

Butler, A., Charlton-Perez, A., Domeisen, D. I., Garfinkel, C., Gerber, E. P., Hitchcock, P., et al. (2019). Sub-seasonal predictability and the stratosphere. In *Sub-seasonal to seasonal prediction* (pp. 223–241). Elsevier. https://doi.org/10.1016/B978-0-12-811714-9.00011-5

Charlton-Perez, A. J., Ferranti, L., & Lee, R. W. (2018). The influence of the stratospheric state on North Atlantic weather regimes. *Quarterly Journal of the Royal Meteorological Society*, *144*(713), 1140–1151. https://doi.org/10.1002/qj.3280

Cover, T. M., & Thomas, J. A. (2005). Entropy, relative entropy, and mutual information. In *Elements of information theory* (pp. 13–55). Wiley. https://doi.org/10.1002/047174882X.ch2

Domeisen, D. I. V., Butler, A. H., Charlton-Perez, A. J., Ayarzagüena, B., Baldwin, M. P., Dunn-Sigouin, E., et al. (2020a). The role of the stratosphere in subseasonal to seasonal prediction: 2. Predictability arising from stratosphere-troposphere coupling. *Journal of Geophysical Research: Atmospheres*, *125*(2), 1–20. https://doi.org/10.1029/2019JD030923

Domeisen, D. I. V., Grams, C. M., & Papritz, L. (2020b). The role of North Atlantic–European weather regimes in the surface impact of sudden stratospheric warming events. *Weather and Climate Dynamics*, *1*(2), 373–388. https://doi.org/10.5194/wcd-1-373-2020

Ferranti, L., Corti, S., & Janousek, M. (2015). Flow-dependent verification of the ECMWF ensemble over the Euro-Atlantic sector. *Quarterly Journal of the Royal Meteorological Society*, *141*(688), 916–924. https://doi.org/10.1002/qj.2411

Fortin, V., Abaza, M., Anctil, F., & Turcotte, R. (2014). Why should ensemble spread match the RMSE of the ensemble mean? *Journal of Hydrometeorology*, *15*(4), 1708–1713. https://doi.org/10.1175/JHM-D-14-0008.1

González-Alemán, J. J., Grams, C. M., Ayarzagüena, B., Zurita-Gotor, P., Domeisen, D. I. V., Gómara, I., et al. (2022). Tropospheric role in the predictability of the surface impact of the 2018 sudden stratospheric warming event. *Geophysical Research Letters*, *49*(1), 1–10. https://doi.org/10.1029/2021GL095464

Grams, C. M., Beerli, R., Pfenninger, S., Staffell, I., & Wernli, H. (2017). Supplement: Balancing Europe's wind-power output through spatial deployment informed by weather regimes. *Nature Climate Change*, *7*(8), 557–562. https://doi.org/10.1038/nclimate3338

Hauser, S., Teubler, F., Riemer, M., Knippertz, P., & Grams, C. M. (2023). Towards a holistic understanding of blocked regime dynamics through a combination of complementary diagnostic perspectives. *Weather and Climate Dynamics*, *4*(2), 399–425. https://doi.org/10.5194/wcd-4-399-2023

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., et al. (2023a). ERA5 hourly data on pressure levels from 1940 to present. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*. https://doi.org/10.24381/cds.bd0915c6

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., et al. (2023b). ERA5 hourly data on single levels from 1940 to present. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*. https://doi.org/10.24381/cds.adbb2d47

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Karpechko, A. Y., Charlton-Perez, A., Balmaseda, M., Tyrrell, N., & Vitart, F. (2018). Predicting sudden stratospheric warming 2018 and its climate impacts with a multimodel ensemble. *Geophysical Research Letters*, *45*(24), 13538–13546. https://doi.org/10.1029/2018GL081091

Kautz, L., Polichtchouk, I., Birner, T., Garny, H., & Pinto, J. G. (2020). Enhanced extended-range predictability of the 2018 late-winter Eurasian cold spell due to the stratosphere. *Quarterly Journal of the Royal Meteorological Society*, *146*(727), 1040–1055. https://doi.org/10.1002/qj.3724

Labitzke, K. (1965). On the mutual relation between stratosphere and troposphere during periods of stratospheric warmings in winter. *Journal of Applied Meteorology*, *4*(1), 91–99. https://doi.org/10.1175/1520-0450(1965)004⟨0091:otmrbs⟩2.0.co;2

Lee, S. H., Furtado, J. C., & Charlton-Perez, A. J. (2019). Wintertime North American weather regimes and the Arctic stratospheric polar vortex. *Geophysical Research Letters*, *46*(24), 14892–14900. https://doi.org/10.1029/2019GL085592

Leutbecher, M., & Palmer, T. (2008). Ensemble forecasting. *Journal of Computational Physics*, *227*(7), 3515–3539. https://doi.org/10.1016/j.jcp.2007.02.014

Mariotti, A., Baggett, C., Barnes, E. A., Becker, E., Butler, A., Collins, D. C., et al. (2020). Windows of opportunity for skillful forecasts subseasonal to seasonal and beyond. *Bulletin of the American Meteorological Society*, *101*(5), E608–E625. https://doi.org/10.1175/BAMS-D-18-0326.1

Matsueda, M., & Palmer, T. N. (2018). Estimates of flow-dependent predictability of wintertime Euro-Atlantic weather regimes in medium-range forecasts. *Quarterly Journal of the Royal Meteorological Society*, *144*(713), 1012–1027. https://doi.org/10.1002/qj.3265

Michel, C., & Rivière, G. (2011). The link between rossby wave breakings and weather regime transitions. *Journal of the Atmospheric Sciences*, *68*(8), 1730–1748. https://doi.org/10.1175/2011JAS3635.1

Osman, M., Beerli, R., Büeler, D., & Grams, C. M. (2023). Multi-model assessment of sub-seasonal predictive skill for year-round Atlantic–European weather regimes. *Quarterly Journal of the Royal Meteorological Society*, *149*(755), 2386–2408. https://doi.org/10.1002/qj.4512

Palmer, T. (2017). The primacy of doubt: Evolution of numerical weather prediction from determinism to probability. *Journal of Advances in Modeling Earth Systems*, *9*(2), 730–734. https://doi.org/10.1002/2017MS000999

Palmer, T. N. (2002). The economic value of ensemble forecasts as a tool for risk assessment: From days to decades. *Quarterly Journal of the Royal Meteorological Society*, *128*(581), 747–774. https://doi.org/10.1256/0035900021643593

Robbins, J., Cunningham, C., Dankers, R., DeGennaro, M., Dolif, G., Duell, R., et al. (2018). *Communication and dissemination of forecasts and engaging user communities*. Elsevier Inc. https://doi.org/10.1016/B978-0-12-811714-9.00019-X

Roberts, C. D., Balmaseda, M. A., Ferranti, L., & Vitart, F. (2023). Euro-Atlantic weather regimes and their modulation by tropospheric and stratospheric teleconnection pathways in ECMWF reforecasts. *Monthly Weather Review*, *151*(10), 2779–2799. https://doi.org/10.1175/MWR-D-22-0346.1

Rupp, P., Loeffel, S., Garny, H., Chen, X., Pinto, J. G., & Birner, T. (2022). Potential links between tropospheric and stratospheric circulation extremes during early 2020. *Journal of Geophysical Research: Atmospheres*, *127*(3), e2021JD035667. https://doi.org/10.1029/2021JD035667

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Slingo, J., & Palmer, T. (2011). Uncertainty in weather and climate prediction. *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences*, *369*(1956), 4751–4767. https://doi.org/10.1098/rsta.2011.0161

Spaeth, J., & Birner, T. (2022). Stratospheric modulation of Arctic Oscillation extremes as represented by extended-range ensemble forecasts. *Weather and Climate Dynamics*, *3*(3), 883–903. https://doi.org/10.5194/wcd-3-883-2022

Spaeth, J., Rupp, P., Garny, H., & Birner, T. (2024). Stratospheric impact on subseasonal forecast uncertainty in the Northern extratropics. *Communications Earth & Environment*, *5*(1), 126. https://doi.org/10.1038/s43247-024-01292-z

Vial, J., Osborn, T. J., & Lott, F. (2013). Sudden stratospheric warmings and tropospheric blockings in a multi-century simulation of the IPSL-CM5A coupled climate model. *Climate Dynamics*, *40*(9–10), 2401–2414. https://doi.org/10.1007/s00382-013-1675-2

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., et al. (2017). The subseasonal to seasonal (S2S) prediction project Database. *Bulletin of the American Meteorological Society*, *98*(1), 163–173. https://doi.org/10.1175/BAMS-D-16-0017.1

Zhu, Y., Toth, Z., Wobus, R., Richardson, D., & Mylne, K. (2002). The economic value of ensemble-based weather forecasts. *Bulletin of the American Meteorological Society*, *83*(1), 73–83. https://doi.org/10.1175/1520-0477(2002)083⟨0073:TEVOEB⟩2.3.CO;2