Stratosphere-Troposphere Coupling and Its Impacts on Northern Winter Subseasonal-to-Seasonal Forecast Uncertainty

Jonas Späth

München 2024

Stratosphere-Troposphere Coupling and Its Impacts on Northern Winter Subseasonal-to-Seasonal Forecast Uncertainty

Jonas Späth

Dissertation
an der Fakultät für Physik
der Ludwig-Maximilians-Universität
München

vorgelegt von
Jonas Späth
aus Bad Tölz

München, den 19. August 2024

Erstgutachter: Prof. Dr. Thomas Birner
Zweitgutachter: Dr. Hella Garny
Tag der mündlichen Prüfung: 30. September 2024

Abstract

Skill of mid-latitude weather predictions largely decays within the first two forecast weeks. Nonetheless, remote coupling with slower-evolving Earth system components can provide predictability beyond the local initial condition memory, at the so-called subseasonal-to-seasonal (S2S) timescale, which spans approximately from two weeks to two months. This thesis delves into a key source of S2S predictability for mid-latitude weather: stratosphere-troposphere coupling, specifically through the strong circumpolar winds in the Arctic stratosphere known as the polar vortex.

Studies of stratosphere-troposphere coupling are often constrained by the scarcity of extreme events in observational records, especially for sudden stratospheric warmings (SSWs), which occur approximately once every second winter. Leveraging a large set of state-of-the-art S2S ensemble forecasts, this thesis improves the statistical characterization of SSWs by employing a sample size roughly 150 times larger than what is available from observations. This approach allows to quantify aspects of stratosphere-troposphere coupling that would be unfeasible to derive from pure observations. Here, this set of SSWs is used to compute the fraction of attributable risk, which demands particularly large statistics. First, it is found that approximately every second split-type SSW and every third displacement-type SSW may be attributable to corresponding wave activity events of tropospheric origin, underscoring that SSW formation relies on suitable conditions in both troposphere and stratosphere. Second, the impact of weak and strong stratospheric polar vortex events on the frequency of large-scale circulation extremes in the troposphere – characterized by the Arctic Oscillation (AO) index – is investigated. The results show that SSWs increase the likelihood of negative AO extremes (exceeding 3 standard deviations) by about 60%, which is expected to also translate into regional weather extremes with potentially high socio-economic impacts. In turn, approximately one fourth of negative AO extremes in winter may be attributable to preceding SSWs.

Furthermore, this thesis reveals that both weak and strong stratospheric polar vortex events affect forecast uncertainty at S2S timescales, as measured by anomalous ensemble spread. Facilitated by the large S2S ensemble dataset, this highlights an important additional aspect of predictability arising from stratosphere-troposphere coupling, which extends beyond the previously recognized shifts in mean weather conditions. Specifically, weak polar vortex events (such as SSWs) are followed by reduced 1000 hPa geopotential height (Z1000) ensemble spread at 3-5 weeks lead time over northern Europe and increased spread over southern Europe, with anomalies reaching up to about 25% relative to clima-

tological ensemble spread. It is shown that these spread anomalies arise from the known southward shift of the North Atlantic storm track following SSWs, which is associated with variations in synoptic-scale variability and extratropical cyclone frequency. Notably, an analysis of ensemble spread of near-surface temperature (T2m) reveals different anomaly patterns compared to those of Z1000: weak polar vortex events are followed by increased ensemble spread over most parts of Europe, indicating an increase of forecast uncertainty.

Finally, this thesis generalizes the analysis of ensemble spread at S2S timescales to the context of North Atlantic-European weather regimes. While each of these regimes is associated with robust patterns of Z1000 and T2m spread anomalies, it is shown that weak polar vortex events also reduce the uncertainty of weather regime forecasts by about 10%, as measured by a novel metric based on information entropy.

Overall, this thesis illustrates how S2S ensemble forecasts can advance the analysis and attribution of extreme events. The applicability of this approach extends beyond stratosphere-troposphere coupling and may offer insight into other high-impact weather phenomena. Additionally, this thesis identified flow-dependent variations in forecast uncertainty at S2S lead times, contributing to advancements of the broader field of S2S predictability.

Zusammenfassung

Die Güte von Wettervorhersagen in den mittleren Breiten nimmt innerhalb der ersten beiden Vorhersagewochen stark ab. Dennoch kann eine gewisse Vorhersagbarkeit auch auf längeren Zeitskalen bestehen, wenn das Wettergeschehen mit sich langsamer verändernden Komponenten des Erdsystems gekoppelt ist. Diese Arbeit untersucht, wie solche Kopplungen zur Vorhersagbarkeit auf der subsaisonalen-bis-saisonalen (S2S) Zeitskala beitragen, was Vorhersagezeiten von etwa zwei Wochen bis zwei Monaten einschließt. Dabei liegt der Fokus auf einer der wichtigsten Quellen für S2S-Vorhersagbarkeit in den mittleren Breiten: der Kopplung zwischen Stratosphäre und Troposphäre, im Zusammenhang mit den starken zirkumpolaren Winden in der arktischen Stratosphäre, dem sogenannten Polarwirbel.

Studien zur Stratosphäre-Troposphäre-Kopplung sind häufig dadurch limitiert, dass Extremereignisse in Beobachtungsdatensätzen nur selten auftreten. Dies trifft insbesondere auf plötzliche Stratosphärenerwärmungen (im Englischen "sudden stratospheric warmings", SSWs) zu, welche in etwa einmal alle zwei Winter vorkommen. Diese Arbeit verbessert die statistische Charakterisierung von SSWs, indem ein umfangreicher Datensatz moderner S2S-Ensemblevorhersagen genutzt wird, wodurch etwa 150-mal mehr SSW-Ereignisse analysiert werden können als in Beobachtungsdaten zur Verfügung stehen. Dieser Ansatz ermöglicht es, Aspekte der Stratosphäre-Troposphäre-Kopplung zu quantifizieren, die sich nicht aus reinen Beobachtungsdaten ableiten lassen. Hier wird dieser SSW-Datensatz verwendet, um den Anteil des attributalen Risikos (im Englischen "fraction of attributable risk") zu berechnen, was besonders große Statistiken erfordert. Die Ergebnisse zeigen, dass sich in etwa jedes zweite Teilungs-SSW und jedes dritte Verschiebungs-SSW auf ein entsprechendes Wellenaktivitätsereignis troposphärischen Ursprungs zurückführen lässt. Dieses Resultat unterstreicht, dass das Auftreten von SSWs sowohl in der Troposphäre als auch in der Stratosphäre geeignete Bedingungen erfordert. Des Weiteren wird der Einfluss schwacher und starker stratosphärischer Polarwirbelereignisse auf die Häufigkeit großräumiger Zirkulationsextreme in der Troposphäre untersucht, hier definiert mittels der Arktischen Oszillation (AO). Die Ergebnisse zeigen, dass SSWs die Wahrscheinlichkeit negativer AO-Extrema (mehr als 3 Standardabweichungen) um etwa 60% erhöhen. Es ist zu erwarten, dass solche Ereignisse auch mit regionalen Wetterextremen einhergehen, welche mit großen sozioökonomischen Auswirkungen verbunden sein können. Darüber hinaus lassen sich etwa ein Viertel aller negativen AO-Extrema im Winter auf vorangehende SSWs zurückführen.

Ferner zeigt diese Arbeit, dass schwache und starke stratosphärische Polarwirbelereignisse

die Unsicherheit von S2S-Vorhersagen beeinflussen, gemessen an der Ensemblestreuung. Dieses Resultat, ermöglicht durch die Verwendung des umfangreichen Vorhersage-Datensatzes, hebt einen wichtigen zusätzlichen Aspekt der Vorhersagbarkeit hervor, der über die bisher bekannte Veränderung mittlerer Wetterbedingungen hinausgeht. So nimmt die Z1000-Ensemblestreuung in 3-5-Wochen-Vorhersagen im Anschluss an schwache Polarwirbelereignisse über Nordeuropa ab, während sie über Südeuropa zunimmt. Die Anomalien erreichen dabei eine Stärke von bis zu 25% relativ zur klimatischen Ensemblestreuung. Erklären lassen sich die Ensemblestreuungs-Anomalien durch die südwärts-Verschiebung der Zugbahn nordatlantischer Stürme, was mit Anomalien in synoptischer Variabilität und Häufigkeit extratropischer Zyklone einhergeht. Bemerkenswerterweise zeigt die Ensemblestreuung in Vorhersagen bodennaher Temperatur (T2m) andere Anomalie-Muster im Vergleich zu Z1000-Vorhersagen: Im Anschluss an schwache Polarwirbelereignisse erhöht sich die T2m-Ensemblestreuung über weiten Teilen Europas, was eine erhöhte Vorhersageunsicherheit impliziert.

Abschließend werden diese Analysen verallgemeinert, indem Variationen in der Ensemblestreuung auf S2S-Zeitskalen im Zusammenhang mit nordatlantisch-europäischen Wetterregimen untersucht werden. Jedes dieser Regime weist robuste Ensemblestreuungs-Anomalien in Z1000- und T2m-Vorhersagen auf. Außerdem nimmt die Unsicherheit von Wetterregime-Vorhersagen im Anschluss an schwache Polarwirbelereignisse um circa 10% ab, gemessen an einer aus der Informationstheorie übernommenen Metrik.

Diese Dissertation zeigt, wie S2S-Ensemblevorhersagen die Analyse und Attribution extremer Wetterereignisse voranbringen können. Dieser Ansatz ist prinzipiell nicht nur auf die Stratosphäre-Troposphäre-Kopplung anwendbar, sondern könnte darüber hinaus Einblicke in andere, bedeutende Wetterphänomene bieten. Des Weiteren zeigt diese Arbeit, wie die Unsicherheit in S2S-Vorhersagen von der großskaligen Zirkulation abhängt und leistet dadurch einen Beitrag zu Fortschritten im Bereich der S2S-Vorhersagbarkeit.

Contents

List of publications

Parts of the results of this thesis have been published or are under review according to the following articles:

- 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
- Rupp, P., Spaeth, J., Garny, H., & Birner, T. (2023). Enhanced Polar Vortex Predictability Following Sudden Stratospheric Warming Events. *Geophysical Research Letters*, *50*(17). https://doi.org/10.1029/2023GL104057
- Spaeth, J., Rupp, P., Garny, H., & Birner, T. (2024a). 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
- Rupp, P., Spaeth, J., Afargan-Gerstman, H., Büeler, D., Sprenger, M., & Birner, T. (2024). The impact of synoptic storm likelihood on European subseasonal forecast uncertainty and their modulation by the stratosphere. *EGUsphere [preprint]*, 1–19. https://doi.org/10.5194/egusphere-2024-1423
- Spaeth, J., Rupp, P., Osman, M., Grams, C. M., & Birner, T. (2024b). Flow-Dependence of Ensemble Spread of Subseasonal Forecasts Explored via North Atlantic-European Weather Regimes. *Geophysical Research Letters*, *51*(14). https://doi.org/10.1029/2024GL109733

Acronyms

AO Arctic Oscillation

BDC Brewer-Dobson circulation

ECMWF European Centre for Medium-Range Weather Forecasts

ENSO El Niño-Southern Oscillation

ERA5 ECMWF Reanalysis 5

FAR fraction of attributable risk

MJO Madden-Julian Oscillation

MSLP mean sea level pressure

NAO North Atlantic Oscillation

NAM Northern Annular Mode

NWP numerical weather prediction

p-SSW predicted sudden stratospheric warming

QBO Quasi-Biennial Oscillation

RPC ratio of predictable components

RMSE root mean squared error

RMSE-SS root mean squared error skill score

SNP signal-to-noise paradox

S2S subseasonal-to-seasonal

SST sea surface temperature

SPV strong polar vortex

SSW sudden stratospheric warming

T2m 2 meter temperature

UKMO UK Met Office

UNSEEN UNprecedented Simulated Extremes using ENsembles

$\mathbf{U}_{60}^{10}$ zonal-mean zonal wind at 60 °N and 10 hPa

Z500 500 hPa geopotential height

Z1000 1000 hPa geopotential height

Chapter 1

Introduction

1.1 Numerical weather prediction

History of weather prediction

Without weather prediction, thousands of flights would be delayed, millions of farmers would struggle to plan their harvests, and countless people would be caught unprepared in the face of extreme weather events. Importantly, successful weather prediction relies on collaborative efforts across diverse scientific disciplines. This collaboration encompasses the development of new measuring instruments, the advancement of mathematical and statistical methods, the continuous improvement of the physical understanding of the Earth system, the implementation of computer algorithms and the communication of forecasts to users. The common objective is to transform knowledge about the Earth system into accurate forecasts of future weather. Based on weather forecasts, it is increasingly possible today to predict extreme weather events, warn against them, and prepare accordingly. In agriculture, weather forecasts assist in determining optimal times for sowing, irrigation, and harvesting. In the energy sector, they are used to estimate upcoming energy supply and demand, thereby playing a crucial role in the transition towards renewable energy.

The endeavour to understand and predict weather is deeply rooted in human history. Early civilizations like the Babylonians (around 650 B.C.) and ancient Greeks[1] (around 350 B.C.), as well as religious texts[2] (cf. Matthew 16:2b-3), used observations and basic principles to make weather predictions, representing early prototypes for what would now be recognized as empirical forecasting. Later, in the course of the 19th century, scientific measurements of the atmosphere were employed, for example, to infer future low or high pressure weather conditions from air pressure tendencies.

In the 20th century, the birth of numerical weather prediction then heralded the era of

[1] For example, Aristotle developed theories about "Winds, their causes and effects" (cf. Chapter 4 in the 2nd book of his treatise "Meteorologica").

[2] "When it is evening, you say, 'It will be fair weather; for the sky is red.' And in the morning, 'It will be stormy today, for the sky is red and threatening."'

modern weather forecasting. Empirical models were replaced by the attempt to solve the physical equations that describe the temporal evolution of air masses in the Earth's atmosphere (Bjerknes, 1904). These equations are based on momentum, mass, and energy conservation in a fluid, with assumptions and approximations appropriate for the properties of the Earth. An approach to solve these equations numerically was first suggested by Lewis Fry Richardson in his article "Weather Prediction By Numerical Process" (Richardson, 1922). Nearly thirty years later, in 1950, a breakthrough was achieved when a team of meteorologists, mathematicians, and programmers – including Jule Charney, Philip Duncan Thompson, Larry Gates, Ragnar Fjørtoft, John von Neumann, and Klara Dan von Neumann – successfully created the first computer-based numerical weather forecast (Charney et al., 1950).

Principles and limits of numerical weather prediction

Numerical weather prediction employs a representation of the state of the Earth's atmosphere on a discrete spatial grid. By numerically solving the governing flow equations, this state is propagated forward in time, time step by time step. Skillful weather forecasts rely on two key elements: an accurate initial state of the Earth system (the initial conditions) and a suitable numerical model. The initial conditions are determined by assimilating observations into a prior forecast, aiming for the best estimate of the current atmospheric state. Using this state and the governing equations, the numerical model then predicts the future evolution of the atmosphere. While processes involving scales smaller than the grid size, such as cloud microphysics, turbulence, or radiation, are not explicitly resolved, they are represented using empirical formulas, known as parameterizations, to capture their interaction with the resolved flow. Fig. 1.1 provides an overview of these basic components involved in numerical weather prediction, from observations and the numerical model to the final forecast.

Despite continuous improvements of weather forecasts over recent decades (Bauer et al., 2015), predictability of the atmosphere remains limited. These limitations stem from the fundamental complexity of the underlying physical equations, characterized by the presence of coupled non-linear terms that make forecasts highly sensitive to initial conditions – a property known as chaos. This sensitivity implies that minor differences in initial conditions at the start of the forecast can lead to major differences across forecasts several days later. In meteorology, this concept is metaphorically referred to as the butterfly effect (Lorenz, 1972), picking up the idea that even the flap of a butterfly's wings can in principle fundamentally change the atmospheric state in the more distant future. However, uncertainties in today's weather forecasts are much larger than butterfly-scale effects: initial condition uncertainty arises from measurement errors and the sparsity of observations, while model uncertainty is primarily related to unresolved processes (cloud microphysics, turbulence, radiation, etc.) and numerical approximations. Today, the skill horizon of weather forecasts extends to about one week, known as the practical limit of predictability, with some differences across regions, seasons, variables and specific cases (Zhang et al., 2019). In

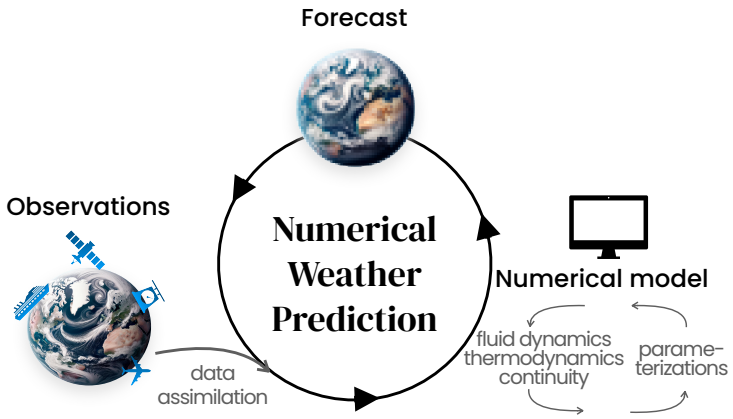

Figure 1.1: Overview of the key components of numerical weather prediction. Initial conditions are determined through data assimilation, combining a prior forecast with latest observations. This state feeds into a numerical model, comprising the governing flow equations (combining fluid dynamics, thermodynamics and continuity) and empirical formulae for subgrid-scale processes (parameterizations), which are then advanced in time to generate the next forecast.

addition to this practical limit, there is also a theoretical limit, known as the intrinsic limit of predictability. While the practical limit could be extended by reducing initial condition and model uncertainties in the future, the intrinsic limit is an inherent property of the atmosphere. The chaotic nature of the atmosphere thereby imposes a fundamental limit on the predictability of weather: even if model uncertainty could be completely eliminated and initial condition uncertainty could be reduced by 99% compared to current levels, the skill horizon of weather forecasts may not extend beyond about two weeks (Lorenz, 1969; Selz, 2019; Selz et al., 2022).

Forecasts across different time horizons

Operational weather forecast systems are usually optimized towards predictions at specific lead time ranges, because the relative importance of different processes can change over lead time. For example, short-term forecasts often only target a specific region, allowing for increased spatial resolution. On the other hand, long-term forecasts may benefit from simulating not only the atmosphere, but also other, more slowly evolving, components of the Earth system. Furthermore, frequent initialization of new forecasts may be more valuable for short-term forecasts, whereas physical constraints such as mass and energy conservation matter more at longer lead times.

Medium-range forecasts, covering lead times of up to approximately 1-2 weeks, may be considered the core of global weather forecasting. Medium-range forecasts generally exhibit a gradual decline in skill over lead time until the practical predictability limit is reached. However, reaching this limit (typically at around one week) does not imply that forecast

skill becomes virtually zero. Instead, skill typically saturates at a fairly low level, which is insufficient for many practical applications.

Nonetheless, recent years and decades have seen an increasing interest in forecasts beyond medium-range lead times, despite the low level of skill. The first attempts for seasonal forecasts date back to the 1970s (Davis, 1976; Davis, 1978). Seasonal forecasts, also referred to as long-range forecasts, target predictions of about one to several months ahead. Seasonal forecasting primarily relies on the idea that the ocean, and importantly sea surface temperatures, may be predictable for much longer lead times than the atmosphere. Additionally, seasonal forecasts are often communicated as temporally or spatially aggregated quantities (e.g., monthly averages instead of daily averages), which helps to smooth out unpredictable low-frequency fluctuations and thus results in higher skill.

This thesis mainly focuses on forecasts at lead times of approximately 3-6 weeks, known as subseasonal-to-seasonal (S2S) forecasts (or extended-range forecasts). S2S forecasts bridge the gap between medium-range and long-range forecasts, targeting lead times beyond two weeks but less than a full season. Forecasting at these lead times poses significant challenges, as the memory of initial conditions has largely vanished, yet the time frame is too short for the full effects of ocean-atmosphere coupling to unfold.

Fig. 1.2 presents forecast skill of 500 hPa geopotential height (Z500) in the Northern extratropics across different lead times, from the medium-, over the extended-, to the long-range. Z500 serves as a metric to evaluate predictions of the large-scale tropospheric flow. Skill is quantified in terms of the root mean squared error skill score (RMSE-SS) (Wilks, 2005, chapter 7.3.3), where RMSE-SS=1 for perfect forecasts (high skill) and RMSE-SS=0 for forecasts that are only as accurate as a multi-year, seasonally evolving, climatological average (low skill). At the medium-range, forecast skill quickly decays from about 0.9 at day 1 to about 0.2 at day 8. Beyond three weeks, daily RMSE-SS approaches zero and even turns slightly negative, implying that a climatological average outperforms the numerical forecast. Forecasts of weekly and monthly means show slightly better skill than daily forecasts. Specifically, RMSE-SS remains above zero for weekly averages up to weeks 3-4 and monthly averages throughout all three months. While these skill scores represent average values and individual cases may exhibit higher skill, the consistently low skill beyond the medium-range highlights the significant scientific and technical challenges associated with S2S and long-range predictions. Given these limitations, it becomes crucial to delve deeper into the complexities of S2S forecasting to understand its potential and limitations. This sets the context for introducing S2S prediction in the following section.

1.2 Subseasonal-to-seasonal prediction

Fig. 1.2 has demonstrated that average skill of subseasonal-to-seasonal (S2S) forecasts for mid-latitude weather in the troposphere is fairly limited. Nevertheless, there has been increasing interest in these forecasts, in particular across sectors that rely on long-term planning, including agriculture, energy supply and medical preparedness. For instance,

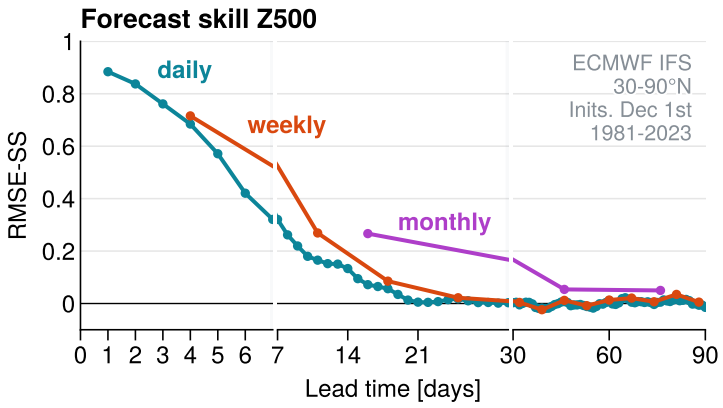

Figure 1.2: Lead time evolution of 500 hPa geopotential height forecast skill in terms of root mean squared error skill score (RMSE-SS) in the Northern extratropics (30° to 90°N latitude). Data is based on 44 seasonal forecasts from 1981 to 2023 initialized on 1 December, respectively, provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) (model cycle 43r1). Note the non-uniform lead time scaling. Scores are computed based on 1) daily instantaneous forecasts for single gridpoints, 2) 7-day averages with 25° longitude and 12.5° latitude rolling mean windows applied, denoted as weekly and 3) 30-day averages with 50° longitude and 25° latitude rolling mean windows applied, denoted as monthly.

farmers use these forecasts to help determine sowing and harvesting times, energy providers use them to anticipate fluctuations in renewable energy production, and public health officials use them to prepare for outbreaks of diseases such as Malaria or Dengue Fever, which are correlated with specific weather conditions.

While this thesis primarily concentrates on S2S time scales, many of these insights equally apply to long-range forecasting.

Predictability across different scales and processes

At early lead times forecast error growth tends to be related to processes that take place at spatial scales too small to be explicitly resolved on the model grid. Over lead time, these errors can spread, amplify and interact with increasingly larger spatial scales (cf. Zhang et al., 2007; Selz and Craig, 2015; Baumgart et al., 2019). Therefore, larger-scale features in the atmosphere tend to be predictable for longer lead times than smaller-scale features (see Fig. 1.3; taken from Buizza et al., 2015). This has important implications for how S2S forecasts should be interpreted: While medium-range forecasts can usually provide skillful predictions for weather at a specific location and at a specific time, skill of S2S forecasts is thus generally limited to temporal and/or spatial averages (recall Fig. 1.2). In this way, unpredictable high-frequency fluctuations average out over time, allowing the more predictable signals associated with larger scales to stand out.

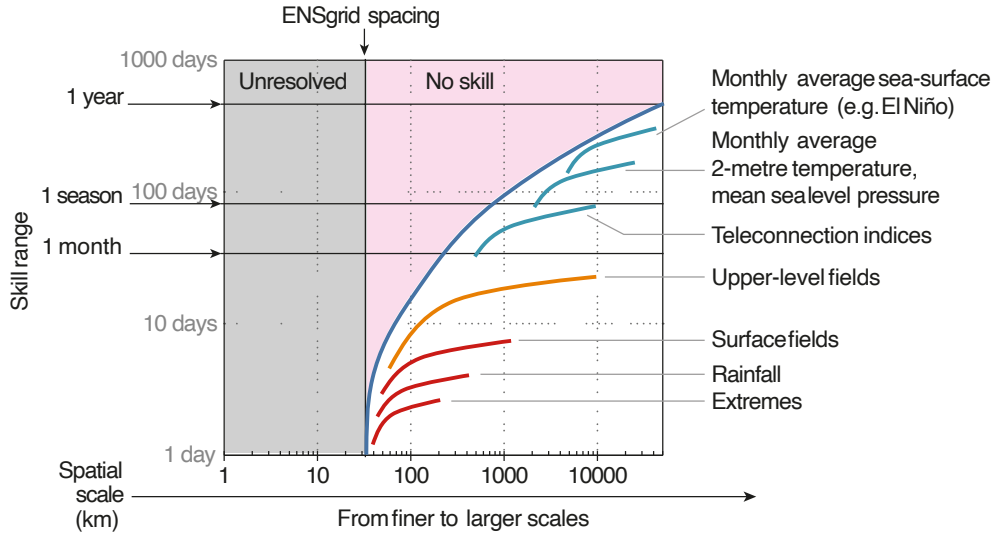

Figure 1.3: Predictability horizons for processes of different spatial scales. In general, larger spatial scales (see right-hand side of the diagram) tend to be predictable for longer lead times (see upper side of the diagram). Processes smaller than the grid spacing of the numerical model are unresolved (gray shading). Taken from Buizza et al. (2015).

Fig. 1.3 links the predictability of different spatial scales to specific processes in the atmosphere. Smaller-scale features, about 100 km in size (e.g., individual troughs, fronts with rain bands and wind gusts), are predictable for about one week, whereas larger-scale features, about 1000 km to 10 000 km in size, can be predictable for a few weeks or even months, such as planetary-scale variations in the jet stream or large-scale spatial averages of some atmospheric variables.

The scale-dependence of predictability implies that different atmospheric vertical layers are associated with different predictability time scales: While skillful predictions of tropospheric weather are limited to about one week, the stratospheric circulation (at about 10 km to 50 km altitude) is characterized by significantly larger spatial and temporal scales (Baldwin et al., 2003, also see chapter 2 of this thesis). Therefore, the stratospheric circulation tends to be predictable for longer lead times than the troposphere, which is central to the themes explored in this thesis.

In addition to the scale-dependence of predictability, other components of the Earth system, such as ocean dynamics, sea ice, or land surface properties can be predictable for longer lead times compared to atmospheric processes. While these phenomena are usually not crucial for medium-range weather forecasts, they can play an important role for S2S forecasts. Accurate representation of the coupling between atmosphere and such additional processes requires increased computational effort. However, it can significantly enhance predictability for the atmosphere at lead times where predictability would otherwise be virtually zero.

From traceable to climatic predictability

One fundamental conceptual distinction between medium-range and S2S forecasts is an altered understanding of predictability. Specifically, one may distinguish "traceable predictability" and "climatic predictability" (Toth and Buizza, 2019). Medium-range forecasts primarily focus on traceable predictability, targeting the prediction of specific, individual weather events at daily or even hourly resolution. For example, consider forecasting the trajectory of a specific cyclone. At early lead times, the cyclone's position and intensity may be known almost perfectly. With proceeding lead time, discrepancies between predicted and actual cyclone track emerge and these errors grow over lead time. In this example, traceable predictability refers to the ability to track the cyclone's evolution from the initial conditions to some later lead time, typically up to about one week, before errors become large and saturated. The intrinsic predictability limit implies that traceable predictability is not infinite even if initial condition and model errors would be extremely small.

In contrast, climatic predictability refers to the ability to make statements about the statistical occurrence of weather events. Rather than predicting the trajectory of a single cyclone, the emphasis shifts to predicting the frequency of such events, often considered as spatial or temporal averages. This form of predictability is also central to climate simulations, where the objective is not to predict exact weather conditions at a specific time but to determine the statistical occurrence of certain weather events, such as heatwaves, extreme storms, or droughts.

S2S forecasts predominantly leverage the principle of climatic predictability, although some large-scale phenomena may still exhibit traceable predictability. As a result, S2S forecasts should not be interpreted deterministically: while medium-range forecasts can often offer insights into specific weather events – like predicting that a particular cyclone over the West Atlantic will bring high wind speeds to Northern Germany next week – S2S forecasts should be interpreted statistically. For instance, they might indicate that the likelihood of storms in Northern Europe during week x is $y\%$ higher than usual for this time of year.

The distinction between traceable and climatic predictability is particularly relevant at S2S timescales, as traceable predictability has mostly vanished for many processes. However, the concept also applies to other timescales. For instance, in medium-range forecasts, the synoptic-scale flow may retain traceable predictability, whereas individual clouds and thunderstorms typically exhibit only climatic predictability.

Importance of initial versus boundary conditions

In S2S forecasts, the shift from traceable to climatic predictability is closely tied to a change in the relative significance of initial and boundary conditions. Medium-range forecasts aim for as accurate initial conditions as possible, exploiting the traceable predictability of the atmosphere during the first 1-2 forecast weeks. Even minor imprecisions can cause forecasts to quickly deviate from the actual atmospheric evolution due to the inherent chaos.

On the other hand, (quasi-)external constraints can act as an effective boundary condition.

The definition of "external" can vary depending on the context, and there are few examples of strictly external processes. For example, greenhouse gas concentrations (or their effect on radiation) may be prescribed in climate simulations and in that case represent a strict boundary condition for the atmospheric circulation. On weather time scales, the distinction between initial and boundary conditions can be less clear, and is often motivated by different timescales associated with specific processes.

In forecasts aimed at predicting tropospheric weather, boundary conditions encompass additional processes coupled to tropospheric weather that evolve over much longer time scales and are thus predictable for longer lead times. For example, ocean and land surface conditions can be considered quasi-external influences in S2S forecasts. Although their evolution is usually part of a prognostic model (and therefore, this represents an initial condition problem on its own), variations in land and ocean conditions tend to be small during the timescale of typical mid-latitude weather events (e.g., low and high pressure systems).

In medium-range forecasts, accurate representation of the evolution of boundary conditions is often less critical, as their influence on the atmosphere tends to evolve over longer timescales. However, in S2S forecasts, the relative importance of (effective) boundary conditions increases compared to that of initial conditions. At such longer lead times, traceable predictability and memory to the initial conditions has mostly vanished, meaning that the initial conditions become increasingly unspecific for a particular forecasted atmospheric state. Moreover, sufficient time has progressed for couplings between the atmosphere and boundary conditions to unfold, which can have major effects on the large-scale atmospheric flow. Boundary conditions can thus be regarded as a (quasi-)external forcing for tropospheric weather, changing the statistics of weather events and hence providing climatic predictability.

Although S2S forecasts are usually targeted at predicting tropospheric weather, they therefore aim to additionally represent many other processes. Importantly, these processes may couple to tropospheric weather, acting as effective boundary condition control and providing climatic predictability (see Fig. 1.4). This may include ocean dynamics (including sea surface temperatures and sea ice), land processes (including moisture exchange with the atmosphere, snow cover and surface stress induced by vegetation) and upper atmospheric dynamics (including the stratospheric circulation or ozone anomalies). More detailed information about such sources of S2S predictability is provided in chapter 2.

1.3 Ensemble forecasting

One key insight of the previous section was that weather forecasts are imperfect in practice, as both initial conditions and the numerical model are associated with uncertainties. For a given forecast, one fixed set of initial conditions and model parameters has to be chosen within these uncertainty bounds. This forecast is deterministic in a sense that based on these choices, a single evolution of the atmosphere is predicted. If the forecast was

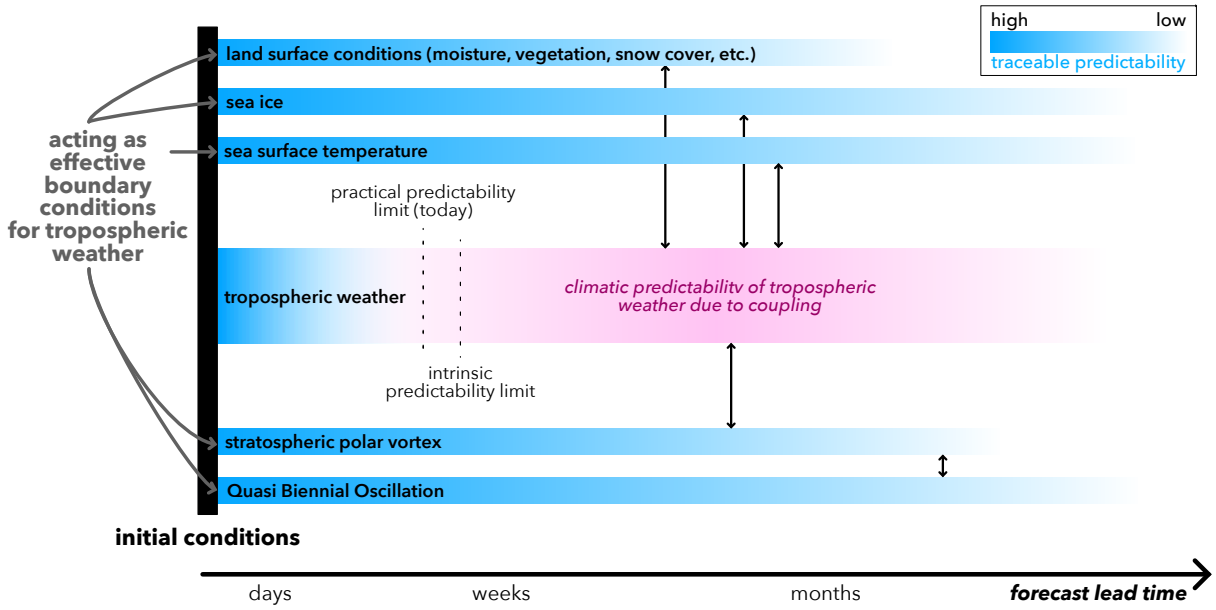

Figure 1.4: Schematic of predictability at different lead times provided by initial and boundary conditions. Traceable predictability is related to accurate initial conditions and decays with lead time. Compared to tropospheric weather, some processes in the Earth system evolve on longer time scales, but they are also coupled to the atmosphere. For predicting tropospheric weather, such processes may therefore be regarded as effective boundary conditions. Their coupling to the atmosphere may change the statistics of tropospheric weather and thereby provide climatic predictability at lead times beyond the traceable predictability limits.

repeated with the exact same initial conditions and model parameters, it would be exactly reproducible, provided that the computing infrastructure does not introduce random errors. However, it follows from the chaotic nature of the atmosphere that another set of initial conditions or model parameters, still within the given uncertainty bounds, could result in a completely different forecast, especially at longer lead times. As a result, deterministic forecasts are sometimes more and sometimes less skillful compared to the real atmosphere evolution, simply by chance. These considerations motivate the use of ensemble forecasts, an approach developed during the late 20th century (Epstein, 1969; Palmer et al., 1992; Molteni et al., 1996) and by now one of the key concepts of operational weather forecasting (Palmer, 2019). Ensembles are used across all forecasting time ranges, and in particular, they play a key role in S2S forecasting.

Instead of a single, deterministic forecast, ensemble forecasting employs a Monte Carlo technique, where different plausible sets of initial conditions and model parameters are randomly sampled within the estimated uncertainty bounds. For each of these sets, a deterministic forecast is produced. Therefore, while each individual ensemble member provides a deterministic forecast, all ensemble members together provide a probabilistic forecast, i.e., a full range of possible forecast scenarios. This is illustrated in Fig. 1.5, which shows a 51-member ensemble forecast for temperature in Munich initialized on January 1st 2018. Up to a few days lead time, all ensemble members predict a quite similar temperature evolution, with differences less than about 2 °C. By day seven, substantial differences between ensemble members have emerged, spanning a range from −7 °C to 4 °C. By day 28, the ensemble distribution is close to the climatological distribution for that time of year, with the most likely temperature being around −1 °C and the average temperature being around −3 °C.

Ensemble forecasts are computationally more expensive compared to a single deterministic forecast and trade-offs have to be made between ensemble size, resolution, initialization frequency, domain size, etc. However, ensembles bring a number of valuable advantages compared to deterministic forecasts, some of which are discussed below.

First, the ensemble mean forecast equals the expected value of the forecast and thus minimizes various skill metrics, including the mean squared error (MSE; even though individual members can be more skillful than the ensemble mean in single cases). Using the ensemble mean is therefore a common way to condense the information contained in a full ensemble back to a single prediction. In many cases, such as for underlying Gaussian distributions, the ensemble mean also represents the most likely forecast scenario.

Furthermore, compared to deterministic forecasts, ensembles provide a forecast distribution and thereby allow for consideration of different forecast scenarios, including an assessment of their likelihood. Importantly, this increases the economic value of ensemble forecasts compared to that of deterministic forecasts (Zhu et al., 2002). As an example, suppose a farmer needs to decide whether to irrigate their crops based on weather forecasts. They have two types of forecasts available: deterministic and ensemble. Based on a deterministic forecast, such as “There will be rain tomorrow.”, the farmer might decide not to irrigate the crops, assuming they will be watered by rain. If the forecast is wrong and it does not rain,

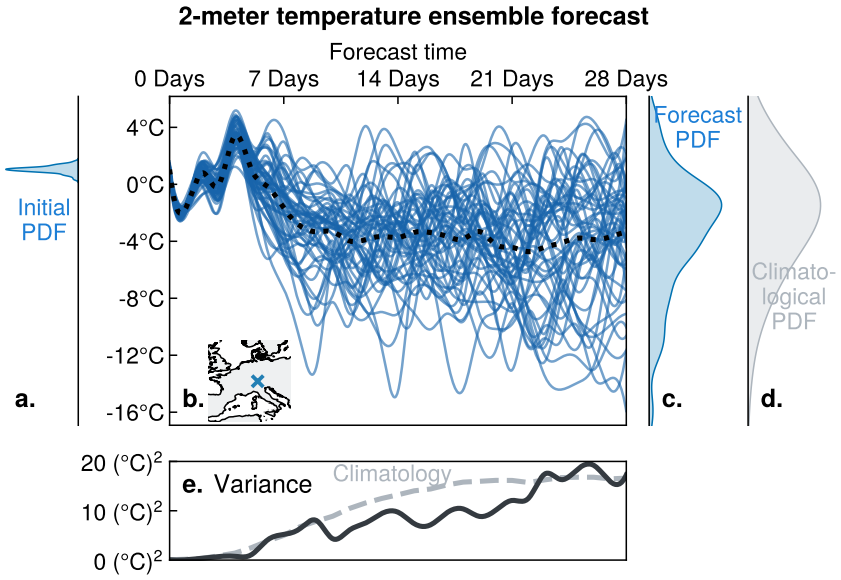

Figure 1.5: Illustration of ensemble forecasting based on a 2-meter temperature forecast for Munich (location indicated in map inset in panel b.), initialized on January 1st 2018. **a.** Ensemble probability density function (PDF) at time of forecast initialization. **b.** Forecasts of individual ensemble members (solid, blue) and ensemble mean (dotted, black). **c.** Ensemble distribution at 28 days lead time. **d.** Climatological distribution. **e.** Lead time evolution of the ensemble variance for the given forecast (black) and averaged over many forecasts (dashed, gray).

the farmer's crops might suffer from lack of water, leading to decreased yield and potential financial loss due to crop damage. In contrast, an ensemble forecast could provide a range of possible outcomes, such as "There is a 60% chance of no rain, 30% chance of light rain, and 10% chance of heavy rain tomorrow." With the ensemble forecast, the farmer can make a more informed decision. They might decide to irrigate the crops if the majority of the forecast scenarios predict no rain, even if there's a chance of light rain. This illustrates how ensembles can help prevent potential damage and financial loss.

Moreover, forecast distributions provided by ensemble forecasts allow assessing risks for extreme weather events. This is particularly important because weather extremes generally have largest impact on society. "Extreme" here refers to the high-impact nature of a weather event, but these events typically also represent statistical extremes in a sense that they happen very rarely and thus correspond to the tails of the climatological distribution. Therefore, large ensembles are required to properly sample forecast probabilities of weather extremes.

An additional important benefit of ensemble forecasts is that they allow assessing forecast uncertainty. A typical way to quantify uncertainty is via the variance across ensemble members. Notably, the ensemble variance is related to the expected value of the squared error of the forecast. Hence, ensembles facilitate estimating their own predictive skill. Figure 1.5d shows that ensemble variance for the Munich temperature forecast is initially small, indicating little uncertainty. Up to about 2 weeks lead time ensemble variance grows until it saturates at a climatological level. While this is a very common qualitative behavior, uncertainty can be anomalously small or large in individual cases. This forms the basis for chapter 5 of this thesis, which identifies atmospheric states under which uncertainty at S2S lead times is systematically smaller or larger compared to the climatological level.

In conclusion, the described advantages make ensembles valuable for forecasts across all lead times, from short- and medium-range to S2S and seasonal prediction. Importantly, S2S forecasts heavily rely on large ensembles: first of all, they aim to exploit climatic predictability, which only allows for conclusions with probabilistic statements. Ensembles are thus key to quantify these probabilities. Second, skill at S2S lead times is generally quite low and only the use of large ensembles allows for predictions that are more skillful than a multi-year climatological average. Finally, uncertainty at such long lead times is generally large, but ensembles make it possible to quantify flow-dependent changes in uncertainty, which can benefit decision-making.

1.4 Scope of the thesis

1.4.1 The stratospheric polar vortex and stratosphere-troposphere coupling

This thesis focuses on S2S prediction of mid-latitude weather in the troposphere during winter. The previous sections have outlined that the intrinsic predictability of mid-

latitude weather is limited to the medium-range, which extends to approximately two weeks. Nonetheless, couplings to more slowly evolving Earth system components can provide climatic predictability also at longer lead times. Such components therefore represent sources of S2S predictability.

For S2S prediction of mid-latitude weather in winter, one of the most important sources of predictability is the circulation in the polar stratosphere. The circulation in the polar stratosphere during winter is characterized by strong circumpolar westerly winds at about 10 km to 50 km altitude, known as the (stratospheric) polar vortex. The polar vortex represents a key source for tropospheric predictability at S2S time scales and this is because of two important factors.

First, polar vortex strength can modulate the tropospheric circulation (Kidston et al., 2015). Although the complete understanding of the underlying mechanisms is still under debate (e.g., see Baldwin et al., 2021), it is generally acknowledged that anomalies in polar vortex strength are linked to temperature anomalies in the polar middle and lower stratosphere that are consistent with changes in tropopause height and associated shrinking or stretching of the tropospheric air mass column in the Arctic. Specifically, a deceleration of the polar vortex is associated with adiabatic warming in the polar lower stratosphere, which is consistent with a tropopause descend. In the troposphere, weak polar vortex events are followed by increased surface pressure over the polar cap, northward of about 60°N. These pressure anomalies project onto the leading mode of planetary-scale variability in the extratropical troposphere, which is known as the Arctic Oscillation (AO) (Thompson and Wallace, 1998). The signal tends to be strongest over the North Atlantic sector (e.g., Charlton-Perez et al., 2018), manifesting in a negative phase of the North Atlantic Oscillation (NAO). A negative NAO phase is linked to a southward shift of the North Atlantic storm track, resulting in modulated frequencies of, e.g., cyclones, cold-air outbreaks and unusual precipitation events (Thompson and Wallace, 1998). Likewise, strong polar vortex episodes are associated with polar lower stratospheric cooling, a lifted tropopause over the Arctic and positive AO and NAO phases in the troposphere.

Second, the polar vortex tends to be predictable for longer lead times than tropospheric weather. Importantly, an exceptionally weak or strong polar vortex can lead to corresponding weakening or strengthening of the circulation in the lowermost stratosphere, where these anomalies can persist throughout several weeks (Baldwin and Dunkerton, 1999). Therefore, skillful prediction of the polar vortex can provide climatic predictability for tropospheric mid-latitude weather (recall Fig. 1.4).

The most prominent phenomena of polar vortex extremes are rapid polar vortex breakdowns, known as sudden stratospheric warmings (SSWs) (Baldwin et al., 2021). In the stratosphere, these events are associated with a strong deceleration of the polar vortex westerlies, which is followed by a reversal of the zonal-mean wind (from westerly to easterly). This deceleration is accompanied by an intense warming of the mid and lower polar stratosphere by several tens of degrees within only few days. In the lowermost stratosphere, SSWs are often followed by long-lived easterly zonal-mean zonal wind anomalies for up to about 2 months. In the troposphere, SSWs are on average followed by a nega-

tive phase of the AO (Baldwin and Dunkerton, 2001). The associated equatorward shift of the tropospheric jet streams has been linked to wet and high-wind anomalies at lower latitudes (e.g., over southern Europe; Ayarzagüena et al., 2018) and to cold anomalies at high latitudes (e.g. over Scandinavia and Siberia; Kolstad et al., 2010; Tomassini et al., 2012; Butler et al., 2017). In conclusion, SSWs are followed by persistent anomalies in the stratospheric circulation, which can couple to the troposphere. Therefore, SSWs are of particular interest for S2S predictions of mid-latitude weather in winter.

Chapters 4 and 5 will investigate the formation of SSWs in greater detail, as well as their surface response and impacts on tropospheric predictability. To provide a brief overview of SSWs, Fig. 1.6 illustrates the dynamics and the surface response of SSWs, based on a specific example and a composite-mean of multiple events. Fig. 1.6a-c show snapshots of the polar vortex using potential vorticity[3] on the 850 K isentropic surface during a SSW that occurred on 12 February 2018. While the polar vortex seems intact and roughly centered over the pole 8 days prior to the SSW date (Fig. 1.6a), the vortex has split into two distinct sub-vortices by the event central date (Fig. 1.6b) and it has been eroded further 8 days later (Fig. 1.6c). The SSW is associated with an intense deceleration of the climatological westerly winds, which are replaced by easterly winds following the event (Fig. 1.6d). Fig. 1.6e shows that SSW events are on average followed by a negative AO phase, indicated by high pressure anomalies over the polar cap and low pressure anomalies in mid-latitudes. Consistent with the negative AO phase, SSWs are on average followed by cold anomalies over northern Eurasia and warm anomalies over the Arctic Sea and southern Asia.

This thesis examines SSWs and strong polar vortex events, with a focus on understanding their influence on subsequent tropospheric extreme events and their contribution to S2S predictability in the troposphere. A key aspect of this work is the use of a large, relatively new set of S2S ensemble forecasts from a state-of-the-art weather forecast model. This extensive dataset enables the application of advanced statistical techniques, allowing for a more robust analysis of stratosphere-troposphere coupling. Moreover, by leveraging this dataset, previously unrecognized aspects of S2S predictability linked to polar vortex extreme events are uncovered. The following section outlines the specific research questions addressed. More background about polar vortex dynamics and also other sources of subseasonal predictability is provided in chapter 2.

[3]Potential vorticity combines absolute vorticity of a fluid parcel with its vorticity due to the rotation of the planet, while taking into account the vertical density profile. Importantly, potential vorticity is conserved on isentropic surfaces for adiabatic frictionless motion. Potential vorticity data was provided by Philip Rupp.

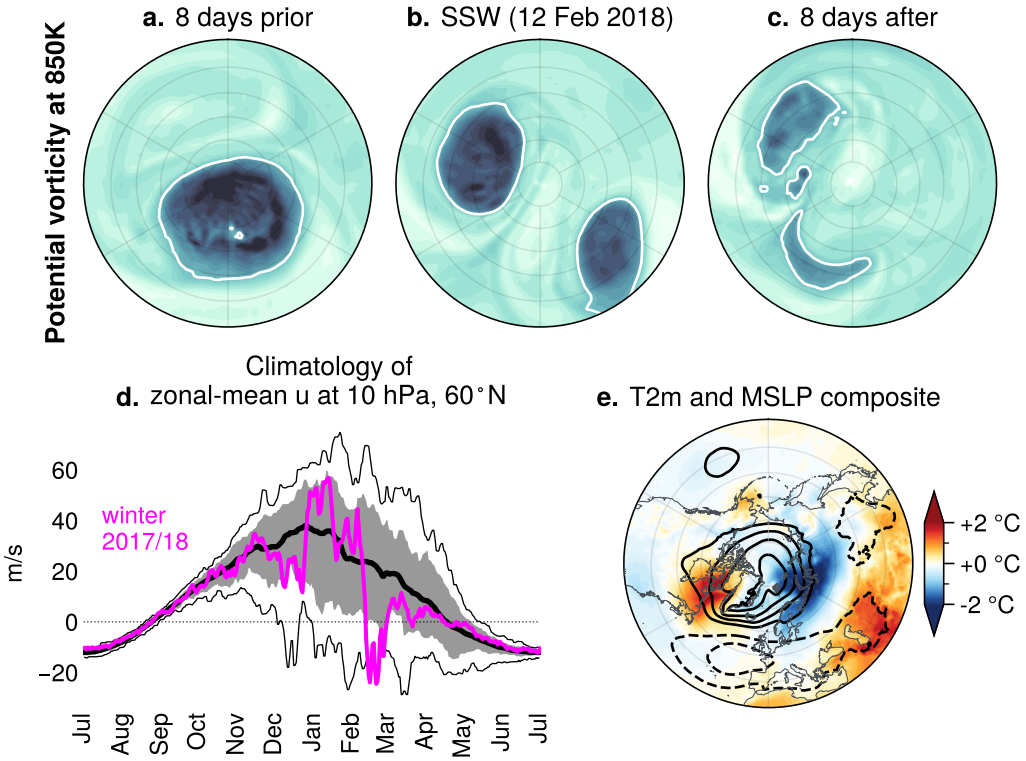

Figure 1.6: Overview of sudden stratospheric warmings. **a-c.** Evolution of the polar vortex during the SSW of 12 February 2018, displayed as stereographic maps of potential vorticity (PV) on the 850 K isentropic surface from 30 °N to 90 °N. PV is plotted using a linear color scale with values ranging from 0 PVU to 1000 PVU with an additional white contour line at 500 PVU. **d.** Climatology of the zonal-mean zonal wind at 60 °N and 10 hPa (U_{60}^{10}), a common metric to measure the polar vortex strength. Black lines and shading denote the multi-year ERA5 climatology (thick line: mean, shading: 10-90 percentiles, thin lines: min/ max). Magenta line shows the polar vortex evolution during the winter 2017/18. The SSW of 12 February 2018 is characterized by an intense deceleration of the zonal-mean zonal wind. **e.** Composite-mean of 2-meter temperature (T2m; shading) anomalies and mean-sea-level pressure (MSLP; contours every 1 hPa, negative values dashed) anomalies during lag days 10 to 60 following SSWs (41 events, based on ERA5 data).

1.4.2 Research questions

(1) How variable is the role of tropospheric wave activity in the formation of SSWs?

SSWs are primarily forced by interactions between the zonal-mean flow and planetary waves (see section 2.2.2). In particular, strong pulses of tropospheric wave activity can trigger SSWs (Matsuno, 1971). In the troposphere, such pulses of wave activity can be associated with persistent anticyclones, known as blockings, in particular over the Eurasian continent (e.g., Garfinkel et al., 2010). However, anomalous wave activity is not always required for SSWs formation, because tropospheric wave activity would already be sufficient under average climatological conditions, i.e., wave activity does not have to be extreme (Plumb, 1981; Birner and Albers, 2017; de la Cámara et al., 2019). Quantifying how important anomalous upward wave activity really is in triggering SSWs is challenging, because SSWs occur only about every other winter and hence, the total number of observed SSWs is relatively small. In the first results part of this thesis, large statistics of SSWs are analyzed, which were predicted in operational S2S ensemble forecasts. The majority of these SSWs were only predicted in the forecasts and never actually occurred in the real atmosphere. Nevertheless, these predicted SSWs are consistent with the model physics and given that SSWs are large-scale phenomena, the assumption may be justified that their representation in the model is reasonably realistic. This approach has in other contexts been referred to as an UNprecedented Simulated Extremes using ENsembles (UNSEEN) approach (Kelder et al., 2020). The UNSEEN approach allows for investigating several thousand SSW events, which would require more than ten thousand years to observe in the real atmosphere. Thereby, the role of tropospheric wave forcing in the formation of SSWs will be revisited with high statistical confidence. Specifically, the focus lies on the following questions: How often may SSWs be attributable to preceding anomalies in upward wave activity? How does the occurrence of tropospheric blockings over Eurasia increase SSW likelihood?

(2) How, and to what extent, can tropospheric circulation extremes be attributed to SSWs?

SSWs tend to be followed by anomalies in the Northern Annular Mode (NAM; Baldwin and Dunkerton, 2001) reaching from the stratosphere to the troposphere. The near-surface NAM is also referred to as Arctic Oscillation (AO) and it condenses information about jet latitude and strength, with implications for pressure, wind, temperature and precipitation. SSWs are on average followed by a negative phase of the AO with a magnitude of about -0.3 standard deviations (e.g., see Fig. 4.8, or Oehrlein et al., 2021). While this mean-response can be derived from observed SSWs, analyzing the full post-SSW AO distribution is challenging due to the scarcity of SSW events. In particular, this holds for quantifying the frequency of AO extremes. However, such changes in extreme event frequency are particularly relevant for the society and economy, because of associated high-impact weather. Based on the UNSEEN approach, several thousand SSW events are employed to

analyze AO extreme occurrence following SSWs, focussing on the following questions: How do SSWs affect AO extreme probability? How often may AO extremes be attributable to preceding SSWs? What is the corresponding impact of strong polar vortex episodes on AO extremes?

(3) How do stratospheric circulation extremes affect tropospheric predictability?

Strong and weak polar vortex events are known to affect mid-latitude weather over the course of several weeks and thereby provide climatic predictability at S2S time scales (Butler et al., 2019). Previous studies have demonstrated that this modulation of mean weather conditions following polar vortex events indeed translates into enhanced skill of S2S forecasts (Sigmond et al., 2013; Tripathi et al., 2015). For example, SSWs are on average followed by lower temperatures over Scandinavia and by a negative NAO phase. However, predictability measures that are not related to such changes in the mean, but to changes in variance have been studied much less. Understanding changes in variance is important, because it measures the uncertainty of a prediction. Here, changes in S2S forecast uncertainty induced by polar vortex extreme events are analyzed. Specifically, the following questions are posed: How do SSWs affect the subsequent forecast uncertainty of the polar vortex? How do polar vortex extremes (including SSWs) affect the subsequent forecast uncertainty of the troposphere? How are tropospheric circulation regimes related to subseasonal forecast uncertainty in general?

1.4.3 Thesis outline

This thesis is structured as follows.

Chapter 2 provides a more detailed background on the sources of S2S predictability for mid-latitude weather in the troposphere. This includes a brief overview about the atmospheric general circulation and, specifically, a theoretical background about polar vortex extreme events and stratosphere-troposphere coupling.

Chapter 3 introduces the datasets used in this study. Different methods for computing S2S forecast anomalies are discussed and a new open-source python library for handling these forecasts is introduced.

Chapters 4 and 5 present the results. In chapter 4, ensemble forecasts are leveraged to boost the statistics of SSWs using the UNSEEN approach. This allows for robust quantification of the tropospheric role in the formation of SSWs via anomalous upward wave propagation. Furthermore, an attribution analysis is presented for weak and strong polar vortex events and subsequent AO extremes. Chapter 5 focuses on how extreme polar vortex states affect the uncertainty in S2S forecasts in the stratosphere and in the troposphere. In addition, a generalized analysis is presented about how different tropospheric large-scale weather regimes are inherently linked to distinct spatial patterns of uncertainty.

Section 6 discusses the main findings, limitations and possible future work.

Chapter 2

Sources of subseasonal-to-seasonal predictability

This thesis primarily investigates the predictability of mid-latitude tropospheric weather during winter at lead times of 2-6 weeks, i.e., at subseasonal-to-seasonal (S2S) time scales. Predictability at these lead times is facilitated by remote couplings across the atmosphere and interactions between the atmosphere and the Earth's surface.

This chapter provides a background on the atmospheric circulation and delves into some of most relevant phenomena to S2S prediction. Special emphasis is placed on the stratospheric polar vortex and its implications for tropospheric weather and forecasting. In addition, teleconnections from the tropics are briefly introduced.

For a comprehensive overview of further sources of S2S predictability, including ocean-atmosphere coupling (e.g., via SST anomalies or sea ice) and land-atmosphere coupling (e.g., via soil moisture or snow cover), the reader is referred to Robertson and Vitart (2019).

2.1 Overview of atmospheric structure and general circulation

In the Earth's atmosphere, density and pressure generally decrease with height. Temperature, on the other hand, follows a more complex vertical profile, which motivates the classification of the atmosphere into five layers: the troposphere (the lowest layer), the stratosphere, the mesosphere, the thermosphere, and the exosphere. Each of these layers has distinct thermodynamic properties and circulation patterns. Most of the dynamics relevant to surface weather occur in the troposphere, but to some extent, also in the stratosphere. These two layers are the main foci of this thesis and will be presented more in detail below.

The troposphere extends up to about 20 km near the equator and about 10 km near the pole. In general, temperature in the troposphere decreases with height, at an average lapse

rate of about $6.5\,\mathrm{K\,km^{-1}}$, primarily because air is warmed by the surface. A local temperature minimum is reached at the top of the troposphere, which can be used to define the tropopause – the boundary between troposphere and stratosphere. In addition to the characteristic vertical temperature profile, the troposphere exhibits a strong meridional temperature gradient resulting from differential solar heating. This meridional gradient, together with the Coriolis effect, forms the basis for two strong westerly jets per hemisphere (see Fig. 2.1). The first jet, called the subtropical jet (STJ), is located near the tropopause at around 30°N/S and results from poleward motion of upper tropospheric air that is deflected eastward by the Coriolis force. The second jet, called the eddy driven jet (EDJ), is located at around 50°N/S and extends from the surface up to the tropopause. It owes its existence to Rossby waves, which are large-scale longitudinal wave patterns of temperature, pressure and wind anomalies (e.g., see review by Wirth et al., 2018). Rossby waves result from angular momentum conservation on a rotating sphere. They are predominantly generated by orography or by baroclinic instability, which is a fluid instability that allows small-amplitude Rossby waves to mutually amplify. In addition to their fundamental impact on the two zonal jets, Rossby waves are crucial for mid-latitude surface weather, manifesting in low- and high-pressure systems and are directly influencing wind, temperature, and precipitation. Additionally, Rossby waves play an important role in redistributing heat and momentum within the atmosphere.

Above the troposphere lies the stratosphere, extending from the tropopause up to about 50 km. Unlike the troposphere, the lapse rate in the stratosphere is negative, meaning that temperature increases with height, which results from UV absorption in the ozone layer. Ozone production, based on photodissociation of oxygen, is dominant from around 10 km to 50 km altitude. At higher altitudes, the atmosphere is too thin and at lower altitudes, much of the UV radiation has already been absorbed. Importantly, the resulting negative lapse rate in the stratosphere creates strongly stable stratification (i.e., strong restoring buoyancy force acting on vertically displaced air parcels). Consequently, convection is suppressed, the flow is baroclinically stable in general, and vertical mixing is much less significant than in the troposphere.

The stratosphere features several large-scale circulation patterns, including the Quasi-Biennial Oscillation (QBO), the Brewer-Dobson-Circulation (BDC) and the polar vortex (also named polar night jet, PNJ). These patterns are illustrated in Figures 2.1 and 2.2, which present vertical cross-sections of zonal-mean zonal wind and temperate during Northern Hemisphere winter.

The QBO dominates the zonal circulation in the tropical stratosphere (see Fig. 2.1) and describes a recurrent downward propagation of westerly and easterly winds at a remarkably regular periodicity of approximately 28 months (Baldwin et al., 2001). The QBO is primarily driven by upward propagating atmospheric waves of a wide range of spatial scales and their interaction with the zonal-mean flow (Lindzen and Holton, 1968; Holton and Lindzen, 1972). Due to its influences on other regions of the atmosphere, the QBO may provide predictability for S2S and seasonal forecasts. However, numerical models often struggle to accurately simulate the QBO, its 28-month periodicity or its teleconnections

(Anstey et al., 2022). These issues are presumably related to the significant contributions of small-scale waves, which are largely unresolved in models.

Furthermore, the stratospheric vertical and meridional circulation is characterized by the BDC (Butchart, 2014). In the tropical stratosphere, the BDC is shaped by wave-induced upwelling near the tropical tropopause, which thereby represents the major entry point of water vapour into the stratosphere. In the extratropical stratosphere, the BDC involves poleward motion and downwelling which is an important process in redistributing water vapour and ozone within the stratosphere.

Finally, the stratospheric circulation in polar regions is characterized by relatively weak and regular easterly winds during summer, and by strong and highly-variable westerly winds during winter. These strong winter westerly winds form the polar vortex, centered around 60°N, 10 hPa (see Fig. 2.1). Due to its predictability over longer lead times compared to the troposphere and its influence on extratropical tropospheric weather, the polar vortex represents an important source of S2S predictability. As such, it is the main focus of this thesis and will therefore be described in more detail in the following section.

2.2 The circulation in the polar stratosphere

2.2.1 The polar vortex

Absence of sunlight and radiative cooling of ozone cool the polar stratosphere to very low temperatures during winter. The resulting meridional temperature gradient is accompanied by strong westerly winds throughout the polar stratosphere, known as the (stratospheric) polar vortex. Following the seasonal evolution of the meridional temperature gradient, the vortex builds up in autumn, becomes strongest in mid-winter and disappears in spring. Despite the regular seasonality of radiative forcing, the polar vortex strength is highly variable, especially in the Northern Hemisphere, from day-to-day to inter-annual timescales, due to Rossby waves and their interaction with the zonal-mean flow. Appendix A provides a brief introduction to the relevant theory of the dynamics of these waves and their impact on polar vortex strength and variability. Here, the focus is more on a phenomenological description of polar vortex extreme events and their coupling to the troposphere.

Polar vortex dynamics are characterized by the continuous interplay of strengthening driven by radiative cooling and weakening driven by the dissipation of upward propagation of Rossby waves. The radiative forcing acts on relatively long time scales of a few weeks (e.g., Andrews et al., 1987), whereas Rossby wave activity exhibits significant day-to-day variability. Rossby waves are primarily generated in the troposphere through orographic forcing and baroclinic instability. According to equation A.7, these waves may propagate vertically into the stratosphere when the background zonal flow is westerly, but weaker than the wavenumber-dependent critical wind speed, U_c (Charney and Drazin, 1961).

These conditions also influence the polar stratospheric circulation during summer: the positive meridional temperature gradient (warmer over the pole than in mid-latitudes)

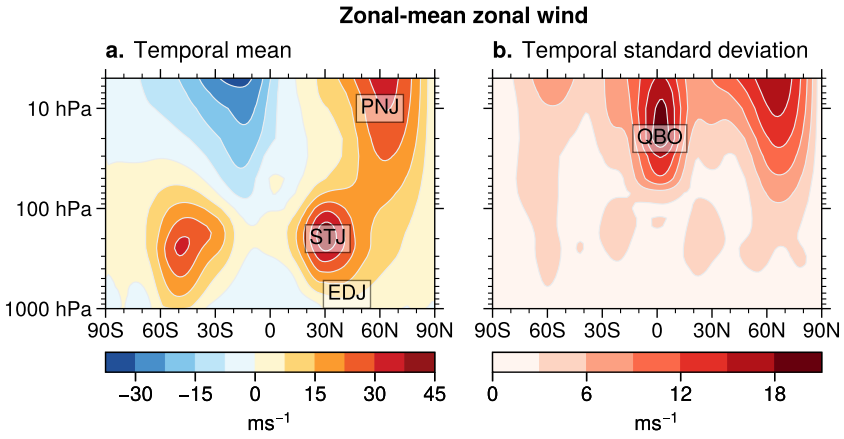

Figure 2.1: Zonal-mean zonal wind (u) during Boreal winter based on 1979-2018 ERA5 reanalysis data. **a.** Temporal average of u and **b.** temporal standard deviation of u (primarily quantifying inter-seasonal and inter-annual variability), based on monthly data. Text boxes denote approximate locations of the subtropical jet (STJ), eddy-driven jet (EDJ), polar vortex/ polar night jet (PNJ) and Quasi-Biennial Oscillation (QBO).

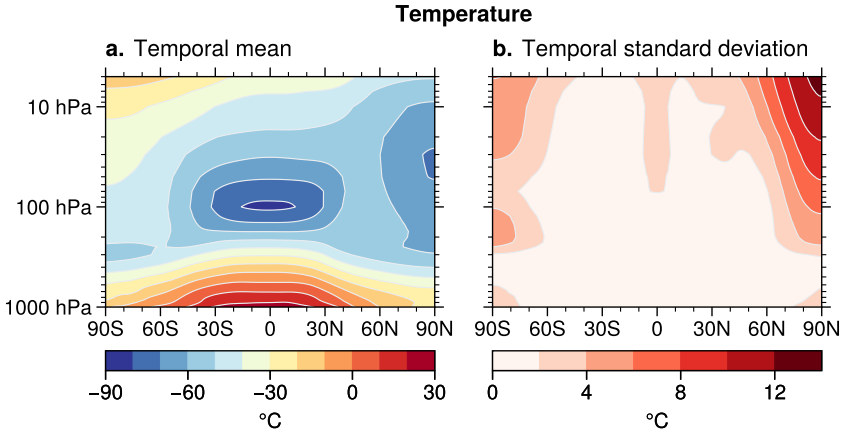

Figure 2.2: As Fig. 2.1, for temperature.

leads to easterly winds, which prevent vertical wave propagation from the troposphere into the stratosphere, resulting in very little variability of the stratospheric summer easterlies (recall Fig. 1.6d).

During winter, as the stratospheric meridional temperature gradient becomes negative, the strong westerly winds of the polar vortex build up. In the Northern Hemisphere, orography leads to substantial planetary wave activity, resulting in a generally weaker polar vortex compared to the Southern Hemisphere. With typical westerly winds of around $20\,\mathrm{m\,s^{-1}}$ to $40\,\mathrm{m\,s^{-1}}$, the vortex becomes susceptible to upward propagation of waves with zonal wavenumber 1 (i.e., one trough and one ridge along one latitude circle; $U_c \approx 32\,\mathrm{m\,s^{-1}}$ for stationary waves) and sometimes zonal wavenumber 2 (i.e., two troughs and two ridges along one latitude circle; $U_c \approx 18\,\mathrm{m\,s^{-1}}$ for stationary waves). The interaction between waves and the mean-flow in the Northern Hemispheric stratosphere not only leads to a weakened but also a much more variable polar vortex. In extreme cases, disruptions can cause complete break downs of the polar vortex. These events are called sudden stratospheric warmings (SSWs), which will be described in more detail in the following section.

2.2.2 Sudden stratospheric warmings

When sunlight returns to the polar stratosphere during spring, the meridional temperature gradient erodes and the polar vortex weakens, until a final breakdown typically occurs around April (in the Northern Hemisphere). However, wave dynamics can induce major disruptions of the polar vortex also during mid-winter, which can affect the circulation in stratosphere and troposphere over the course of several weeks. These rapid deceleration events, first discovered by Richard Scherhag in 1951 (Scherhag, 1952), are called sudden stratospheric warmings (SSWs) (Baldwin et al., 2021), named for the drastic temperature increase of up to about 50 K within only a few days. In the Northern Hemisphere, 33 of the winters between 1960/61 and 2022/23 experienced a SSW ($\approx 52\%$ or once every 1.91 winters), with 7 of those winters experiencing two SSWs ($\approx 7\%$ or once every 9 winters), resulting in a total of 40 SSWs (0.63 per year).

SSWs are fundamentally driven by the interaction of planetary waves and the mean-flow (Matsuno, 1971; Geisler, 1974; Holton and Mass, 1976; Plumb, 1981). Planetary Rossby waves, predominantly generated in the troposphere, can propagate up into the stratosphere, as described in the previous section. Upward propagating Rossby waves generally transport heat poleward. Under conservative propagation, this is compensated by adiabatic cooling due to the induced upwelling, resulting in no net change of temperature. However, non-conservative effects, including those due to non-linear wave breaking and dissipation, transfer the Rossby wave's intrinsic eastward momentum to the zonal-mean flow. This deceleration of westerlies is associated with a poleward Coriolis torque, resulting in convergence over the pole. This induces upwelling anomalies above and downwelling anomalies below, leading to a relative adiabatic cooling of the upper stratosphere and warming of the lower- to mid-stratosphere.

In the aftermath of SSWs, the climatological westerlies in the polar stratosphere are re-

placed by easterly zonal winds. These anomalies can persist for several weeks, particularly in the lower stratosphere where radiative time scales are longest (Baldwin et al., 2003). This change prevents further upward propagation of waves, thereby suppressing wave activity in the mid-stratosphere.

2.2.3 Stratosphere-troposphere coupling

Extreme states of the polar vortex, including SSWs and strong polar vortex episodes, can be followed by persistent anomalies of tropospheric weather over the course of several weeks (Kidston et al., 2015).

Specifically, SSWs are followed by increased near-surface pressure over the polar cap, poleward of approximately 65 °N. A comprehensive theory for this downward response is still lacking (Baldwin et al., 2021). Qualitatively, the increase in near-surface pressure can be explained by the convergence of air over the polar cap at stratospheric levels during SSWs. However, the accumulated air mass in the polar stratosphere can not explain the full magnitude of the signal near the surface.

From a potential vorticity perspective, SSWs are associated with negative potential vorticity anomalies in the polar stratosphere, which have remote effects on the troposphere. However, this perspective also falls short in explaining the complete downward response to SSWs, as the tropospheric response would be expected to decrease monotonically from the tropopause to the surface (Ambaum and Hoskins, 2002). In contrast, both observations and model studies indicate that the post-SSW signal is stronger near the surface than in the mid-troposphere.

It is generally acknowledged that important missing links in these perspectives are tropospheric feedbacks between the longitudinally uniform response and atmospheric waves, including tropospheric eddies and planetary waves. These feedbacks likely play a critical role in the downward response to polar vortex extremes.

Specifically, the lower stratosphere acts as an upper boundary condition for the troposphere and stratospheric anomalies can thereby modulate the characteristics of tropospheric wave dynamics, including their interaction with the zonal-mean flow, resulting in an amplification of the tropospheric signal (Domeisen et al., 2013; Hitchcock and Simpson, 2016; Smith and Scott, 2016; Rupp and Birner, 2021).

To trace impacts of SSWs and strong polar vortex events on the large-scale circulation from the stratosphere to the troposphere, one suitable diagnostic are Northern Annular Mode (NAM) indices (Baldwin and Dunkerton, 1999). The NAM represents the dominant mode of variability of daily geopotential height anomalies in the Northern Hemisphere extratropics. The NAM indices are constructed based on an Empirical Orthogonal Function (EOF) and at each vertical level the NAM index corresponds to the first principal component time series. The sign convention is such that a positive NAM index corresponds to a strengthened and poleward-shifted zonal circulation, while a negative NAM index indicates a weakened and equatorward-shifted zonal circulation.

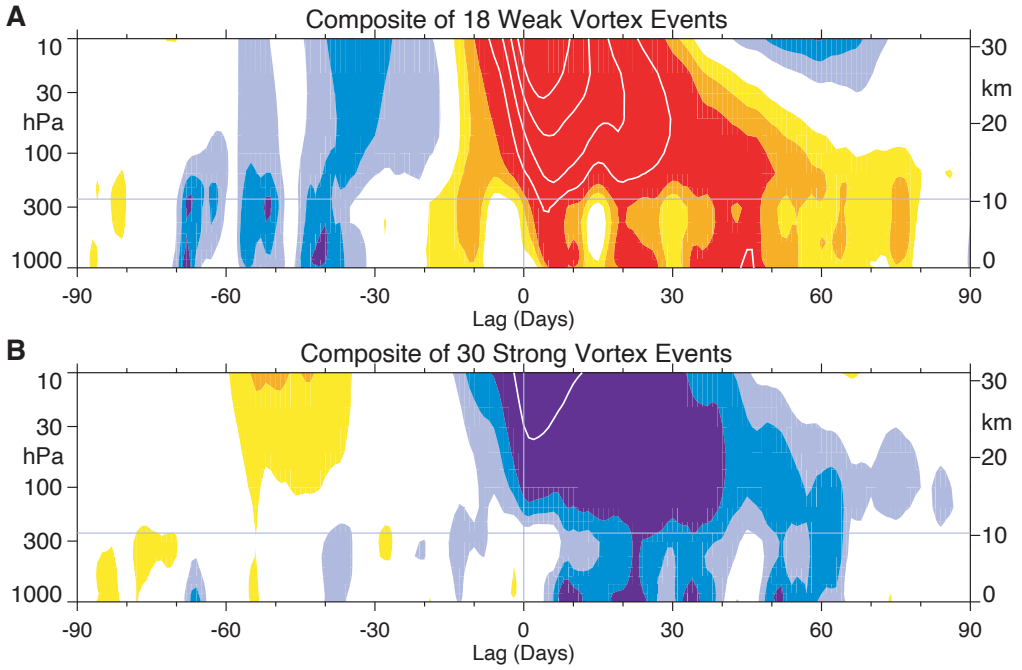

Figure 2.3: Northern Annular Mode (NAM) composite-mean of (A) weak and (B) strong polar vortex events. Events are defined according to -3 and $+1.5$ standard deviations of the 10-hPa NAM index. **Red:** negative NAM values **Blue:** positive NAM values. Colored contour intervals correspond to 0.25 standard deviations, with additional white contour lines every 0.5 standard deviations. Thin horizontal line denotes the approximate height of the tropopause. Taken from Baldwin and Dunkerton (2001).

Fig. 2.3 presents a landmark figure of stratosphere-troposphere coupling by Baldwin and Dunkerton (2001), showing a NAM composite for weak and strong polar vortex events. Weak polar vortex events, such as SSWs, represent a weakening of the zonal circulation and therefore manifest in a negative NAM in the stratosphere. In the aftermath of such events, the average NAM index remains negative in the lowermost stratosphere and the troposphere over the course of approximately two months.

Near the surface, the negative NAM indicates the increase in surface pressure over the polar cap. This increase results in a reduced meridional pressure gradient, leading to a weakened and southward shifted eddy-driven jet. In addition, these changes are associated with anomalous near-surface weather, such as an increased risk for persist high-pressure systems and cold-air outbreaks on the poleward flank of the jet, as well as increased storm frequency around its new central latitude. Studies have shown that this signal tends to be strongest over the Atlantic (associated with a negative phase of the NAO) and enhances the likelihood of cold air outbreaks over Eurasia (Thompson et al., 2002; Lehtonen and Karpechko, 2016; Afargan-Gerstman et al., 2020; Kautz et al., 2020; Huang et al., 2021).

Likewise, strong polar vortex events manifest in a positive NAM. In the troposphere, these events tend to be followed by a positive NAO, associated with a poleward jet shift and greater likelihood for storm series (Domeisen et al., 2020c; Rupp et al., 2022).

From this perspective stratospheric conditions may provide predictability on S2S time scales. Indeed, past literature has demonstrated improved forecast skill for ensemble forecasts initialized around SSWs (Sigmond et al., 2013) or strong polar vortex episodes (Tripathi et al., 2015). However, the employed skill measures such as the correlation skill score (CSS) primarily rely on the ensemble mean signal (e.g., the above mentioned negative NAO state following SSWs). Section 5.2 will demonstrate that polar vortex extreme states are additionally followed by anomalies in ensemble spread, which manifests in skill scores such as the mean squared error.

2.3 Teleconnections from the tropics

In addition to the stratosphere, atmospheric teleconnections between the extra-tropical troposphere and the tropics can provide predictability at S2S and seasonal lead times. The following subsections introduce the most relevant of these teleconnections.

El-Niño-Southern-Oscillation (ENSO)

The El Niño-Southern Oscillation (ENSO) (Timmermann et al., 2018) is a coupled atmosphere-ocean phenomenon occurring in the tropical Pacific. Climatologically, trade winds push warm water at the ocean surface from the eastern towards the western equatorial Pacific, creating a longitudinal sea surface temperature (SST) gradient. This results in higher SSTs in the western Pacific, which lead to enhanced convection and precipitation in this region,

while the eastern Pacific remains substantially cooler and experiences less convection and precipitation.

Situations of weakened trade winds and a reversal of the climatological SST pattern represent events that are known as El Niño. During El Niño, the western Pacific experiences anomalously low SSTs, while the eastern Pacific exhibits anomalously high SSTs. This reversal not only leads to significant precipitation anomalies in the equatorial Pacific region but also affects the global climate system. El Niño events occur approximately once every 2 to 7 years and typically last for about 9 to 12 months.

Conversely, events where the climatological patterns of SSTs and precipitation over the equatorial Pacific intensify are referred to as La Niña. La Niña and El Niño represent opposite phases of ENSO. Among the various indices used to characterize ENSO, this thesis employs the ENSO 3.4 index (Trenberth, 1997), which is based on standardized SST anomalies over the equatorial Pacific and one of the most commonly used metrics.

ENSO represents a significant source of atmospheric predictability at S2S and seasonal timescales (Horel and Wallace, 1981; Taschetto et al., 2020, November). Specifically, ENSO-induced anomalies in tropical heating can trigger Rossby wave trains that influence weather in the extratropics. For instance, during El Niño events, the Aleutian Low tends to intensify, which has important implications for weather over North America. Additionally, ENSO has been linked to anomalies over the North Atlantic and Europe, although these signal are generally weaker than those over the North Pacific (Brönnimann, 2007; Scaife et al., 2017).

Beyond its tropospheric teleconnections, ENSO has been linked to anomalous stratospheric polar vortex states through various mechanisms (Domeisen et al., 2019, and references therein). Specifically, increased wave activity during El Niño events is associated with a weaker polar vortex by accelerating the QBO, strengthening the BDC, and enhancing vertical wave propagation in the extratropics. Consequently, ENSO can affect climate, particularly over the North Atlantic and Europe, also through a stratospheric pathway.

Madden-Julian-Oscillation (MJO)

The Madden-Julian Oscillation (MJO) (Madden and Julian, 1971; Zhang, 2005) describes an eastward-propagating precipitation pattern in the tropics with a periodicity of about 30 to 60 days. The state of the MJO is often represented in a 2D phase space, spanned by the first two principal components of an Empirical-Orthogonal-Function analysis of zonal wind and precipitation (or a proxy thereof) in the tropics (e.g., Wheeler and Hendon, 2004). These two principal components are used to compute an MJO amplitude and phase, often discretized into eight characteristic phases. In the first phase, anomalous precipitation occurs over the Indian ocean, where convection is triggered by warm SSTs, leading to lower-level convergence, upper-level divergent outflow, and latent heat release. Although a complete theory for the MJO mechanism is still lacking, coupling to atmospheric tropical waves, including equatorial Rossby waves and Kelvin waves, appears to be play an important role in the organization and eastward-propagation of this convection pattern.

Rainfall anomalies associated with the MJO are usually strongest over the Indian ocean (phases 1-3) and the Maritime continent (phases 4-5), and become substantially weaker when propagating over the Pacific (phases 5-7) and Atlantic (phase 7-8).

While its implications on weather and climate are strongest in the tropics, the MJO can excite Rossby waves that propagate into the extratropics. Specifically, when MJO-induced convection reaches the eastern Indian Ocean and western Pacific, this favors a negative phase of the Pacific North Atlantic (PNA), leading to temperature, wind and rainfall anomalies over North America (Mori and Watanabe, 2008). Over the North Atlantic and Europe, the response to the MJO projects onto the NAO: enhanced convection over the Eastern Indian Ocean tends to be followed by a negative NAO phase about two weeks later, and vice versa (Cassou, 2008).

Due to its intraseasonal timescale and its teleconnections with the extratropics, the MJO is considered an important source of S2S predictability (Woolnough, 2019). However, current atmospheric models still have some issues with a representing the MJO, including its initiation, propagation and extratropical response (e.g., Vitart, 2017). These deficiencies are presumably related to the involved complex interactions between small-scale processes related to convection and large-scale processes related to atmospheric waves. Therefore, it can be assumed that the full potential predictability offered by the MJO has not yet been fully realized.

Quasi-Biennial-Oscillation (QBO)

The Quasi-Biennial Oscillation (QBO) is an alternating pattern of westerly and easterly zonal winds in the tropical stratosphere with a periodicity of approximately 28 months. By influencing the circulation in the tropical troposphere and in the polar stratosphere, the QBO can contribute to S2S and seasonal predictability, provided that the QBO itself and its teleconnections are realistically simulated in models (Anstey et al., 2022).

Specifically, QBO-induced easterly winds near the the tropical tropopause tend to enhance convection and strengthen the MJO (Yoo and Son, 2016), though the underlying mechanisms remain debated. In the extratropics, the QBO can modulate Rossby wave propagation, which affects the strength of the polar vortex. Specifically, QBO-induced easterly winds in the lower tropical stratosphere tend to suppress wave propagation from mid-latitudes into the tropical stratosphere, leading to enhanced wave propagation into the polar stratosphere, which tends to weaken the polar vortex (Holton and Tan, 1980).

Chapter 3

Data and Methodology

3.1 Data

3.1.1 Subseasonal-to-seasonal ensemble forecasts

This thesis primarily uses operational subseasonal-to-seasonal (S2S) ensemble forecasts provided through the S2S prediction project (Vitart et al., 2017). The project is a joint initiative of the World Weather Research Programme (WWRP) and the World Climate Research Programme (WCRP). Several leading weather services contribute S2S forecasts to a database hosted by the European Centre for Medium-Range Weather Forecasts (ECMWF). These forecasts slightly vary in terms of their specific configurations, including initialization frequency, ensemble size, hindcast generation, resolution, and output data.

The majority of the analyses in this thesis are based on ECMWF forecasts, with some analyses also conducted using forecasts by the UK Met Office (UKMO) to test the robustness of the results. While high forecast skill is not essential for the scientific objectives of this thesis, the results depend on the model's sufficiently accurate simulation of the atmosphere, including the general circulation, teleconnections, wave dynamics, and related phenomena. For instance, a well-resolved stratosphere is essential for studying stratosphere-troposphere coupling (Domeisen et al., 2020c). Additionally, several analyses are based on the assumption that the forecasts are reliable, meaning they correctly reflect the likelihoods of possible real-atmosphere evolutions given the initial conditions and the numerical model (see section 3.3).

Realtime forecasts and hindcasts

All contributing centers produce both realtime forecasts and hindcasts. Realtime forecasts are the operational forecasts and initialized from the latest available analysis, whereas hindcasts use the same model configuration but are initialized from past years' reanalyses. Hindcasts are essential for model calibration, for example, allowing to adjust raw forecasts for mean model biases. Additionally, hindcasts are used to compute a model climatology.

In this thesis, hindcasts are further used to increase the forecast sample size. While realtime forecasts are available from the start of the S2S prediction project (around 2015), hindcasts extend back to the late 1990s.

The two weather services whose forecasts are used in this thesis (primarily ECMWF and partially UKMO) generate hindcasts using the so-called "on-the-fly" method. With this approach, hindcasts are co-generated at the same time (or, for UKMO, at similar times) as the realtime forecasts. In contrast, some other services use the "fixed" hindcast approach, producing a complete set of hindcasts spanning multiple years and seasons all at once when a new model cycle is released.

Forecast setup at ECWMF

The following description of ECMWF forecasts refers to the model cycles used in this study (CY43R3 - CY47R1). Note that since version CY48 (introduced in June 2023), realtime forecasts are produced once per day with an ensemble size of 101. From version CY47R2 onwards (introduced in May 2021), the model is run at a higher vertical resolution of 137 levels.

ECMWF produces realtime forecasts (51 ensemble members each) twice per week, on Mondays and Thursdays. For each realtime forecast, a corresponding set of 20 hindcasts (11 ensemble members each) is produced. These hindcasts are initialized from the reanalysis of the same date of the 20 previous years. For example, hindcasts for the 2018-10-18 realtime forecast use initial conditions from 1998-10-18, 1999-10-18, ..., 2017-10-18.

The model operates on a spectral grid at T639L91 resolution (about 16 km horizontal grid spacing, 91 vertical levels) for the first 15 days. After day 15, it uses a T319L91 resolution (about 32 km horizontal grid spacing). The atmospheric model is coupled to an ocean model (0.25 degree resolution, 75 vertical levels, 5 ensemble members), a wave model, and a sea-ice model. Hindcasts use initial conditions from ERA5 reanalysis – except for those corresponding to realtime initializations before 2019-06-11, which use ERA-Interim – while the operational analysis is obtained from a high-resolution forecast (T1279L137) and a 4D-Var data assimilation algorithm. Ensemble members are generated by adding pairs of positive and negative perturbations to the (re-)analysis, obtained via singular vectors (representing the fastest growing perturbations), and ensemble data assimilation techniques. More details can be found at https://confluence.ecmwf.int/display/S2S/ECMWF+model+description.

Forecasts used this thesis

The majority of the analyses in this thesis employ S2S forecasts by ECMWF produced during winters 2017/18 to 2020/21, spanning from 16 November to 22 February each year. During these winters, there are 114 realtime forecasts (51 ensemble members each) and $114 \cdot 20 = 2280$ corresponding hindcasts (11 ensemble members each), totaling $114 + 2280 =$

2394 ensemble forecasts, or 30 894 single forecasts. Each forecast covers lead times of 0-46 days, resulting in approximately 1.5 million simulated winter days, which is roughly equivalent to 10 thousand winters, although these forecasts are not fully independent. Note that due to model updates, this set of forecasts originates from four different model cycles (CY43R3, CY45R1, CY46R1, CY47R1); however, the resulting uncertainties are assumed to be minor compared to the benefit of a larger sample.

3.1.2 ERA5 reanalysis

In general, reanalyses are produced by assimilating observations into a state-of-the-art weather forecast model, aiming to represent the best estimate of the atmospheric state on a full spatial and temporal grid. This study uses ERA5 (Hersbach et al., 2020), the latest reanalysis dataset provided by ECMWF. For analyses of large-scale dynamics ERA5 may be considered a representation of the real atmosphere (e.g., ERA-interim – the predecessor of ERA5 – has been demonstrated to represent a suitable dataset to study stratosphere-troposphere; Fujiwara et al., 2022, January). Furthermore, ERA5 is here used to evaluate forecast skill.

3.2 Defining sudden stratospheric warming and strong polar vortex events

Various definitions for SSWs have been employed in the literature (Butler et al., 2015). This thesis adopts the definition by Charlton and Polvani (2007) due to its relatively low data requirements and widespread usage.

For technical reasons, slightly different definitions of weak and strong polar vortex events are used in this thesis. While the SSW definition for ERA5 reanalysis data follows Charlton and Polvani (2007), adjusted definitions are used for the forecast data. Moreover, relaxed weak and strong polar vortex definitions are employed in chapter 5 to increase the sample size.

SSWs following Charlton and Polvani (2007)

This definition is based on daily means of zonal-mean zonal wind at 60 °N and 10 hPa (U_{60}^{10}). SSWs are defined during extended winter (November to March) as the first day when U_{60}^{10} becomes easterly. The threshold of $0\,\mathrm{m\,s^{-1}}$ is dynamically motivated by its implications on vertical Rossby wave propagation (Charney and Drazin, 1961, see equation A.7). Once a SSW is detected, a new event may occur 20 days later at the earliest, to avoid multiple counts of a single dynamical event. Final warmings are excluded, defined as SSWs where U_{60}^{10} does not return to westerly for 20 subsequent days before 30 April.

Predicted SSWs in S2S forecasts

In the ECMWF S2S forecasts, not all criteria of the Charlton and Polvani (2007) SSW definition can be checked due to the finite lead time. Therefore, a slightly relaxed definition is employed to diagnose predicted SSW in the these forecasts, which forms the basis for SSW analyses in chapter 4.

Predicted SSWs (p-SSW) are defined as the first day when U_{60}^{10} in a forecast becomes easterly. If multiple candidates for a p-SSW occur within one forecast, only the first event is considered. Importantly, p-SSWs require $U_{60}^{10} > 0\,\mathrm{m\,s^{-1}}$ between lead time days 0 to 9. This ensures that the polar vortex is intact at the beginning of the forecast and avoids the need to consider the U_{60}^{10} evolution in reanalysis data for the event definition. In addition, this ensures a degree of statistical independence, as events occurring at earlier lead times are likely to be similarly simulated across ensemble members. No explicit criterion is applied to exclude final warmings.

Accordingly, predicted strong polar vortex events (p-SPVs) are defined as the first day when U_{60}^{10} exceeds $47\,\mathrm{m\,s^{-1}}$. This threshold approximately describes the "opposite" quantile (Q91%) of the U_{60}^{10} winter season climatology compared to $0\,\mathrm{m\,s^{-1}}$ (Q9%).

Weak and strong polar vortex episodes

Weak and strong polar vortex episodes are used in chapter 5 to categorize S2S forecasts during the winter season (initializations from 16 November to 22 February) based on the polar vortex strength at initialization time. To simplify and achieve a larger sample size, weak polar vortex initializations are defined as forecasts where U_{60}^{10} in ERA5 at the time of forecast initialization is easterly. Strong polar vortex initializations are defined accordingly, where U_{60}^{10} exceeds a certain threshold. This threshold is chosen such that strong polar vortex initializations comprise an equal number of forecasts as weak polar vortex initializations. For the analysis in section 5.1, this threshold is $51.78\,\mathrm{m\,s^{-1}}$.

3.3 Reliability of ensemble forecasts

The initial conditions for a weather forecast are never perfectly known, but associated with an uncertainty. Ensemble forecasts provide a powerful framework to reflect these uncertainties (plus model uncertainties). In particular, the first two moments of the ensemble distribution, the ensemble mean and the ensemble variance, have important mathematical properties that will become relevant in the results part of this thesis. These properties generally assume that the numerical model represents a sufficiently complete and accurate representation of atmospheric physics. This translates into assuming that ensemble members are statistically indistinguishable from the observations. This implies that in a sufficiently large ensemble, there would always be an ensemble member that is close to the observations (but a priori, it is unknown which members are more or less close).

The ensemble mean corresponds to the expected value of the predicted quantity. Moreover, the ensemble mean forecast has a smaller squared forecast error than any individual ensemble member when averaged over many cases. Therefore, the ensemble mean is a common way to condense the full information contained in an ensemble.

The ensemble variance is a measure of forecast uncertainty. Importantly, the ensemble variance is directly related to the expected value of the squared error of the ensemble mean forecast. Let $Y^{(m)}$ with $m = 1, ..., M$ denote a set of M observations and $X^{(m,n)}$ with $n = 1, ..., N$ denote corresponding forecasts, with n indicating single ensemble members and N the ensemble size. $\bar{X}^{(m)} = \frac{1}{N}\sum_{n=1}^{N} X^{(m,n)}$ denotes the ensemble mean of the m'th forecast. The squared error of the ensemble mean of the forecasts, averaged over all cases, then equals the average ensemble variance, except for a factor $\frac{N+1}{N}$ (see Fortin et al., 2014, for a full derivation):

$$\frac{1}{M}\sum_{m=1}^{M}\underbrace{\left(Y^{(m)} - \bar{X}^{(m)}\right)^2}_{\text{squared error}} = \left(\frac{N+1}{N}\right)\frac{1}{M}\sum_{m=1}^{M}\underbrace{\frac{1}{N-1}\sum_{n=1}^{N}\left(X^{(m,n)} - \bar{X}^{(m)}\right)^2}_{\text{ensemble spread}}. \tag{3.1}$$

The equality may be anticipated by recognizing that the ensemble mean of the forecasts, $\bar{X}^{(m)}$, is subtracted 1) from the observation on the left-hand side (yielding the squared error) and 2) from individual ensemble members on the right-hand side (yielding the ensemble variance). When observations and individual ensemble members are statistically indistinguishable, the squares of these differences become equal when averaged over many cases, i.e., over sufficiently large M. An additional factor arises because the ensemble variance is computed using $N - 1$ degrees of freedom (one degree of freedom is dropped for computing the ensemble mean), whereas the calculation of the squared error is based on N degrees of freedom.

This spread-error relation is derived under the assumption that the model is reliable (i.e., has no errors), which implies that the ensemble spread reflects the actual uncertainty of the future atmospheric state under presence of initial condition uncertainty. In turn, spread-error comparisons are often used to test whether equation 3.1 holds, and therefore to assess whether the model is actually reliable. In so-called reliability diagrams, forecasts are grouped according to their ensemble spread. For each group, the mean ensemble spread is then plotted against the mean forecast error, and if the model is reliable, these averages should fall on the one-to-one line.

Such reliability diagrams are presented for the S2S ECMWF forecasts later in this section. However, before that, some synthetically generated random data are considered, highlighting intricacies of the spread-error relation, especially in cases of small ensemble sizes (such as in the ECMWF hindcasts, where $N = 11$). The reliability diagrams of these synthetic data experiments should guide the interpretation of the reliability diagrams of the actual S2S forecasts later.

Setup of a toy model to assess the spread-error relation

The toy model uses artificially generated observations and ensemble forecasts by randomly drawing from a specified normal distribution. By construction, these forecasts are reliable. Reliability diagrams are considered for 6 experiments, reflecting different methods to bin forecasts, different underlying distributions, and different ensemble sizes.

For each experiment, $M = 5000$ observations are randomly drawn from a normal distribution of variance σ^2_{true}. For each observation, an ensemble forecast comprising N ensemble members is generated, with each ensemble member being drawn from the same normal distribution of variance σ^2_{true}. In the real atmosphere, σ^2_{true} would represent the uncertainty of a prediction produced by a perfect model. This perfect model prediction is not deterministic, because of initial condition uncertainty. In this toy model, the forecasts are reliable, because there is no difference in the generation of observations and ensemble members. This setup facilitates analyses of how different situations manifest in a reliability diagram (see Fig. 3.1).

Experiment a: excellent spread-skill relation

Across the 5000 observation-forecast pairs, σ^2_{true} is chosen to be uniformly distributed between 0.5 to 2: $\sigma^2_{\text{true}} \sim \mathcal{U}(0.5, 2)$. Thereby, different flow situations are mimicked, in which initial condition uncertainty results in different levels of forecast uncertainty, i.e., different σ^2_{true}. The ensemble size is set to $N = 1000$. Forecasts are grouped based on eight quantiles of ensemble variance (i.e., each group is equally populated).

Fig. 3.1a shows that the average ensemble spread aligns well with the RMSE. Note that compared to equation 3.1, the square root of both the squared error and the ensemble variance is plotted, which is a common way to ensure the same units as the forecast quantity. Taking the square root does not change the interpretation of the results.

Experiment b: grouping based on forecast error results in bad spread-skill relation

This experiment uses an equal setting of parameters as experiment a, except that forecasts are grouped based on eight quantiles of forecast error, instead of quantiles of forecast spread. Fig. 3.1b shows that average spread and error of these eight groups do not align on the one-to-one line. Why does grouping based on error give qualitatively different results as grouping based on spread, as in experiment a? The spread-error relation holds if forecasts and observations are drawn from an underlying distribution with the same variance, σ^2_{true}. Ensemble spread is a good estimator of σ^2_{true}, at least for large ensemble sizes. Conversely, a single forecast error is not a robust estimate of σ^2_{true}, because sampling uncertainty is exceedingly large. Grouping based on forecast error, as done here, can therefore result in very different groups compared to grouping based on σ^2_{true}. Consequently, average error and spread are not aligned on the one-to-one line in this experiment.

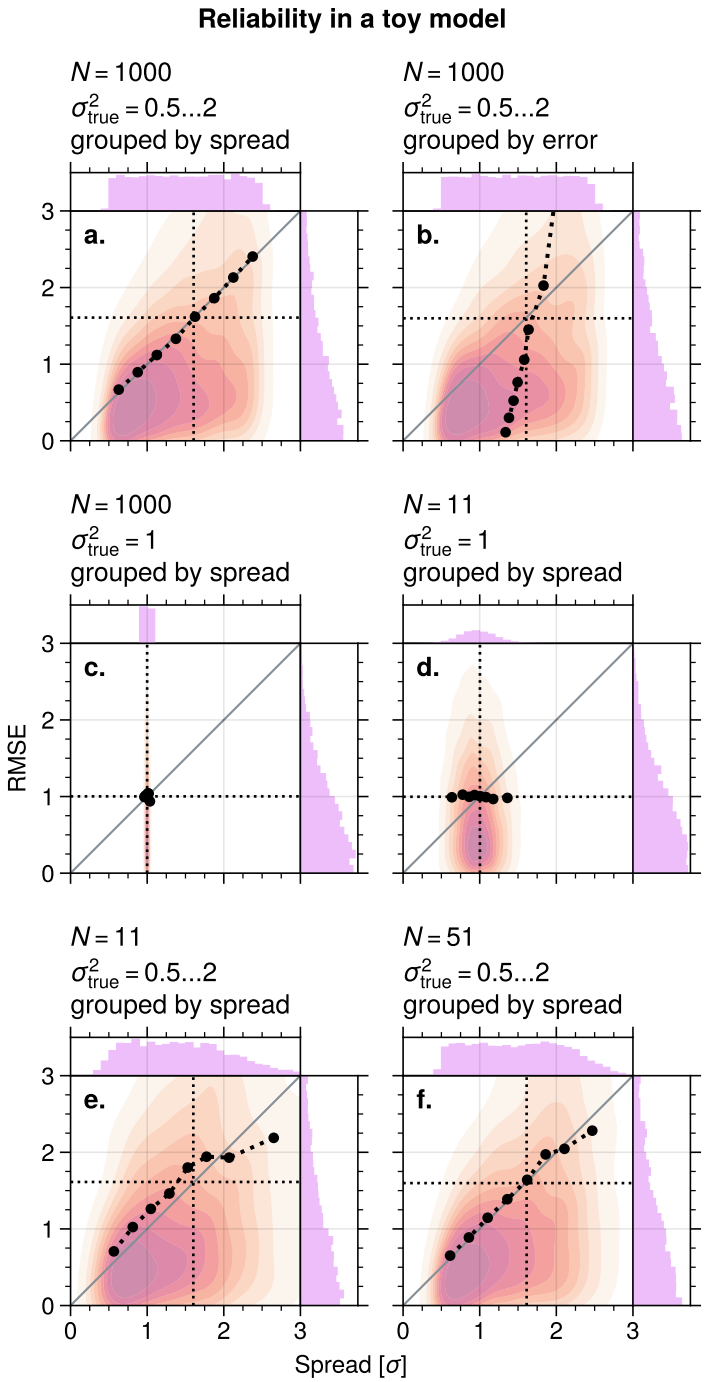

Figure 3.1: Illustration of the spread-skill relation in a toy model, based on synthetic data. Each panel presents one experiment, with specifications indicated on top. N is the ensemble size, σ^2_{true} is the variance of the underlying distribution from which ensemble members and observations are drawn. **x-axis:** square-root of the average ensemble variance (units of standard deviations, scaled by a factor of $\frac{N+1}{N}$), **y-axis:** root-mean-squared-error (dimensionless). Pale shading displays the 2-dimensional distribution of spread-RMSE pairs (one pair for each ensemble forecast). Black dots mark average spread and RMSE of eight equally populated forecasts that are grouped by ensemble spread (panel b: by RMSE). Pink histograms present the 1D distributions of spread and error. Thin black dotted horizontal and vertical lines indicate the total mean spread and error.

Experiment c: $\sigma^2_{\mathbf{true}} = \text{const.}$ results in constant spread and error, given large ensemble sizes

This experiment uses an ensemble size of $N = 1000$, but unlike experiments a and b, σ^2_{true} is fixed at $\sigma^2_{\text{true}} = 1$. Forecasts are grouped based on quantiles of ensemble spread.

Fig. 3.1c shows that while the forecast error distribution is still fairly wide, ensemble spread is always very close to 1. Therefore, mean spread and error of the eight quantile groups are all very close together, at values of error = 1, spread = 1.

This situation is not very realistic, because in the real atmosphere, different initial conditions can generally result in forecasts associated with different levels of uncertainty. Nonetheless, this experiment may demonstrate that the one-to-one line in a reliability diagram results from different values of σ^2_{true}.

Experiment d: $\sigma^2_{\mathbf{true}} = \text{const.}$ results in no spread-error relation in case of small ensemble sizes

This experiment uses the same setup as in experiment c, but the ensemble size is reduced to $N = 11$.

Fig. 3.1d shows that mean error and spread of the eight quantile groups do not align along the one-to-one line, but follow a horizontal line at error = 1. This means that variations of spread cannot indicate variations of error. This is because σ^2_{true} is kept constant, and hence, variations in spread only result from sampling uncertainty and do not reflect variations in σ^2_{true}.

Experiments e and f: potential spread-error-relation in S2S forecasts (11 and 51 ensemble members)

Figs. 3.1e and 3.1f present the same setup as in experiment a, where one-to-one alignment of average spread and error was nearly perfect, but instead of an ensemble size of $N = 1000$, the ensemble sizes of the S2S realtime forecasts ($N = 51$) and hindcasts ($N = 11$) are used. These cases illustrate how the reliability diagram would look like if the model was in principle reliable (ensemble members interchangeable with observations), which is always the case in this toy model. Note that this experiment still assumes $\sigma^2_{\text{true}} \sim \mathcal{U} \in (0.5, 2)$, while the distribution of σ^2_{true} is unknown in the real atmosphere.

In conclusion, the toy model highlights several key points:

- The ensemble spread in individual forecasts can be prone to substantial sampling errors, especially when the ensemble size is small.
- When grouping a sufficiently large set of forecasts into bins according to the ensemble spread, the average spread in each group approximately aligns with the corresponding average RMSE. However, deviations from the one-to-one line remain when the

ensemble size is small (generally appearing to be less than 25% in 11-member ensembles and less than 5% in 51-member ensembles). This is consistent with (Craig et al., 2022), who found that approximately 40 to 100 ensemble members are required for the ensemble standard deviation to follow an asymptotic convergence law.

- The total average spread matches the total average RMSE. While this requires a sufficiently large set of forecasts, the correspondence is independent of the ensemble size.

Now that these synthetic experiments explored the spread-error relation in a controlled setup, the following paragraphs analyze reliability of 1000-hPa-geopotential height (Z1000; instantaneous daily values at 0 UTC) and 2-meter-temperature (T2m; daily means) in the S2S ECMWF forecasts. These results are part of the supplementary material of Spaeth et al. (2024b).

Fig. 3.2 presents the spatial distribution of Z1000 average ensemble variances and squared errors at 7-13 days lead time. The squared error is computed as the squared difference between the ensemble mean forecast anomalies (with respect to forecast model climatology) and ERA5 anomalies (with respect to ERA5 climatology). The errors are computed grid-point wise based on daily fields and then averaged over lead time and over cases. At 7-13 days lead time, the average ensemble variance (scaled by a factor of $\frac{N+1}{N}$) is slightly smaller than the average mean squared error, indicating under-dispersion/ over-confidence. Qualitative structures are in good agreement with a maximum in variance around the storm track from Newfoundland over Iceland to Northern Europe (Figs. 3.2a-c). On average, at an exemplary grid point within the North Atlantic storm track (60 °N, 10 °E) the ensemble is slightly under-dispersive at the short- and medium-range and tends to be slightly over-dispersive at lead time beyond about five weeks (Fig. 3.2d). Furthermore, a reliability diagram shows forecasts binned into ten quantiles according to the ensemble spread averaged over 7-13 days lead time (Fig. 3.2e). There is a high correlation between mean spread and mean squared error across the quantiles.

In light of the toy model results (specifically, Fig. 3.1e), the deviations from the one-to-one line likely result from the relatively small ensemble size of the hindcasts (11 members), but could additionally be due to model errors.

Fig. 3.3 shows the corresponding reliability assessment for T2m. Climatological variance (scaled by a factor of $\frac{N+1}{N}$) and squared error are larger over land than over the sea and are larger towards the pole (Figs. 3.3a,b). The ensemble at 7 to 13 days lead time tends to be under-dispersive at high latitudes (Figs. 3.3c). At an exemplary grid point at 50 °N, 10 °E, the ensemble tends to be slightly under-dispersive in the first forecast week and somewhat over-dispersive beyond the second forecast week (Fig. 3.3d). Anomalies of variance and error at this grid point are fairly well aligned, despite some deviations from the one-to-one line (Fig. 3.3e), which may result from a small ensemble size and from model errors.

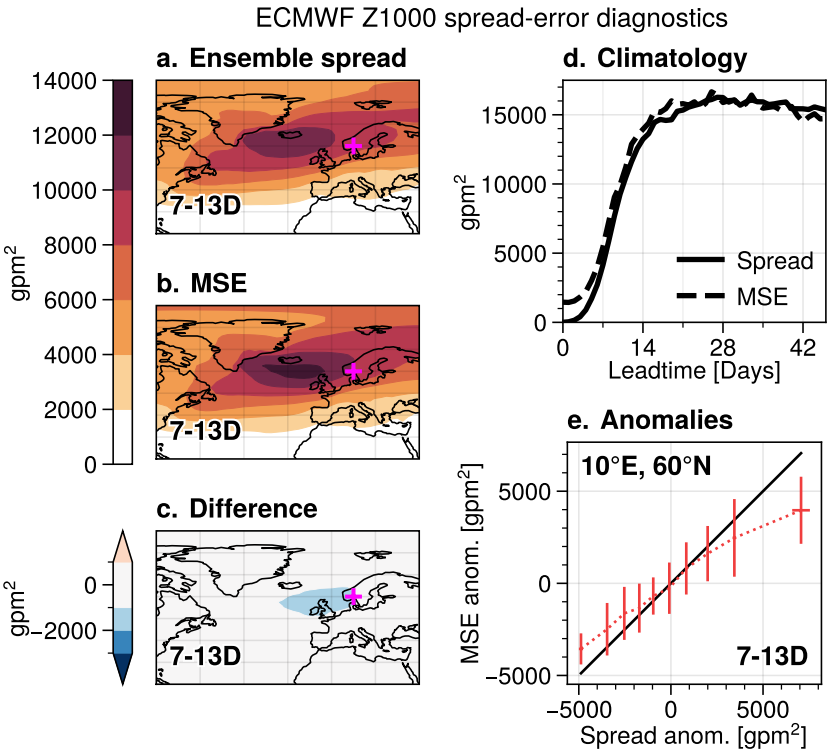

Figure 3.2: Reliability assessment of 1000 hPa geopotential height (Z1000) in terms of ensemble variance and mean squared error, based on 1197 ensemble forecasts (model cycles CY46R1 and CY46R1 with initializations from November 16 to February 22). (a.) Mean ensemble spread (here equal to the ensemble variance), (b.) mean squared error and (c.) their difference (spread minus error), each averaged over 7-13 days lead time. (d.) Climatological evolution of ensemble spread (solid line) and mean squared error (dashed line) over lead time at 60 °N, 10 °E (location indicated by magenta cross in panels a.-c.). (e.) Corresponding anomalies of ensemble spread and mean squared error at 7-13 days lead time, both binned into ten quantiles according to anomalous spread. Red bars indicate 95% confidence intervals of the respective quantile averages, derived from resampling. Deviations from the one-to-one line can arise from a small ensemble size and additionally from model errors. In all panels, ensemble spread is scaled by a factor (N+1)/N, where N is the ensemble size (see equation 3.1).

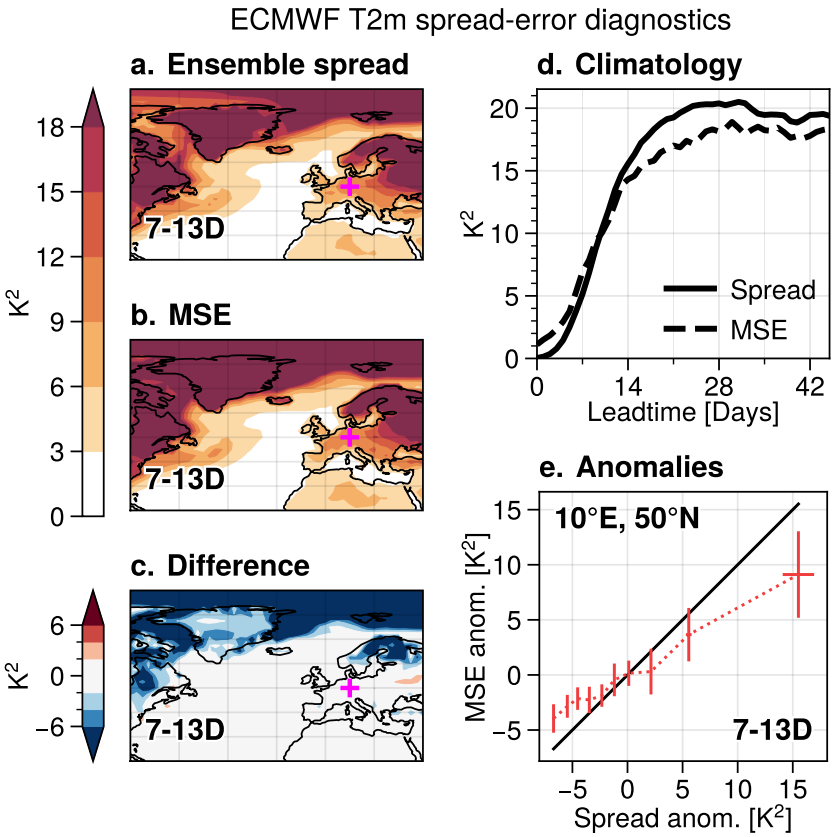

Figure 3.3: As Fig. 3.2, but for 2-meter-temperature. Panels d, e refer to an exemplary grid point at 50 °N, 10 °E.

3.4 Forecast deseasonalization

Many atmospheric quantities follow a seasonal cycle tied to the evolution of incoming solar radiation. To analyze variations beyond this seasonal cycle, its climatological mean is often subtracted from the data to obtain anomalies. Several methods exist to compute the climatological seasonal cycle.

One method is to compute a daily climatology by averaging the values for each day of the year across many years. This requires a long-term record, typically spanning at least a few decades, to ensure that high-frequency weather fluctuations occurring in individual years are smoothed out. Alternatively, a monthly climatology can be computed by averaging the values for each month across all years. Daily climatologies can then be interpolated from this monthly climatology. Another approach is to fit Fourier modes to the data, which may be suitable for quantities with a relatively harmonic seasonal cycle.

In this thesis, anomalies are also computed for forecast data. However, deriving a multi-year climatology for forecasts presents additional challenges due to the finite forecast lead time and the fact that model climatologies can be lead time dependent as a result of model drifts. To address these issues, weather centers contributing to the S2S database provide hindcasts in addition to realtime forecasts. For a given realtime forecast, ECMWF provides a set of 20 hindcast ensembles, which are forecasts initialized from the initial conditions of the 20 previous years. These hindcasts are used to compute inter-annual averages, forming a climatology.

The following provides a detailed description of how these climatologies can be computed.

For illustration, the climatology of the polar vortex strength is considered (Fig. 3.4), quantified by the zonal-mean zonal wind at 60 °N and 10 hPa (U_{60}^{10}), from mid-November to mid-April. For comparison to the forecast data, the climatology based on ERA5 reanalysis data is also shown, computed as inter-annual averages for a given day of year with an additional 7-day moving average applied. The polar vortex climatology reaches its maximum strength around late December to early January and then continuously weakens until spring.

The climatology for the forecast data is computed for each realtime initialization separately. The presented period includes 29 initializations from 2020-11-16 to 2021-02-22. For each of these initializations, a climatology is computed by averaging over the 20 corresponding hindcasts and over ensemble members.

While there is overall fair agreement between the ERA5 and the S2S forecast climatologies, some discrepancies exist, partly arising from model drifts over lead time. These discrepancies can be as large as $10\,\mathrm{m\,s^{-1}}$, though all climatologies still fall within the inter-annual inner-quartiles of ERA5. Insufficient sampling likely also contributes to differences between climatologies of different initialization dates and between ERA5 and forecast climatologies.

The fact that S2S forecast climatologies show differences between different initializations motivates averaging the climatologies for neighboring initializations. For instance, the climatology for the 2020-12-07 initialization may become more statistically robust by additionally including hindcasts from the 2020-12-03 and 2020-12-10 initializations. However,

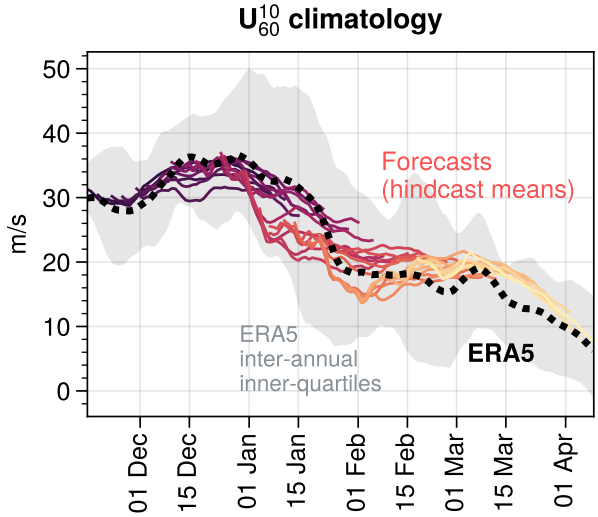

Figure 3.4: Illustration of forecast climatologies for the polar vortex strength as measured by the zonal-mean zonal wind at 10 hPa, 60 °N. Each colored thin line represents an average of 20 hindcast sets (one for each year) that corresponds to one realtime forecast (realtime initializations from 2020-11-16 to 2021-02-22). Thick dotted line denotes the climatological mean from 2000-2019 ERA5 data, with a 7-day-running-mean applied. Shading indicates the corresponding inter-annual inner-quartiles, i.e., year-to-year variability.

this raises questions about how many neighboring initializations should be included and whether hindcasts of neighboring initializations should be averaged according to the same day of the year or the same lead time.

Fig. 3.5 illustrates how resulting anomalies differ between various forecast climatology calculations. The example focuses on the realtime forecast initialized on 2020-12-07, for which different climatologies are constructed and forecast anomalies are computed accordingly. Four methods are compared:

a. **Seasonal average:** All hindcasts of a winter period, here chosen to include initializations from 16 November to 22 February, are combined. Averages are computed for each lead time day including all hindcasts and all ensemble members.

b. **Single initialization:** Only hindcasts corresponding to the realtime forecast (initialized on 7 December 2020) are included. Averages are computed over hindcasts and over ensemble members for each lead time day.

c. **Neighboring initializations average:** The climatology is computed as in (b), but including all hindcasts corresponding to the realtime forecasts initialized within 7 December 2020 $\pm$ 7 days. Hence, instead of 20 hindcast ensembles, the average is computed over $5 \cdot 20 = 100$ hindcast ensembles.

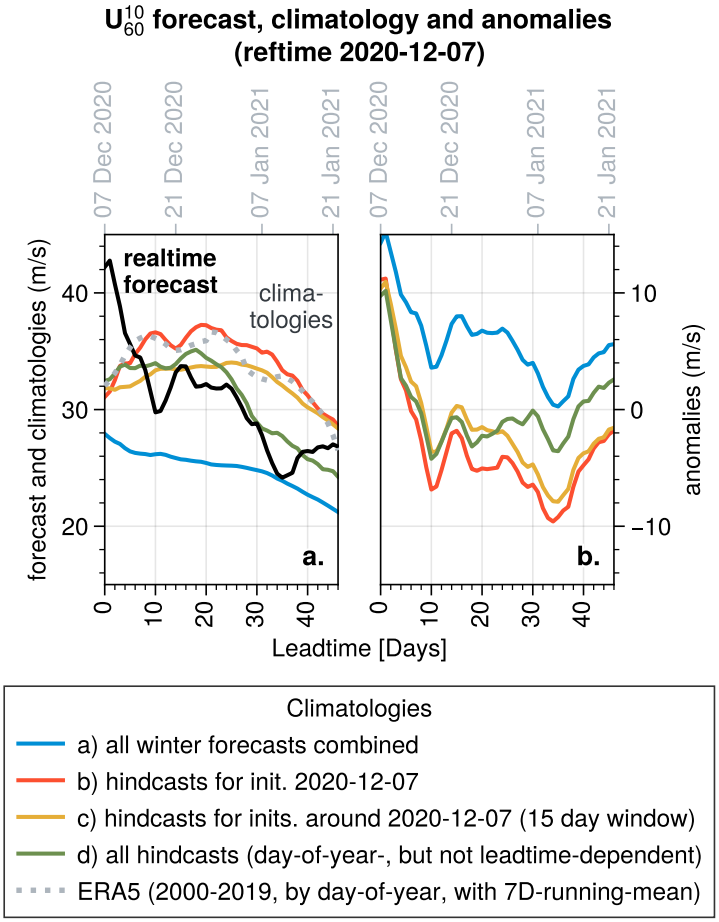

Figure 3.5: Comparison of climatology calculation methods, exemplary for deseasonalizing the 2020-12-07 realtime forecast of U_{60}^{10}. **a** Realtime ensemble mean forecast (thick black line) and five different climatologies (see legend and text for details). **b** Forecast anomalies as obtained by subtracting the climatology from the full field.

d. **Day-of-year average:** The climatology is based on all hindcasts corresponding to the realtime forecast initialized in winter 2020/21. The average is computed over all initialization-leadtime-ensemble member-combinations for a given day-of-year. The climatology for a given day therefore includes different lead times and model drifts are not corrected for.

Climatology (a) shows the weakest wind speeds throughout the forecast among the four methods. The weaker wind speeds arise because forecasts that are initialized more than two months later were included, during which the polar vortex is already much weaker. Therefore, method (a) does not appear to be suitable, particularly if the quantity of interest follows a distinct seasonal cycle.

Climatology (d) differs from climatologies (b) and (c), as well as from the ERA5 climatol-

ogy, especially at longer forecast lead times. This discrepancy arises because method (d) mixes different lead times, while the polar vortex strength shows considerable lead time dependence, particularly in January (see Fig. 3.4). Note that this lead time dependence could also result from insufficient sampling. Generally, this method does not appear to be suitable due to substantial lead time-dependent forecast drifts (i.e., biases) in some cases. However, method (d) includes a large number of data points, which can be beneficial when analyzing higher moments or extremes of the climatological distribution.

Climatology (b) uses only the 20 hindcast ensembles that correspond to the realtime forecast of interest. This method tends to be computationally cheaper than the other methods. While it includes fewer data points, it may be suitable when inter-annual variations are not overly strong and the climatology is expected to converge based on the included 20 hindcasts ensembles with 11 members each. Note that this method cannot be applied to a realtime/ hindcast setup where realtime forecasts are produced at a higher frequency than hindcast forecasts, such as at the UK MetOffice (realtime forecasts: daily, hindcasts: 4 per month).

Climatology (c) follows the calculation of climatology (b), but it also includes forecasts from "neighboring" initializations. For the 2020-12-07 forecast, additional forecasts are included from 2020-11-30 to 2020-12-14 (5 in total). A smaller window size ensures that seasonal variations within the included forecasts are small, whereas a larger window size includes more data and is thus more statistically robust. This method is used throughout this thesis, with a window size of ± 7 days, unless stated otherwise.

The four climatology computation methods can result in quantitative differences in practical analysis outcomes. To illustrate this, composite-mean 2 meter temperature (T2m) forecast anomalies are considered four weeks following SSWs. The analysis includes hindcast ensemble forecasts that are initialized between 3 days before to 3 days after SSWs, with SSWs being identified in ERA5 following the event definition by Charlton and Polvani (2007). This approach results in 94 ensemble forecasts initialized around 14 SSWs from winters 1998/99 to 2019/20.

Figure 3.6 presents the T2m anomalies for the four different underlying climatology methods and for ERA5. The ERA5 climatology is computed by averaging days with the same day of year, using the period 1998 to 2019. For the analysis of post-SSW T2m, ERA5 anomalies are sampled to match the same days as the S2S forecasts. However, the ERA5 anomalies are likely affected by sampling uncertainty due to only few SSWs and should therefore not be treated as a ground truth. While sampling in the S2S forecasts is approximately 10 times better, due to the 11 ensemble members (compared to a single "ensemble member" in ERA5), differences still arise depending on the method. The comparison reveals that anomalies computed using the climatology method (a), which is based on a seasonal average, show the largest deviation from other methods. This is likely because the seasonal temperature evolution is not captured. In contrast, differences between the other composites, which are based on climatology methods (b), (c) and (d), are minor. This result suggests that when multiple forecasts are combined and averaged, the choice of the underlying climatology method may not impact the outcome drastically.

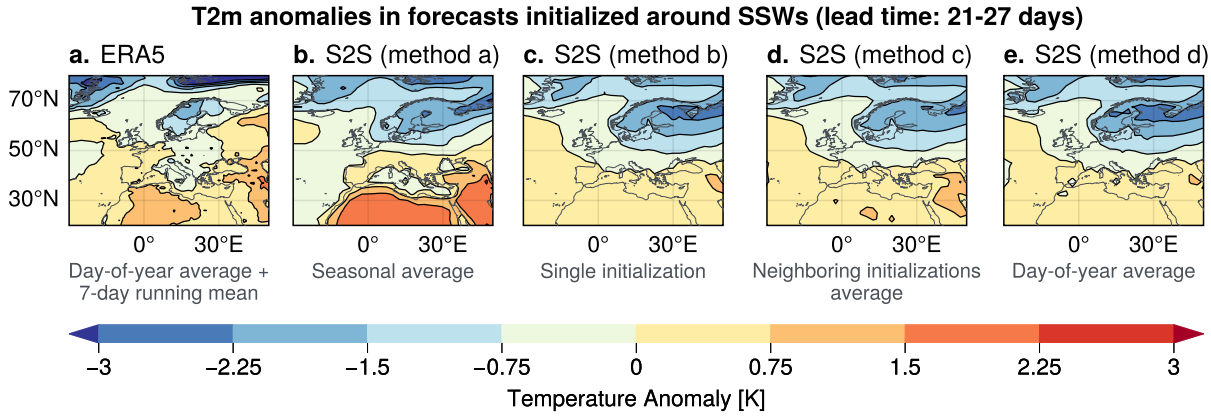

Figure 3.6: Composite-mean of 2 meter temperature anomalies in forecasts that are initialized between 3 days before to 3 days after SSWs (identified in ERA5 data) at lead times of 21 d to 27 d. **a.** ERA5 reanalysis data, with days sampled to match the S2S forecasts. **b.-e.** S2S forecast data, with different methods to compute the climatology (see text for details).

	Mean differences				Mean abs. diff.				Q95% of abs. diff.			
	a	**b**	**c**	**d**	**a**	**b**	**c**	**d**	**a**	**b**	**c**	**d**
method a	0	0.19	0.14	0.28	0	1.21	1.03	1.14	0	4.77	4.03	4.69
method b	−0.19	0	−0.05	0.09	1.21	0	0.3	0.29	4.77	0	1.09	0.94
method c	−0.14	0.05	0	0.14	1.03	0.3	0	0.3	4.03	1.09	0	1.08
method d	−0.28	−0.09	−0.14	0	1.14	0.29	0.3	0	4.69	0.94	1.08	0

unit: Kelvin

Figure 3.7: Comparison of 2-meter temperature forecast anomalies for different climatology calculation methods. These forecasts comprise 94 ensembles initialized around SSWs in ERA5 (tolerance: ±3 days relative to the SSW, see Fig. 3.6). Mean differences, mean absolute differences and the 95th percentile of absolute differences (unit: Kelvin) are computed for individual forecasts, lead times and grid points. Differences are computed by subtracting the anomaly obtained via the method given in each column from the anomaly obtained via the method given in each row. The climatology methods are as follows: a) seasonal average, b) single initialization, c) neighboring initializations average, d) day-of-year average (see Fig. 3.5).

However, substantial differences can arise for individual forecasts, lead times and grid points, as shown in Figure 3.7. Compared to method (c), 5% of the individual anomaly data points deviate by more than 1 Kelvin when using methods (b) or (d), and by more than 4 Kelvin when using method (a).

Finally, standardized anomalies can be computed by dividing anomalies by the inter-annual standard deviation. It is important to note that the inter-annual standard deviation must be computed based on individual ensemble members, not the ensemble mean. This is because the ensemble mean always converges towards climatology at long lead times, resulting in a too small inter-annual standard deviation.

3.5 `s2stools`: a new Python library for S2S forecast data processing

This section describes `s2stools`, a Python library developed to deal with the following challenges in processing S2S forecast data.

First, data size is considerable, particularly because of the hindcasts produced alongside each realtime initialization. For instance, a single variable on one vertical level for the Northern Hemisphere, with a $2.5° \times 2.5°$ horizontal grid, with daily output (0-46 days lead time), requires about 130 MB of disk space for one initialization. This includes 51 realtime ensemble members and $20 \cdot 11$ hindcast ensemble members. Analyzing multiple vertical levels, variables and initialization dates results in data sizes that often exceed the available RAM capacity.

Second, the forecast data structure is complex, with multiple coordinates linked to the time dimension. These coordinates include initialization time, lead time, valid time, hindcast year, and potentially more. This structure makes vectorized operations, which are desired in Python, challenging.

To streamline the processing of S2S forecasts and facilitate efficient computations, the Python package `s2stools` was developed along with this thesis. `s2stools` provides a collection of tools for structuring S2S forecast data and performing common calculations. The package is fully based on `xarray` (Hoyer and Hamman, 2017), a leading python library for working with multi-dimensional scientific datasets. This also enables all `s2stools` functionalities to be used with `dask` (Dask Development Team, 2016), a Python library facilitating parallelized computations that can scale to large clusters in high-performance computing environments.

Below are some of the main functionalities provided by `s2stools`. Comprehensive documentation can be found at https://s2stools.readthedocs.io/en/latest/. The code is available at https://github.com/jonas-spaeth/s2stools/.

Opening multiple netcdf files of s2s forecasts

ECMWF S2S forecasts downloaded from the S2S database[1] usually comprise four netcdf files per realtime initialization: realtime control forecast (1 member, 47 days of forecast time), realtime perturbed forecasts (50 members), hindcast control forecasts (1 members) and hindcast control forecasts (10 members, 20·47 days of forecast time, corresponding to the 20 different hindcasts years).

`s2stools` provides a user-friendly way of combining these four file types, optionally for multiple initialization dates, all at once, using the `s2stools.process.s2sparser`:

```
import xarray as xr  # Hoyer and Hamman, 2017
import s2stools      # developed for this thesis

path = "data/raw/s2s/ecmwf/u_10hpa_60/*2020*.nc"
ds_s2s = xr.open_mfdataset(path, preprocess=s2stools.process.s2sparser)
print(ds_s2s)
```

```
<xarray.Dataset>
Dimensions:    (leadtime: 47, longitude: 432, latitude: 1,
number: 51, reftime: 29, hc_year: 21)
Coordinates:
* leadtime   (leadtime) timedelta64[ns] 0 days ... 46 days
* longitude  (longitude) float32 -180.0 -179.0 ... 179.0
* latitude   (latitude) float32 60.0
* number     (number) int64 0 1 2 3 4 5 ... 46 47 48 49 50
* reftime    (reftime) datetime64[ns] 2020-01-02 ... 20...
* hc_year    (hc_year) int64 -20 -19 -18 -17 ... -2 -1 0
validtime  (reftime, leadtime, hc_year) datetime64[ns] ...
Data variables:
u          (reftime, latitude, longitude,
leadtime, hc_year, number) float32
```

where the control member (`number=0`) has been combined with the other ensemble members (`number=1...51`) and the original dimension `time` has been split into three dimensions: `reftime` (the initialization date of the corresponding realtime forecast), `hc_year` (integer values indicating the hindcast year relative to the year of the realtime forecast), and `leadtime`. Additionally, the new coordinate `validtime` provides the actual date for which the forecast is made.

Deseasonalization

`s2stools` facilitates straight-forward calculation of a leadtime- and initialization-dependent climatology, following method (c) described in section 3.4:

```
ds_s2s_clim = s2stools.clim.climatology(ds_s2s, window_size=15, ndays_clim_filter=7)
print(ds_s2s_clim)
```

```
<xarray.Dataset>
Dimensions:    (reftime: 29, leadtime: 47, longitude: 432,
latitude: 1)
Coordinates:
* reftime    (reftime) datetime64[s] 2020-01-02 ... 202...
* leadtime   (leadtime) timedelta64[ns] 0 days ... 46 days
```

[1] https://apps.ecmwf.int/datasets/data/s2s-realtime-instantaneous-accum-ecmf

```
* longitude  (longitude) float32 -180.0 -179.0 ... 179.0
* latitude   (latitude) float32 60.0
Data variables:
u          (reftime, latitude, longitude, leadtime) float32
```

where `window_size` refers to the total number of days spanning the window in which forecasts are included for the climatology and `ndays_clim_filter` applies a centred running mean smoothing. Anomalies can then be computed using `ds_s2s - ds_s2s_clim`.

Combining reanalysis and forecast data

A common task in dealing with forecast data is a comparison against a verification dataset, in this case ERA5. It can be technically challenging to compare the continuous ERA5 time series with the forecasts, which are irregularly initialized and have finite lead time. `s2stools` provides a simple interface for comparing reanalysis with forecast data:

```
path = "data/raw/era5/u60_10hpa.nc"
ds_era5 = xr.open_dataset(path)
print(ds_era5)
```

```
<xarray.Dataset>
Dimensions:    (time: 5092, longitude: 144)
Coordinates:
* time       (time) datetime64[ns] 1979-01-01 ... 2020-...
* longitude  (longitude) float32 -180.0 -177.5 ... 177.5
Data variables:
u          (time, longitude) float32 ...
```

```
ds_combined = s2stools.process.combine_s2s_and_reanalysis(
ds_s2s, ds_era5, ensfc=False
)
print(ds_combined)
```

```
Renaming reanalysis variables by adding _verif (for verification)
<xarray.Dataset>
Dimensions:    (leadtime: 47, longitude: 432, latitude: 1,
number: 51, reftime: 29, hc_year: 21)
Coordinates:
* leadtime   (leadtime) timedelta64[ns] 0 days ... 46 days
* longitude  (longitude) float32 -180.0 -179.0 ... 179.0
* latitude   (latitude) float32 60.0
* number     (number) int64 0 1 2 3 4 5 ... 46 47 48 49 50
* reftime    (reftime) datetime64[ns] 2020-01-02 ... 20...
* hc_year    (hc_year) int64 -20 -19 -18 -17 ... -2 -1 0
validtime  (reftime, leadtime, hc_year) datetime64[ns] ...
plev       int64 ...
Data variables:
u          (reftime, latitude, longitude, leadtime, hc_year, number) float32 ...
u_verif    (reftime, leadtime, hc_year, longitude) float32 ...
```

where the reanalysis field has been been projected onto the s2s forecast grid. Note that `u` in the reanalysis dataset has been renamed to `u_verif` due to duplicate variable names in the two datasets.

Additional functionalities

`s2stools` offers a range of additional features, including:

- creating event-based composites (see also the Python package `xevent`[2] developed along this thesis)
- proper formatting of plots with a lead time axis (addresses the issue of timedelta objects having nanosecond precision in `xarray`)
- adding the ECMWF model cycle (e.g., "CY47R1") as a coordinate to a dataset
- an interface for online access to commonly used teleconnection indices including QBO, MJO and ENSO.

More information is provided in the documentation: https://s2stools.readthedocs.io/en/latest/.

[2]Code is available at https://gitlab.lrz.de/ru35hub/xevent, documentation is available at https://ru35hub.pages.gitlab.lrz.de/xevent/

Chapter 4

UNSEEN sudden stratospheric warmings: using large ensembles to boost the statistics of extreme events

4.1 Application of the UNprecendented Simulated Extremes using ENsembles (UNSEEN) approach to S2S forecasts and sudden stratospheric warmings

Extreme events are captivating to study due to their potential implications for high-impact weather. However, by definition, they occur rarely, which undermines the statistical robustness of the analyses. This issue also pertains to extreme polar vortex events, including SSWs. It is known that SSWs can affect the tropospheric circulation and lead to surface weather extremes (Butler and Domeisen, 2021), but the observational record comprises only a few dozen events (ERA5: 42 SSWs from 1958 to 2023). While this sample size may suffice to robustly determine the mean and, perhaps, the variance of a distribution affected by SSWs, modulations of higher moments and of extremes remain highly uncertain.

To address this challenge, some studies have used climate simulations (e.g., Charlton-Perez et al., 2008; Bancalá et al., 2012; Hitchcock and Simpson, 2014), which provide on the order of 100 to 1000 winters' worth of data for deriving SSW statistics. Other studies have conducted simulations using idealized models (e.g., Jucker et al., 2014; Lindgren et al., 2018; Finkel et al., 2023b). While this approach may provide large sample sizes, the validity of results hinges on the assumption that these models accurately capture the essential dynamics of the stratosphere and its coupling with the troposphere.

Here, a novel approach to studying rare events is presented, which has in similar contexts been referred to as the UNprecedented Simulated Extreme ENsemble (UNSEEN) approach. This method assumes that individual ensemble members of a state-of-the-art weather forecast model each represent possible evolutions of the real atmosphere. There-

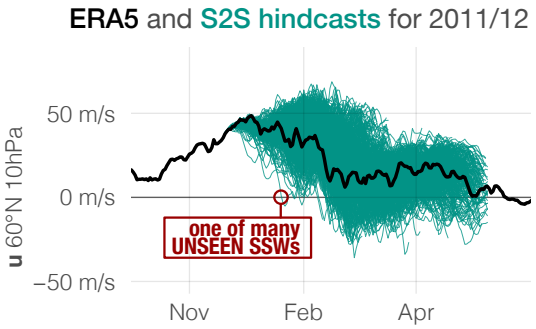

Figure 4.1: Illustration of the UNprecedented Simulated Extreme ENsemble (UNSEEN) approach to SSWs. ERA5 evolution of the zonal-mean zonal wind at 60 °N and 10 hPa (U_{60}^{10}), for winter 2011/12 (black) and corresponding S2S ECMWF hindcasts (teal, related to realtime forecasts of 2017-2021) for mid-November to end-February initializations. While no major SSW occurred in ERA5 during this winter, many ensemble members predicted SSWs, some of which simulate up to 30 m s^{-1} easterly winds.

fore, events simulated within these ensemble members can be used as surrogates of real events, even if the majority of these events have not occurred in the real atmosphere. If the dynamics and physics of ensemble members are indistinguishable from those of the real atmosphere, then ensemble forecast data can provide substantially more data then observations or reanalyses. This abundance of data allows for a more robust analysis of rare events, potentially aiding our understanding of extreme weather phenomena, such as SSWs (see Fig. 4.1).

The UNSEEN approach was initially proposed by van den Brink et al. (2004) and has since then been applied to various contexts, such as precipitation extremes using climate simulations (Thompson et al., 2017) and seasonal forecasts (Kelder et al., 2020). The Master's thesis preceding this PhD thesis was among the first studies to apply the UNSEEN approach to polar vortex extremes (Spaeth, 2020). This groundwork has been further developed, and the outcomes are presented in this chapter. Section 4.3 builds upon Bauer (2022) and Sirbescu (2023), studying the attribution of SSWs to preceding tropospheric heat flux and blocking events. Section 4.4 analyzes the relation between SSWs and subsequent AO extreme events (published in Spaeth and Birner, 2022).

To date, other studies have applied the UNSEEN approach to derive frequency and seasonal distributions of SSWs (Finkel et al., 2023a), estimate the impact of ENSO on SSWs (Ineson et al., 2023), and quantify the NAO response to different SSW types (Bett et al., 2023).

Data for identifying UNSEEN SSWs

Here, 2394 ensemble forecasts from ECMWF's extended-range forecasts are used, comprising 114 realtime forecasts (51 members each) and 2280 hindcasts (11 members each). This results in a total of 30 894 ensemble members, each of which provides a 47 day time series

of the atmospheric evolution. Details are described in section 3.1.1.

Identification of UNSEEN SSWs

We define specific events and then scan each forecast for the occurrence of such an event. SSWs are defined as days when U_{60}^{10} transitions from positive to negative, indicating a breakdown of the polar vortex. If multiple SSWs occur within one forecast, only the first event is considered. Note that, by definition, all identified events are predicted events that may or may not actually occur in the real atmosphere. To highlight this aspect, and to avoid confusion with actual real-atmospheric events, the events identified in the forecasts are denoted with a "p"-prefix, where "p" stands for "predicted" (alternatively, it could be interpreted as "potential" for some aspects). Please note that the term "p-events" is used interchangeably with UNSEEN events in this thesis. However, we continue to use "p-events" in Section 4.4 to maintain consistency with our published article.

For early lead times, when ensemble spread is small, different ensemble members are not statistically independent. Therefore, we analyze only those events that occur at or after a forecast lead time of 10 days. It is assumed that by this point, initial condition-memory has sufficiently vanished, ensuring that individual forecasts are sufficiently independent. Furthermore, the use of hindcasts ensures that different phases of low-frequency variability are sampled, such as related to the El-Niño-Southern-Oscillation, the Maddan-Julian-Oscillation or sea ice variations. Finally, to ensure that both negative and positive lags can be evaluated for each event, only events predicted between 10 and 36 days lead time are considered. This allows for the examination of the lag period ± 10 days for all events, though fewer events contribute to the composite at lags longer than this period.

In order to ensure an intact westerly polar vortex at the start of the forecast, p-SSWs are only identified in forecasts predicting positive U_{60}^{10} throughout the these first 10 days. Following this event definition, we identify 6 101 p-SSWs in the ECMWF forecasts.

Statistics of p-SSWs

Figure 4.2 provides an overview of the distribution of ECMWF p-SSWs as a function of the year, forecast lead time and calendar month. p-SSWs are found in all winter seasons considered, with the absolute numbers indicating how many events each winter season contributes to the analysis.

The largest absolute number of events is identified in the winter season 2017/18, which includes the most forecasts (realtime 2017/18 plus hindcasts related to initializations from 2018/19 to 2020/21). Several factors contribute to the varying number of events between the different years. These factors include internal dynamic variability, climate variability on interannual timescales, and a slightly varying number of underlying forecasts due to the realtime/ hindcast prediction setup.

A forecast that is initialized with a strong polar vortex tends to maintain a strong polar vortex and predict fewer SSWs compared to a forecast with an initially weak polar vor-

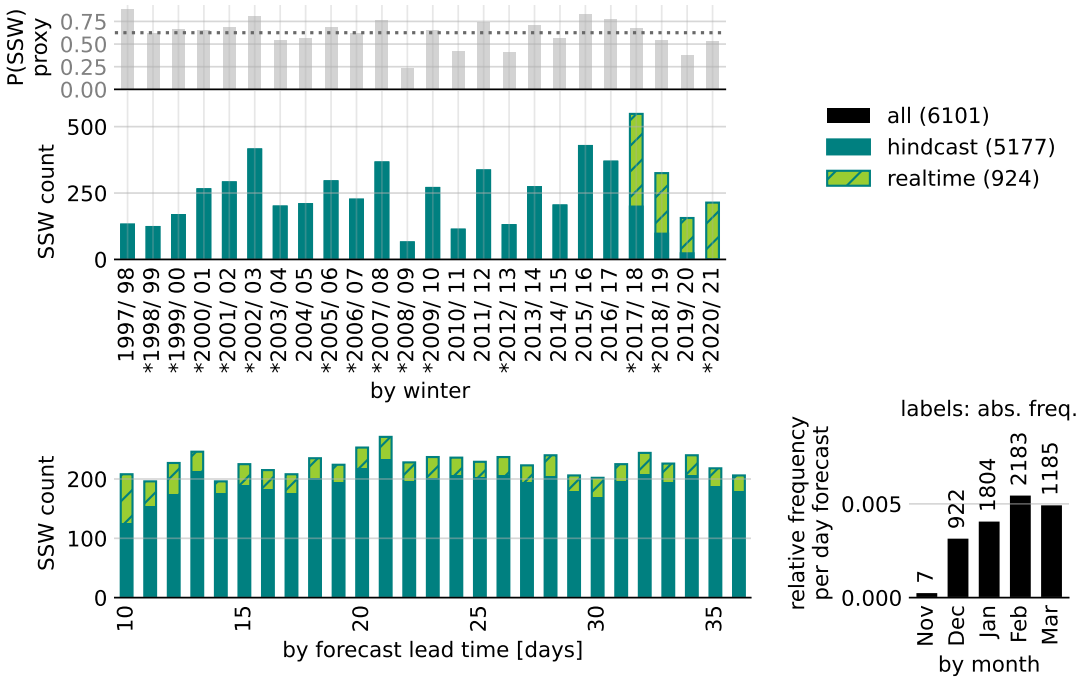

Figure 4.2: Distribution of analyzed p-SSWs in ECMWF forecasts. Absolute event counts (center left) and seasonal probability proxy (top left), grouped by winter. Asterisks denote years with real atmosphere SSWs. Grouped by forecast lead time (bottom left) and by month (bottom right).

tex. Moreover, forecasts that do not start with ten consecutive days of positive U_{60}^{10} are discarded by default. Thus, if the polar vortex in the real atmosphere is already easterly at the initialization time or is predicted to become easterly within the first ten days, such forecasts will not contribute any events to the analysis. This can be illustrated by the example of the 2009'th SSW (24 January 2009, see Butler et al., 2017). The event had low predictability at lead times longer than 8 days (Karpechko, 2018). Before the event, between end of December 2008 and mid-January 2009, the polar vortex was exceptionally strong, resulting in marginal SSW probability in the forecasts and suggesting that the event itself was unlikely given the prevailing dynamics[1]. As a result, 2008/09 shows the lowest number of p-SSWs: In the first winter half up to initialization dates around mid-January, hardly any events were predicted due to the relatively strong polar vortex. Later, forecasts predicting the real atmosphere SSW only did so at less than +10 day lead time, such that those events are discarded. Later initializations up to mid-February are excluded, because these do not predict persistent positive U_{60}^{10} within the first 10 days lead time, due to the preceding SSW. As a result, winter season 2008/ 09 contributes only 64 p-SSWs to the analysis.

Figure 4.2 also presents a normalized SSW probability, which serves as a proxy for P(SSW). From observations, the annual probability of SSWs can be derived by normalizing the number of winters with SSWs with the total number of winters. In the S2S model framework, it is however less straightforward to compute the frequency of SSWs per winter, as the maximum leadtime is shorter than a winter period and many forecasts overlap. In this study, we compute a proxy for the model's seasonal SSW probability based on the number of SSWs per forecast day, as described in the following:

For each winter season i, forecasts with initialization dates between mid-November and mid-February are analyzed, resulting in a total of $\tilde{N} = \sum_i \tilde{N}_i$ forecast runs (counting ensemble members separately). We search for p-SSWs only in forecasts that have solely positive U_{60}^{10} within the first 10 days after initialization, resulting in $N = \sum_i N_i$ forecasts ($N \leq \tilde{N}$). We find E_i p-SSWs events in the winter seasons, respectively, and group those by daily leadtime (similar to Fig. 4.2, bottom left panel), yielding $E_{i,d}$ p-SSWs in winter i at leadtime $+d$ days. As $E_{i,d}$ is approximately constant over leadtime, we compute the average number of p-SSWs in winter i per day leadtime: $E_i = \overline{E_{i,d}}$, where the overbar denotes the mean over lead times. Hence, the probability that a random forecast in winter i at a random leadtime shows a p-SSW is $p_{i,daily} = \frac{E_i}{N_i}$. The probability of no SSW for an entire winter (≈ 135 days from mid-November to end of March) is therefore $(1-p_{i,daily})^{135}$. Finally, the probability of at least one SSW in winter i becomes: $p_i = 1 - (1 - p_{i,daily})^{135}$, as presented in Fig. 4.2 (top left panel). The model's average seasonal SSW probability yields 63%. This is consistent with the number of observed SSWs in reanalyses that is roughly 6 per decade (Butler et al., 2015). More details are provided in appendix B of Spaeth and Birner (2022).

[1]This also seems consistent with the interpretation of this event to fall under the category of self-induced resonance, which requires conditions (vortex geometry etc.) to be "just right" (see discussion in Albers and Birner, 2014).

Finally, the p-SSWs frequency shows a maximum in February (bottom right panel in Fig. 4.2). While the overall seasonal distribution of SSWs seems consistent with reanalyses (e.g, Ayarzagüena et al., 2019), previous literature has documented minor inconsistencies across different leadtimes, highlighting that SSW probability in models could be slightly affected by model drifts (Lawrence et al., 2022).

Northern Annular Mode composite of p-SSWs

Baldwin and Dunkerton (2001) demonstrated that Northern Annular Mode (NAM) indices effectively trace circulation anomalies associated with both weak and strong polar events from the stratosphere to the troposphere (see Fig. 2.3), based on reanalysis data. Following the same approach, Figure 4.3 presents a corresponding NAM composite for 2264 p-SSW and 2129 predicted strong polar vortex (p-SPV) events, based on S2S forecast data[2].

Note that Baldwin and Dunkerton (2001) used an empirical orthogonal function (EOF) analysis to compute the NAM indices, whereas the analysis shown here uses negative standardized anomalies of polar cap geopotential height, based on an area-weighted average of anomalies poleward of 60 °N. In addition, Baldwin and Dunkerton (2001) used polar vortex event definitions based on NAM thresholds at 10 hPa, whereas the analysis shown here uses the p-SSW and p-SPV definitions described above, based on thresholds for U_{60}^{10} . Given the finite lead time of the S2S forecasts, the maximum lag in this analysis is restricted to ±36 days relative to the events' central dates.

The NAM composite based on predicted events in the forecasts is qualitatively very similar to the original version in Baldwin and Dunkerton (2001). However, the evolution of the anomalies is substantially smoother due to the larger sample sizes. Additionally, tropospheric anomalies appear somewhat weaker (around 0.4 standard deviations) compared to the version in Baldwin and Dunkerton (2001) (approximately between 0.2 and 1 standard deviations).

In addition to the established downward impact following the stratospheric events, oppositely signed anomalies are observed a few weeks prior to the events. Similar precursor signals are present in the original study by Baldwin and Dunkerton (2001), but the small sample size left uncertainty about their statistical robustness.

[2]The NAM composite is based on realtime forecasts and hindcasts from the winters of 2021/22 and 2022/23 (November 16 to February 22, respectively). While other analyses in this thesis use earlier winter periods, geopotential height data for those analyses were only downloaded at the 1000 hPa level. For recent analyses, multiple vertical levels were downloaded, but only for the latest winters. The NAM composite presented here uses this new data, however, the results are assumed to be insensitive to the specific realtime initialization dates.

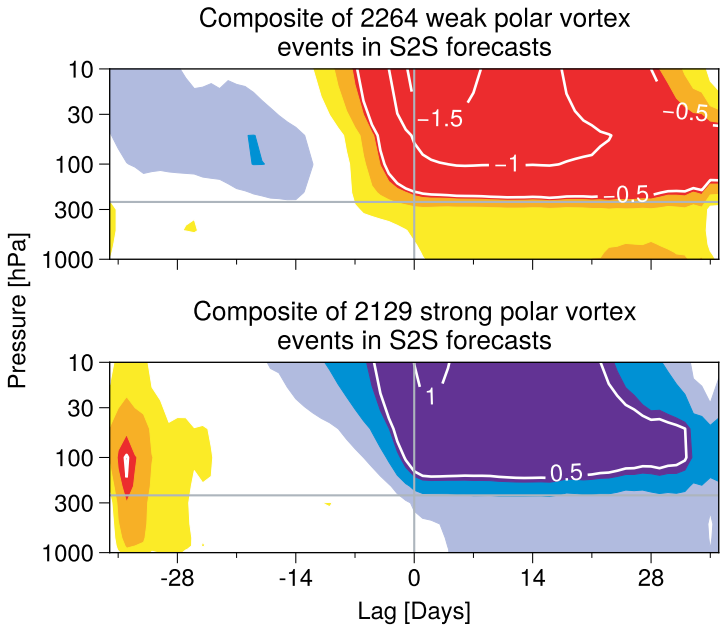

Figure 4.3: Composite-mean of NAM indices for predicted polar vortex events in S2S forecasts. **a.** Weak polar vortex events, based on U_{60}^{10} reversals (from westerly to easterly). **b.** Strong polar vortex events, based on U_{60}^{10} exceeding $50\,\mathrm{m\,s^{-1}}$. The NAM indices are computed as negative standardized anomalies of polar cap averaged (60 °N to 90 °N) geopotential height. Blue shading denotes positive NAM values, red shading denotes negative NAM values (contours intervals are -0.45, -0.3, -0.15, 0.15, 0.3, 0.45). White lines indicate additional contour levels with 0.5 spacing. Horizontal thin line shows the 250 hPa level, the approximate location of the tropopause. Visualization follows Baldwin and Dunkerton (2001), however, note that the figures presented here use finer contour intervals.

4.2 Relative probability increase and fraction of attributable risk

This section introduces the main statistical diagnostics that will be applied to polar vortex events and their coupling to the troposphere in the subsequent sections.

Facilitated by the UNSEEN approach, the large sample sizes of SSWs and strong polar vortex events enable a statistical characterization of certain aspects of stratosphere-troposphere coupling that would be difficult to quantify from observational data alone. In this chapter, a statistical framework based on conditional probabilities is applied to analyze UNSEEN events in the stratosphere and troposphere. Specifically, diagnostics are presented to quantify the modulated likelihood of an event, Y, given the presence or absence of a preceding event, X.

For example, as will be shown in section 4.4, the probability of negative Arctic Oscillation extremes increases following SSWs. This scenario is somewhat akin to the situation in climate attribution science, where one aims to quantify the increased risk of an extreme event due to anthropogenic climate change (e.g., National Academies of Sciences Engineering and Medicine, 2016, July), or to the situation in epidemiology, where one aims to quantify the increased risk of contracting a disease given an exposure to a particular factor (e.g., smoking in the case of lung cancer; Peto, 2000). In such situations, the additional risk due to exposure can be quantified using the **relative risk increase (RRI)**:

$$\mathrm{RRI} = \frac{P(Y|X)}{P(Y|\neg X)} - 1, \tag{4.1}$$

where $P(Y|X)$ is the probability of event Y given event X (the so-called "risk among the exposed"), and $P(Y|\neg X)$ is the probability of event Y given the absence of event X (the so-called "risk among the unexposed").

In climate attribution science "exposure" may be thought of as "under the influence of anthropogenic climate change", whereas lack of exposure (denominator in the RRI equation) may be thought of as "without the influence of climate change" (e.g., based on pre-industrial control climate). In the case of stratosphere-troposphere coupling, exposure may be thought of as "given that a stratospheric extreme occurred", although the assumption that the stratosphere acts as an external influence may be more severe, due to two-way couplings, compared to the influence of greenhouse gasses in climate simulations.

In general, the lack of exposure has to be evaluated with care. For instance, if "no X" is observed on day t_0, and an event Y occurs later on day t_Y, Y cannot simply be categorized as "unexposed" based on the absence of X at t_0, because an X event may have occurred between t_0 and t_Y.

To properly evaluate $P(Y|\neg X)$, a time window needs to be defined to check whether X occurred at any time within this period preceding Y. Therefore, a subscript wt denotes that events Y are preceded by X within a specific time window: $P(Y \mid X_{wt})$.

Computing the denominator of equation 4.1, $P(Y|\neg X)$, can be challenging in practice because it requires a composite of $\neg X$ "events", which is less straight-forward compared

to a composite of Y. To avoid this issue, the probability $P(Y|X)$ in equation 4.1 can alternatively be normalized by the climatological occurrence of Y, $P(Y)$, which includes both exposed and unexposed cases. To avoid confusion, this modified RRI will be referred to as **relative probability increase (RPI)**:

$$\mathrm{RPI} = \frac{P(Y_{wt}|X)}{P(Y_{wt})} - 1. \tag{4.2}$$

Here, $P(Y_{wt}|X)$ represents the probability to observe Y within time t following X. A negative RPI indicates that Y becomes less likely following X. Conversely, a larger positive RPI suggests that Y becomes more likely following X, indicating that X serves as a better predictor for Y.

While RRI and RPI quantify the increased likelihood of Y based on the occurrence of X, the fraction of attributable risk (FAR) provides additional insights into their relationship. FAR quantifies the fraction of events Y that are attributable to preceding events X. It is bounded by -1 and 1, where negative values indicate that X reduces the probability of Y.

FAR depends on $P(Y|\neg X_{\mathrm{wt}})$, which can be challenging to evaluate directly. To address this issue, the conditioning can be swapped: instead of $P(Y|\neg X_{\mathrm{wt}})$, we evaluate $P(\neg X_{\mathrm{wt}}|Y)$, which can be derived from a composite-analysis of Y (and similar for $P(X_{\mathrm{wt}}|Y)$). Using Bayes' theorem, the conditional probabilities $P(\neg X_{\mathrm{wt}}|Y)$ and $P(X_{\mathrm{wt}}|Y)$ can be transformed into $P(Y|\neg X_{\mathrm{wt}})$ and $P(Y|X_{\mathrm{wt}})$, respectively. These probabilities allow calculation of FAR.

Importantly, we distinguish two types of FAR. First, the **fraction of attributable risk among the exposed ($\mathrm{FAR_e}$)** is defined as:

$$\mathrm{FAR_e} = \frac{P(Y|X_{\mathrm{wt}}) - P(Y|\neg X_{\mathrm{wt}})}{P(Y|X_{\mathrm{wt}})} = 1 - \frac{P(X_{\mathrm{wt}})}{P(\neg X_{\mathrm{wt}})}\frac{P(\neg X_{\mathrm{wt}}|Y)}{P(X_{\mathrm{wt}}|Y)}, \tag{4.3}$$

which quantifies the fraction of events Y *preceded by* X that are attributable to X. Second, the **fraction of attributable risk among the population ($\mathrm{FAR_p}$)** is defined as:

$$\mathrm{FAR_p} = \frac{P(Y) - P(Y|\neg X_{\mathrm{wt}})}{P(Y)} = 1 - \frac{P(\neg X_{\mathrm{wt}}|Y)}{P(\neg X_{\mathrm{wt}})}, \tag{4.4}$$

which quantifies the fraction of all events Y (including exposed and unexposed) that are attributable to preceding X. RRI, $\mathrm{FAR_e}$ and $\mathrm{FAR_p}$ provide complementary perspectives on the statistical relation between X and Y.

Section 4.3 addresses the attribution of SSWs (Y) to preceding eddy heat flux events (X). Section 4.4 addresses the attribution of AO extremes (Y) to preceding SSWs (X).

4.3 Attribution of sudden stratospheric warmings to tropospheric planetary wave forcing

SSWs are predominantly driven by dissipating planetary Rossby waves (see section 2.2.2). Early studies of SSWs emphasized that this wave forcing typically originates in the tropo-

sphere, where wave activity is generally much larger than in the stratosphere (Matsuno, 1971; Holton and Mass, 1976).

Phenomena linked to increased wave activity may thus serve as predictors for weak polar vortex states across various timescales, from medium-range, to subseasonal and seasonal forecasts. Notable examples of these phenomena include high pressure anomalies over Siberia and the Ural Mountains (Martius et al., 2009; Karpechko et al., 2018), reduced sea ice over the Barents and Kara Seas (Kim et al., 2014), increased snow cover over Eurasia (Cohen et al., 2014; Kretschmer et al., 2016), and El Niño events (Domeisen et al., 2019).

However, while it is clear that planetary waves of tropospheric origin play a crucial role in the formation of SSWs, there is ongoing debate about the importance of anomalously large tropospheric wave activity. This is because climatological wave activity may already be sufficiently large to drive SSWs in the Northern Hemisphere. Following this line of reasoning, it has been suggested that the state of the stratospheric mean flow stratospheric circulation itself may play an equally important role in determining whether tropospheric wave pulses can propagate into the stratosphere and disrupt the polar vortex (Plumb, 1981; Palmer, 1981; McIntyre, 1982; Albers and Birner, 2014; Birner and Albers, 2017; de la Cámara et al., 2017; de la Cámara et al., 2019).

Quantitatively addressing this question remains challenging because wave activity is highly variable and SSWs are relatively rare events. Consequently, robust analyses require a much larger sample size of SSWs than what is available from observations. The UNSEEN framework presents a promising approach to revisit this question regarding SSW formation and associated tropospheric wave activity anomalies, using several thousand p-SSWs within the framework of the ECMWF S2S model.

Sirbescu (2023) recently showed that on average p-SSWs, identified in the ECMWF S2S forecasts, are preceded by near-surface geopotential height anomalies that enhance the climatological planetary waves, leading to anomalous vertical wave propagation into the stratosphere. Here, this work is extended by statistically attributing these p-SSWs to preceding eddy heat flux events at different vertical levels, using the fraction of attributable risk framework.

Moreover, Bauer (2022) showed that p-SSWs, identified in the ECMWF S2S forecasts, are preceded by increased frequency of high pressure systems over Eurasia, called Ural blockings. Consistent with this precursor signal, Ural blockings increase the likelihood for subsequent weak polar vortex events. Here, these findings are combined with those of Sirbescu (2023), by linking Ural blockings to increased planetary wave activity, which manifests in positive eddy heat flux anomalies reaching from the troposphere into the stratosphere.

Specifically, this section addresses the following research questions:

1. How can SSWs be statistically attributed to preceding anomalies in tropospheric wave activity?

2. Ural blockings in the troposphere have been suggested as precursors for SSWs. How

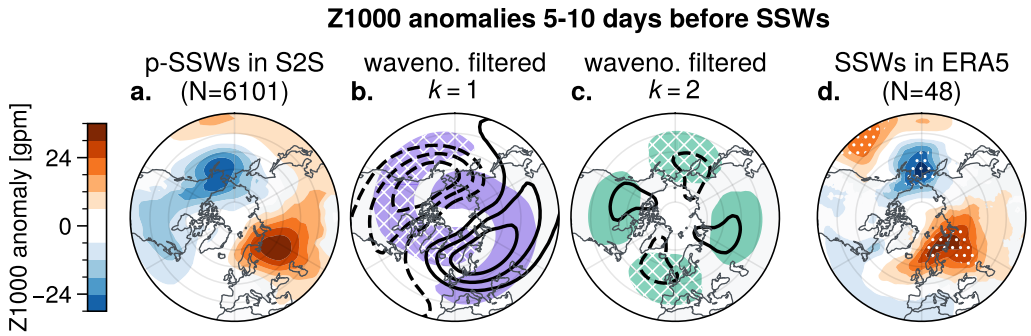

Figure 4.4: 1000 hPa geopotential height (Z1000) anomalies averaged over 5-10 days prior to SSWs. **a.** UNSEEN SSW (p-SSWs) composite-mean based on 6101 events. **b.** As panel a, but truncated to zonal wavenumber 1. Black contour lines denote the wave-1 component of the anomaly field (contours each 3 gpm; zero line omitted). Shading denotes the wave-1 component of the model winter climatology (values below −12 gpm shaded and hatched, values above 12 gpm shaded). **c.** As panel b, but for zonal wavenumber 2. **d.** As panel a, but for 48 SSWs in ERA5, with statistically significant anomalies indicated by white stippling ($p < 0.05$ based on a two-tailed one-sample Student's t-test).

are Ural blockings linked to anomalies in planetary wave activity? Additionally, how is the probability of SSWs modulated by Ural blockings?

Near-surface geopotential height precursors to SSWs

To investigate mean precursor signals of the tropospheric circulation prior to SSWs, Fig. 4.4a presents 1000 hPa geopotential height (Z1000) anomalies for a composite of 6101 p-SSWs, averaged over the period from −10 d to −5 d relative to the SSW date. Consistent with previous studies (Garfinkel et al., 2010; Kretschmer et al., 2017; Köhler et al., 2023; Sirbescu, 2023), positive Z1000 anomalies are observed over Eurasia, while negative anomalies are observed over the Pacific and North America. These signatures also agree with those based on ERA5 data (Fig. 4.4d).

The fact that only planetary waves, typically of zonal wavenumbers 1 and 2, can propagate into the stratosphere (Charney and Drazin, 1961) motivates a spectral decomposition of the observed Z1000 signal into zonal wavenumbers. The results show that the wave-1 and wave-2 components of the total anomaly field project onto the climatological wave-1 and wave-2 patterns, thereby amplifying these waves (Figs. 4.4b,c). In regions where wave-1 and wave-2 anomalies destructively interfere, such as over the Atlantic and East Asia, the full Z1000 anomaly field is close to zero. The largest positive Z1000 anomalies are observed over Eurasia, where the wave-1 and wave-2 contributions interfere constructively. Similar findings were reported by Martius et al. (2009) (based on reanalysis data), Garfinkel et al. (2010) (based climate model simulations) and Sirbescu (2023) (based on S2S forecasts).

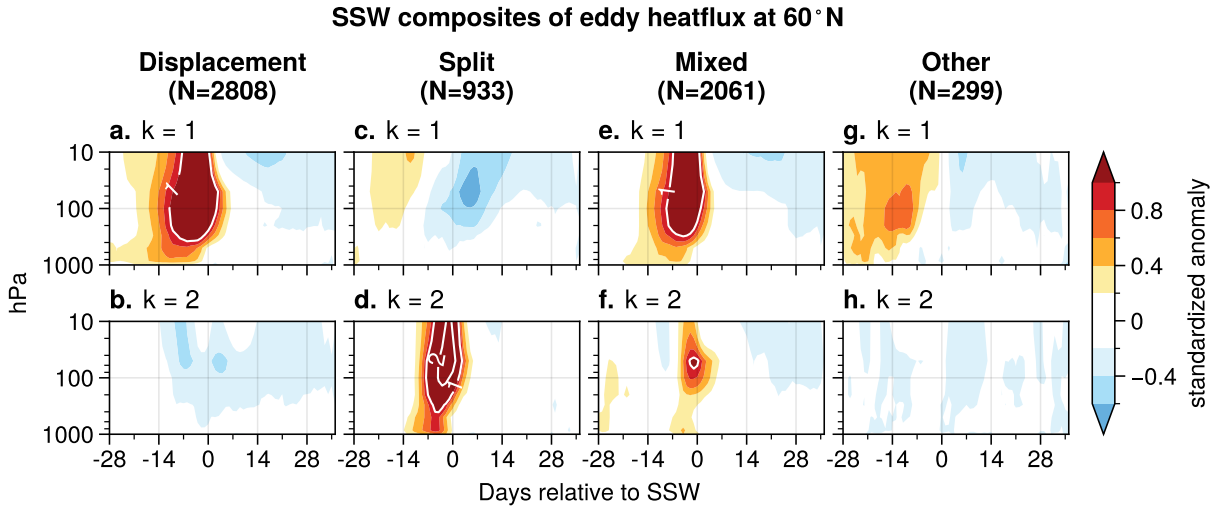

Figure 4.5: Composite-mean of 60 °N eddy heat flux standardized anomalies, for **a., b.** displacement **c.,d.** split, **e.,f.** mixed and **g.,h.** remaining events, for zonal wavenumber $k = 1$ (top row) and zonal wavenumber $k = 2$ (bottom row). White lines provide additional contour levels every 1σ.

Composite-mean anomalies of eddy heat flux during SSWs

The strengthened planetary waves 1 and 2 prior to SSWs are expected to coincide with anomalous vertical wave propagation. As a proxy for vertical wave propagation, the eddy heat flux ($\overline{v't'}$) is used, where v and t are the meridional wind and temperature, $(\cdot)'$ denotes deviations from the zonal-mean and $\overline{(\cdot)}$ denotes a zonal average. The eddy heat flux is proportional to the vertical component of wave activity flux (Vallis, 2017, June, chapter 10: "Waves, Mean-Flows, and their Interaction"). The vertical propagation is assessed for zonal wave-1 ($v't' \,|_{k=1}$) and wave-2 ($v't' \,|_{k=2}$) separately by first applying a Fourier transform to v' and t', and then computing the covariance. In the analyses, the eddy heat flux is displayed as standardized anomalies: the inter-annual mean is subtracted, and the result is divided by the inter-annual standard deviation.

Before moving on to an attribution of SSWs to preceding heat flux anomalies, Fig. 4.5 presents the time-lag composite-mean of standardized eddy heat flux anomalies at 60 °N during SSWs, similar to Sirbescu (2023). The analysis is shown separately for displacement- and split-type SSWs, as substantial differences between $v't' \,|_{k=1}$ and $v't' \,|_{k=2}$ in the troposphere are expected. Displacement and split SSWs are here defined similarly to de la Cámara et al. (2019), based on standardized eddy heat flux anomalies at 50 hPa. Displacement SSWs are defined as SSWs where $v't' \,|_{k=1} > 1$ standard deviation (σ) at any day within ±5 d relative to the SSW and $v't' \,|_{k=2} < 1\sigma$ throughout this period. Split SSWs are defined accordingly, where $v't' \,|_{k=1} < 1\sigma$ and $v't' \,|_{k=2} > 1\sigma$. Out of the 6101 SSWs, 2808 classify as (pure) displacement, 933 as (pure) split and 2061 as mixed events (where both wavenumbers 1 and 2 exceed the thresholds).

Displacement SSWs are, on average, preceded by positive $v't' \,|_{k=1}$ anomalies for about three weeks (up to 0.4σ near the surface and up to 1.5σ in the stratosphere). Similarly, split SSWs are preceded by positive $v't' \,|_{k=2}$ anomalies for about ten days (nearly 1σ near the surface and exceeding 2σ in the stratosphere). The larger magnitude of wave-2 anomalies compared to wave-1 could result from the fact that the climatological wave-2 is weaker than wave-1 in the stratosphere. Therefore, wave-2 forcing needs to be larger in an anomaly sense to induce a similar level of mean-flow deceleration compared to wave-1.

Moreover, wave-2 anomalies appear more instantaneous, in agreement with de la Cámara et al. (2019), whereas wave-1 anomalies seem to build up more continuously over a longer period. Especially for wave-2, there is only short time-lag between tropospheric and stratospheric anomalies, a potential hint for resonance behavior (e.g., Plumb, 1981; Matthewman and Esler, 2011; Albers and Birner, 2014).

The composite of mixed SSWs, exhibiting anomalous $v't' \,|_{k=1}$ and $v't' \,|_{k=2}$, shares characteristics of both displacement and split events, as expected (Fig. 4.5). Finally, the remaining SSWs that do not qualify for any of categories show only weak anomalies in general. However, there are positive $v't' \,|_{k=1}$ anomalies at all levels more than one week before the events. This suggests that even though these events did not exhibit anomalous $v't' \,|_{k=1}$ between lags of -5 to $+5$ days relative to the SSW, the polar vortex already weakened before this periods.

All events are followed by reduced $\overline{v't'}$ in the stratosphere. This is presumably related to zonal-mean easterly winds in the stratosphere following SSWs, which suppress upward wave propagation according to quasi-geostrophic theory (Charney and Drazin, 1961).

Towards attribution of SSWs to preceding tropospheric heat flux events

The composite-mean analysis shows that prior to SSWs, vertical propagation of planetary waves from the troposphere into the stratosphere is enhanced. However, this average behavior cannot determine whether anomalous vertical wave propagation is observed ahead of all SSWs. If we instead considered the total number of SSWs preceded by anomalous heat flux in the troposphere, this count would include cases where a tropospheric wave pulse and a subsequent SSW co-occur fortuitously, without any causal link. The fraction of attributable risk among the population ($\mathrm{FAR_p}$) provides a framework to account for such cases and offers an estimate of how many SSWs are attributable to preceding heat flux events. Here, SSWs are defined to be preceded by a heat flux event if $v't' \,|_{k=1}$ or $v't' \,|_{k=2}$ exceed 1σ on any day within 0 d to 10 d before the SSW. Applying equation 4.4 to this example, $\mathrm{FAR_p}$ becomes

$$\mathrm{FAR_p} = 1 - \frac{P(\text{a given SSW is not preceded by a preceding heat flux event})}{P(\text{no heat flux event})}. \qquad (4.5)$$

Fig. 4.6 presents the $\mathrm{FAR_p}$ of SSWs to preceding anomalous eddy heat flux. To isolate their distinct characteristics, the focus is here on pure displacement events (and their attribution to wave-1 heat flux; $\mathrm{FAR_p} \,|_{k=1}$) and pure split events (and their attribution

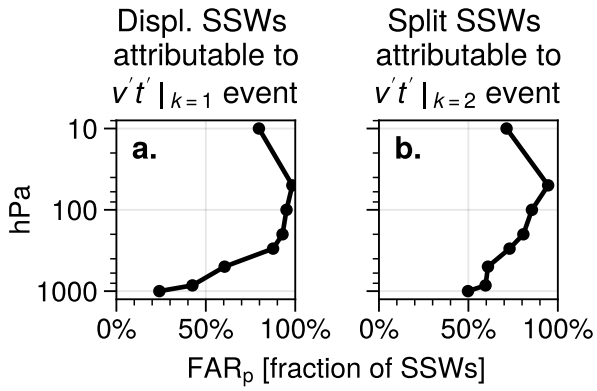

Figure 4.6: Fraction of attributable risk analysis. **a.** Fraction of displacement SSWs attributable to preceding events of anomalous wave-1 eddy heat flux at 60 °N ($v't' \mid_{k=1}$), on different vertical levels. **b.** As panel a, for split SSWs attributable to preceding wave-2 eddy heat flux ($v't' \mid_{k=2}$).

to wave-2 heat flux; $\mathrm{FAR_p} \mid_{k=2}$). Mixed events are more often attributable to wave-1 heat flux events than to wave-2 heatflux events (not shown), suggesting that mixed events may be more similar to displacement events that are then accompanied by higher wavenumbers during the polar vortex breakdown.

Both $\mathrm{FAR_p} \mid_{k=1}$ and $\mathrm{FAR_p} \mid_{k=2}$ are close to 1 near the 50 hPa level, because displacement and split SSWs are here defined based on anomalous heat flux at this level. In the troposphere, the values decrease to about 1/3 for displacement and about 1/2 for split SSWs, implying that about every third displacement SSW is attributable to preceding wave-1 eddy heat flux events in the troposphere, whereas about every second split SSW is attributable to preceding wave-2 eddy heat flux events in the troposphere.

These numbers are of a similar order of magnitude as those reported by de la Cámara et al. (2019), although their study focuses on the frequency of tropospheric wave events prior to weak vortex events, without accounting for fortuitously co-occurring events. They found approximately 40% of displacement-type weak polar vortex events to be preceded by tropospheric wave 1 events, and approximately 30% of split-type weak polar vortex events to be preceded by tropospheric wave 2 events. In contrast, this study, based on S2S forecast data, suggests a stronger link between split-type events and tropospheric wave-2 activity. This discrepancy may arise from different event definitions (de la Cámara et al., 2019, used deceleration of U_{60}^{10} to define weak polar vortex event), thresholds (de la Cámara et al., 2019, used 2σ compared to 1σ to define wave events), or attribution methods (whether or not accounting for climatological SSW and wave events occurrence).

Modulation of SSW probability by Ural blockings

Fig. 4.4 showed that on average, SSWs are preceded by anticyclonic anomalies over central Eurasia. On synoptic scales, such anomalies can manifest as Ural blockings. Ural blockings may thus take an important role in increasing planetary wave activity in the troposphere, thereby increasing the likelihood for SSWs. In the following, the role of Ural blockings is analyzed with respect to their influence on planetary wave activity and associated SSW probability. As before, events are identified in the forecasts, following the UNSEEN approach.

Ural blockings are identified in the S2S forecasts following Bauer (2022), defined as days where the Ural-Z1000 index exceeds 2σ. This index is defined as the standardized anomaly of the area-weighted spatial average of Z1000 over 45° to 72.5°E and 50° to 70°N. Only events occurring after a lead time of 9 days are considered to ensure a certain degree of statistical independence. In Fig. 4.7a, the Z1000 composite-mean anomalies on the central day of Ural blockings are presented. Ural blockings are accompanied by increased eddy heat flux, lasting about 3 weeks for wavenumber 1 and approximately 1 week for wavenumber 2 (Figs. 4.7b,c). The anomalous heat flux corresponds to increased vertical wave propagation into the stratosphere. Consistently, the daily probability for a reversal of the zonal-mean zonal wind at 60°N, 10 hPa increases from about 0.8% before Ural blockings (corresponding to a climatological probability) to about 1.8% following the events (Fig. 4.7d). Ural blockings may thus provide predictability for weak polar vortex events, as proposed by Karpechko et al. (2017).

Discussion

The composite-mean analyses of SSWs with respect to their tropospheric Z1000 precursor signals are in good agreement with existing literature (e.g., Garfinkel et al., 2010; Kretschmer et al., 2017; Domeisen et al., 2020a; Köhler et al., 2023). This provides further confidence that the S2S model can realistically represent relevant aspects of upward stratosphere-troposphere coupling. This motivated using the large sample of UNSEEN SSWs also to compute more sophisticated statistical diagnostics, which cannot be derived from reanalysis datasets. Specifically, the approach was applied to investigate two research questions:

1. *How can SSWs be statistically attributed to preceding anomalies in tropospheric wave activity?*

   The fraction of attributable risk among the population (FAR_p) was computed to quantify how many SSWs may be attributable to preceding anomalous wave activity in the troposphere. This was done separately for displacement and split SSWs, following a simple classification based on eddy heat flux anomalies at 50 hPa.

   Due to the limited vertical resolution of the S2S model output, vertical wave propagation was quantified using the eddy heat flux at 60 °N as a proxy, which is proportional

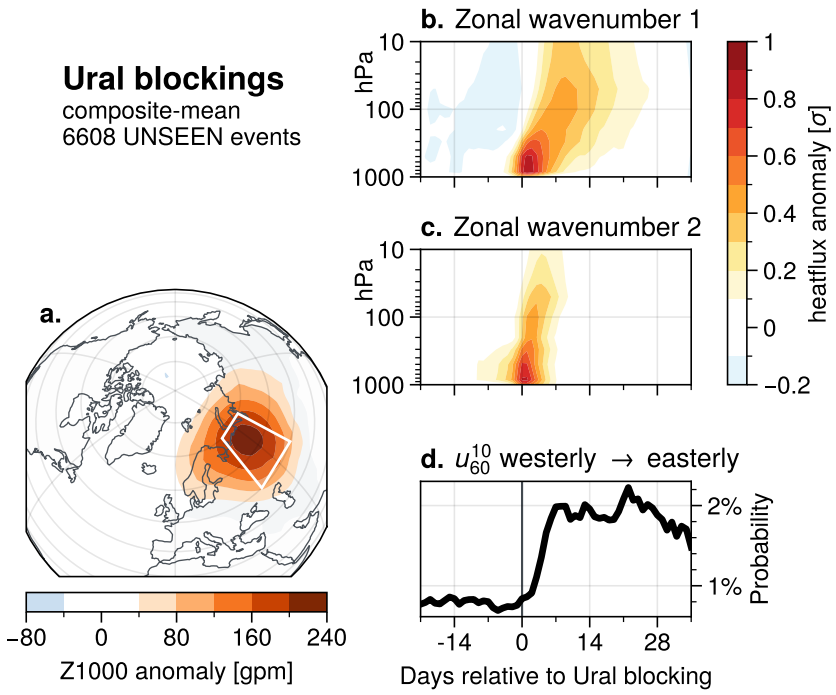

Figure 4.7: Composite analysis of 6608 UNSEEN Ural blockings in the S2S forecasts. **a.** Composite-mean 1000 hPa geopotential height anomalies. White box indicates the region based on which the events are defined. **b.** Composite-mean eddy heat flux standardized anomalies at 60 °N for zonal wavenumber 1. **c.** As panel b, but for zonal wavenumber 2. **d.** Daily probability for U_{60}^{10} to transition from westerly to easterly.

to the vertical component of the quasi-geostrophic Eliassan-Palm (EP) flux. Given the substantial data size of the S2S forecasts, the analyses are here restricted to this single latitude (representing over 100 GB of data). While this single latitude is expected to reasonably capture the average characteristics of upward coupling dynamics, some aspects of the tropospheric response to SSWs may be missed, as the tropospheric jets tend to shift equatorward following the event.

Consistent with de la Cámara et al. (2019), a significant fraction of SSWs was found to be *not* statistically attributable to preceding anomalous eddy heat flux (exceeding 1σ) in the troposphere, with 2/3 of displacement and 1/2 of split SSWs falling into this category. These findings underscore the importance of the stratosphere's "susceptibility" to upward propagating tropospheric waves.

In particular, potential vorticity gradients are expected to play a critical role in determining wave propagation conditions. These gradients are not only influenced by planetary waves but also by smaller-scale gravity waves (Albers and Birner, 2014; Polichtchouk et al., 2023; Wicker et al., 2023). Consequently, these stratospheric preconditioning aspects might be more challenging to predict by forecast models than the prediction of planetary wave activity in the troposphere. If this hypothesis holds true, then SSW predictability might be more strongly governed by the accurate prediction of features and gradients in the stratosphere than by the prediction of tropospheric dynamics.

2. *Ural blockings in the troposphere have been suggested as precursors for SSWs. How are Ural blockings linked to anomalies in planetary waves? And how is the probability for SSWs modulated in the aftermath of Ural blockings?*

   A spectral decomposition of the Z1000 anomaly field prior to SSWs, similar to Martius et al. (2009) and Garfinkel et al. (2010), revealed that there is only one region where increased wave-1 and wave-2 amplitudes constructively interfere locally – near the Ural mountains, over Eurasia. Consistently, a composite analysis of 6608 UNSEEN Ural blockings revealed that these events are associated with positive wave-1 and wave-2 eddy heat flux anomalies extending from the troposphere into the stratosphere. The increased wave activity leads to a weakening of the polar vortex, approximately doubling the probability for a zonal-mean zonal wind reversal at 10 hPa, 60 °N from westerly to easterly (consistent with Bauer, 2022).

   If conditioning on wave events in the lower stratosphere rather than tropospheric events, a previous study by Jucker and Reichler (2018) (based on climate model simulations) found that the probability of sudden stratospheric warmings (SSWs) increases by a factor of about 5 following such wave events. This highlights that not all Ural blockings are associated with wave anomalies that extend into the stratosphere.

Finally, one question that arose during the analyses concerns the role of the vertical tilt of planetary wave-1, during SSWs. Prior to SSWs, the wave-1 anomaly exhibits a maximum over Eurasia, amplifying the climatological wave-1 (Fig. 4.4b). However, Sirbescu (2023)

demonstrated that wave-1 anomalies undergo a longitudinal phase shift of approximately 90°, from few days before to few days after the SSW. After SSWs, the phase of the wave-1 anomaly then aligns with the negative NAO pattern. The relationship between the tropospheric and stratospheric phases of wave-1 remains unclear: the phase change in the troposphere could correspond to a phase shift in the stratosphere or, alternatively, arise from a change in the vertical tilt of the wave. Analyzing the vertical tilt of wave-1 during SSWs could potentially offer further insights into the downward response to these events.

4.4 Attribution of Arctic Oscillation extremes to polar vortex variability

In this section, we continue to analyze UNSEEN SSWs (p-SSWs), with the focus shifting from upward-coupling to downward coupling. Specifically, we investigate the impact of weak and strong polar vortex states on subsequent large-scale circulation extreme events in the troposphere – quantified via the Arctic Oscillation index. This work is published in Spaeth and Birner (2022).

Motivation

Numerous studies show that extreme polar vortex events tend to be followed by tropospheric anomalies that project onto the Arctic Oscillation (AO), the leading mode of day-to-day variability in the northern extratropics (Thompson and Wallace, 1998; Baldwin and Dunkerton, 1999; Limpasuvan and Hartmann, 2000). Specifically, strong polar vortex (SPV) episodes are associated with a subsequent positive AO index, whereas weak polar vortex states, including SSWs, are associated with a subsequent negative AO index (e.g., Baldwin and Dunkerton, 2001; Lawrence et al., 2020; Rupp et al., 2022). Consistent with the local implications of a negative AO index, SSWs can for example lead to cold spells in Northern Europe and increased storminess over Southern Europe (Domeisen and Butler, 2020, and references herein).

Whether it is generally valid that SSWs and also SPV episodes lead to a subsequently more likely occurrence of AO extremes (and associated local extremes) is difficult to analyze because the statistical links are weak in each case, i.e., not each SSW/SPV event is followed by an AO extreme. Therefore, very large samples of SSW and SPV events are needed to quantify the subsequent risk increase of AO extremes. However, reanalyses data only cover about 40-70 years, depending on the data set, and thus about 30-40 SSWs – too few to robustly determine conditional probabilities (e.g., given a stratospheric extreme event, how likely is a following tropospheric extreme event).

Hence, the UNSEEN approach offers a powerful framework to address this issue, since it can provide a large sample of events. In the following analysis, we aim to improve the characterization of coupled stratospheric and tropospheric circulation extremes based on UNSEEN SSWs (or p-SSWs), which are predicted in the S2S ensemble forecasts.

The results of this section are published in 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.

In addition to ECMWF forecasts, the results are presented for UKMO forecasts, which may guide the interpretation of model uncertainty. Importantly, both these models have been demonstrated in previous studies to have a realistic representation of stratosphere-troposphere coupling (Domeisen et al., 2020c; Domeisen et al., 2020a).

We will compute statistical measures that combine conditional and base rate probabilities for stratospheric and AO extreme (co-)occurrences, in order to address our following research questions:

1. By how much is the probability of persistently positive or negative AO phases increased following stratospheric polar vortex extremes?
2. By how much is the probability of subsequent AO extremes increased following stratospheric polar vortex extremes?
3. What fraction of AO extremes may be attributable to preceding stratospheric polar vortex extremes?

To illustrate which AO extremes are classified as "attributable", consider the following scenarios where a stratospheric event is followed by an AO extreme: In relation to the AO extreme the stratospheric extreme may
(a) represent a necessary and sufficient cause
(b) represent one among multiple contributory causes
(c) be caused by a confounding factor, which also causes the AO extreme
(d) not be causal
In scenario a, the AO extreme is attributable to the preceding stratospheric event, whereas it is not attributable in scenario d. In scenario b, disentanglement of different contributory factors is difficult. Each involved process can but does not need to be also a necessary cause. (Consider for example a situation where an AO extreme would have occurred also without a preceding stratospheric extreme, but the stratospheric extreme resulted in a stronger or earlier manifestation.) In this study, we do neither aim to disentangle the multiple involved pathways a-c nor to provide a rigorous quantification of causality (which is itself ambiguous in a complex system). Instead, we estimate how many AO extremes may be attributable to the stratospheric extreme, which refers to the fraction that would have statistically not occurred without the stratospheric event. Importantly, scenario c shows that "without the stratospheric event" requires to also remove any confounding factors. The analysis follows an observational approach (which is based on post-hoc computation of conditional probabilities) rather than a counterfactual approach (which is based on active interventions in the system; Pearl, 2009, September, see section 4.4.6 for a more detailed interpretation of the results with respect to causality). However, even without disentangling scenarios a, b and c, the observational approach provides relevant and practical insights into the statistical association between and the importance of stratospheric and subsequent AO extremes.

The study begins with a composite-mean analysis, validating that the mean AO response to SSWs and SPVs agrees between S2S models and ERA5 reanalysis (section 4.4.1). This

The results of this section are published in 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.

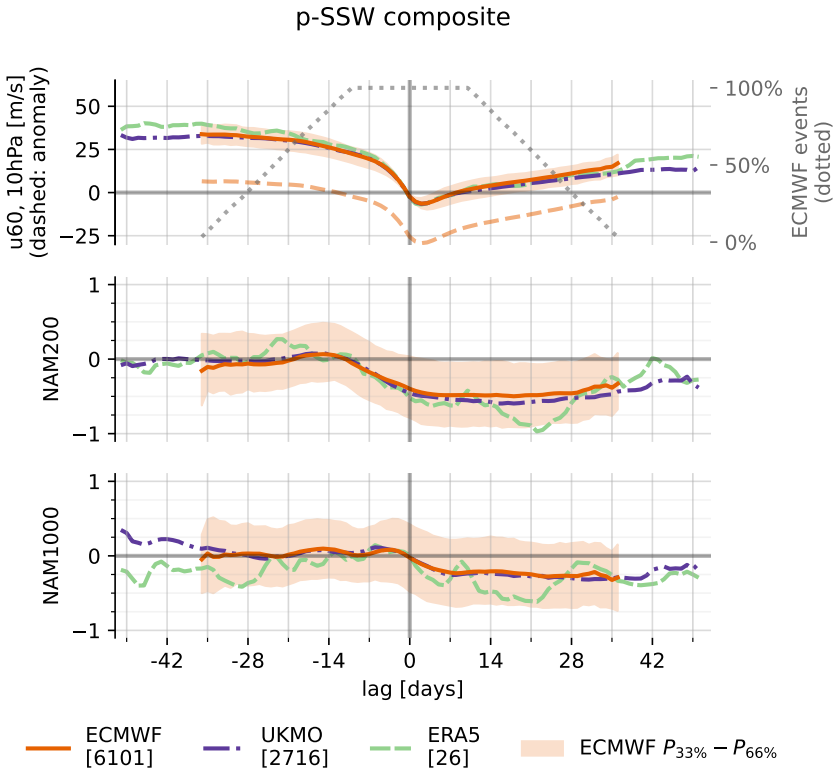

Figure 4.8: Lagged composite evolution of U_{60}^{10} (top panel), NAM200 (middle panel) and NAM1000 (=AO, bottom panel) relative to p-SSWs (ECMWF, UKMO) and SSWs (ERA5). It is presented the mean across all ECMWF events (orange, solid), the 33rd to 66th percentiles across all ECMWF events (orange, shaded), the mean of all UKMO events (purple, dash-dotted) and the mean across all ERA5 events (green, dashed). In the top panel further denoted are the average U_{60}^{10} anomalies (orange, dashed) and the relative number of contributing events to the composite in the ECMWF model (gray, dotted). Square brackets denote the total number of events, for each dataset.

motivates section 4.4.2, where the persistence of negative AO phases is investigated, and section 4.4.3, where the probability of AO extremes following predicted SSWs is analyzed. Conversely, section 4.4.4 shows how often predicted AO extremes are preceded by predicted SSWs and how many AO extremes may be attributable to preceding SSWs. Section 4.4.5 reveals in a similar fashion the statistical relation between predicted strong polar vortex events and predicted positive AO extremes, before the key findings are discussed and summarized in section 4.4.6.

The results of this section are published in 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.

4.4.1 Evaluation of stratosphere-troposphere coupling based on predicted SSWs

To provide a baseline for our more detailed statistical analyses in the following sections, we first evaluate the general behavior of stratosphere-troposphere coupling based on p-SSW events in the S2S data. To do so we focus on the lag-composite evolution of the AO index relative to p-SSWs compared to real-atmospheric SSWs from ERA5. In addition, we show the NAM index at 200hPa (short: NAM200) because the lower stratosphere has been found to play an important role in stratosphere-troposphere coupling (e.g, Karpechko et al., 2017; White et al., 2020).

Figure 4.8 shows the evolution of U_{60}^{10} (top), NAM200 (center) and AO (bottom, equal to the NAM1000) during SSWs, averaged over all events, separately for ECMWF and UKMO. In addition to the composite mean, the 33rd to 66th percentiles across all ECMWF events on the respective lag day are shown. By construction, 100% of all events (ECMWF: 6 101, UKMO: 2 716) contribute to lags within ±10 days relative to the SSW. For larger positive or negative lags, some forecasts have reached their maximum forecast lead time or have not yet been initialized. Therefore, the number of events drops off, which makes the statistics less robust: For the ECMWF model, the number of contributing events falls below 20% for lags smaller than −31 and larger than +31 days (UKMO: smaller than −44 and larger than +39 days).

By construction, U_{60}^{10} transitions from westerly to easterly at lag 0. Anomalies of U_{60}^{10} are slightly positive ahead of −14 days lag, which we interpret as an indication for vortex preconditioning (McIntyre, 1982; Albers and Birner, 2014; Jucker and Reichler, 2018). The anomalies become negative within the second week prior to the event central date. The largest average negative anomalies occur only few days after the event central day (lag +2 days: $-6\mathrm{ms}^{-1}$). Afterwards, the vortex reestablishes and the average anomalies reach zero again after approximately 35 days. Consistent with, e.g., Baldwin and Dunkerton (2001), both NAM200 and AO are negative following the event. The shift of the NAM200 happens earlier (at lag day −11) and the timing aligns well with the weakening of the polar vortex at 10hPa. The NAM200 anomaly is also more pronounced (≈ -0.5) compared to the AO (≈ -0.3). Interestingly, the AO distribution is slightly shifted toward positive values in the week prior to the central date, which is also robust for other diagnostics like the 10th, 30th, 70th and 90th percentiles (not shown). At long positive lag times, the NAM indices at 200 and 1000hPa are still negative (ECMWF: lag +36 days, UKMO: lag +51 days), but the trend goes to weaker negative values again.

Overall, the results are in agreement with ERA5 and previous literature and especially the evolution of U_{60}^{10} is remarkably similar. The negative NAM response at 200 hPa and 1000 hPa seems to be slightly stronger in the reanalysis, however, it is also noisier due to the smaller sample size.

The results of this section are published in 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.

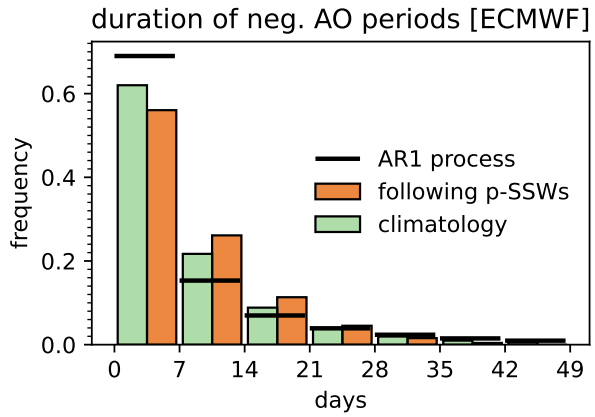

Figure 4.9: Histogram of the duration of negative AO periods, quantified by the number of consecutive days of AO < 0 and binned by 7-day chunks. Periods following ECMWF p-SSWs (orange bars, right half) are compared to the ECMWF model's climatology (green bars, left half) and a random first order auto-regressive model of the same 1-day-lag-autocorrelation as the AO in ERA5 (black, horizontal lines). ERA5 climatology is not shown, but agrees very well with the ECMWF forecast climatology.

4.4.2 Persistence of negative AO phases

Before focussing on the impact of p-SSWs on extreme *magnitudes* of the AO index, we investigate anomalies in the *persistence* of negative AO phases.

Figure 4.9 presents a histogram of the duration of predicted negative AO phases in the ECMWF model, binned into 7 day chunks. The duration is defined as the number of consecutive days with negative AO. The climatology serves as a reference including all 30 894 ECMWF forecasts used for this study. With approximately 62%, most phases of negative AO are shorter than 8 days. As another reference, a first order autoregressive model was set up with zero mean and standard deviation of 1, which may serve as a baseline. Its 1-day-autocorrelation is chosen to match the ERA5 AO time series and for robustness, it is estimated by averaging the lag-1-autocorrelation and the square-root of the lag-2-autocorrelation, yielding 0.91. ECWMF (S2S) and ERA5 agree very well in terms of climatology and lag-1-autocorrelation (not shown). However, the AO climatology shows short negative phases (≤ 7 days) less often and long negatives phases (≥ 8 days) more often compared to the AR1 process, indicating an AR1 process cannot reproduce AO variability.

In addition, the diagnostic is presented for periods following p-SSWs. Here, the AO index is analyzed between lag day +1 relative to the event date and the maximum available lag time, which ranges between +10 and +36 days, depending on the forecast lead time when the event happens. Similar to the reference climatology, this diagnostic also underestimates the occurrence of long negative AO periods as the forecasts have finite maximum lead time. Nevertheless, periods following SSWs show a reduced frequency of shorter and an increased frequency of longer negative AO periods, compared to the climatology (and thus also to

The results of this section are published in 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.

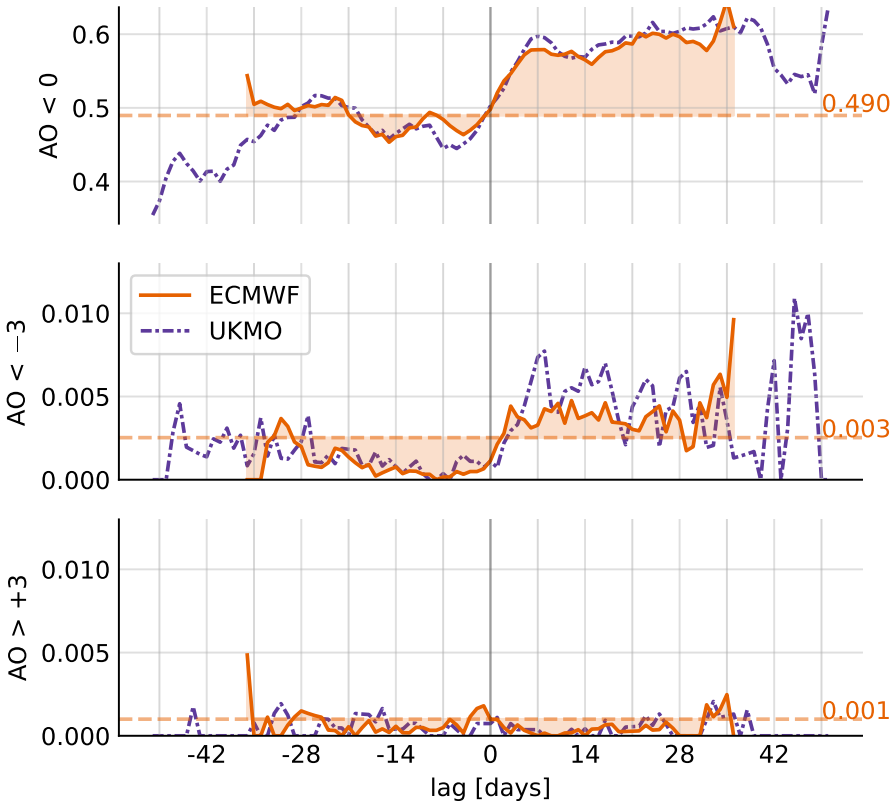

Figure 4.10: Daily probabilities of AO<0 (top panel), AO< −3 (middle panel) and AO>+3 (bottom panel) relative to p-SSWs, quantified by the fraction of events fulfilling the respective condition, separately for ECMWF (orange, solid) and UKMO (purple, dash-dotted). Day 0 corresponds to the p-SSW central date. In addition, probabilities are compared to the corresponding daily ECMWF climatology (dashed horizontal lines).

the AR1 process): For instance, 38% of negative AO periods are longer than 7 days in the climatology, whereas this probability rises to 44% following p-SSWs, which corresponds to a relative increase of 16%.

Sampling uncertainties turn out to be negligible within 95% confidence intervals. A similar analysis based on UKMO data shows very good quantitative agreement (not shown), which further confirms the robustness of the results.

4.4.3 Modulated probability of AO extremes

It is known that SSWs shift the subsequent AO distribution (see Fig. 4.8). This also implies an increased daily probability of negative and a reduced probability of positive AO extremes compared to their respective climatological probabilities. Fig. 4.10 shows the probabilities of negative (< 0), extremely negative (< -3) and extremely positive ($> +3$) AO values on a particular lag day t relative to the SSW central date. Mathematically,

The results of this section are published in 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.

these probabilities can be written as $P(AO \mid SSW)$. Per construction, lag day 0 describes the SSW central day. At each lag day, the probabilities are computed by normalizing the number of events fulfilling the respective condition with the total number of available events at the respective lag day (which decreases for large positive and negative lags).

In addition, the overall daily probabilities of AO< 0, AO< -3 and AO$> +3$ are presented, providing climatological baselines $P(AO)$, which are independent of lag time. In any forecast, AO events occur at each day with probabilities of about 49.0% for AO < 0, about 0.3% for $AO < -3$ and about 0.1% for $AO > +3$. Asymmetry between positive and negative values arises from the AO distribution that is not perfectly Gaussian (skewness: -0.13).

The fraction of events in the p-SSW composite that have negative AO values fluctuates around $P(AO^{-0} \mid SSW) = 50\%$ at negative lags with only small deviations from the climatology. Within the first week following the event, this fraction increases and appears to saturate around 60%. Consequently, in the period following a p-SSW, a negative AO is, at each day, approximately 50% more likely compared to a positive AO (60% vs. 40%). The results are consistent between ECMWF and UKMO during the ± 4 week period where the composites for both models consist of more than 30% of all events.

Extremely negative AO values in the dataset appear with a climatological probability that is similar to what would be expected for a (one-sided) 3-sigma-event of a standard normal distribution (0.27%). At negative lags, they occur overall less frequent compared to climatology. In contrast, around lag 0, the probability increases and persists at $P(AO^{-3} \mid SSW) \approx 0.40\%$ for more than four weeks. The increase appears to be larger in the UKMO model, however due to fewer events the diagnostic is also noisier. The fraction of events with extremely positive AO values is smaller compared to climatology throughout the entire lag period. This is largely consistent between the models from ECMWF and UKMO. ERA5 (not shown) overall reveals higher probabilities of negative AO values following SSWs, $P(AO^{-0} \mid SSW)$. However, large uncertainties (95%-CI $\approx [45\%; 85\%]$) in ERA5 make it difficult to distinguish whether observed differences arise from sampling errors in the reanalysis or from imperfect models. The ERA5 baseline probabilities of AO extremes are modestly lower compared to the S2S models[3] ($P^{ERA5}(AO^{-3}) = 0.06\%$; $P^{ERA5}(AO^{+3}) = 0.02\%$) and not a single AO$^{\pm 3}$ extreme event occurred within a four week period following a real atmosphere SSW, resulting in $P^{ERA5}(AO^{\pm 3} \mid SSW) = 0$, likely due to the very limited sample size.

An altered probability of extreme AO events may be of higher socio-economic relevance than a small shift of the mean. However, the absolute daily probabilities of extremely negative AO events are still small even though the relative increase given the p-SSWs is indeed considerable. In practice, the relevant question might not be how much the probability increases on only one specific day following a p-SSW. It may be more relevant

[3]Note that we have standardized the AO in ERA5 such that the inter-annual standard deviation is 1, similar to the deseasonalization that is applied to the S2S forecasts. The lower baseline probabilities are consistent with a non-zero kurtosis of the AO distribution in ERA5 of ~ -0.3 (ECMWF: ~ 0.0, UKMO: ~ 0.1).

The results of this section are published in 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.

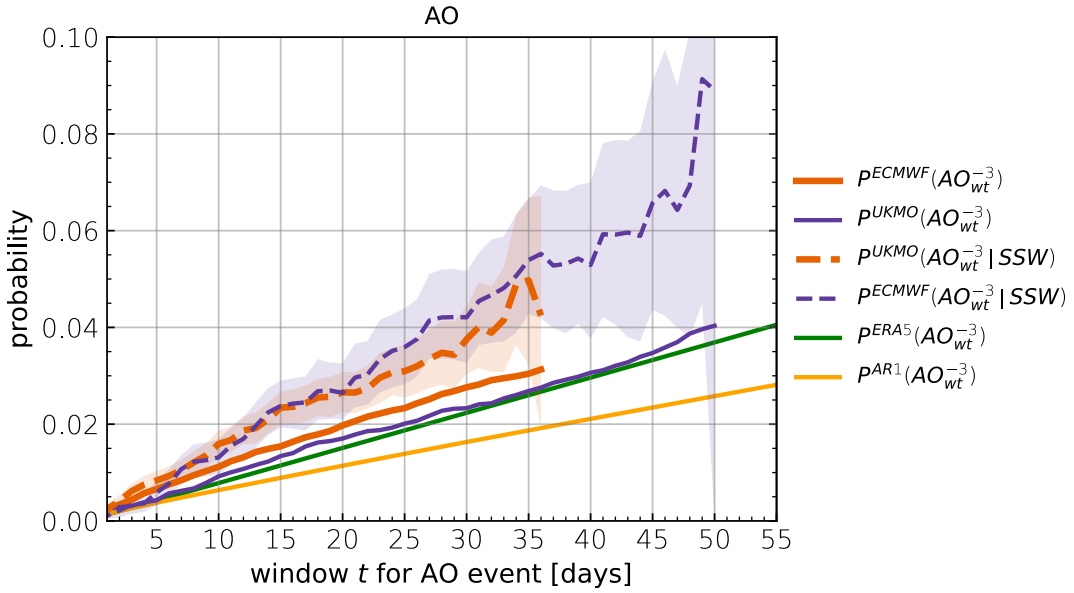

Figure 4.11: Probabilities of at least one AO^{-3} event within a window of time t following p-SSWs (dashed, mean incl. 95%-confidence-interval) are compared to climatology (solid), separately for ECMWF (orange) and UKMO (purple). In addition, the climatologies for ERA5 (green) and a random first-order auto-regressive model of the same 1-day-autocorrelation (yellow) are presented.

The results of this section are published in 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.

to quantify the increased risk for an extreme AO within a given time period following a p-SSW.

Figure 4.11 therefore shows the probability of at least one AO^{-3} extreme between day 1 and day t as a function of t. We compare the period following p-SSWs, $P(AO^{-3}_{wt} \mid SSW)$ to the respective model climatologies, $P(AO^{-3})$, the ERA5 climatology and an AR1 process of the same autocorrelation as the AO index in ERA5. Confidence intervals were obtained for $P(AO^{-3}_{wt} \mid SSW)$ by bootstrap sampling all SSW events. For ECMWF and UKMO climatology, probabilities were sampled from lead time +10 days[4] to lead time +10+t days within all forecasts. Similarly, baseline probabilities of ERA5 and the AR1-process are obtained by sampling from all days t_0 of the time series to day $t_0 + t$, respectively.

Clearly, all probabilities increase with t, as the time window for finding at least one AO^{-3} extreme gets wider. However, with increasing t, also fewer events contribute to the composite due to the finite forecast lead time, leading to larger sampling errors. The results show that p-SSWs are consistently leading to an increased time-integrated risk of AO^{-3} events. For example, the probability in the ECMWF forecasts of at least one AO extreme within 30 days following the event is 3.8%, compared to 2.9% for its climatology. Overall, p-SSWs seem to affect the probability more in the UKMO model, as the probability following p-SSWs is higher and the climatological baseline is also lower compared to the ECMWF model. The baseline in ERA5 is slightly lower than in the ECMWF model, but agrees well with the UKMO climatology. All probabilities range considerably higher than the probability of a one-sided 3-sigma event for the AR1-process and as before, this is a result of the negative skewness of the AO distribution.

Generally, all probabilities appear approximately linear in t, but it should be noted that the linear regime only holds for small enough t as the probability will approach 1 and saturate in the limit of very large t. Furthermore, it is expected that for much larger t (which cannot be evaluated here, due to the finite maximum forecast lead time), the effect of a p-SSW increasing the subsequent extreme AO^{-} probability diminishes and the climatology will approach the one for p-SSWs.

Based on the presented probabilities, the probability increase of at least one AO event within time t following SSWs can be estimated *relative* to the climatological baseline. Following equation 4.2 in section 4.2, the relative probability increase (RPI) is computed as

$$\mathrm{RPI} = \frac{P(AO_{wt} \mid SSW)}{P(AO_{wt})} - 1. \tag{4.6}$$

An RPI larger than 0 corresponds to an increase of AO probability following SSWs, while negative values describe a probability decrease. This ratio is a function of the length of the time window t (see supplement Fig. S2 of Spaeth and Birner, 2022). In the limit of large t, where the SSW influence becomes negligible, it is expected to approach 1, such that the relative probability increase approaches 0. However, for medium time windows t that correspond to a typical timescale of stratosphere-troposphere coupling, the relative

[4]We choose 10 days as we also start to search for p-SSWs at lead time day 10, however, this choice is arbitrary and the resulting climatology is not very sensitive to this choice.

The results of this section are published in 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.

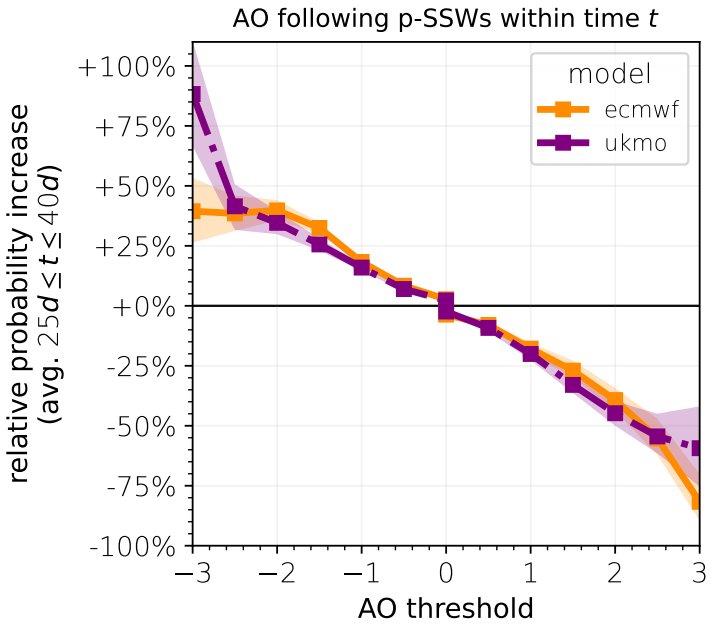

Figure 4.12: Probability increase (in percent) for at least one negative (positive) p-AO extreme below (above) threshold following p-SSWs within a certain period t, relative to climatology, averaged over 25d$\leq$ t $\leq$ 40d, separately for ECMWF (orange, solid) and UKMO (purple, dash-dotted).

The results of this section are published in 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.

probability shows a wide plateau. This motivates the calculation of the relative probability increase averaged over the plateau, which is estimated to correspond to 25 days $\leq t \leq$ 40 days. The resulting relative probability increase (Fig. 4.12) provides an estimate for the extent to which p-SSWs increase the probability of p-AO extreme events – not limited to a specific lag day, but time-integrated and thus independent of t. Note that the measure is relative to the climatology, which also includes AO extremes that occur following SSWs. The diagnostic can therefore be interpreted as the relative probability modulation of at least one $AO^{\pm}$ event within a certain time period following the occurrence of a SSW, relative to the baseline probability where the stratospheric state is unknown.

The relative probability increase of AO events around 0 (e.g., at least one day below/ above 0) is very small, as these events are already almost certain, even in the climatological reference. Both models show a gradual increase of relative probability of more negative AO thresholds (e.g., $\sim +35\%$ for AO< -2) and a gradual decrease for more positive AO thresholds ($\sim -40\%$ for AO$> +2$), which is consistent with a shift of the distribution toward more negative values. Quantitative differences in the results between the models are observed for AO thresholds of ± 3. Indeed, sampling uncertainties become considerable for thresholds greater than 2 standard deviations as well, as indicated by 95% confidence intervals that are obtained via bootstrap sampling among all SSW events. However, model discrepancies reach beyond the indicated confidence intervals, which will be briefly discussed in section 4.4.6.

4.4.4 Toward attribution of predicted AO extremes to preceding SSWs

The last section focused on given p-SSWs and subsequent statistical signatures in AO extremes within a period t: $P(AO_{wt} \mid SSW)$. It was shown that AO^- extremes are significantly more likely following a SSW.

In this section, we aim to evaluate the alternative question: How many AO^- events may statistically be attributable to preceding SSWs? Similar to section 4.3, we address this question using the fraction of attributable risk (FAR). Note that we neglect potential common drivers of both AO and stratospheric extremes, such as due to tropical teleconnections. Consequently our analyses of FAR may overestimate the part that is solely due to the stratosphere. Nevertheless, they serve to quantify the statistical association between stratospheric extremes and the AO, as well as quantify the predictive skill due to the stratosphere.

Fraction of attributable risk among the exposed

First, we focus on the FAR among the exposed (FAR_e), which quantifies how many AO extremes that are preceded by a SSW are attributable to that SSW. FAR_e quantifies how many AO extremes that are preceded by a SSW ("exposed" category) are attributable to that SSW. An FAR_e of 0 means that the probability of finding an AO^- extreme is

The results of this section are published in 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.

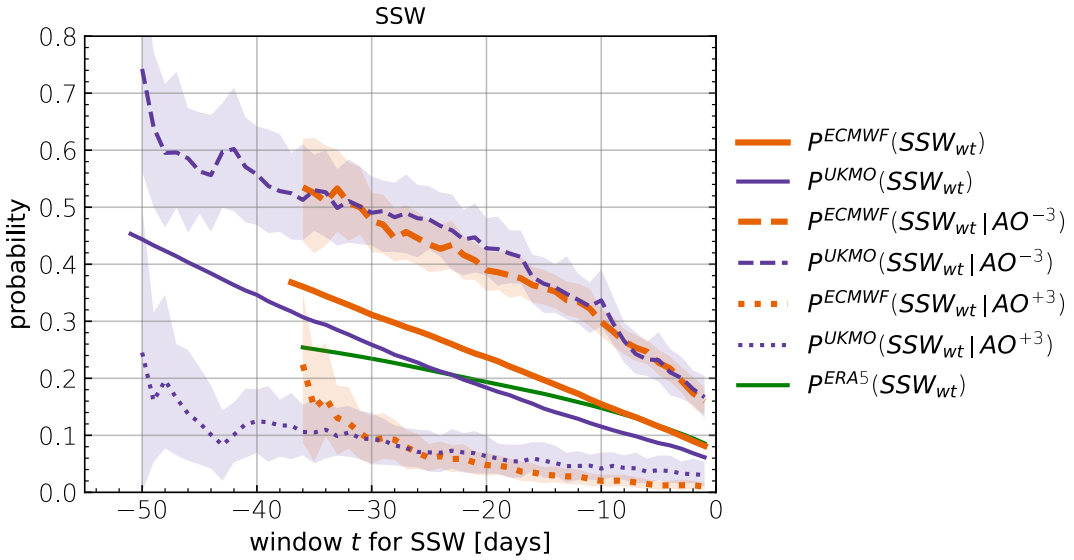

Figure 4.13: Probabilities of at least one day $U_{60}^{10} < 0$ within day t and day -1 relative to day 0, where day 0 is either a randomly sampled day (solid) , an AO^{-3} extreme event (dashed), or an AO^{+3} extreme event (dotted). S2S ECMWF (orange), S2S UKMO (purple) and ERA5 (green).

independent of exposure to a preceding SSW. Likewise, an FAR_e of 1 means that AO^- extremes do not happen without exposure to a preceding SSW.

Using the second re-write of equation 4.3 (section 4.2), we compute FAR_e based on probabilities that condition on AO extremes:

$$FAR_e = 1 - \frac{P(SSW_{wt})}{P(\neg SSW_{wt})} \frac{P(\neg SSW_{wt} \mid AO^-)}{P(SSW_{wt} \mid AO^-)} \tag{4.7}$$

Note that whether a particular AO^- extreme occurs with or without preceding SSWs requires choosing a time window (t) for the potential exposure (e.g., was a given AO extreme preceded by a SSW within the preceding 30 days or not).

In total, the ECMWF analysis is based on 752 AO^{-3} and 486 AO^{+3} events, where asymmetry arises from non-zero skewness of the AO distribution (UKMO: 299 and 186 events).

Fig. 4.13 shows the probability that $AO^{\pm 3}$ events are preceded by at least one day of negative U_{60}^{10} within time t, corresponding to $P(SSW_{wt} \mid AO^{\pm 3})$. For example, the probability of p-SSW occurrence within 30 days preceding AO^{-3} extremes is close to 0.5 in both models, whereas it is around 0.1 preceding AO^{+3} extremes. 95% confidence intervals, which were derived by bootstrap resampling all AO events, confirm that the diagnostics get less robust for larger time windows, due to fewer available events contributing to the AO composite. The probabilities of the extremes to be *not* preceded by at least one day of negative U_{60}^{10} are given by $P(\neg SSW_{wt} \mid AO^{\pm 3}) = 1 - P(SSW_{wt} \mid AO^{\pm 3})$.

We can use the estimated probabilities $P(SSW_{wt} \mid AO^{\pm 3})$, together with the baseline

The results of this section are published in 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.

climatology for SSWs within a time period ($P(SSW_{wt})$) to evaluate $\mathrm{FAR_e}$. Our estimates of FAR_e are shown in Fig. 4.14a as a function of time window t, for two AO event thresholds (–2 and –3). We find that these estimates are not a strong function of the chosen time window. Fig. 4.14b summarizes the FAR_e averaged over time windows of 25 to 40 days: For example, based on the ECMWF forecasts we estimate that on average about 50% of all AO^{-3} events that are preceded by a SSW may statistically be attributable to that SSW. For the UKMO forecasts this value is slightly higher (∼60%). For AO^{-2} events these percentages are somewhat smaller but overall similar between the models. Boxplots reveal that associated sampling uncertainties are generally small, but larger for AO^{-3} events.

Fraction of attributable risk among the population

Second, we additionally compute the FAR among the population ($\mathrm{FAR_p}$), which quantifies how many AO extremes (from the entire population) are attributable to SSWs. In this case one is interested in quantifying the fraction of AO^- extremes that occur in addition to those that are "unexposed" (were not preceded by a SSW). The corresponding FAR among the population is computed following equation 4.4 of section 4.2, as

$$\mathrm{FAR_p} = 1 - \frac{P(\neg SSW_{wt} \mid AO)}{P(\neg SSW_{wt})}. \tag{4.8}$$

$\mathrm{FAR_p}$ quantifies the fraction of AO extremes that may statistically be attributable to a preceding SSW. For example, an $\mathrm{FAR_p}$ of 0 means that SSWs do not increase the probability of AO extremes, whereas an $\mathrm{FAR_p}$ of 1 means that all AO extremes may be attributable to a preceding SSW within time t. The same caveats about common drivers as for FAR_e should be kept in mind.

Figure 4.14c shows our estimates of $\mathrm{FAR_p}$ as a function of time window t, similar as for FAR_e. As expected, estimates of FAR_p are generally lower than for $\mathrm{FAR_e}$: the likelihood of any AO extreme to be attributable to a SSW that may or may not have happened before the AO extreme should be much smaller than that of an AO extreme that was indeed preceded by a SSW. $\mathrm{FAR_p}$ increases somewhat with t for small t, but tends to saturate for windows longer than about 2 weeks. For AO^{-2} events both models saturate near 0.2, whereas for AO^{-3} events they show slightly larger $\mathrm{FAR_p}$ of around 0.25-0.3. Overall our estimates therefore suggest that between 20-30% of AO^- extremes may statistically be attributable to a preceding SSW (within 2-6 weeks). Fig. 4.14d summarizes the FAR_p averaged over time windows of 25 to 40 days. Despite the lower number of contributing events for larger time windows, associated sampling uncertainties are small (e.g., 95% confidence intervals for $\mathrm{FAR_p}$ in ECMWF for AO^{-3}: $[21\%; 28\%]$).

4.4.5 Strong polar vortex events and associated AO extremes

The previous sections revealed that SSWs increase the probability of subsequent AO^- extremes and that a significant fraction of AO^- extremes may be attributable to preceding

The results of this section are published in 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.

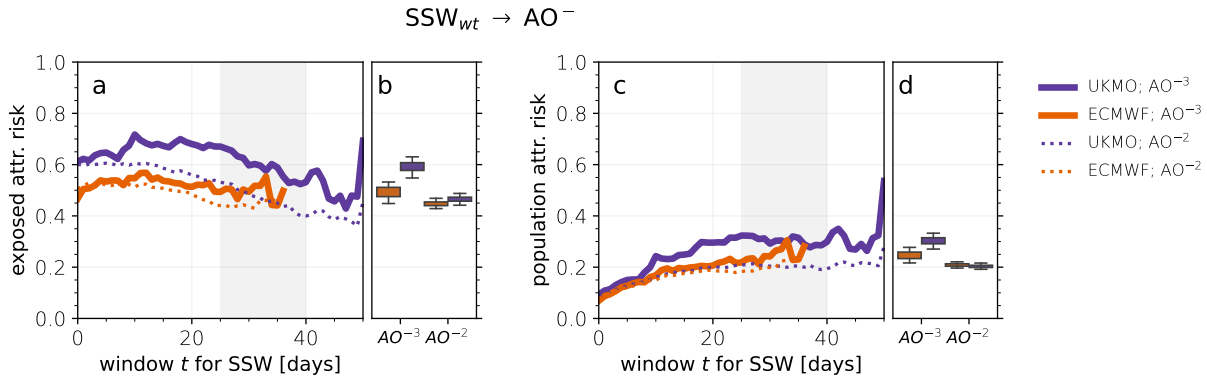

Figure 4.14: Left: Fraction of AO^{-2} (dotted) and AO^{-3} (solid) extremes that are preceded by a SSW within time t that may be attributable to the SSW (fraction of attributable risk among the exposed/ FAR_e, panel a). Boxplots (quartiles 1 to 3 and 95% confidence intervals, obtained via bootstrap resampling) show FAR_e averaged over time windows 25 to 40 days (gray shaded), as function of AO threshold (panel b). Right: Fraction of all AO^{-2} and AO^{-3} extremes that may be attributable to a preceding SSW within time t (fraction of attributable risk among the population/ FAR_p, panel c). Boxplots (as in panel b) show FAR_p averaged over time windows 25 to 40 days (panel d). Note that for larger t, fewer events contribute to the diagnostics, hence, observed fluctuations for long time windows t are likely related to sampling uncertainty. UKMO (purple) and ECMWF (orange).

SSWs. In the following, we summarize an analogous analysis for the statistical relationship between strong polar vortex events (SPVs) and AO^+ extremes.

The composite-mean evolution of p-SPVs (Fig. 4.15) reveals that U_{60}^{10} anomalies are of opposite sign, somewhat weaker in magnitude, but otherwise qualitatively similar to p-SSWs (lag 0: $\sim +20$ ms^{-1} for p-SPVs; ~ -30 ms^{-1} for p-SSWs, cf. Fig. 4.8). Both S2S models agree very well in this respect. Moreover, for negative lags, there is little difference compared to a corresponding composite based on ERA5 data, but for positive lags, U_{60}^{10} is slightly stronger in ERA5. The NAM response at 200hPa and 1000hPa (=AO) is qualitatively similar for p-SPVs and p-SSWs (with opposite sign), but the anomalies are again slightly weaker for p-SPVs, which is consistent with the weaker U_{60}^{10} anomalies (lag 21: +0.35 at 200hPa, +0.25 at 1000hPa). It is interesting that the NAM200 seems to react later to p-SPVs than to p-SSWs: While the index for p-SSWs starts to shift significantly to negative values already at lag -10 on average, a shift to positive NAM200 values for p-SPVs is observed only from lag -5 on. As with p-SSWs, the evolution of the NAM at 200hPa and 1000hPa relative to p-SPVs is less robust in ERA5 due to the smaller sample size, however, the anomalies tend to be slightly more pronounced than in the two S2S models. Overall, the composite-mean evolution of p-SPVs in the ECMWF and UKMO models appear to be consistent with real-atmosphere SPVs (as revealed by reanalysis data), as well as with previous studies (e.g., Baldwin and Dunkerton, 2001).

Following the same methodology as for p-SSWs, we use the large event sample sizes to

The results of this section are published in 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.

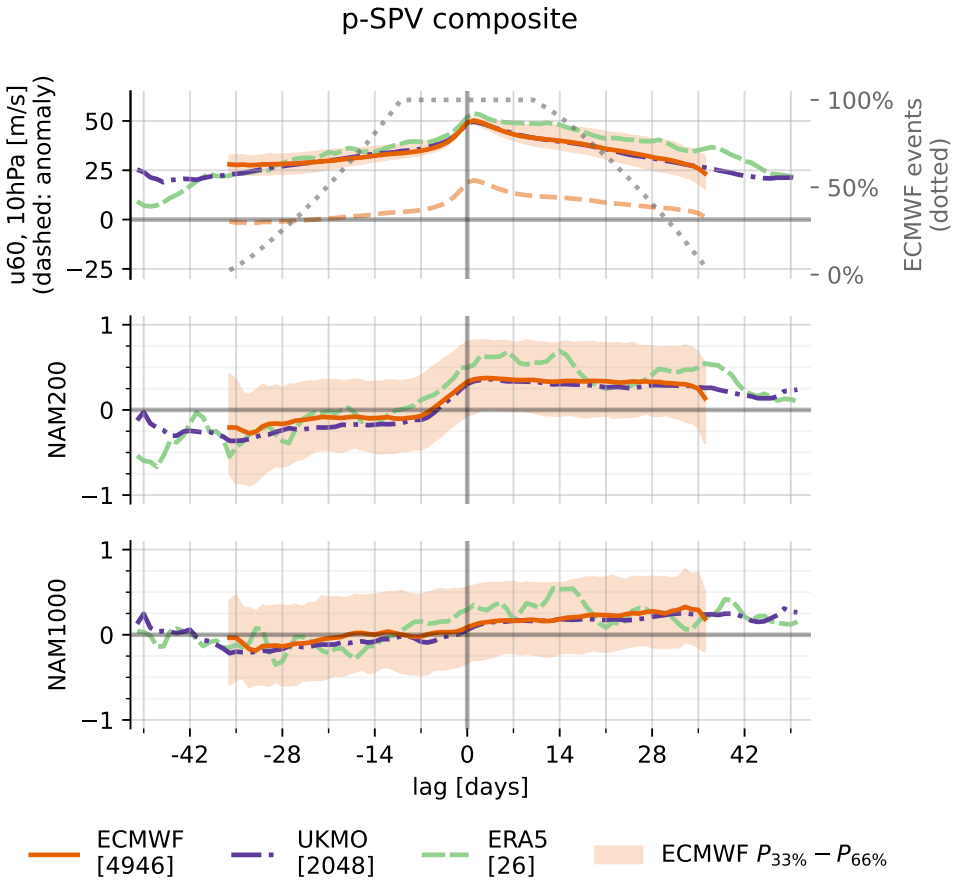

Figure 4.15: As in Fig. 4.8, for p-SPVs.

The results of this section are published in 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.

quantify the statistical relation between p-SPVs and subsequent AO^+ extremes. First, we quantify the relative probability increase for at least one AO extreme after a given p-SPV within a certain time. Second, we analyze how many AO^+ extremes may be attributable to preceding p-SPVs.

Figure 4.16 shows the relative probability increase of AO extremes following SPVs relative to climatology as a function of the AO threshold, for both S2S models and averaged over time windows 25 days $\leq t \leq$ 40 days:

$$\mathrm{RPI} = \frac{P(AO_{wt} \mid SPV)}{P(AO_{wt})} - 1 \tag{4.9}$$

Consistent with the positive shift of the AO distribution following SPVs, the risk gradually increases for positive AO extremes, whereas it gradually decreases for negative AO extremes. For extreme thresholds of up to 2 standard deviations, the relative probability change appears to be of similar magnitude compared to periods following SSWs (≈30-40%, see Fig. 4.12). Larger thresholds reveal a reduced probability change compared to SSWs, however, 95% confidence intervals mark increasing sampling uncertainty, especially for AO^{+3}_{wt} events.

Figure 4.17 shows our estimates of the fraction of positive AO extremes that may be attributable to a preceding p-SPV within a time period t:

$$\mathrm{FAR_e} = 1 - \frac{P(SPV_{wt})}{P(\neg SPV_{wt})} \frac{P(\neg SPV_{wt} \mid AO^-)}{P(SPV_{wt} \mid AO^-)} \tag{4.10}$$

$$\mathrm{FAR_p} = 1 - \frac{P(\neg SPV_{wt} \mid AO)}{P(\neg SPV_{wt})} \tag{4.11}$$

where $\mathrm{FAR_p}$ and $\mathrm{FAR_p}$ denote the fraction of attributable risk among exposed and population, as in section 4.4.4 for SSWs and AO^- events. Among all AO^{+3} events that are preceded by at least one SPV event within four weeks, about 55% (UKMO) to 65% (ECMWF) may be attributable to the SPV (Figs. 4.17a, 4.17b). However, significant sensitivities to the exact time window are observed, as well as differences between the models. One problem is the strong seasonal dependence of SPV events, as most events occur in December when the polar vortex is generally strongest. AO extremes that happen later in the winter have therefore a smaller probability to be preceded by a SPV event within a short time window than AO extremes that occur in December or January. AO^{+2} events reveal a fraction of attributable risk among the exposed to preceding SPVs of around 40% to 55%, similar to SSWs and AO^{-2} events.

Finally, the fraction of all AO^+ extremes that may be attributable to preceding SPVs is slightly larger but similar to that for AO^- extremes and SSWs, with a population attributable risk of around one quarter for AO^{+2} and around one third for AO^{+3} extremes for preceding time windows of 25 to 40 days (Figs. 4.17c, 4.17d).

More detailed analyses that apply the diagnostics presented in Fig. 4.9, Fig. 4.10 and Fig. 4.11 to positive AO extremes and p-SPVs can be found in the supplement of Spaeth and Birner (2022).

The results of this section are published in 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.

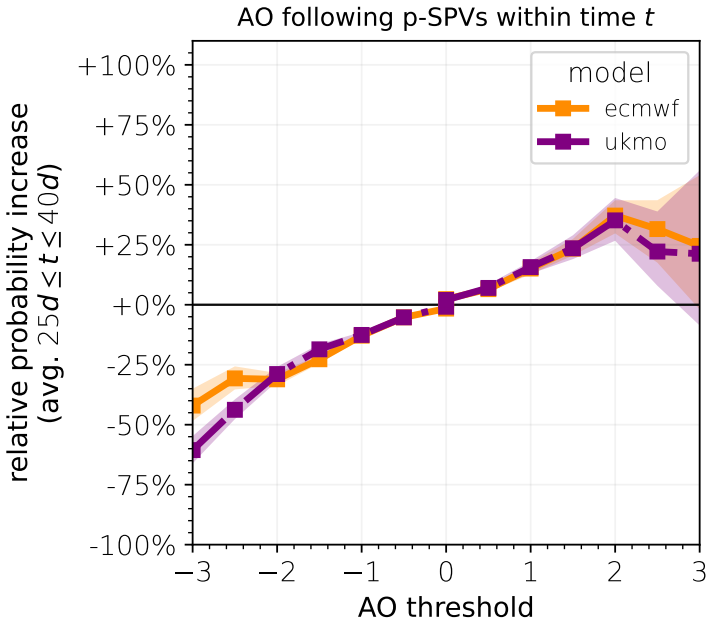

Figure 4.16: As in Fig. 4.12, for p-SPVs and subsequent AO extremes within time t.

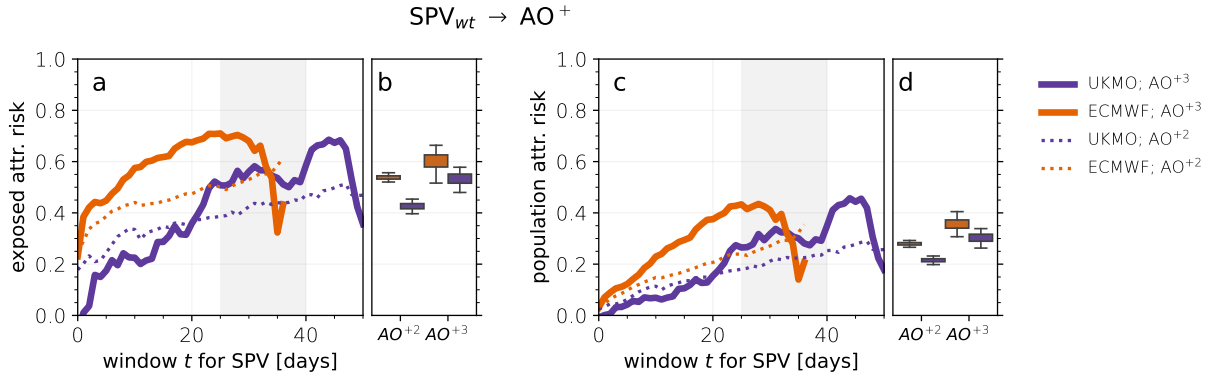

Figure 4.17: As in Fig. 4.14, for positive AO extremes that may be attributable to preceding SPV events within time t.

The results of this section are published in 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.

4.4.6 Conclusions

Our results, based on a large number of extended-range ensemble forecasts, provide further evidence for stratospheric modulation of large-scale weather patterns near the surface, broadly consistent with previous results (Domeisen and Butler, 2020, and references therein). Previous studies generally suffer from relatively small available sample sizes, which hampers estimation of robust statistical relationships between stratospheric and tropospheric extremes (= rare events). In this study, by analyzing extended-range forecast periods around predicted extreme events (e.g., p-SSWs), we effectively boost the available sample size by more than a factor of 100 and are therefore in the position to obtain robust estimates in response to our research questions:

1. *By how much is the probability of persistently positive or negative AO phases increased following stratospheric polar vortex extremes?*

   Climatologically, 38% of negative AO phases (days with consecutive AO$<$ 0) are longer than 7 days. Following p-SSWs, this is increased to 44%, which corresponds to a relative increase of 16%.

   Following p-SPVs, the probability of positive AO phases that last longer than 7 days is increased from 40% to 44%.

2. *By how much is the probability of subsequent AO extremes increased following stratospheric polar vortex extremes?*

   Following p-SSWs, the probability of subsequent negative AO extremes increases whereas it decreases for positive AO extremes. For instance, AO^{-3} events are about 40% (ECMWF forecasts) to about 80% (UKMO forecasts) more likely following p-SSWs. However, the absolute probabilities are still low, i.e., only 3.5% of SSWs are followed by AO^{-3} within four weeks, based on ECMWF forecasts (UKMO: 4%).

   Following p-SPVs, the probability of AO^{+3} is increased by about 25% relative to climatology, whereas AO^{-3} occur about 40% (ECMWF) to 60% (UKMO) less often.

3. *What fraction of AO extremes may be attributable to preceding stratospheric polar vortex extremes?*

   About 50% (ECMWF) to 60% (UKMO) of AO^{-3} extremes that occur following a SSW may be attributable to that SSW (fraction of attributable risk among the exposed). 20-30% of all AO^{-3} events may be attributable to preceding SSWs (fraction of attributable risk among the population). "Attributable" does not necessarily imply strict causality (see discussion below), but here refers to the fraction of SSW-AO^{-} co-occurrences in addition to fortuitously aligned events.

While our stratospheric event definitions are based on absolute thresholds of the zonal-mean zonal wind, the tropospheric response is quantified via standardized anomalies of averaged geopotential. The construction of an appropriate corresponding climatology is

The results of this section are published in 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.

crucial, in particular for the analysis of extreme events. However, it is also not unambiguous. Standardized anomalies are computed by normalizing differences from a population mean with the population standard deviation (taking into account seasonal variations). As the population is usually finite, any additional data point may change the population mean and will change the population standard deviation, resulting in a small adjustment of all previous (standardized) data points. On the one hand, the effect is negligible in the limit of a large population. On the other hand, it is generally larger when the additional data point is an outlier with respect to the previous distribution. For this study, S2S forecasts were deseasonalized using the available hindcasts. The assumption is that these hindcasts sufficiently sample different kinds of variability, such that a) extreme events that occurred in individual years do not significantly distort the population distribution and thereby also the population mean and standard deviation and that b) the constructed population is robust across different initialization dates (e.g., a given event that is equally predicted at two different leadtimes corresponds to a the same standardized event in both model integrations).

Do the analyses of modulated probabilities allow conclusions about causal links between stratospheric and tropospheric circulation extremes?

A definition of (probabilistic) causality is provided by Pearl (2009, September):

$$P(\text{effect} \mid do(\text{cause})) > P(\text{effect} \mid \ do(\neg\text{cause})), \tag{4.12}$$

where the *do* operator denotes an intervention that forces the occurrence or not-occurrence of the cause[5]. In the atmosphere, such controlled situations can usually only be simulated using numerical model experiments. In this study, a post-hoc analysis of an existing dataset is presented. No interventions are performed and therefore, no strict causal relations can be inferred following the provided definition. Instead, conditional probabilities are computed, which Pearl (2009, September) calls a predictive or observational approach, e.g.:

$$P(AO^- \mid SSW) > P(AO^- \mid \ \neg SSW). \tag{4.13}$$

Our knowledge of coupled stratosphere-troposphere dynamics suggests that a causal connection does in principle exist[6]. This connection manifests in observed conditional probabilities, which may, however, be modulated also by further possibly involved pathways.

First, conditional probabilities may in practice overestimate the (direct) causal link between stratospheric and AO extreme due to the existence of confounding factors (see scenario *c* listed in the introduction). For example, MJO may lead to modified risk of AO extremes (Barnes et al., 2019) while at the same time modifying the likelihood of SSWs (Garfinkel et al., 2012). On the other hand, the dynamical coupling between the MJO and the AO may involve a stratospheric pathway (Garfinkel et al., 2014) and in such cases the stratosphere

[5]This definition relies on counterfactual dependence, i.e., if there had not been the cause, then there would not have been the effect (and if there had been the cause, then there would have been the effect).

[6]It is important to keep in mind that the coupling is, in general, mutual and causality works in both directions (even though, as always, any cause has to precede the effect).

The results of this section are published in 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.

does represent a causal driver of AO modulations. Similar arguments hold for impacts due to climate variability, such as Arctic sea ice concentrations (Kretschmer et al., 2016) and the ENSO (Domeisen et al., 2019). Causal pathways may in such cases be disentangled using a causal inference-based network (Adams et al., 2021). We have carried out preliminary analyses using such a framework to distinguish causal pathways during different ENSO phases, which suggest that the direct pathway *polar vortex* $\rightarrow$ *AO extremes* is significantly stronger than those via ENSO. A detailed analysis of these pathways is left for future work.

However, even if common drivers can be neglected the statistical nature of inferred fraction of attributable risk can only quantify an *effective* causality in the following sense. Assume, for the moment, that all SSWs cause an AO^- extreme, but AO^- extremes additionally occur due to internal tropospheric variability. In this case some of the observed AO^- extremes may have happened due to internal tropospheric variability alone while additionally be forced/enhanced by a preceding SSW (see scenario *b* listed in the introduction). A probability analysis (e.g., estimating the FAR among the population) will then always underestimate the actual causal link and can only reveal an effective causality. This also represents a limitation of the binary classification (AO extreme / no AO extreme).

Despite these caveats, conditional probabilities may provide useful insights. The conversion into statistical metrics such as RPI and FAR may thereby facilitate the practically relevant interpretation. For example, RPI of AO extremes due to the prior occurrence of a stratospheric extreme does serve to quantify the state of the stratosphere as a predictor of subsequent AO extremes, which may be of practical value regardless of its underlying causal nature. Furthermore, FAR provides an estimate of how many AO extremes would statistically be expected less without preceding stratospheric events, when keeping in mind that "without a preceding stratospheric event" would require to remove also confounding factors.

How should the observed differences between ECMWF and UKMO model be interpreted? Overall, our analyses show that the probability modulation of AO extremes up to about two standard deviations given preceding stratospheric extremes are similar between the ECMWF and the UKMO model. AO extremes of three standard deviations, i.e., $AO < -3$ and $AO > +3$ reveal discrepancies between the models. Our bootstrapping approach, e.g., for the relative probability increase (Fig. 4.12), shows that especially analyses based on UKMO forecasts become less robust. However, the observed discrepancies cannot be solely attributed to sampling uncertainty, given that they exist also beyond the respective 95% confidence intervals. Which model better represents the dynamics of the real atmosphere is difficult to assess, as the observational record is too short to allow for robust, similar analyses. Potential causes of the observed differences are numerous, involving differences in wave-mean flow feedbacks or external forcings, e.g., from the tropics. Augier and Lindborg (2013) show that the eddy kinetic energy spectrum in the ECMWF model is still in parts unrealistic and that the model may be too dissipative even at large scales, clearly indicating that models are unable to reproduce real-atmosphere dynamics perfectly ac-

The results of this section are published in 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.

curate. Lawrence et al. (2022) investigate biases in different S2S models and find, inter alia, a modest cold bias in the ECMWF and a modest warm bias in the UKMO model in the extra-tropical lower stratosphere. As the lower stratosphere has been shown to play an important role in stratosphere-troposphere coupling, we speculate that occurrences of tropospheric extremes following stratospheric circulation anomalies are sensitive to temperature biases in this region. However, a detailed analysis would be beyond the scope of this study.

In general, we note that any two different imperfect models, will likely always reveal quantitative differences in the analysis of extreme events for a sufficiently strict extreme threshold. In the present study, we find such differences, e.g., for the relative risk, at a threshold of around three standard deviations. It is possible that more data are needed to conclusively attribute the differences to particular dynamical processes. Nevertheless, we argue that our analyses, even at a threshold of 3 standard deviations and given the associated uncertainties, are able to provide insightful quantitative estimates; especially as no obvious a priori estimate exists even for the order of magnitude of the investigated probability metrics.

In addition to the particular points already mentioned, future work should address the question, how much of the predicted surface impact following predicted stratospheric extremes, i.e., following p-SSWs and p-SPVs, can be explained by the AO. Lastly, we conclude that the analysis of *predicted* events offers potential for improved statistical characterization of other atmospheric extreme events, provided that the forecast model is capable of truthfully representing the event of interest.

The results of this section are published in 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.

Chapter 5

Forecast uncertainty in subseasonal-to-seasonal forecasts

Uncertainty is the only certainty there is, and knowing how to live with insecurity is the only security.
John Allen Paulos, * 4th July 1945, Professor of Mathematics and Author

This chapter explores flow-dependent modulations of uncertainty in subseasonal-to-seasonal forecasts, with a main focus on stratosphere-troposphere coupling.

Section 5.1 analyzes forecast uncertainty of the polar vortex strength and shows that this uncertainty reduces following SSWs, implying enhanced predictability. This motivates the question of whether SSWs also reduce uncertainty in tropospheric forecasts, which is addressed in **Section 5.2**. Indeed, the results reveal robust modulations of uncertainty in 1000 hPa geopotential height forecasts. Motivated by the hypothesis that these signatures are driven by anomalies in storm frequency, **Section 5.3** assesses the relation between individual storms and ensemble spread more systematically.

In general, forecast uncertainty depends on the large-scale flow, with stratosphere-troposphere coupling being one example. However, the influence of the large-scale flow on forecast uncertainty can be generalized, as shown in **Section 5.4**: eight North Atlantic-European weather regimes are associated with distinct signatures in forecast uncertainty of 1000 hPa geopotential height and 2-meter temperature, respectively. Importantly, the uncertainty of these weather regimes themselves can be influenced by teleconnections, such as the polar vortex, which rounds off this chapter.

5.1 Reduced polar vortex forecast uncertainty following sudden stratospheric warmings

This section is based on Rupp et al. (2023). Philip Rupp had lead the analyses. I had contributed to the design and interpretation of the results and I had produced the corresponding figures, which are re-printed here. All figure captions are copied from the published article and were written by Philip Rupp. Compared to the published article I adapted the text below at many places to better fit this thesis.

The circulation in the stratosphere is generally more predictable than in the troposphere (e.g., Ambaum and Hoskins, 2002; Baldwin et al., 2003; Domeisen et al., 2020c). Nonetheless, stratospheric forecasts still exhibit substantial forecast uncertainty, as indicated by ensemble variance. The primary source of variability (and consequently, ensemble spread) in the winter polar stratosphere are planetary waves and their interaction with the mean flow (Baldwin and Holton, 1988). Certain conditions, however, are linked to persistent anomalies in variability. From a forecasting perspective, reduced variability can enhance forecast skill. This section demonstrates that such conditions can be established by SSWs.

Specifically, it will be shown that forecasts initialized around SSWs exhibit long-lasting anomalies in the lowermost stratosphere, affecting planetary wave activity and thereby variability, and predictability. Note that the same S2S forecast data is used as in Chapter 4. However, the following analyses do not follow the UNSEEN approach (that is, composites are not based on events *predicted* in the forecasts), but forecasts are instead clustered according to the state of the polar vortex in the real atmosphere at the time of initialization.

Fig. 5.1a shows that U_{60}^{10} ensemble spread at S2S timescales reduces by more than one third in forecasts initialized around SSWs (central date $\pm$ 3 days) compared to climatology. Even forecasts initialized 10 days before a SSW exhibit reduced spread, although anomalies are weaker. From Fig. 5.1b, it is evident that the reduced spread in these SSW-initializations is not simply attributable to seasonal variations in both spread and SSW frequency: anomalies of ensemble variance reach up to minus one standard deviation at lead times of about five to six weeks. The negative spread anomalies of the zonal-mean zonal wind at 60 °N (U_{60}) are largest at 10 hPa, but also propagate downward to the lower stratosphere around 200 hPa.

Reduced spread reflects a narrower range of possible evolutions of the real-atmosphere. Consequently, the ensemble mean forecast in such situations tends to be closer to the real-atmosphere evolution, resulting in reduced error (recall the discussion of reliability in Section 3.3). Figure 5.2 confirms that the decreased U_{60} ensemble spread in SSW initializations is associated with reduced absolute error. This improvement in skill is expected to be more pronounced for strong SSWs compared to weak SSWs. Therefore, we show error anomalies separately for these two subsets.

Strong (weak) SSWs are defined based on the lower (upper) tercile of U_{60}^{10} among all SSWs, averaged over lags of +7 to +13 days relative to the SSW central date. Errors are computed as the absolute difference between the ensemble mean forecast and ERA5. SSW

The results of this section are published in Rupp, P., Spaeth, J., Garny, H., & Birner, T. (2023). Enhanced Polar Vortex Predictability Following Sudden Stratospheric Warming Events. *Geophysical Research Letters*, 50(17).

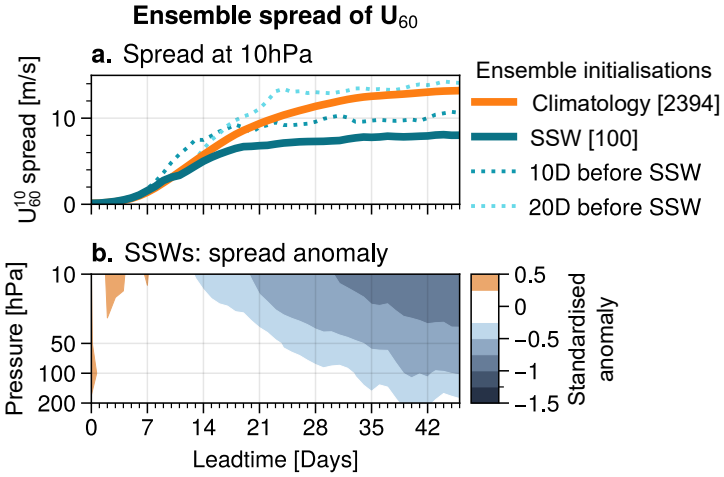

Figure 5.1: **a.** Evolution of U_{60}^{10} ensemble spread in terms of square root of the average ensemble variance over all forecast initializations (climatology) and over SSW initializations. Dotted lines show the spread evolution of forecasts initialized 10 or 20 days ahead of a SSW. Numbers in square brackets indicate the number of ensemble forecasts within the respective group. **b.** Standardized (w.r.t. the leadtime, season and model-version dependent, climatological distribution of ensemble spread) anomaly of the U_{60} ensemble spread within SSW forecast initializations compared to climatological initializations. All shown anomalies are significant ($p < 0.05$).

The results of this section are published in Rupp, P., Spaeth, J., Garny, H., & Birner, T. (2023). Enhanced Polar Vortex Predictability Following Sudden Stratospheric Warming Events. *Geophysical Research Letters*, 50(17).

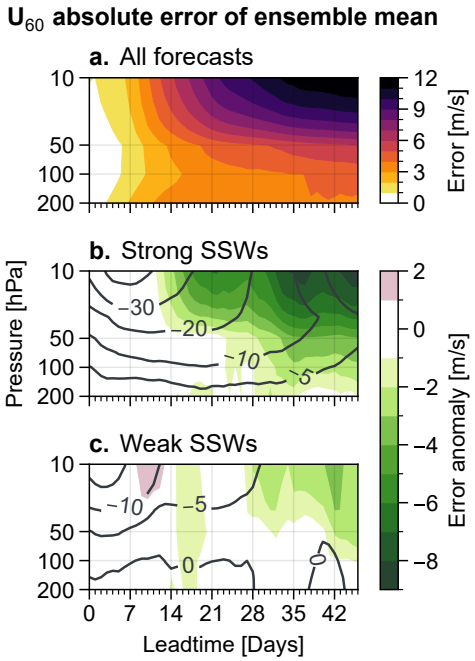

Figure 5.2: **a.** Absolute error of the ensemble mean of U60 anomalies w.r.t. reanalysis evolution averaged over all forecasts (climatology). **b. and c.** The average anomaly of the absolute error for SSW initializations corresponding to strong/medium and weak SSWs (see Section 2.3), respectively. Contour lines show composite mean U_{60} anomalies (in $\mathrm{m\,s^{-1}}$).

initializations are associated with reduced error from about forecast week two onward, consistent with the reduced spread shown in Figure 5.1. The error anomaly signal is most pronounced at 10 hPa (up to $-9\,\mathrm{m\,s^{-1}}$ for strong, up to $-4\,\mathrm{m\,s^{-1}}$ for weak SSWs), although climatological errors are generally larger in the middle compared to the lower stratosphere (Fig. 5.2a). The alignment of reduced spread with reduced error indicates reliability of the model, suggesting that the reduced range of possible scenarios reflects actual atmospheric conditions rather than model limitations.

What is the reason for the reduced spread and error in forecasts initialized around SSWs? SSWs tend to be followed by persistent zonal-mean easterly winds in the middle and lower stratosphere. Importantly, such easterly winds suppress upward propagation of Rossby waves (Charney and Drazin, 1961). Consequently, heat flux, a proxy for vertical Rossby wave propagation, for zonal wave number 1 is anomalously small following most SSWs (Fig. 5.3a).

Compared to the climatological distributions of U_{60}^{10} and heat flux, SSWs form a distinct cluster characterized by negative U_{60}^{10} and negative heat flux anomalies. Within that cluster, there appears to be a somewhat linear relation between the two variables. However, this linear relationship does not necessarily extend beyond SSWs, as strong positive U_{60}^{10} is

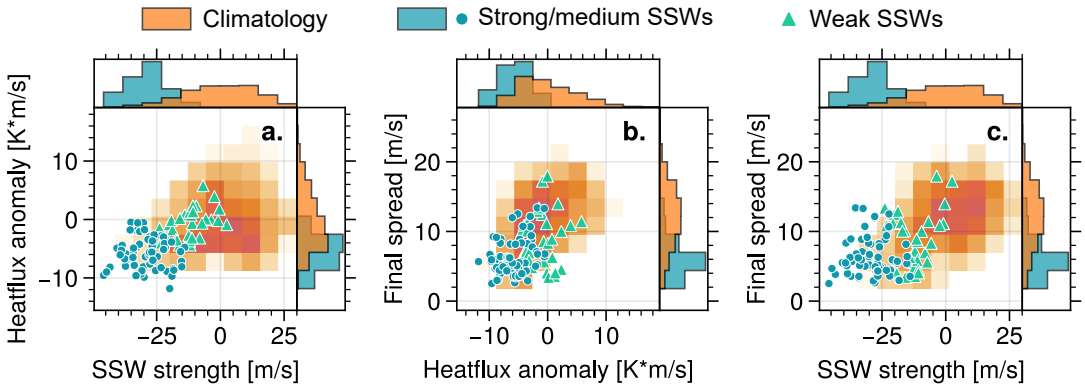

Figure 5.3: Scatter plots of different metrics calculated from all forecast initializations (climatology) and SSW ensemble initializations. Metrics are polar vortex strength (U_{60}^{10} anomaly averaged over days +7 to +13), heat flux anomaly (wave number 1 eddy heat flux at 100 hPa averaged over days +14 to +34) and final spread (square root of the mean ensemble variance of U_{60}^{10} averaged over days +35 to +41). SSW initializations are divided into strong/medium and weak cases (see Section 2.3). The shading denotes the bivariate distributions derived from all available forecasts. Histograms show univariate and normalized distributions for the climatological and strong/medium SSW initializations.

associated with slightly negative heat flux anomalies, as well. In such cases, the causality might be reversed, with negative negative heat flux anomalies potentially leading to a decrease in the deceleration of U_{60}^{10} .

Figure 5.3a has confirmed that SSWs, especially strong ones, are followed by reduced heat flux. Figure 5.3b relates these heat flux anomalies to the ensemble spread at lead times of six weeks. SSWs initializations that are followed by negative heat flux anomalies exhibit smaller final U_{60}^{10} ensemble spread. Furthermore, there is a general relationship between heat flux anomalies and ensemble spread, not limited to SSWs. We interpret this relation as follows. Upward propagating planetary waves, quantified by the eddy heat flux, represent the primary source of variability in the stratosphere. The precise strength and timing of strong planetary wave pulses, as well as their impact on the mean-flow are challenging to predict. At longer lead times, these events are simulated differently across ensemble members, resulting in larger ensemble spread. SSW initializations are particularly interesting because the easterly winds in the lower stratosphere suppress upward propagation of planetary waves. This suppression manifests in negative heat flux anomalies and leads to smaller ensemble spread (see Fig. 5.3c).

The reduced heat flux following SSWs decreases U_{60}^{10} ensemble spread and thereby improves polar vortex predictability. We now explore an additional implication of these persistent negative heat flux anomalies. As a result of reduced mid-stratospheric wave activity, the polar vortex becomes stronger, driven by radiative cooling, and less disturbed following SSWs. This leads to a robust delay of the final warming compared to years that did not experience a mid-winter SSW (Fig. 5.4).

The results of this section are published in Rupp, P., Spaeth, J., Garny, H., & Birner, T. (2023). Enhanced Polar Vortex Predictability Following Sudden Stratospheric Warming Events. *Geophysical Research Letters*, 50(17).

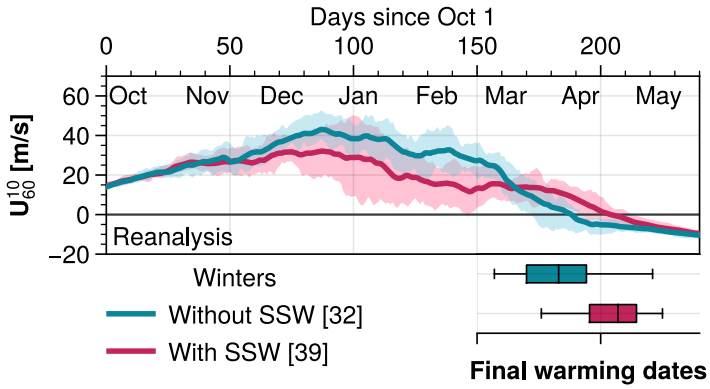

Figure 5.4: Climatological evolution of U_{60}^{10} in reanalysis data, separated for winter with (39 winters) and without (32 winters) a major SSW. Solid lines denote the mean, shading denotes inner-quartiles of the distributions. Box plots show the distributions of final warming dates (black lines: median; box-edges: inner-quartiles; whiskers: minima and maxima).

In ERA5, 32 out of 71 winters did not experience a mid-winter SSW, with a median final warming occurring on the 2nd of April. In contrast, the 39 winters with a mid-winter SSW saw the median final warming delayed by more than 3 weeks, to the 26th of April. With reduced wave activity in the mid-stratosphere, there is also less wave absorption and dissipation, resulting in reduced westward forcing acting on the zonal mean flow. Consequently, winters with a major SSW consistently exhibit a delayed final warming.

5.2 Stratospheric impact on extended-range forecast uncertainty in the Northern extratropics

The previous section revealed that ensemble spread of the polar vortex strength reduces following SSW. However, it remains unclear whether this reduced spread in the stratosphere also translates into reduced ensemble spread in the troposphere. This section will address this question. The results are published in Spaeth et al. (2024a).

The effects of extreme polar vortex states on subsequent mean surface weather are widely recognized. For instance, abrupt breakdowns of the polar vortex, such as associated with SSW (Baldwin et al., 2021), tend to be followed by a persistent equatorward jet shift in the troposphere for up to two months (Baldwin and Dunkerton, 2001). This signal tends to be strongest over the Atlantic (associated with a negative phase of the NAO) and enhances the likelihood of cold air outbreaks over Eurasia (Thompson et al., 2002; Lehtonen and Karpechko, 2016; Afargan-Gerstman et al., 2020; Kautz et al., 2020; Huang et al., 2021). Likewise, events of extremely strong polar vortex tend to be followed by a poleward jet shift with greater likelihood for storm series (Domeisen et al., 2020c; Rupp et al., 2022).

From this perspective stratospheric conditions may provide a source of tropospheric pre-

The results of this section are published in 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.

dictability on S2S time scales. Indeed, past literature has demonstrated improved forecast skill for ensemble forecasts initialized around SSW (Sigmond et al., 2013) or strong polar vortex episodes (Tripathi et al., 2015). However, the employed skill measures, such as the correlation skill score (CSS), primarily rely on the ensemble mean signal (e.g., the aforementioned negative NAO state following SSW).

Another, perhaps equally important, aspect of predictability is the ensemble spread signal, as demonstrated in the previous section. The ensemble spread may be thought of as measuring the degree of confidence in the ensemble mean signal, thus indicating whether the flow becomes more predictable. Notably, the response of tropospheric ensemble spread to stratospheric extreme events has not been studied and this is the main focus of the analyses below.

We cluster the ECWMF S2S ensemble forecasts (see section 3.1.1) into categories of weak, moderate and strong initial polar vortex states. Regions are diagnosed that experience anomalously low or high spread in the troposphere following polar vortex extremes. In particular, we identify regions of near-zero ensemble mean signal that nevertheless exhibit enhanced or reduced intrinsic predictability due to spread anomalies.

Before presenting the results, the methodological approach is described more in detail, including the clustering of the ensembles and the use of eddy activity as a proxy for synoptic-scale wave activity.

Clustering of ensembles based on the polar vortex

Each ensemble forecast is labeled as weak, moderate or strong vortex, depending on U_{60}^{10} in ERA5 reanalysis at initial time. Thereby, the effect of initial conditions on forecast evolution is isolated (though few cases exist where the polar vortex becomes strong or weak only during the forecast period). For weak vortex forecasts, U_{60}^{10} is required to be negative. Following a standard SSW definition (Charlton and Polvani, 2007), the forecasts in this group are thus likely to have experienced a SSW shortly before initialization. Out of all 2394 ensemble forecasts, 169 qualify as weak vortex initializations. Similarly, the 169 forecasts with the strongest initial polar vortex are labelled as strong vortex initializations. The remaining forecasts form the moderate vortex category.

Overall, forecasts sample winters from 1997 to 2021. The distribution of weak and strong vortex initializations over different years is provided in the supplementary Fig. S1a of Spaeth et al. (2024a).

Eddy Activity

Daily atmospheric variability in the extratropics is characterized by synoptic-scale weather systems (decorrelation timescales on the order of few days, Williams et al., 2017). These eddies can be quantified via quasi-geostrophic wave activity, which is proportional to potential vorticity variance around some basic state, which separates eddies from the mean

flow. Upon integration over the full domain, wave activity fulfills a conservation law for adiabatic, inviscid dynamics (Vallis, 2017, June).

As a proxy for wave activity, we define the time- and space-dependent eddy activity, $\mathcal{A}$, a simplified measure for synoptic-scale waves, which corresponds to potential vorticity variance:

$$\mathcal{A} = \left(\mathrm{PV320K} - \overline{\mathrm{PV320K}}\right)^2, \tag{5.1}$$

where PV320K is the potential vorticity at the 320 K isentrope. Mean-flow and synoptic-scale eddies are separated using simple time-filtering, based on a 7-day centered running mean at each grid point, denoted by $\overline{\mathrm{PV320K}}$. Planetary-scale perturbations typically have longer time scales than seven days and are therefore filtered out. Variability that is associated with periods shorter than one day do not contribute as well, because data is used only at daily resolution. Therefore, the eddy activity essentially corresponds to a temporal bandpass filtered potential vorticity variance. We have tested that the results are not very sensitive to choices of basic state timescales between 5 and 14 days (not shown) and Gibbs phenomena have near-zero impact on results as tested by multiple iterative application of the 7-day moving average (not shown). Large values of eddy activity can for example be associated with Rossby wave breaking.

5.2.1 Forecast uncertainty following polar vortex events

We use 1000 hPa geopotential height (Z1000) as a metric to quantify the anomalous surface circulation in weak or strong vortex initializations. The results are based on daily diagnostics that are averaged over lead time days 14 to 34. Figure 5.5 illustrates that weak vortex initializations are on average followed by a negative NAO pattern in the ensemble mean, consistent with many previous studies. Positive anomalies (=high pressure) are observed over the polar cap northward of about 60 °N, negative anomalies are observed over the Euroatlantic sector. The positive anomalies over the Pacific seem in conflict with the expected negative Arctic Oscillation (AO) signature. They likely represent ENSO signatures that do not fully average out due to the limited hindcast period and associated sampling bias. This is supported by the average ENSO index at initial time, which yields −0.36 for weak and +0.23 for strong vortex initializations.

How is the negative NAO signal linked to anomalies in forecast spread? Fig. 5.5 reveals a pronounced reduction of ensemble spread, and therefore enhanced forecast confidence over the North Atlantic jet exit region, centered around Iceland to Scandinavia. The observed negative anomalies of ensemble variance of up to ∼3000 gpm^2 correspond to about -0.5σ (standard deviations, relative to the climatological spread distribution) or to a reduction by roughly 20%. In the extra-tropics this is the largest signal in terms of standardized anomalies. Interestingly, the regions of anomalous Z1000 spread differ from those of anomalous Z1000 ensemble mean. For example, the region around Denmark shows essentially zero ensemble mean response, but is nevertheless characterized by enhanced forecast confidence due to reduced ensemble spread.

The results of this section are published in 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.

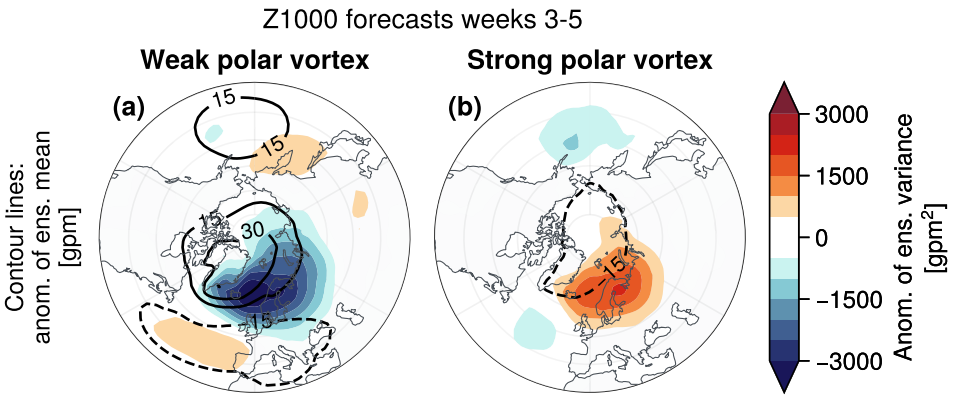

Figure 5.5: Ensemble statistics of 1000 hPa geopotential height (Z1000) in S2S ECMWF forecasts with a weak (panel a, average over 169 cases) and strong (panel b, average over 169 cases) initial polar vortex. Contours denote the Z1000 ensemble mean (15 gpm intervals). Shading shows lead-time- and season-aware Z1000 anomalies of ensemble variance.

Furthermore, weak vortex initializations reveal positive spread anomalies over the Atlantic around 40 °N and over the North-West Pacific. While enhanced spread translates into reduced forecast confidence, the magnitude of the signals is weaker in comparison to the negative spread anomalies over North Eurasia.

Forecasts initialized with a strong polar vortex qualitatively show similar patterns of mean and spread anomalies compared to weak vortex initializations, with opposite sign. Most pronounced is an increase in forecast spread over Northern Eurasia, where anomalies correspond to about +0.3 standard deviations or to an increase by roughly 15%.

Most forecasts neither fall into the weak or strong vortex category by construction. The associated composite-mean anomalies of mean and spread are close to zero (not shown). An equivalent analysis based on UKMO forecast data (1568 ensemble forecasts with lead times up to +60 days) reveals overall very similar results (see supplementary Fig. S3 of Spaeth et al., 2024a).

5.2.2 Dynamical link between ensemble mean and spread

What causes the observed spatial patterns of ensemble mean and spread? To help answer this question we focus on the Atlantic sector, as signals over the Pacific are weaker and more strongly affected by ENSO variability.

Our guiding hypothesis is as follows. We generally expect largest ensemble spread to occur in regions of strongest synoptic-scale storm development, primarily due to the unpredictable nature of exact storm location and strength at subseasonal lead times. Following stratospheric extreme events, the Atlantic jet undergoes latitudinal shifts that are associated with similar shifts in storm tracks (Baldwin and Dunkerton, 2001) and these should manifest in associated anomalies in ensemble spread.

The results of this section are published in 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.

To illustrate effects of polar vortex variability on the North Atlantic eddy-driven jet, Figures 5.6a and 5.6c present mean and variance anomalies of horizontal winds at 850 hPa (zonal wind: U850, meridional wind: V850). In alignment with the negative (positive) NAO phase, the composite-mean westerly zonal winds at latitudes northward of about 50 °N weaken (strengthen) over the Atlantic, corresponding to the latitudinal jet shift. Weak polar vortex forecasts reveal anomalously low ensemble variance (i.e., high confidence) of zonal wind over Iceland to Scandinavia. Positive ensemble variance anomalies of up to 0.4σ are observed over Canada's East coast and the Canary basin. Overall, the ensemble spread anomalies of U850 spatially align well with the spread anomalies of Z1000.

Over the North Atlantic, the strong meridional potential vorticity (PV) gradient in the upper troposphere acts as a waveguide for Rossby waves (Wirth et al., 2018). Climatologically, the largest meridional gradient of PV at 320K (PV320K) over the Atlantic is located near 50 °N (see Figs. 5.6b, d). As Rossby waves propagate eastward and encounter regions of smaller PV gradient, e.g., from the Atlantic to Europe, they typically undergo a nonlinear decay stage. This Rossby wave breaking can be associated with the occurrence of extreme weather such as strong cyclones (Chang, 2005; Wirth and Eichhorn, 2014; Gómara et al., 2014; Tamarin-Brodsky and Harnik, 2023).

During the negative NAO phase the region of largest Rossby wave activity and breaking is shifted southward along with the strong meridional gradient of PV320K. Northern Europe then experiences less synoptic variability, whereas the opposite holds for corresponding regions southward of ∼45 °N.

To illustrate regions of anomalous Rossby wave activity, we define the time and space dependent eddy activity $\mathcal{A}$. Eddies are separated from the mean flow using time-filtering, based on a 7-day running mean, which is applied gridpoint-wise to PV320K. Eddy activity then corresponds to the time variance of PV320K relative to that 7-day background state and thereby serves as a proxy for synoptic-scale wave activity. Figure 5.6b presents deseasonalized anomalies of eddy activity averaged over weak polar vortex initializations. Negative anomalies are observed over Iceland to Scandinavia, corresponding to reduced synoptic variability. Positive anomalies appear over Canada to Greenland and southward of 50 °N over the entire Atlantic sector from the US to Europe. This is in agreement with the known equatorward jet shift that is indicated by the 200 hPa zonal wind weakening (strengthening) at the poleward (equatorward) flanks of maximum climatological meridional PV320K gradient. Overall, the anomaly patterns of eddy activity align well with ensemble variance anomalies in Z1000 (Fig. 5.5a) and U,V850 (Fig. 5.6a). The agreement supports the hypothesis that ensemble spread at subseasonal lead times is associated with synoptic variability. In order to test model sensitivity, we have computed eddy activity anomalies following weak vortex events based on ERA5 data and found good qualitative agreement (not shown).

Strong polar vortex initializations show overall similar spatial patterns of opposite sign for U,V850 ensemble mean, U850 ensemble variance anomalies, U200 ensemble mean anomalies and eddy activity anomalies (Figs. 5.6c,d). The positive NAO phase relates to a northward jet shift. For example, over Northern Europe this results in increased storminess (see

The results of this section are published in 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.

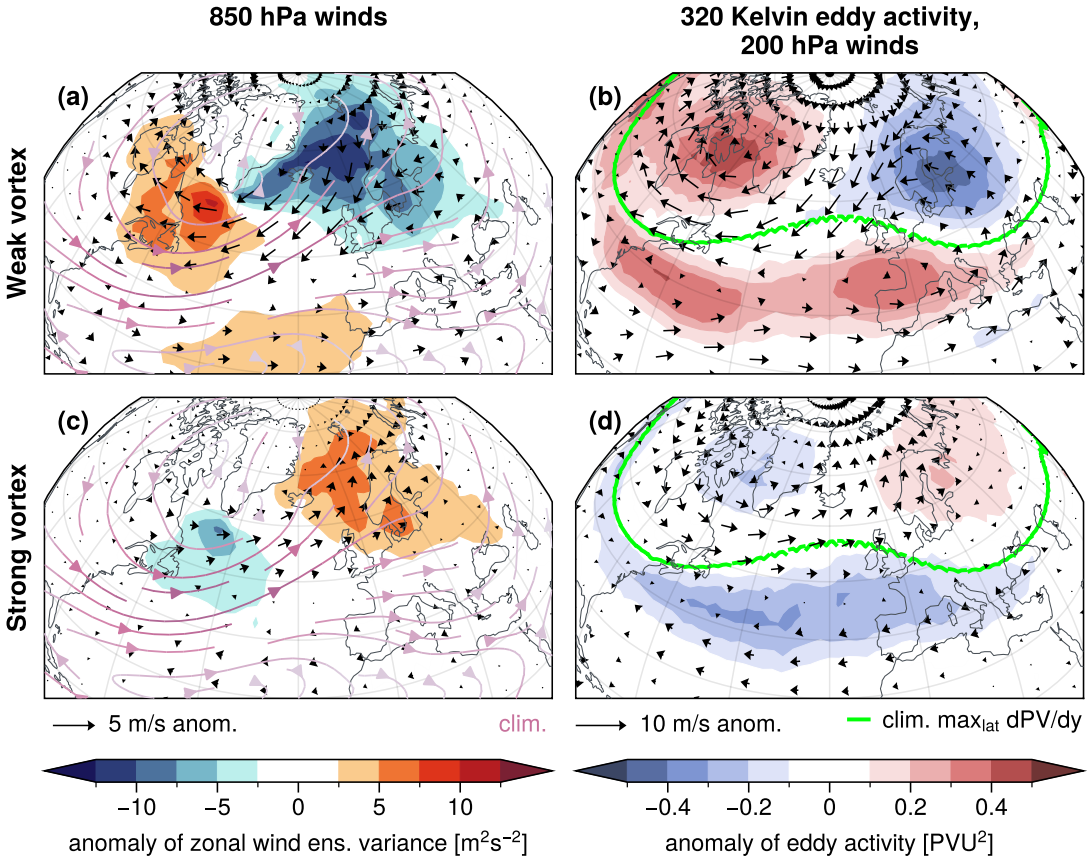

Figure 5.6: Weeks 3-5 composite-mean of weak (panels a, b; 169 cases) and strong (panels c, d; 169 cases) polar vortex forecasts. 850 hPa horizontal wind diagnostics (panels a, c) for Mid-November to End-March climatology (purple streamlines; larger values marked darker), anomalies of ensemble mean (black arrows) and anomalies of ensemble variance (shading). Upper troposphere diagnostics (panels b, d) in terms of 200 hPa horizontal wind anomalies (black arrows), anomalies of eddy activity at 320 Kelvin isentrope as a measure for synoptic-scale Rossby waves (shading; see text for details) and Mid-November to End-March climatological latitude of the maximum meridional potential vorticity gradient at 320 Kelvin isentrope.

The results of this section are published in 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.

larger U850) and increased synoptic variability (see positive eddy activity anomalies) that translates into anomalously high forecast uncertainty (see positive U850 and also Z1000 spread anomalies).

In addition to a latitudinal dipole of synoptic variability (e.g., decreased spread over Northern Europe and increased spread over the Mid-Atlantic following a weak polar vortex), a variance anomaly signal is observed over Newfoundland. Weak vortex initializations show increased spread of U850 and, to some extent, of Z1000. This signal is associated with smaller magnitudes of standardized anomalies compared to that over Northern Europe ($+0.2\sigma$ versus -0.4σ). It is spatially consistent with anomalies in eddy activity, though it cannot be explained by a shift in the jet exit region. It is correlated with the NAO (not shown) and seems to be associated with increased storm activity over the West Atlantic during the negative NAO phase (see Fig. 5.9 in Sec. 5.3, consistent with Afargan-Gerstman et al., 2024). Associated with a negative NAO, blocking episodes around Greenland and Northern Europe potentially hinder the eastward progression of storms originating from around Newfoundland (Pinto et al., 2009; Hauser et al., 2023a). Tamarin-Brodsky and Harnik (2023) argued that the increased cyclone occurrence over the West Atlantic and the increased blocking occurrence to the East are dynamically linked to cyclonic Rossby wave breaking events.

5.2.3 Discussion

Extreme stratospheric polar vortex events tend to be followed by persistent tropospheric circulation anomalies over the course of several weeks, with associated modulations of tropospheric predictability at subseasonal timescales. Previous work has focused on the ensemble mean response as a measure of such a window of forecast opportunity. Our results highlight an additional aspect of stratospheric impacts on subseasonal predictability: modifications to regional eddy activity cause robust changes in ensemble spread (forecast uncertainty) that directly translate into modified predictability. Notably, such spread-related changes in predictability are not simply co-aligned with ensemble mean-related changes in predictability and reveal modified predictability even if the ensemble mean signal is zero (such as over Denmark in Fig. 5.5).

It should be noted that in general both mean and spread modulations affect predictability, but conclusions about forecast skill depend on the evaluation metric and, ultimately, the underlying question. For example, shifts in the mean of the NAM distribution are observed following weak or strong polar vortex events. Such modulations of the ensemble mean are reflected in certain forecast skill metrics, such as the correlation skill score (CSS) or the anomaly correlation coefficient (ACC, Sigmond et al., 2013; Tripathi et al., 2015). In practice, such forecast skill metrics will be strongly sensitive to a given model's ability to truthfully represent stratosphere-troposphere coupling (Domeisen et al., 2020a). However, even in a hypothetical model that includes a perfect representation of stratosphere-troposphere coupling, and therefore a perfect representation of the ensemble mean response to stratospheric extreme events, predictability will be intrinsically limited by chaos due to

The results of this section are published in 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.

unpredictable weather fluctuations. It is the ensemble spread that measures the amount of unpredictable fluctuations. In this sense the flow may be described as being intrinsically more or less predictable in situations where the ensemble spread is smaller or larger, respectively. Our results have demonstrated that stratospheric extreme events lead to robust modulations of such intrinsic predictability.

While the CSS or ACC are strongly sensitive to the ensemble mean signal, the ensemble spread strongly affects another frequently used forecast skill metric: the mean squared error (MSE, i.e., the mean of the squared difference between the ensemble mean forecast and the observations). Notably, in a perfect model the MSE is independent of the ensemble mean, but only depends on spread. For imperfect models, MSE is additionally affected by model biases. Nevertheless, provided that the model bias contribution is small compared to the ensemble spread contribution (perhaps a reasonable assumption for the ECMWF system used here), ensemble spread anomalies are still expected to align with anomalies in MSE. Our results would then suggest weak polar vortex initializations to be associated with reduced MSE over Northern Europe and increased MSE over the Mid Atlantic. Indeed, composite-mean anomalies of Z1000 MSE for weak polar vortex initializations align closely with those of spread anomalies, although the fields are more patchy (see supplementary Fig. S2 of Spaeth et al., 2024a). Given the relatively small number of weak/ strong vortex events in the real atmosphere, more events would be needed to robustly determine changes in the forecast error. Under the assumption mentioned above, ensemble spread anomaly diagnostics facilitate statistically robust estimates of where forecast errors increase or decrease. Large ensembles can hereby, to some extent, make up for the scarcity of verification dates in the real atmosphere.

Given the substantial stratospheric impact on intrinsic subseasonal predictability as measured by the ensemble spread, other known teleconnection patterns may show similar modulations of forecast uncertainty at subseasonal time scales. Indeed, different initial states of well-known teleconnections of tropical origin, including ENSO[1], MJO[2], and QBO[3], all show robust anomalies in Z1000 ensemble spread averaged over weeks 3-5 (Fig. 5.7). Of these, ENSO shows spread anomalies of similar strength compared to those following extreme stratospheric polar vortex states, whereas MJO and QBO show much smaller anomalies. Note that some of these signatures of tropical teleconnections may be facilitated through the stratosphere (Garfinkel and Hartmann, 2007; Garfinkel et al., 2012; Lee et al., 2019a), which limits interpretations of their relative importance for tropospheric predictability. In an attempt to circumvent this issue, we have tested clustering the results in Fig. 5.7 by conditioning on moderate states of the polar vortex (see supplementary Fig. S5 of Spaeth

[1]The ENSO 3.4 index is used to characterize the ENSO state. Data was downloaded from the National Oceanic and Atmospheric Administration: https://psl.noaa.gov/gcos_wgsp/Timeseries/Nino34/.

[2]The MJO is characterized based on the real-time multivariate MJO index (RMM, Wheeler and Hendon, 2004) and the data is downloaded from the Columbia University of New York: https://iridl.ldeo.columbia.edu/SOURCES/.BoM/.MJO/.RMM/.

[3]The QBO phase is here defined based on the observational data of the zonal wind over Singapore, provided by the Freie Universitaet Berlin: https://www.geo.fu-berlin.de/met/ag/strat/produkte/qbo/qbo.dat. A positive (negative) zonal wind at 50 hPa here defines the westerly (easterly) phase of the QBO.

The results of this section are published in 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.

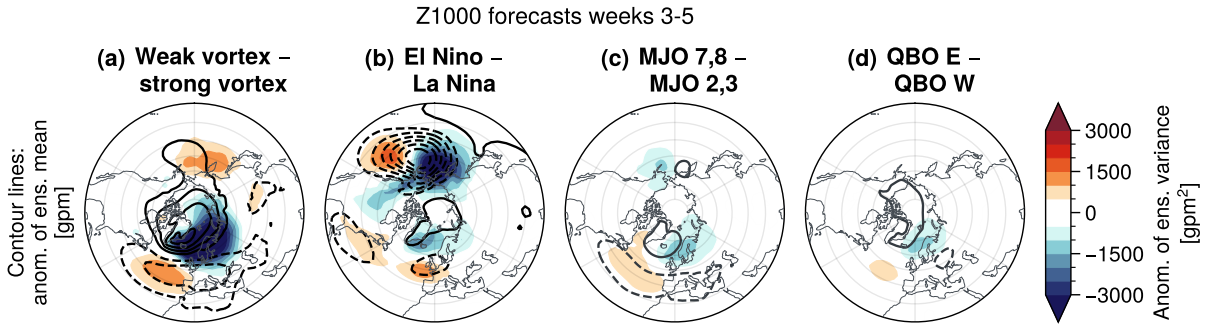

Figure 5.7: Difference of anomalies of Z1000 ensemble mean (black contours every 15 gpm, negative values dashed) and ensemble variance (shading) between forecasts initialized in (a) strong versus weak vortex conditions, (b) El Niño (ENSO 3.4 index above 1; 334 cases) versus La Niña (ENSO 3.4 index below −1; 412 cases) conditions, (c) MJO phases 7,8 (497 cases) versus phases 2,3 (434 cases), where the MJO amplitude exceeds 0.75 and (d) westerly (822 cases) versus easterly (1572 cases) QBO phase (defined by the zonal wind at 50 hPa). MJO and QBO are displayed with an additional Z1000 mean contour at ±10 gpm (in gray) due to smaller signals.

et al., 2024a). While QBO-related tropospheric signals seem to be mediated through stratospheric anomalies, the impact of ENSO appears to be relatively unaffected by the conditioning, indicating that its tropospheric pathway might be dominant.

Figure 5.7 suggests that over the Pacific, Z1000 mean and spread anomalies are dominated by ENSO variability that projects strongly onto a Pacific-North American (PNA) pattern at subseasonal time scales. Forecasts initialized during El Niño are followed by a strengthened Aleutian low, which is associated with negative spread anomalies on its western and positive spread anomalies on its eastern flank. In general, the signals over the Pacific appear to be more sensitive to ENSO than polar vortex variability. Therefore, the available S2S data do not allow fully comprehensive analyses of whether near-surface signatures of polar vortex variability is more NAO or more AO (Ambaum et al., 2001). The limited time period from 1997 to 2021 is presumably not long enough to properly sample, e.g., sudden warmings under all different ENSO influences and a longer reforecast period would be desirable. Over the Atlantic, signals associated with ENSO variability are generally weaker. During El Niño, Z1000 exhibits negative mean and positive spread anomalies over Great Britain and also upstream, at the West coast of North America, a potential hint for increased cyclogenesis there.

In comparison to polar vortex and ENSO variability, there are only small signals observed following different MJO phases. Consistent with previous literature (L'Heureux and Higgins, 2008), we find that MJO phases 7 and 8 are followed by a negative NAO. Conditioning on a moderate polar vortex leads to slight weakening of the signals (see supplementary Fig. S5b of Spaeth et al., 2024a).

Similarly, the influence of different QBO phases on Z1000 mean and spread is relatively

The results of this section are published in 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.

weak. Forecasts initialized during an easterly QBO phase result in slight negative Z1000 spread anomalies over Northern Europe and positive spread anomalies over the Mid-Atlantic. The signals reduce further when conditioning on a moderate polar vortex (supplementary Fig. S5c of Spaeth et al., 2024a), suggesting that the observed anomalies are mediated via more weak polar vortex states in the easterly QBO phase, in alignment with previous work (Holton and Tan, 1980).

5.2.4 Conclusion

Based on large sets of extended-range ensemble forecasts our results demonstrate that forecast uncertainty as measured by the ensemble spread offers added value over the ensemble mean for characterizing subseasonal predictability. Importantly, reductions in ensemble spread may indicate enhanced predictability associated with reduced forecast uncertainty in regions where the ensemble mean response is weak, i.e., in regions where the ensemble mean itself does not represent a source of predictability. Strong surface signatures of ensemble spread anomalies are found following weak or strong stratospheric polar vortex events. Consistent with NAO-related modulations of jet latitude and strength and the associated variations in synoptic storm activity, spread anomalies are not in phase with the mean anomalies, but appear where the storm track is altered, which occurs especially in the jet exit region.

In general, spread anomalies facilitate a statistically robust estimate of forecast error patterns as they can be computed without a verification dataset. Applying the developed diagnostics to remote tropical forcings revealed that extended-range forecast errors over the Atlantic are at first order affected by the long-lasting NAO signals, such as introduced by polar vortex extreme events. In contrast, forecast errors over the Pacific are dominated by PNA-variability that is strongly influenced by ENSO variability.

5.3 Impact of extratropical cyclones on geopotential height forecast uncertainty

The previous section revealed that weak and strong polar vortex states modulate the ensemble spread in subseasonal forecasts of Z1000. These signals are consistent with latitudinal shifts in the North Atlantic storm track and corresponding variations in synoptic-scale wave activity.

One prominent manifestation of synoptic-scale wave activity are extra-tropical cyclones (here referred to as “storms”; not to be confused with meso-scale wind extremes). Therefore, we anticipate that forecasts, regions, and lead times with increased storm frequency will be associated with larger ensemble spread. Moreover, ensembles in which one or several members simulate a strong storm are expected to show larger spread at the storm’s location and time.

The results of this section are under review at Weather and Climate Dynamics: Rupp, P., Spaeth, J., Afargan-Gerstman, H., Büeler, D., Sprenger, M., & Birner, T. (2024). The impact of synoptic storm likelihood on European subseasonal forecast uncertainty and their modulation by the stratosphere. *EGUsphere [preprint]*, 1–19.

The following analyses delve deeper into relationship between storm occurrence and increased ensemble spread at S2S lead times. It is crucial to note that this study does not evaluate predictions of individual storms that occur in the real atmosphere, as these are generally unpredictable beyond medium-range lead times. Instead, the focus is on understanding how processes acting as effective boundary conditions, such as the stratosphere, affect the overall frequency of storms and thereby lead to variations in ensemble spread.

The results presented in this section are part of a study led by Philip Rupp (preprint: Rupp et al., 2024). The following analyses highlight my main contributions to this study.

Storm tracks

To investigate ensemble spread in the neighborhood of storms, storm tracks in the S2S forecasts are analyzed. We use the same storm tracks as Büeler et al. (2024), with data provided by Michael Sprenger, Dominik Büeler and Hilla Afargan-Gerstman. Storms are identified based on a feature-track algorithm, detecting closed contours in mean sea level pressure (MSLP), enclosing one or several local MSLP minima (Wernli and Schwierz, 2006; Sprenger et al., 2017). A time-dependent spatial storm mask is then defined via the area enclosed by the outermost closed contour of a storm. Furthermore, the corresponding storm centre is defined as the location of minimum MSLP within this closed contour and the storm strength as the value of this MSLP minimum (Pmin). Storm tracks (paths of the corresponding storm centres) are computed based on 6-hourly data, however, only daily values are used to match the available Z1000 data in the S2S forecasts. To neglect weak and short-lived storms, only storms are considered with a total lifetime of at least 36 h and peak storm strength of Pmin < 985 hPa. In addition, only storms with a storm centre over the North Atlantic-European sector are considered ($-80\,°$E to $40\,°$E, $30\,°$N to $80\,°$N).

The following analyses are based on S2S ECMWF ensemble forecasts comprising 25 realtime forecasts with 51 members each, initialized from 1st December to 28th February 2021 and, additionally, 25×20 corresponding 10-member hindcasts (the unperturbed control member is omitted). Note that this dataset represents a subset of the forecasts used in other sections of this thesis, as the storm track data was only available for these initializations.

Storm-centred composite of ensemble spread anomalies

Fig. 5.8a presents a storm-centred composite of Z1000 anomalies. Note that section 5.2 investigated ensemble mean anomalies, whereas the anomalies here only include ensemble members with storms. In addition, the figure displays ensemble spread anomalies, computed as relative deviations from a ensemble spread climatology (based on method a. presented in Sec. 3.4). For each specific storm identified in a single ensemble member, the Z1000 anomalies of that member are included in the composite, while the corresponding ensemble spread is determined from the entire ensemble at the time and location of the storm in that single member.

The results of this section are under review at Weather and Climate Dynamics: Rupp, P., Spaeth, J., Afargan-Gerstman, H., Büeler, D., Sprenger, M., & Birner, T. (2024). The impact of synoptic storm likelihood on European subseasonal forecast uncertainty and their modulation by the stratosphere. *EGUsphere [preprint]*, 1–19.

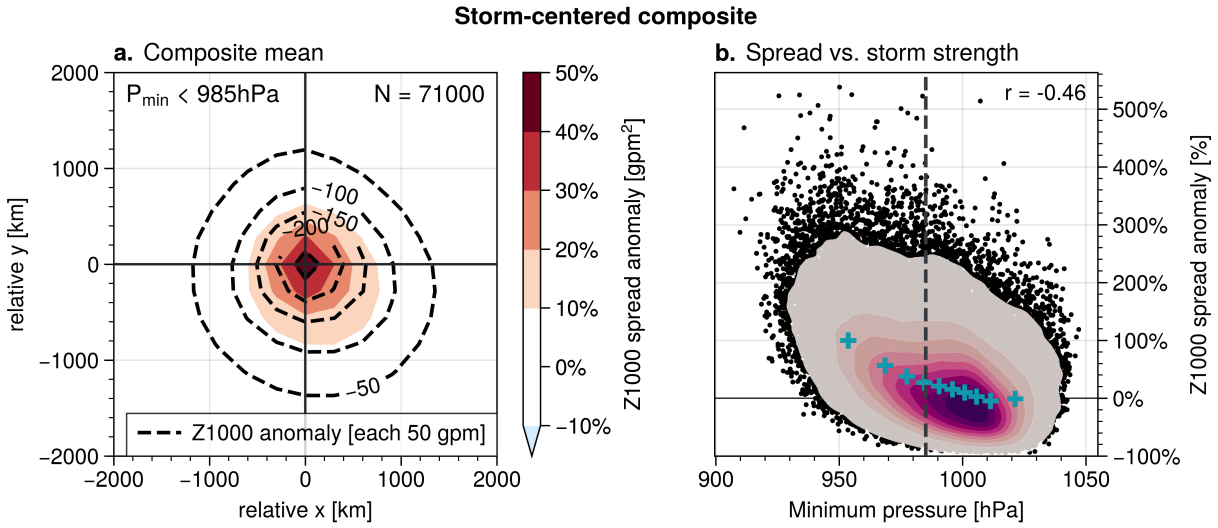

Figure 5.8: **a.** Storm-centred composite mean of 1000 hPa geopotential height (Z1000) mean anomalies (black contour lines) and ensemble spread anomalies (shading). The mean anomalies show average Z1000 anomalies of members that predict a storm, on the central lead day of the storm. The ensemble spread anomalies are computed among all members of an ensemble, where one member predicts a storm, on the storm central day. Only storms with a minimum mean sea level pressure (MSLP) below 985 hPa are included (71 000 storms). **b.** Scatter plot of Z1000 ensemble spread anomalies (y-axis) versus minimum MSLP of the storm (x-axis). Black dots represent individual storms (pearson correlation coefficient $r = -0.46$) and shading shows the 2D-histogram. Blue crosses indicate the mean values of MSLP and Z1000 spread for 10 storm groups, given by 10 MSLP quantiles, i.e., these groups are equally populated.

The results of this section are under review at Weather and Climate Dynamics: Rupp, P., Spaeth, J., Afargan-Gerstman, H., Büeler, D., Sprenger, M., & Birner, T. (2024). The impact of synoptic storm likelihood on European subseasonal forecast uncertainty and their modulation by the stratosphere. *EGUsphere [preprint]*, 1–19.

The storms are characterized by negative Z1000 anomalies, with values peaking around −250 gpm in the storm centre. Z1000 ensemble spread in the storm centre increases by up to 50% relative to climatological ensemble spread, supporting the initial hypothesis that storms in individual ensemble members can substantially increase the ensemble spread computed across all members.

Fig. 5.8b shows that particularly strong storms, characterized by low MSLP values, are associated with increased ensemble spread. For instance, the strongest 10% of the storms in terms of Pmin are linked to an increase in ensemble spread by approximately 100%. In contrast, weak storms, with Pmin>1000 hPa, are not associated with anomalous spread.

Spatial distribution of storms under weak vortex conditions

To link the storm-induced increase in ensemble spread to the anomalous ensemble spread patterns following weak polar vortex states (see Sec. 5.2), we analyze the spatial distribution of storms under these conditions. Weak vortex initializations are classified as the 20% of forecasts with the lowest initial U_{60}^{10} .

Figure 5.9 shows 2D-histograms of the spatial distribution of storms. Storm frequency is computed by dividing the North Atlantic-European domain (−80 °E to 40 °E, 30 °N to 82.5 °N) into 7.5° × 7.5° horizontal bins and counting the number of storm centres within each bin. The frequency is then normalized by the total number of storms. Normalized frequencies are computed for the locations of full tracks (single storms count multiple times, once per day, following their track), genesis locations (based on the first day of the storm track), and cyclolysis locations (based on the last day of the storm track).

Climatologically, North Atlantic-European storm tracks are primarily distributed from eastern North America, through Southern Greenland and Iceland, to the Norwegian Sea (Fig. 5.9a). The main regions of cyclogenesis are observed over eastern North America and the southern tip of Greenland, whereas cyclolysis occurs predominantly over the southern tip of Greenland and the Norwegian Sea.

Forecasts initialized during weak polar vortex states show a southward shift of the North Atlantic storm track, as evidenced by a latitudinal shift of track, cyclogenesis and cyclolysis locations (Figs. 5.9b,d,f). Consistent with these changes, the median storm track latitude shifts southward (Fig. 5.9b), although this shift is only a few degrees. Following weak polar vortex states, fewer storms reach the Norwegian Sea, while more storms reach the mainland of southern and central Europe.

Discussion

We identified two critical aspects of the relationship between weak polar vortex states, ensemble spread, and storms. The previous section showed that weak polar vortex states lead to reduced Z1000 ensemble spread over northern Europe and increased Z1000 ensemble spread over southern Europe. This section demonstrated that these changes are consistent with a southward shift of the North Atlantic storm track. Importantly, the occurrence of

The results of this section are under review at Weather and Climate Dynamics: Rupp, P., Spaeth, J., Afargan-Gerstman, H., Büeler, D., Sprenger, M., & Birner, T. (2024). The impact of synoptic storm likelihood on European subseasonal forecast uncertainty and their modulation by the stratosphere. *EGUsphere [preprint]*, 1–19.

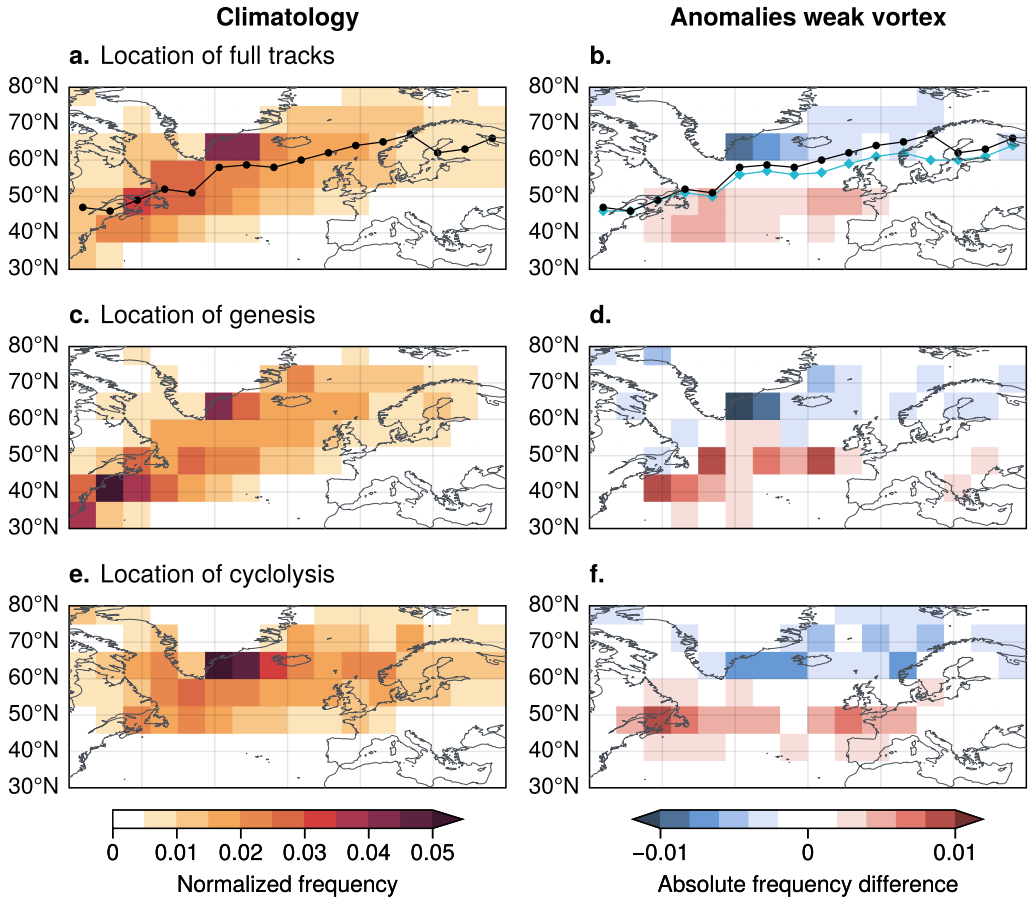

Figure 5.9: 2D-histograms of the climatological spatial distribution of **a.** storm tracks, **c.** storm genesis, **c.** and cyclolysis (final location of storm track). The frequencies are normalized such that the sum over all bins equals one (see text for details). **b., d., f.:** Corresponding 2D-histograms of anomalies in the spatial distribution in weak vortex forecasts.

The results of this section are under review at Weather and Climate Dynamics: Rupp, P., Spaeth, J., Afargan-Gerstman, H., Büeler, D., Sprenger, M., & Birner, T. (2024). The impact of synoptic storm likelihood on European subseasonal forecast uncertainty and their modulation by the stratosphere. *EGUsphere [preprint]*, 1–19.

storms in individual ensemble members is linked to increased spread in the ensemble. Note that this ensemble spread is computed across all 10 members, including the member with the storm and the 9 other members. As a sanity check, we tested computing the spread only across the 9 members that do not predict the storm and found near-zero Z1000 mean and spread anomalies (not shown), which provides further confidence that the spread increase is attributable to the "storm-member".

The relative ensemble spread increase caused by a single storm depends on the ensemble size. For instance, a single storm leads to a larger relative increase in an 11-member ensemble compared to a 51-member ensemble. However, in larger ensembles, it is likely that multiple members simulate storms simultaneously, making the overall contribution of storms to the full ensemble spread insensitive to the ensemble size (see Rupp et al., 2024, for a more detailed discussion).

Finally, it is important to note that the anomalies in spatial storm distributions are not perfectly aligned with the ensemble spread anomalies following weak polar vortex states. This suggests that storms are just one factor contributing to the reduced spread. Additionally, anomalies in Rossby wave activity – unrelated to well-defined storms but associated with the general occurrence of troughs and ridges – likely contribute to reduced synoptic-scale wave activity. Together, these factors may account for the observed ensemble spread anomalies following weak polar vortex states.

5.4 Flow-dependence of ensemble spread of subseasonal forecasts explored via North Atlantic-European weather regimes

Based on the Z1000 ensemble spread anomalies identified in section 5.2 and their link to the occurrence of storms discussed in section 5.3, this section explores three additional aspects of subseasonal forecast uncertainty. This work is published in Spaeth et al. (2024b).

1. Generalization of ensemble spread anomalies to large-scale flow patterns

Section 5.2 demonstrated that weak polar vortex states are followed by anomalies in Z1000 ensemble spread. Given that these anomalies were linked to variations in synoptic-scale wave activity, including anomalies in storm frequency (see section 5.3), we anticipate that these spread anomalies are not exclusively linked to polar vortex extreme events. Specifically, any teleconnection projecting onto the North Atlantic Oscillation is expected to be followed by similar spread anomalies at S2S timescales. Moreover, teleconnections that project onto a different large-scale flow pattern may be linked to different spread anomalies, depending on the associated variations of synoptic-scale wave activity.

To investigate the relation between the large-scale flow and ensemble spread anomalies more systematically, we employ North Atlantic-European weather regimes. Weather regimes have successfully been employed by previous studies to analyze flow-dependence of forecast skill (Ferranti et al., 2015; Matsueda and Palmer, 2018; Büeler et al., 2021) as well as the

tropospheric response to stratospheric circulation anomalies (Charlton-Perez et al., 2018; Beerli and Grams, 2019; Lee et al., 2019b; Domeisen et al., 2020b; Roberts et al., 2023).

Here, we use the weather regimes introduced by Grams et al. (2017) and condition on each of these regimes to analyze the associated ensemble spread anomalies. These results may guide future analyses of teleconnections and their impact on ensemble spread, by decomposing the single pathway *teleconnection* → *spread* into two pathways: *teleconnection* → *weather regime* and *weather regime* → *spread*.

2. Diagnosing ensemble spread anomalies of near-surface temperature

The previous two sections focused on forecast uncertainty in Z1000 forecasts. To investigate whether other atmospheric variables follow similar patterns of ensemble spread anomalies, this section also includes analyses of 2 meter temperature (T2m). Interestingly, the results will show that T2m ensemble spread in subseasonal forecasts is governed by a different process compared to Z1000 ensemble spread, which results in different spatial patterns of spread anomalies.

3. Uncertainty of the large-scale flow regime

Separate consideration of *teleconnection* → *weather regime* and *weather regime* → *spread* motivates the question whether polar vortex extremes also modulate the uncertainty of weather regimes. However, the weather regime is characterized by a categorical variable, i.e., each day is assigned one out of eight weather regimes. Therefore, it is not possible to compute ensemble variance as a metric for forecast uncertainty, as for T2m and Z1000. Instead, we propose information entropy as a measure for weather regime uncertainty. Specifically, we compute a climatological regime entropy and show that polar vortex extremes are followed by changes in entropy relative to this climatology.

Before presenting the results, the employed S2S forecasts, the weather regimes and the regime entropy are introduced in more detail.

S2S forecast data

We analyze 1197 S2S ensemble forecasts provided by ECMWF, comprising 57 51-member realtime forecasts from winters 2019/20 and 2020/21 and 1140 11-member hindcasts from winters 1999/2000 to 2019/20. Note that this dataset represents a subset of the forecasts used in different sections of this thesis, as weather regime data was only available for these forecasts.

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 °N−90 °N) adapted to ERA5 (Hauser et al., 2023b) are used in this study (data provided by Christian M. Grams).

The weather regime 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 reanalyses (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 in ECMWF S2S forecasts (model cycles CY46R1 and CY47R1) are used as in Osman et al. (2023), with data provided by Marisol Osman. 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.

Definition of regime entropy to quantify forecast uncertainty of weather regimes

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 information entropy H (hereafter also referred to as regime entropy). 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, i.e., $p_{wr} = \frac{n_{wr}}{N}$. The regime entropy (based on information theory, see Shannon, 1948; Cover and Thomas, 2005, September) 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:

$$H = \frac{-\sum_{wr} p_{wr} \cdot \ln(p_{wr})}{\ln(8)}. \tag{5.2}$$

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

The results of this section are published in 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(14).

members (realtime) can split exactly equally among the 8 regimes, e.g., $H \approx 0.82$ given 11 members and a uniform background distribution (empirically derived using synthetic data).

5.4.1 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 Z1000 ensemble variance are computed and averaged over 7 to 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 5.10 shows that Z1000 ensemble variance anomalies are largest for the two regimes that are closely linked to NAO variability (cf. Ferranti et al., 2015; Grams et al., 2017; Beerli and Grams, 2019): During the Greenland Blocking regime (associated with NAO− conditions) Z1000 ensemble spread is reduced by up to 1800 gpm^2 near 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 1000 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 Sections 5.2 and 5.3, showing that occurrence of individual storms contributes 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 5.11 reveals that T2m ensemble variance anomaly patterns are qualitatively different compared to Z1000, suggesting that they involve different mechanisms. In general, T2m

The results of this section are published in 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(14).

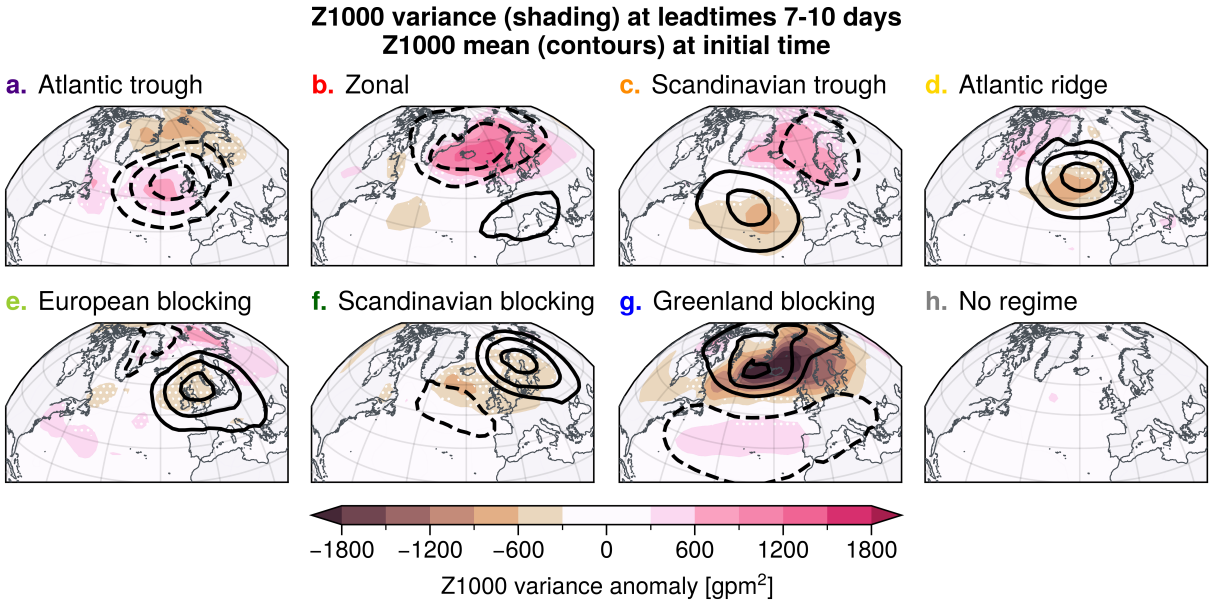

Figure 5.10: Composite mean of Z1000 ensemble variance anomalies (color-shaded), grouped by the weather regime at initialization time, averaged over lead times 7-10 days. Z1000 ensemble mean anomalies at initialization time denoted by black contours (50 gpm spacing, negative values dashed). White stippling indicates regions where color-shaded Z1000 ensemble variance anomalies are not statistically robust (based on the 95% confidence level using bootstrapping with 2000 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 results of this section are published in 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(14).

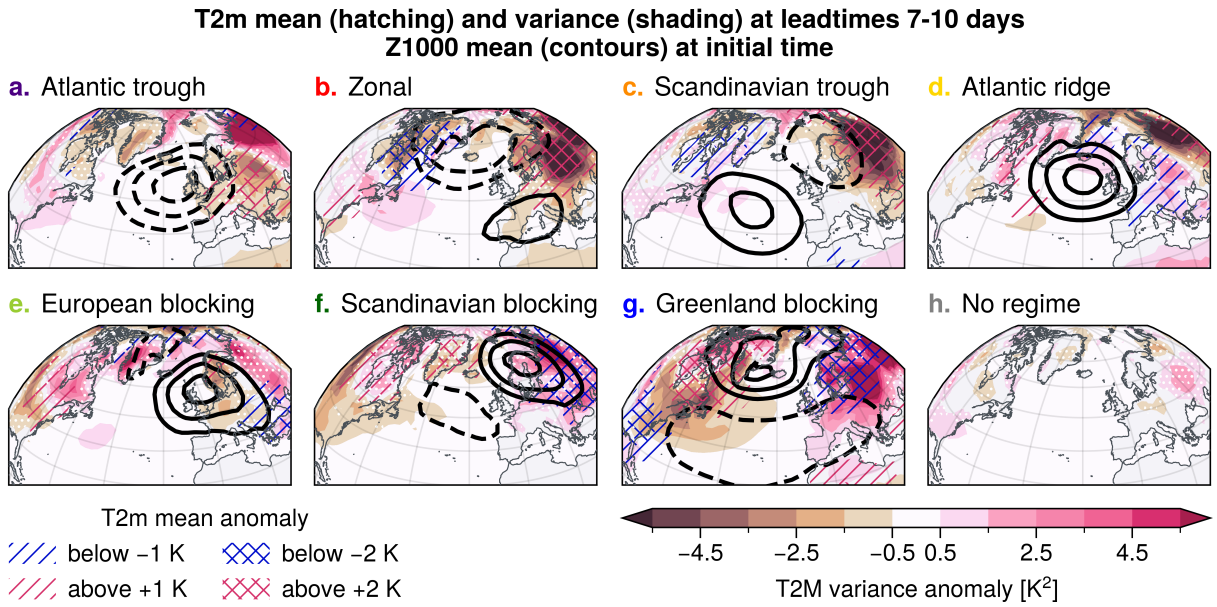

Figure 5.11: Same as Fig. 5.10, but with color-shading indicating anomalies of 2 meter temperature (T2m) variance. White stippling indicates regions where color-shaded T2m ensemble variance anomalies are not statistically robust (based on the 95% confidence level using bootstrapping with 2000 samples). 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 Fig. 5.10.

The results of this section are published in 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(14).

variance is smaller over the ocean than over land (see supplementary Fig. S2 of Spaeth et al., 2024b, 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 $-3\,\mathrm{K}$) and increased T2m variance (up to $7\,\mathrm{K}^2$, 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.

Our results are consistent with the hypothesis that T2m ensemble variance anomalies are predominantly driven by advection, i.e., 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 Fig. S3 of Osman et al., 2023). Over Europe, Atlantic Ridge and Atlantic Trough are associated with opposite flow directions (easterly versus westerly) and hence opposite temperature anomalies (cold versus mild), which can additionally contribute to large T2m ensemble spread following Greenland blocking (see supplementary text S3 for a comparison of Greenland blocking transitions to Atlantic Ridge versus 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 to 10 days lead time that were presented in figures 5.10 and 5.11. However, as we will show below, a substantial part is resulting from the inherent "intra-regime variance", i.e., the variance across a single given weather regime.

Figure 5.12 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 5.10 and 5.11, 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 5.10, 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

The results of this section are published in 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(14).

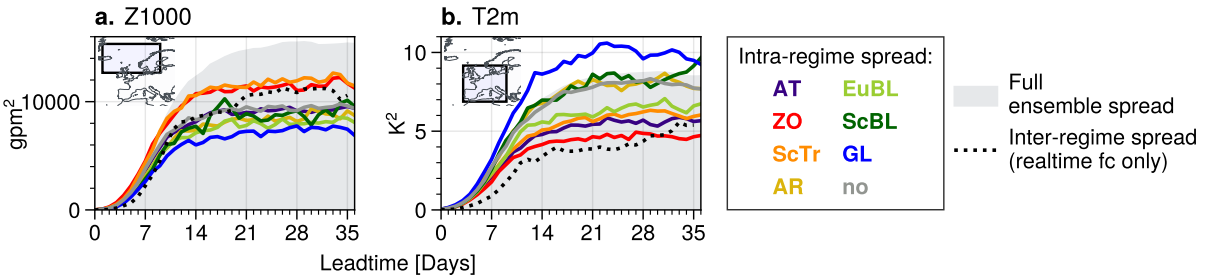

Figure 5.12: Evolution of intra-regime ensemble variance of **a.** 1000 hPa geopotential height over northern Europe and **b.** 2-meter temperature over central Europe (55-75°N/ −20-20°E and 35-60°N/ −10-20°E, respectively, indicated in map insets). For each lead time and ensemble forecast, the intra-regime 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, i.e., the full ensemble variance. Dotted black lines denote the average inter-regime spread (realtime forecasts only), i.e., the variance between the regime-grouped averages within an ensemble.

T2m ensemble variance over Europe, which on average is twice as large during Greenland blocking than during Zonal. In fact, T2m intra-regime variance during Greenland Blocking is even larger (saturating at $10\,\mathrm{K}^2$) than the average total ensemble spread (saturating at $8\,\mathrm{K}^2$).

This regime-dependence of forecast uncertainty is in good agreement with Figures 5.10 and 5.11, 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.

5.4.2 Uncertainty of weather regimes under different stratospheric initial conditions

In section 5.4.1 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. 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.

The results of this section are published in 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(14).

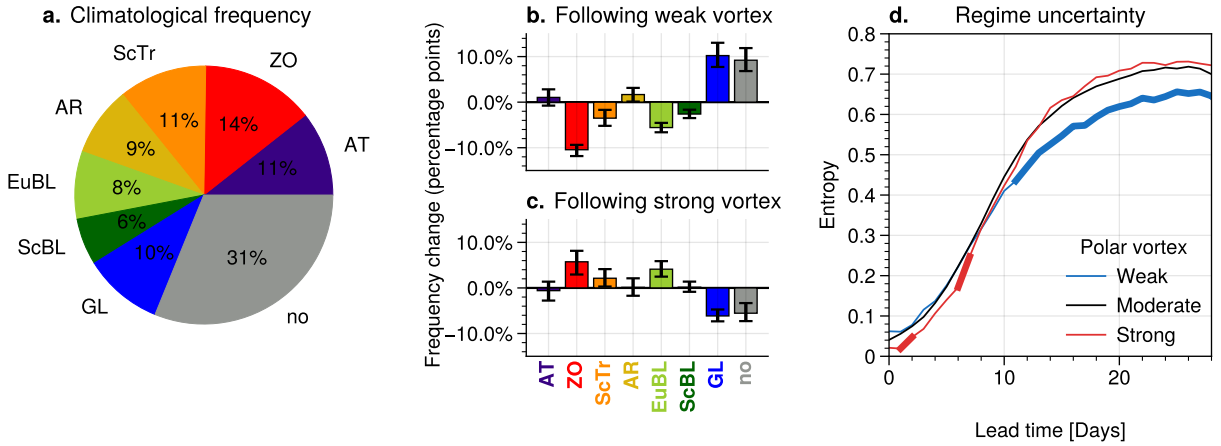

Figure 5.13: **a.** Climatological frequency of weather regimes in ECMWF S2S forecasts averaged over lead times 14-27 days. **a/b.** Percentage point changes in predicted regime frequencies following **b.** weak and **c.** strong stratospheric polar vortex conditions. Error bars indicate 95% confidence 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.

Figure 5.13 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[4]. Here, weak, moderate and strong polar vortex conditions are classified based on U_{60}^{10} in ERA5 (weak: $U_{60}^{10} < 0\,\mathrm{ms}^{-1}$, resulting in 86 forecasts; strong: $U_{60}^{10} > 51.82, \mathrm{ms}^{-1}$, ensuring the same number of forecasts as for weak; moderate vortex: 1025 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 Fig. 6 in Beerli and 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".

[4]The final computation of the regime frequencies and the regime entropy was carried out by Marisol Osman.

The results of this section are published in 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(14).

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. Figure 5.13d 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.

5.4.3 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 at the beginning of this section.

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 Figs. 5.10, 5.11) or based on the regime at each day, separately (see Fig. 5.12). 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.

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 Fig. 5.13). From this perspective, weak polar vortex episodes,

The results of this section are published in 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(14).

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, i.e., 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 et al. (2020a) 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 of Spaeth et al., 2024b). 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 of Spaeth et al. (2024b) briefly discusses how the February 2018 SSW (e.g., Karpechko et al., 2018; Kautz et al., 2020; González-Alemán et al., 2022) was associated with reduced regime uncertainty, reduced Z1000 uncertainty and increased T2m uncertainty in forecasts.

The results of this section are published in 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(14).

Chapter 6

Conclusions, Discussion, Outlook

6.1 Answers to research questions

(1) How variable is the role of tropospheric wave activity in the formation of SSWs?

Approximately one third of displacement-type SSWs and one half of split-type SSWs are attributable to preceding wave events of tropospheric origin. That is, approximately one half to two thirds of SSWs are not attributable to anomalously strong preceding tropospheric wave activity.

Planetary waves and their non-linear interaction with the zonal-mean flow play a critical role in the formation of SSWs. However, quantifying the influence of anomalous tropospheric planetary wave activity on SSWs using reanalysis data is challenging due to the scarcity of SSW events and the highly variable nature of wave activity. While climate and idealized models allow for a larger sample size of events, their inherent simplifications – such as limited spatial resolution – introduce uncertainties regarding the fidelity of simulated atmospheric processes.

In this study, 6101 SWs were identified in ECMWF subseasonal-to-seasonal ensemble forecasts, leveraging a state-of-the-art weather forecast model with a large sample size of extreme events. This is referred to as an UNSEEN (UNprecedented Simulated Extremes using ENsembles) approach. In addition to SSWs, wave events were defined for zonal wavenumbers 1 and 2, based on standardized anomalies of eddy heat flux. Conditional probabilities were computed to quantify the joint occurrence of SSWs and preceding wave events.

The results revealed that out of 2808 displacement SSWs, approximately one-third can be attributed to preceding anomalous wavenumber 1 wave activity near the surface. Likewise, out of 933 split SSWs, about half are attributable to preceding anomalous wavenumber 2 wave activity near the surface.

Increased planetary wave activity near the surface often manifests in the formation of high pressure systems over the Ural mountains, here referred to as Ural blockings. 6608 Ural

blockings were identified in the forecasts, which showed that the associated increase in vertical wave propagation approximately doubles the subsequent SSW probability. Hence, skillful prediction of Ural blockings may enhance stratospheric predictability.

(2) How, and to what extent, can tropospheric circulation extremes be attributed to SSWs?

To attribute tropospheric circulation extremes to preceding SSWs, the fraction of attributable risk framework was employed, which is suitable for such analysis but requires large statistics. Leveraging a large dataset of state-of-the-art S2S ensemble forecasts, it was found that approximately every fourth large-scale tropospheric circulation extreme in winter may be attributable to preceding SSWs.

While it is well-established that SSWs are followed by tropospheric *mean* anomalies that project onto a negative phase of the Arctic Oscillation (AO), quantifying the modulation of AO *extremes* remains challenging due to the limited observational record.

Using an UNSEEN approach, 6101 SSWs were identified in ECMWF and 2716 in UKMO S2S ensemble forecasts. Conditional probabilities were computed to quantify the joint occurrence of SSWs and subsequent AO extreme events.

The results showed that SSWs increase the subsequent risk for negative AO extremes (defined as deviations beyond 3 standard deviations) by approximately $40\% - 80\%$. Likewise, the risk for positive AO extremes reduces. Furthermore, AO extremes that follow a SSW can be attributed to that SSW in about $50\% - 60\%$ of cases. Moreover, approximately $\frac{1}{3}$ of all AO extremes during the extended winter season are attributable to preceding SSWs.

(3) How do stratospheric circulation extremes affect tropospheric predictability?

This thesis revealed that polar vortex extremes affect forecast uncertainty at S2S timescales, highlighting an important aspect of predictability associated with stratosphere-troposphere coupling, which extends beyond the previously recognized shifts in mean weather conditions.

Previous research has established that weak and strong polar vortex events enhance predictability by inducing mean shifts in distributions, such as related to the North Atlantic Oscillation. This thesis extends this understanding by demonstrating that polar vortex events also influence the variance of multiple distributions, quantified through ensemble variance (or "ensemble spread"). Ensemble spread is an important aspect of predictability as it quantifies forecast uncertainty and corresponds to the expected value of the mean squared error of the ensemble mean forecast.

In the stratosphere, SSWs lead to a decrease in ensemble spread of the polar vortex strength, measured by the zonal-mean zonal wind at 60 °N, 10 hPa. The decrease is ex-

plained by suppressed upward wave activity as a result of easterly zonal-mean winds in the lower stratosphere following SSWs.

Furthermore, the results show that weak and strong polar vortex events lead to modulated ensemble spread in the troposphere as well, in terms of 1000 hPa geopotential height forecasts. Anomalous ensemble spread arises from latitudinal shifts of the North Atlantic storm track and associated changes in synoptic-scale wave activity. Specifically, weak polar vortex events are followed by reduced spread over northern Europe by up to 25% relative to climatological spread, whereas spread slightly increases over southern Europe.

To assess the relation between synoptic-scale variability and ensemble spread more in detail, a storm-centered composite is computed, based on nearly one hundred thousand North Atlantic cyclones in the S2S forecasts at lead times of 20-46 days. The results show that at the location where a single ensemble member predicts a storm, spread across the entire ensemble is increased by more than 40% relative to climatological spread.

Therefore, anomalies in ensemble spread are not only expected following polar vortex events, but more generally, in response to the large-scale flow, and associated variations in synoptic variability. To highlight this aspect, the large-scale flow was characterized using a set of eight established North Atlantic-European weather regimes, and their relation to ensemble mean and spread anomalies at 2-6 weeks lead time was studied. In particular, the tropospheric response to weak and strong polar vortex events can be explained by their impact on Greenland blocking (NAO−) and Zonal (NAO+) frequency, respectively.

Moreover, regime entropy was introduced to quantify whether not only uncertainty of near-surface local weather, but also of the large-scale weather regime is modulated by polar vortex events. Consistent with the long-lasting increase in Greenland blocking frequency, weak polar vortex events are followed by reduced entropy compared to climatology, indicating reduced uncertainty/ enhanced predictability. In contrast, entropy increases following strong polar vortex events, which primarily results from a decrease in “no regime” frequency. No regime climatologically occurs most often, hence, a decrease in frequency tends to equalize the frequencies among the eight weather regimes, which increases entropy.

6.2 Limitations

This section discusses the major limitations of the results presented in this thesis. Specifically, two key areas are addressed: First, Section 6.2.1 examines various factors that may support or challenge the applicability of conclusions drawn from forecast model outputs to the real atmosphere. Second, Section 6.2.2 discusses different sources of sampling uncertainty that may be inherent to the analyses.

6.2.1 Perfect model assumption

This thesis is largely based on subseasonal-to-seasonal forecasts provided by the ECMWF. These forecasts not only serve to evaluate forecast skill, but also allow for deeper insights

into atmospheric dynamics and the intrinsic limits of predictability. However, the insights gained from these analyses are valid for the real atmosphere only if the model sufficiently represents the complexity of the relevant aspects of atmospheric physics. Therefore, the interpretation of the results is based on a perfect model assumption. The following paragraphs discuss where and why a perfect model is implicitly assumed, and highlight aspects that may support or challenge this assumption.

Model representation of teleconnections

The analyses of UNSEEN SSWs, SPVs, Ural blockings, planetary wave events and AO extremes in the S2S forecasts (chapter 4) implicitly rely on a perfect model assumption. The assumption posits that individual ensemble members simulate events that could have also occurred in the real atmosphere. This approach facilitates far larger sample sizes of extreme events compared to observations, which substantially reduces sampling uncertainty. For instance, ERA5 covers 33 SSW events between 1960 and 2023, while the S2S ECMWF forecasts used in this study provide 6101 SSWs, with the potential for even more by including more initialization dates or models.

However, this advantage comes at the cost of increased model uncertainty. Specifically, verifying whether the model accurately simulates extreme events is challenging, as these events are rare in observations and therefore subject to sampling errors (Oehrlein et al., 2021).

In general, stratosphere-troposphere coupling in the ECMWF model seems to be consistent with reanalyses (Fig. 4.8; Schwartz and Garfinkel, 2020; Domeisen et al., 2020a), although various S2S models underestimate the downward coupling, in particular after strong polar vortex events (Garfinkel et al., 2024). Notably, high-top models (such as ECMWF) generally exhibit fewer biases in the stratosphere (Lawrence et al., 2022) and are therefore better suited for studying UNSEEN polar vortex extreme events.

Some other teleconnections tend to be less accurately represented in models. Lawrence et al. (2022) showed that the QBO amplitude unrealistically decays over lead time in S2S models, and their teleconnection to the polar vortex (Holton and Tan, 1980) is underestimated. Furthermore, Garfinkel et al. (2022) found that the extratropical response to ENSO is generally too weak in S2S models, primarily due to underestimation of Rossby wave amplitudes, and errors in the North Pacific jet stream, which affects the propagation of Rossby waves into the extratropics. Similarly, while models have made progress in predicting the MJO (Stan et al., 2022), the extratropical response to MJO events is still underestimated (Garfinkel et al., 2022).

Given these shortcomings, the UNSEEN approach may be less justified when applied to QBO, ENSO and MJO teleconnections compared to stratosphere-troposphere coupling. Nevertheless, future advancements in our physical understanding and computational capabilities are expected to improve models. Such improvements will allow the UNSEEN approach to produce results with reduced model uncertainty compared to the models available today.

The underestimation of various teleconnections indicates that today's models may not fully exploit the potential predictability associated with these phenomena. Furthermore, this underestimation can contribute to the so-called signal-to-noise paradox, a phenomenon primarily observed in seasonal forecasts (Scaife and Smith, 2018). The paradox arises when the ensemble mean exhibits a higher correlation with observations than with individual ensemble members. This appears counterintuitive because, in a perfect model where an ensemble member is statistically indistinguishable from observations, these correlations should be equal. Given that models are imperfect, one might expect higher correlations of the ensemble mean with individual ensemble members (i.e., within the model world) than with the real world. Surprisingly, in several applications, the opposite is observed, and the reasons for this are still debated (Weisheimer et al., 2024). Appendix B demonstrates, using a simple toy model, that underestimation of the strength of teleconnections can lead to the signal-to-noise paradox. Therefore, the underestimation of ENSO, QBO and MJO teleconnections in S2S models might contribute to a signal-to-noise paradox in S2S forecasts. The complexity of this issue is exacerbated by the findings in Chapter 5, which reveal correlations between the ensemble mean (signal) and ensemble spread (noise). Future work may address how this correlation affects the signal-to-noise ratio and how it manifests in the signal-to-noise paradox.

Model representation of uncertainty

The analyses of forecast uncertainty, primarily quantified through ensemble variance (chapter 5), also rely on a perfect model assumption. In addition to the assumption that ensemble members each represent possible real atmospheric evolutions, it is assumed that variations in forecast uncertainty reflect actual variations in the predictability of the atmosphere.

In general, forecast uncertainty arises from initial condition and model uncertainty. An accurate representation of initial condition uncertainty is arguably most critical at shorter lead times, i.e., at short- and medium-range time scales. Nonetheless, the following briefly discusses some key aspects of initial condition uncertainty to highlight how reliability at such shorter timescales may differ from reliability at subseasonal-to-seasonal timescales.

Conceptually, uncertain initial conditions can be thought of as a high-dimensional probability density function (PDF) in the atmospheric phase space. ECMWF estimates this PDF using ensemble data assimilation (EDA), which generates an ensemble of initial conditions based on a prior forecasts and new observations (Leutbecher and Palmer, 2008; Rodwell et al., 2018). For each ensemble member, a deterministic forecast is produced, incorporating model uncertainty through stochastic paramaterizations. The resulting ensemble forecast represents a finite sample of the PDF of the future atmospheric state, considering both initial condition and model uncertainties.

In practice, EDA (representing initial condition uncertainty) and stochastic parameterizations (representing model uncertainty) still tend to generate too little ensemble spread (Palmer et al., 2006; Leutbecher and Palmer, 2008), indicated by the average spread being smaller than the average error (see section 3.3). To address this, ECMWF adds additional

perturbations to the initial conditions that ensure maximum divergence of ensemble members within the first few days (Leutbecher and Palmer, 2008). This approach helps account for missing uncertainties but also highlights the imperfections of the forecasts.

At short- and medium-range timescales, it is critical that the uncertainty reflects the characteristic rapid growth of forecast errors, which is associated with the decay of traceable predictability. In contrast, errors tend to be roughly saturated at subseasonal-to-seasonal time scales. Importantly, the amplitude at which spread and errors saturate depends more on an accurate atmospheric energy spectrum (i.e., realistic variability at different spatial scales) rather than on initial condition and model uncertainty.

At subseasonal-to-seasonal timescales, deviations from climatological spread can arise for two reasons. First, anomalous ensemble spread can result from sampling errors, especially when the ensemble size is small. However, these sampling errors tend to compensate over many cases (except for situations where spread is very large or small, especially with a small ensemble size; see Fig. 3.1f). Second, anomalous ensemble spread can result from quasi-external forcings, related to teleconnections, that provide climatic predictability at these timescales.

Importantly, Figures 3.2 and 3.3 showed that such anomalies in ensemble spread at subseasonal-to-seasonal align well with anomalies in error, suggesting that forecasts can reflect flow-dependent variations in intrinsic atmospheric predictability. For practical applications, the most significant of these situations are likely those with reduced ensemble spread, indicating enhanced predictability and a "window of opportunity" for subseasonal-to-seasonal prediction (Mariotti et al., 2020).

While the relatively good spread-error relation provides some evidence that many relevant sources of subseasonal predictability are accurately represented in the model, future analyses will need to verify whether this relationship holds when forecasts are subset according to specific teleconnections. Given the aforementioned underestimation of the strength of some teleconnections (such as ENSO, QBO, MJO) it is plausible that their influence on intrinsic atmospheric predictability is underestimated as well.

6.2.2 Sampling errors

Finite ensemble size

The ensemble sizes of the employed forecasts are 11 for the hindcasts and 51 for the realtime forecasts[1]. While analyzing extreme events, such as SSW, requires a large overall sample size, this sample may comprise either few forecasts with a large ensemble size or many forecasts with a small ensemble size. Therefore, the analysis of UNSEEN events is not restricted to a minimum ensemble size.

Similarly, sample variance provides an *unbiased* estimate for ensemble variance. Consequently, in the limit of many cases, the average ensemble spread does not depend on the

[1]Note that the latest cycle of the ECMWF model now uses 101 members for realtime forecasts.

underlying ensemble size. However, as discussed in section 3.3, the amplitude of both large negative and large positive spread anomalies is overestimated when the ensemble size is small. In conclusion, while the relatively small ensemble size in the hindcasts ($N = 11$) may introduce minor quantitative mismatches between spread and error, it does not generally limit the analysis of qualitative structures of ensemble spread anomalies.

Low-frequency climate variability

The hindcast period of ECMWF forecasts spans 20 years. While this extensive dataset provides excellent sampling of high-frequency atmospheric variability, it offers limited sampling of low-frequency climate variability, such as those related to ENSO and QBO. Notably, when characteristics of high-frequency variability are modulated by underlying low-frequency variability, this poses additional challenges. For illustration, the ENSO influence on SSWs is considered. Depending on their mutual statistical and physical links, SSW composites might be compromised by insufficient sampling of ENSO, such as linked to one of the following considerations.

1. Assume that SSWs were unrelated to ENSO. When computing SSW composites, residuals of ENSO variability may still remain due to insufficient sampling of different ENSO phases. In regions where the atmospheric response to ENSO is much larger than to SSWs, the ENSO-induced signal might outweigh the SSW signal, potentially leading to the false attribution of the signal to SSWs instead of ENSO.

2. Assume that the frequency of SSWs depends on the ENSO phase, although the dynamics of a specific SSW does not depend on whether it is induced by ENSO or not. Due to the limited hindcast period, the number of SSWs under El Niño conditions relative to all SSW is likely different in forecasts (approximately from 1997 to 2021) and reanalyses (approximately from 1950 on, depending on the dataset). As a result, the response to SSWs is influenced by ENSO in both S2S forecasts and ERA5, but the magnitude of the relative influence is likely not the same, which makes a comparison between S2S and ERA5 SSWs challenging.

3. This issue becomes even more severe if the dynamics of SSWs were inherently different when forced by an El Niño event compared to SSWs occurring during neutral ENSO. Given a different relative number of El Niño-induced SSWs in S2S forecasts and ERA5, the SSW composites in the two datasets would then suggest differences in SSW dynamics between S2S and ERA5. It would be difficult to attribute these differences to either model deficiencies or ENSO influence.

4. Finally, even if ENSO is sampled sufficiently well, resulting in a similar distribution of ENSO phases during SSWs in forecasts and ERA5, non-additive influences of other teleconnections, such as the QBO, could complicate SSW analyses. If SSWs during El Niño appeared differently under easterly versus westerly QBO conditions, it would not be enough to sample different ENSO and QBO phases independently over the 20-year-period. All ENSO-QBO phase combinations would need to be sampled, necessitating an even longer hindcast period.

The results presented in this thesis do not conclusively assess the severity of insufficient sampling of low-frequency climate variability. One region likely affected by ENSO-related sampling errors is the North Pacific, where ENSO is the dominant mode of climate variability and residuals (see case 1) can overwhelm the generic SSW response (Polvani et al., 2017).

Insufficient sampling of low-frequency climate variability affects both mean (related to chapter 4) and variance (related to chapter 5) analyses. Addressing these sampling issues, such as through appropriate sub-sampling and statistical testing (see Deser et al., 2018; Oehrlein et al., 2021), would be valuable for future research. Additionally, extended hindcast periods would be beneficial.

6.3 Categorizing sources of variance anomalies

Understanding anomalies in variance requires identifying and understanding the sources of climatological variance. These sources, when altered in their strength, location, or general behavior, can lead to variance anomalies. However, the ways in which sources of variance may actually change are diverse. Table 6.1 broadly categorizes the modulation of variance sources due to spatial shifts, shielding, intensification, presence of strong gradients and bifurcation events.

6.4 Scale-dependence of ensemble spread

Temporal scales

Chapter 5 examined uncertainty in subseasonal-to-seasonal forecasts through ensemble variance (and regime entropy) based on daily values. However, S2S forecasts are often evaluated using temporal averages to even out unpredictable daily fluctuations and thereby enhance predictable signals. This raises the question whether the ensemble spread anomalies following weak and strong polar vortex events (see Sec. 5.2) are sensitive to the underlying temporal resolution.

To check this point, Z1000 ensemble spread was additionally computed based on temporal averages. That is, a running mean over lead time was applied to the raw Z1000 forecasts (window sizes from 2 days to 30 days were tested). Based on these temporally smoothed forecasts, the ensemble spread was computed. Ensemble spread anomalies were then obtained by subtracting the corresponding ensemble spread climatology, where the climatology was re-computed for each window size, respectively.

The results revealed that the modulations of Z1000 ensemble spread following weak polar vortex events are robust across the tested underlying temporal resolutions, from daily values (as employed in chapter 5) to weekly and monthly averages (not shown). However, absolute anomalies are smaller in magnitude when spread anomalies of time-averaged

Modulation of variance sources due to	**Z1000**	**T2m**	**U6010**	**Other examples**
spatial shifts	latitudinal shifts in the storm track associated with the NAO (see sections 5.2 and 5.4)	circulation anomalies can advect maritime air (small variance) to continents, and vice versa (see section 5.4)	latitudinal shifts of the polar vortex edge	Shifts of precipitation over South America caused by El Niño likely affect precipitation variance
shielding	persistent atmospheric blockings can shield a region from cyclone-induced variability (see sections 5.2, 5.3 and 5.4)	increased cloud cover tends to reduce diurnal temperature variability	easterly winds in the lower stratosphere suppress vertical wave propagation into the mid-stratosphere (see section 5.1)	
intensification	strengthening of the North Atlantic storm track in response to a strong polar vortex 5.2	reduced cloud cover tends to increase diurnal temperature variability	increased upward propagation of planetary waves (e.g., associated with atmospheric blockings, see section 4.3)	
presence of **strong gradients**	strong meridional gradients along the North Atlantic eddy-driven jet	sharp gradients along weather fronts	strong gradients along the polar vortex edge	sharp gradients near the tropopause (e.g., affecting potential vorticity, potential temperature, water vapour, etc)
bifurcation events in the atmosphere	The Greenland blocking regime often transitions into Atlantic Trough or Atlantic Ridge (see section 5.4)	Foehn breakthrough in mountain valleys associated with strong and rapid temperature increase	reflection versus absorption of a planetary wave pulse in the stratosphere	melting of sea ice (e.g., affecting albedo, radiation, and temperature)

Table 6.1: Selection of mechanisms inducing variance anomalies, with specific examples provided for 1000 hPa geopotential height (Z1000), 2 m temperature (T2m), and zonal-mean zonal wind at 60 °N, 10 hPa (U6010). Referenced examples are discussed in the respective results sections, while others are hypothetical and not directly studied in this work.

quantities are considered as expected, because climatological spread is generally larger for daily values compared to weekly or monthly averages. Notably, relative spread anomalies (evaluated as the relative deviation from climatology) appear to be fairly insensitive to the underlying temporal resolution. For example, section 5.2 showed that ensemble spread of daily Z1000 forecasts decreases by about 20% over northern Europe following weak polar vortex events. Similarly, ensemble spread of *weekly averaged* or *monthly averaged* Z1000 forecasts decreases by about 25%-30% in these locations.

The fact that these spread anomalies are roughly insensitive to the underlying temporal resolution suggests that polar vortex events modulate variability across different temporal scales. On the one hand, this includes highly transient phenomena, such as individual cyclones (see Sec. 5.3), on the other hand, relatively persistent phenomena, such as variations of the North Atlantic jet exit region, likely also contribute.

Spatial scales

To link spread anomaly signals, such as those observed following weak polar vortex events, to particular physical processes, one can consider a decomposition of ensemble variance into contributions from different spatial scales. However, this approach encounters challenges due to covariances between variability at different scales, as will be discussed in the following.

Figure 6.1b presents ensemble variance of Z1000 as a function of longitude, decomposed into contributions from different zonal wavenumbers. This is achieved by applying a Fourier transform along the zonal dimension, filtering the Fourier coefficients for specific wavenumbers, applying an inverse Fourier transform to the filtered result to return to physical space, and then calculating the variance across the ensemble dimension.

The results show that the sum of ensemble variances for individual zonal wavenumbers does not equal the ensemble variance of the full field. This discrepancy arises due to covariances between different wavenumbers, which can be positive (increasing the total spread), such as over the Atlantic and Pacific, or negative (reducing the total spread).

For instance, as discussed in section 5.2, extreme polar vortex events are followed by a mean tropospheric response primarily associated with planetary scales (wavenumber 1 and 2). However, this planetary-scale mean response also induces changes in synoptic-scale variability. Consequently, the mean and variance response to polar vortex events reflect anomalies in variability across different, yet correlated, spatial scales.

When a zonal average is applied, covariances average out to zero since zonal wavenumbers form an orthogonal basis. As a result, the time-mean zonal-mean ensemble variance can indeed be expressed as the sum of variances related to individual zonal wavenumbers (Fig. 6.1a), unlike ensemble variance at a specific longitude. At a lead time of 28 days, variability associated with zonal wavenumber 1 shows the largest contribution to the total ensemble spread, followed by zonal wavenumber 2, and higher wavenumbers.

In contrast, within the first forecast week, the relative contributions of zonal wavenumbers 1 and 2 are smaller than that of waves 4-7 and 8-20 (Fig. 6.2b). This is due to

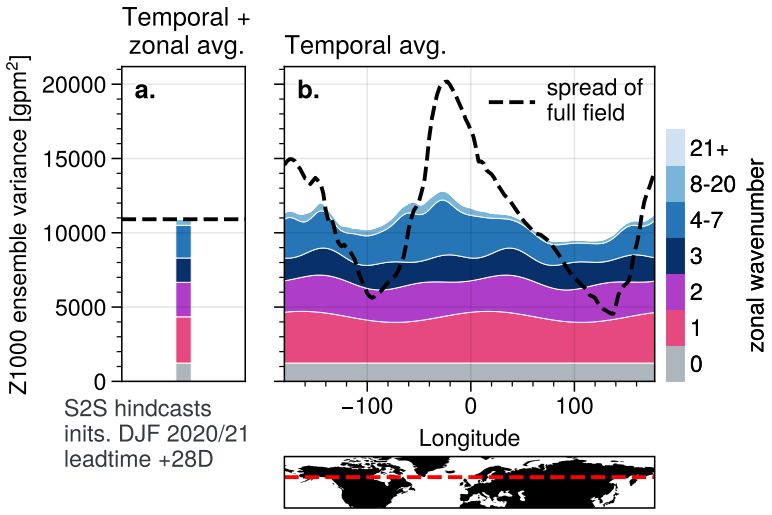

Figure 6.1: 1000 hPa geopotential height ensemble variance at 60 °N at 28 d lead time decomposed by zonal wavenumber (see text for details). **a.** Ensemble variance averaged over longitude and forecasts. **b.** Zonally resolved Ensemble variance, averaged over forecasts. The sum over ensemble variance contributions from different zonal wavenumbers only adds up to the ensemble variance of the full field (black dashed line) if a zonal average is applied.

planetary waves exhibiting longer predictability, and their phase and amplitude are not yet de-correlated between ensemble members. Consequently, the spread related to larger scales saturates at later lead times: zonal wavenumbers 0 and 1 reach 80% of their saturated ensemble variance level after 16 days on average, wavenumber 2 after 15 days, wavenumber 3 after 13 days, wavenumbers 4-7 after 11 days, wavenumbers 8-20 after 8 days and wavenumbers higher than 20 after 7 days (Fig. 6.2c). Note that the saturation level is here computed by averaging lead times of 30-46 days. In principle, it is expected that variance continues to slightly increase further at longer lead times due to low-frequency climate variability. However, an analysis of inter-annual variance reveals that this increase is small (wavenumber 0: +2.9%, wavenumber 1: +1.8%, wavenumber 2: +1.7%, higher wavenumbers: less than +1%; not shown).

Note that theoretical estimates for predictability time scales can also be derived from turbulence theory (Lorenz, 1969; Lilly, 1972). These estimates are based on scaling arguments where it is assumed that initial errors at small scales are gradually transferred to larger scales at a rate determined by the eddy turnover time (Vallis, 2017, June, Ch. 11.4.2). Some of these estimates (e.g., 16.8 days by Lorenz, 1969, for wavenumber k=1) are remarkably close to the results shown here, though the experimental setups substantially differ.

Scale-dependent error growth and resulting predictability limits have predominantly been studied to understand the intrinsic limits of mid-latitude weather prediction at medium-range timescales (e.g. Lorenz, 1969; Rotunno and Snyder, 2008; Zhang et al., 2019; Selz,

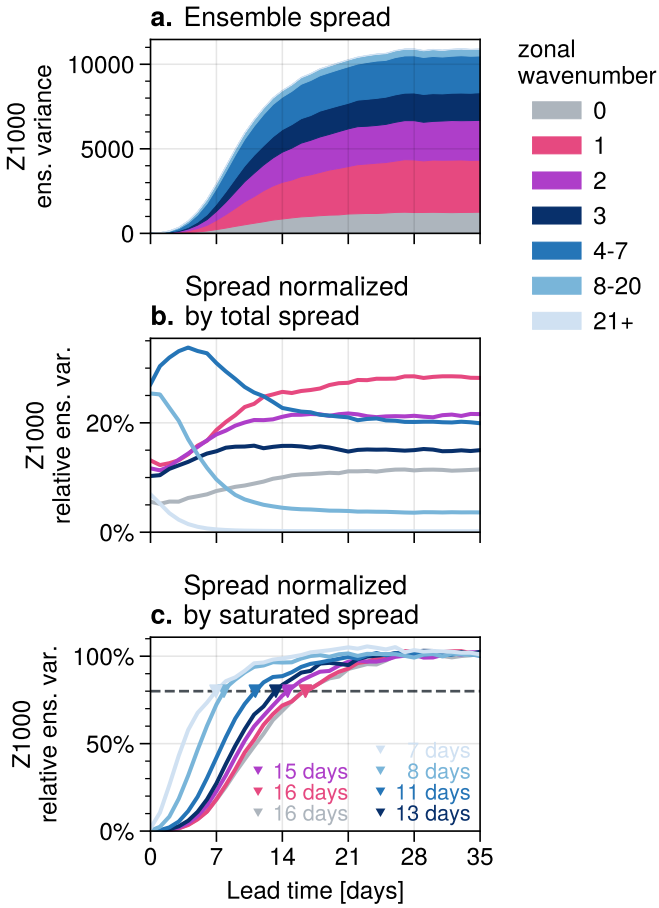

Figure 6.2: Leadtime evolution of 1000 hPa geopotential height (Z1000) ensemble variance decomposed by zonal wavenumber, averaged over longitude and over different forecasts. **a.** Absolute ensemble variance contributions by zonal wavenumber. **b.** Relative ensemble variance contributions to the total spread at a given lead time day by zonal wavenumber. **c.** Ensemble variance normalized by the saturation level of variance by zonal wavenumber. The saturation level is computed as an average over 30 to 46 days lead time, for each wavenumber separately. The horizontal dashed line indicates the 80% level, whereas labels and triangular markers denote the lead time at which this level is reached.

2019; Selz et al., 2022). However, a systematic assessment of variances and covariances across different spatial scales, including their modulation by teleconnections, could also shed light on S2S prediction.

Future research could explore whether "windows-of-opportunity" for S2S prediction arise from slower upscale error growth, lower error saturation levels, or both. Additionally, different teleconnections may provide predictability at different spatial scales, although this thesis highlighted that different spatial scales are inherently coupled (e.g., planetary-scale variability associated with the NAO is coupled to synoptic-scale variability associated with baroclinic eddies).

6.5 Prospects and pathways for further study

This thesis examined forecast uncertainty at S2S timescales, focusing on stratosphere-troposphere coupling during winter in the extratropics of the Northern Hemisphere. The following paragraphs discuss potential applications of these findings to related topics, highlighting opportunities for future research.

Forecast uncertainty associated with the summer circulation

While chapter 5 investigated forecast uncertainty during Northern Hemisphere winter, it is anticipated that similar flow-dependent variations in uncertainty exist during the summer season as well.

Given their year-round definition, the North Atlantic-European weather regimes employed in section 5.4 may serve as a suitable starting point.

Forecast uncertainty of near-surface geopotential height is expected to correlate with cyclone frequency also during summer, although differences may arise due to the northward-shift of the storm track and the general decreases in storm frequency and intensity.

Furthermore, summer weather in the extratropics is partly characterized by convection and associated precipitation or thunderstorms. While convection is usually only predictable on short lead times (hours to a few days), large-scale weather regimes may influence probabilities for convection in specific regions. Situations where the large-scale flow is more predictable than usual could therefore also provide some climatic predictability for forecasts of convection at subseasonal lead times.

Forecast uncertainty in a warming climate

Previous literature has discussed that climate change may not only manifests in distribution shifts but can also affect the variance and higher moments (e.g., Allan and Soden, 2008; Boehm and Thompson, 2023; Shaw and Miyawaki, 2024). For example, alongside general warming (i.e., a shift in the temperature distribution), changes in near-surface

temperature variance have been suggested. During Boreal winter, temperature variability may decrease, particularly at high latitudes (Schneider et al., 2015; Tamarin-Brodsky et al., 2019). This phenomenon is attributed to stronger warming in polar regions, known as Arctic amplification, which leads to weaker horizontal temperature gradients and consequently reduced variance. Likewise, temperature variance could increase during summer in certain regions, such as Europe, partly due to a drier climate and reduced soil moisture (Schär et al., 2004; Fischer and Schär, 2009).

Such changes in variance are crucial because they imply changes in the frequency of extreme events. Additionally, changes in variance could influence the intrinsic predictability of the atmosphere, leading to changes in ensemble spread.

The list below presents a few climate change aspects and hypothesizes their potential impacts on subseasonal-to-seasonal forecast uncertainty. Future research could test these hypotheses through controlled experiments of subseasonal-to-seasonal ensemble forecasts under current versus future climate conditions.

- While future changes in the strength of the polar vortex strength remain uncertain (Karpechko et al., 2022), any changes are likely to influence the mean tropospheric circulation. Based on the same principles as discussed for subseasonal-to-seasonal timescales (section 5.2), such changes would presumably be accompanied by changes in tropospheric variability and ensemble spread.

- Under global warming, land generally warms faster than the sea, leading to changes in horizontal temperature gradients. Specifically, the zonal land-sea temperature contrast may decrease during winter, reducing temperature variability, and increase during summer, increasing temperature variability.

- Melting sea ice may intensify ocean-atmosphere coupling, which could reduce near-surface temperature variability.

- The increase in absolute humidity in a warmer climate, consistent with the Clausius-Clapeyron relation, increases the frequency of extreme precipitation events, which may also increase precipitation variance.

Uncertainty in data-driven forecast models

Recent advancements in data-driven weather models utilizing machine learning methods have shown that these models can generate forecasts with skill levels comparable to or even surpassing those of traditional physics-based general circulation models (Price et al., 2023; Chen et al., 2023; Lang et al., 2024a). Initially, these models focused on deterministic forecasts at medium-range lead times. However, efforts are now underway to extend their capabilities to probabilistic forecasts (Li et al., 2024b; Zhong et al., 2024; Lang et al., 2024b) and to forecasts with longer lead times (Li et al., 2024a).

With respect to probabilistic forecasting, it will be crucial to assess whether data-driven models can provide reliable estimates of uncertainty. In general though, they offer the potential for generating large ensemble sizes at low costs, which may be useful for both applications and research.

In the realm of subseasonal-to-seasonal and seasonal forecasting, a key question is whether data-driven models can learn processes relevant for these timescales based on reanalysis data only. So far, data-driven models have been successful primarily in medium-range forecasting, where relevant processes – such as fronts, Rossby waves, baroclinic instability, etc. – are presumably sampled well by reanalysis datasets. In contrast, subseasonal-to-seasonal and seasonal forecasts rely on lower-frequency processes whose sampling is only limited in reanalyses. Therefore, representing those processes accurately, including coupling between different low-frequency phenomena (e.g., QBO + ENSO), as well as their coupling to the extratropics, may pose additional challenges for data-driven approaches.

Appendix A

Vertically propagating Rossby waves

Rossby waves are atmospheric modes that get excited primarily by baroclinic instability and flow over (planetary-scale) orography, and arise from angular momentum conservation in a fluid on a rotating sphere. Rossby waves play a critical role in shaping the large-scale circulation of the Earth and contribute to the redistribution of momentum and energy across the atmosphere.

In the context of this thesis, a key aspect of Rossby waves is their vertical propagation, which represents a major source of variability for the polar stratospheric circulation. Specifically, Rossby waves are associated with intrinsic westward momentum, resulting from their intrinsic westward phase speed. Upon breaking and dissipation, Rossby waves transfer westward momentum to the mean flow, acting to decelerate the eastward zonal-mean zonal wind. This mechanism plays a key role in the formation of sudden stratospheric warmings (SSWs).

To motivate some of these aspects the following briefly sketches the theoretical background of vertically propagating Rossby waves. This overview follows the course "Advanced Atmospheric Dynamics" (winter semester 2021/22 at Ludwig-Maximilians-University of Munich) by Prof. Thomas Birner.

Angular momentum conservation in a rotating fluid can be formulated in terms of potential vorticity, P, which combines the relative vorticity of an air parcel, ζ_θ, with the planetary vorticity, f, and the vertical density structure, $\sigma \equiv -\frac{\partial_\theta p}{g}$. θ denotes potential temperature, p denotes pressure, and g denotes the gravitational constant. Under adiabatic motion, in an inviscid fluid, potential vorticity is materially conserved:

$$\frac{\mathrm{D}}{\mathrm{D}t} P = 0, \quad \text{where} \quad \frac{\mathrm{D}}{\mathrm{D}t} = \partial_t + \vec{v} \cdot \vec{\nabla} \quad \text{and} \quad P = (\zeta_\theta + f)/\sigma, \tag{A.1}$$

with $\vec{v}$ being the 3-dimensional wind vector. Note that ζ_θ and $\vec{v}$ are evaluated in isentropic coordinates, i.e., horizontal derivatives are evaluated along surfaces of constant potential temperature and the vertical velocity equals diabatic heating (which is zero for adiabatic motion).

In general, equation A.1 is non-linear, because, for example, both P and $\frac{\mathrm{D}}{\mathrm{D}t}$ depend on $\vec{v}$ and general solutions can only be computed numerically. However, under certain assumptions, analytical solutions can be derived, which can provide additional insights into the underlying dynamics.

In quasi-geostrophic (QG) theory, the flow is assumed to be almost in geostrophic balance (i.e., the pressure gradient force is almost balanced by the Coriolis force). Specifically, the Rossby number (a non-dimensional number measuring the effect of $\frac{\mathrm{D}}{\mathrm{D}t}$ relative to the Coriolis effect) is assumed to be small, and deviations from geostrophic balance are on the order of the Rossby number. This implies that ageostrophic wind components are small compared to geostrophic components: $\vec{v}_\mathrm{a} \ll \vec{v_\mathrm{g}}$. Moreover, this manifests in ageostrophic flow components getting advected only by geostrophic winds. In addition, latitudinal variations (denoted as y) in the Coriolis parameter are linearized (a so-called β-plane approximation) and assumed to be small: $f \approx f_0 + \beta y$, where $|\beta y| \ll f_0$ and $f_0 = \text{const.}$

Linearization of equation A.1 under these approximations yields a conservation law for QG potential vorticity, q (here evaluated in log-p coordinates):

$$(\partial_t + U\partial_x)\, q' + v'\partial_y Q = 0, \tag{A.2}$$

where q, u (the eastward wind) and v (the northward wind) have been decomposed into a zonally uniform basic state Q, U, V (of zero'th order) and perturbations, q', u', v' (of order Rossby number), respectively. To obtain an equation for only one prognostic variable (so far: q' and v'), a geostrophic streamfunction, ψ, can be introduced that fulfills $u_\mathrm{g} = -\partial_y\psi$ and $v_\mathrm{g} = \partial_x\psi$. QG potential vorticity can then be expressed in terms of this geostrophic streamfunction:

$$q' = \vec{\nabla}^2\psi' + \rho_R^{-1}\partial_Z(\rho_R\gamma\partial_Z\psi'), \tag{A.3}$$

where ψ has been decomposed into a basic state and a perturbation term, $\psi = \Psi + \psi'$, ρ_R is a height-dependent reference density profile and γ is a constant factor that measures the effect of rotation (i.e., horizontal motion) relative to the effect of stratification (i.e., vertical motion). Under suitable boundary conditions, equation A.3 can be inverted to compute the streamfunction (and the associated balanced dynamical fields; i.e., winds, temperature and pressure) from a given potential vorticity field – a principle known as invertibility.

Moreover, the meridional gradient of the basic state QG potential vorticity, $\partial_y Q$, in equation A.2 can be expressed as a function of the latitudinal variations of planetary vorticity, β, and the basic state zonal wind, U:

$$\partial_y Q = \beta - \partial_{yy}U - \rho_R^{-1}\partial_Z(\rho_R\gamma\partial_Z U). \tag{A.4}$$

Plugging q' and $\partial_y Q$ into equation A.2 yields a wave equation for ψ'. Assuming a channel geometry in y (i.e., $\psi' = 0$ at the latitudinal boundaries) leads to solutions of the following form:

$$\psi' = \Re\{\Psi(Z)\exp[ik(x-ct)]\sin ly\}, \tag{A.5}$$

with zonal and meridional wavenumbers, k and l, zonal phase speed, c and a vertical structure function, $\Psi(Z)$. Using $\rho_R \sim \exp(-Z/H)$, $\gamma = \text{const.}$ and $\partial_y Q = \beta = \text{const.}$, the

vertical structure takes wave solutions:

$$\Psi(Z) \sim \Re\{\frac{Z}{2H}\exp(imZ)\}, \tag{A.6}$$

where m corresponds to the vertical wavenumber. From plugging $\Psi(Z)$ into equations A.5 and A.2, it follows that these waves are only vertically propagating (i.e., $m^2 > 0$), if:

$$\begin{aligned} 0 < U - c < U_c \equiv \beta\left(\frac{\gamma}{4H^2 + k^2}\right): &\quad \text{vertical propagation} \\ \text{else}: &\quad \text{vertical evanescence,} \end{aligned} \tag{A.7}$$

where $l = 0$ was assumed and U_c is the critical zonal flow. Importantly, stationary waves ($c = 0$) can propagate vertically only in a background flow that satisfies $U > 0$ (i.e., westerlies) and $U < U_c(k)$ (Charney and Drazin, 1961).

When conditions for vertical propagation are satisfied, the wave amplitude is expected to increase exponentially with height (i.e., $\Psi \sim \mathrm{e}^{Z/(2H)}$) as energy conservation holds and density decreases exponentially with height. In the real atmosphere, vertical amplitude growth is typically less than exponential though, because linear theory and the implied assumption that perturbations from the basic state are small, breaks down. Therefore, waves of large amplitudes tend to break and dissipate, thereby transferring westward momentum to the zonal-mean flow, acting to decelerate the zonal mean westerly winds.

The properties of the derived solutions for upward propagating Rossby waves have fundamental implications for polar vortex dynamics. Some of them are listed here:

- Vertical Rossby wave propagation into the polar stratosphere is only possible during winter, as $U < 0\,\mathrm{m\,s^{-1}}$ during summer. As a result, variability of the summer easterlies in the polar stratosphere is much smaller than the variability of the winter westerlies.

- The critical zonal flow, U_c, increases with the wavelength of the waves. Specifically, for waves with zonal wavenumbers $s > 2$ (i.e., for waves with more than two minima or maxima along one latitude circle), U_c is typically smaller than the zonal flow U. Consequently, it is predominantly planetary-scale waves ($s = 1$ and $s = 2$) that can freely propagate into the stratosphere.

- Rossby wave breaking can decelerate the zonal mean flow in the polar stratosphere, which can in extreme cases lead to SSWs.

- SSWs occur almost solely in the Northern Hemisphere, as tropospheric wave activity induced by planetary-scale orography and the associated deceleration of the stratospheric zonal-mean westerly flow are much larger in the Northern Hemisphere than in the Southern Hemisphere.

Appendix B

The signal-to-noise paradox in a toy model

Atmospheric phenomena are associated with different predictability time scales. For example, the NAO is modulated via the stratosphere polar vortex on monthly time scales. However, the NAO is also modulated by individual cyclones and anti-cyclones over the North Atlantic and these may be predictable only for about a week.

At a given lead time, a prediction, say, of the NAO, can thus conceptually be decomposed into a predictable component ("signal", e.g., arising from stratosphere-troposphere coupling) and a non-predictable component ("noise", e.g., arising from individual cyclones and anticyclones). The variability of the signal relative to the variability of the noise is measured by the signal-to-noise ratio, and can vary across lead times, seasons, cases, etc. In ensemble forecasts the signal can be estimated via the ensemble mean. This assumes that all ensemble members share the same signal, whereas the noise is uncorrelated between the members. Averaging over the ensemble dimension isolates the signal, since the noise has a mean of zero. Moreover, the noise can be measured by the ensemble spread, i.e., the variance across the ensemble members.

A conceptual decomposition into signal and noise can also be applied to observations. However, an observational time series inherently contains both signal and noise, making it challenging to separate these components based solely on observations. In a perfect model, where individual ensemble members are statistically indistinguishable from the observations, the signal would be the same in the model and the observations. Furthermore, even though the noise is uncorrelated, its variability would be the same in the model and the observations as well. As a result, in a perfect model, the correlation between the ensemble mean and the observations ($r_{\text{ens.mean, obs}}$) would equal the correlation between the ensemble mean and individual ensemble members ($r_{\text{ens.mean, members}}$). The ratio of the square of these two correlations is called the ratio of predictable components (RPC):

$$\text{RCP} = \frac{r^2_{\text{ens.mean, obs}}}{r^2_{\text{ens.mean, members}}} \tag{B.1}$$

Consequently, in a perfect model, the RPC equals 1.

In practice, models are not perfect, and characteristics of signal and noise may differ between model and observations. Therefore, one might expect the ensemble mean to exhibit a higher correlation with individual members (i.e., with the model world) than with observations (i.e., with the real world), resulting in RPC<1. Surprisingly, the opposite has been observed in seasonal forecasts in some cases, leading to counter-intuitive situations in which $r^2_{\text{ens.mean, obs}} > r^2_{\text{ens.mean, members}}$, resulting in RPC>1. Such situations are described by the so-called signal-to-noise paradox (Scaife and Smith, 2018).

A recent article has summarized this puzzle, discussed various possible reasons and highlighted that it is not a "simple-to-understand nor a simple-to-fix problem" (Weisheimer et al., 2024). Although the signal-to-noise paradox is not the primary focus of this thesis, the presented results may still be influenced by its effects. For instance, underestimation of teleconnection strengths is a common issue in S2S forecast models (see discussion in section 6.2.1). Previous studies have identified this underestimation in seasonal forecasts, linking it to the signal-to-noise paradox (Siegert et al., 2016; Falkena et al., 2022; Williams et al., 2023). However, interpreting these results can be complex, as underestimation may be phase-dependent (e.g., affecting only the negative NAO phase; Falkena et al., 2022), or results may be compromised by sampling uncertainty, especially when the signal is weak or the hindcast period is short (Weisheimer et al., 2024). Therefore, to illustrate how underestimation of teleconnection strengths can manifest in the signal-to-noise paradox, a toy model is presented here that isolates the necessary mathematical ingredients in a straightforward and controlled framework.

The following hypothesis is tested: if the tropospheric response to stratospheric extreme events was larger in observations than in the model, then the signal-to-noise paradox becomes apparent. For this purpose, a synthetic observational time series with 20 000 time steps is generated. These observations consist of a pre-specified signal drawn from a Gaussian distribution with a temporal variance of 0.4 (dimensions are not important). Uncorrelated noise with a temporal variance of 0.6 is added to the signal. Note that the amplitude of the signal relative to the amplitude of the noise could be somewhat weaker at S2S timescales, which does, however, not affect the qualitative conclusions drawn from this analysis. In addition to the observational time series, ensemble data are generated, where each ensemble member consists of the same signal as the observations plus an uncorrelated noise, also with a temporal variance of 0.6.

Fig. B.1 presents three cases, each with different parameters chosen for the signal and noise in observations and ensembles. For each case, the general setup, the RPC and the two correlations, $r^2_{\text{ens.mean, obs}}$ and $r^2_{\text{ens.mean, members}}$, are shown as function of the ensemble size. Cases where RCP > 1 exhibit the signal-to-noise paradox (SNP).

Case 1: No SNP

The first synthetic case uses parameters as described, i.e., with a signal that has variance of 0.4 and a noise that has variance of 0.6, in both observations and ensemble members.

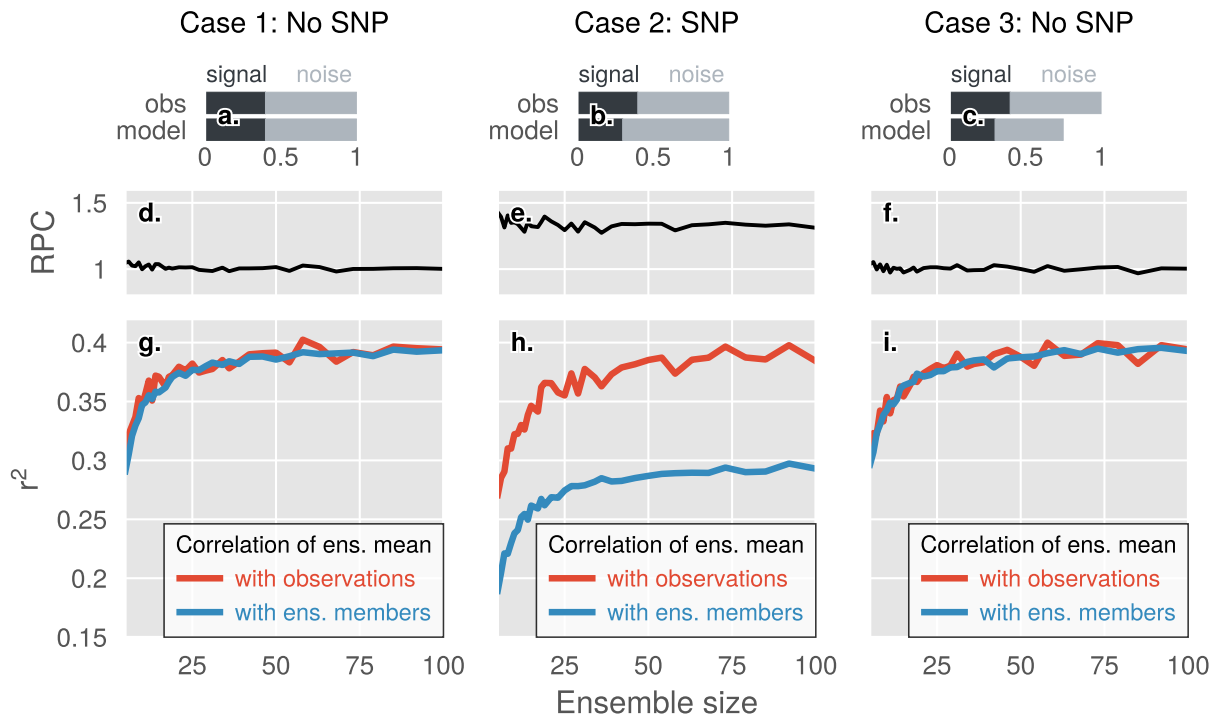

Figure B.1: Toy model setup to investigate the signal-to-noise paradox. Each column presents one synthetic case. **a-c.** Magnitude of signal and noise in the (synthetical) observational time series and the (synthetical) model ensemble forecasts, quantified by their temporal variance. By construction, the signal is identical in observations and in all ensemble members (up to a factor, which is 1 in the first case and is 0.75 in the second and third case). By construction, the noise is uncorrelated. **d-f.** Ratio of predictable components (RPC; see text for details). RCP > 1 indicates presence of the signal-to-noise paradox. **g-f.** r^2 for ensemble mean vs. observations and for ensemble mean vs. individual ensemble members, as a function of the ensemble size.

Ensemble members are generated identically to the observations, hence a perfect-model assumption is implicit. The results show that RPC is very close to 1, except for sampling uncertainty. The correlations of the ensemble mean with observations and with individual ensemble members increase with ensemble size and approach 0.4 (which is equal to the temporal variance of the signal) in the limit of large ensemble sizes.

Case 2: SNP

In a second synthetic case, the observational time series remains the same as in the first case. However, the signal in the ensemble members is weaker than the signal in observations, scaled by a factor of 0.75. The temporal variance of the ensemble mean thus equals 0.3. To ensure that ensemble members have the same overall temporal variance as observations (equal to 1), the temporal variance of the noise in the ensemble members

is increased from 0.6 to 0.7. The results show that the RCP in this case is substantially larger than 1, with RCP = $1/0.75 = \frac{4}{3}$. Consistently, $r^2_{\text{ens.mean, members}}$ approaches 0.3, whereas $r^2_{\text{ens.mean, obs}}$ approaches 0.4 in the limit of large ensemble sizes. This difference is due to the different signal-to-noise ratios in observations and ensemble members. On the one hand, the correlation $r^2_{\text{ens.mean, obs}}$ remains the same as in case 1, because observations are unchanged and the ensemble mean is only scaled by a constant factor, to which the correlation coefficient is insensitive. On the other hand, $r^2_{\text{ens.mean, members}}$ decreases as the ensemble mean explains less variance of the temporal evolution of the individual ensemble members due to the increased noise. As pointed out in Weisheimer et al., 2024, using the regression coefficient instead of the correlation coefficient would reveal that the signal is smaller in the model than observations.

Case 3: No SNP

In a third synthetic case, it is shown that RPC = 1 can be achieved even when the signal in the model is underestimated. The same observational time series is used as in the first and second cases. As in the second case, the signal in the model is scaled by a factor of 0.75 compared to the signal in observations. In this case, also the model noise is scaled by a factor of 0.75. The overall temporal variance in the ensemble members is thus smaller as in observations (0.75 versus 1). This results in model and observations exhibiting the same signal-to-noise ratio, because both signal and noise are scaled by the same factor. Consequently, this case does not exhibit the SNP.

In conclusion, these toy model experiments demonstrate a simple possible explanation for the presence of the SNP. Underestimation of the signal in models can lead to RPC > 1, as discussed by previous studies (Siegert et al., 2016; Falkena et al., 2022; Williams et al., 2023). In the context of this thesis, an important underestimated teleconnection might be the QBO-NAO relationship via the polar vortex, which might therefore contribute to a SNP in S2S forecasts. Compared to other studies, the approach presented here is not based on parameters derived from actual forecasts but rather illustrates, in a highly simplified framework, the key element of the signal-to-noise paradox: different signal-to-noise ratios in ensemble members and in observations.

Moreover, chapter 5.2 reveals that ensemble mean anomalies can be associated with changes in ensemble spread. This correlation between the signal and the amplitude of the noise was not incorporated into the above toy model. Future work could explore how this correlation impacts the signal-to-noise ratio and its relation to the signal-to-noise paradox. The presented toy model setup could serve as a useful test bed for such analyses.

List of Figures

List of Tables

Software

The analyses presented in this thesis were conducted using Python versions 3.8, 3.9, 3.10, 3.11 and 3.12. Some of the most important external libraries used include:

- xarray (https://docs.xarray.dev/en/stable/)
- pandas (https://pandas.pydata.org/docs/)
- dask (https://docs.dask.org/en/stable/)
- matplotlib (https://matplotlib.org/stable/index.html)
- numpy (https://numpy.org/doc/stable/)
- ProPlot (https://proplot.readthedocs.io/en/stable/)

Some of my own software is publicly available, including `s2stools` and `xevent` (see section 3.5). In addition, software used for the analyses of section 5.2 is available at https://gitlab.physik.uni-muenchen.de/Jonas.Spaeth/stratospheric-impact-on-forecast-uncertainty.

Bibliography

Adams, S. V., Arribas, A., Prudden, R., Robinson, N., Saggioro, E., & Shepherd, T. G. (2021). Quantifying Causal Pathways of Teleconnections. *Bulletin of the American Meteorological Society*, *102*(12), E2247–E2263. https://doi.org/10.1175/BAMS-D-20-0117.1

Afargan-Gerstman, H., Büeler, D., Wulff, C. O., Sprenger, M., & Domeisen, D. I. V. (2024). Stratospheric influence on the winter North Atlantic storm track in subseasonal reforecasts. *Weather and Climate Dynamics*, *5*(1), 231–249. https://doi.org/10.5194/wcd-5-231-2024

Afargan-Gerstman, H., Polkova, I., Papritz, L., Ruggieri, P., King, M. P., Athanasiadis, P. J., Baehr, J., & Domeisen, D. I. V. (2020). Stratospheric influence on North Atlantic marine cold air outbreaks following sudden stratospheric warming events. *Weather and Climate Dynamics*, *1*(2), 541–553. https://doi.org/10.5194/wcd-1-541-2020

Albers, J. R., & Birner, T. (2014). Vortex Preconditioning due to Planetary and Gravity Waves prior to Sudden Stratospheric Warmings. *Journal of the Atmospheric Sciences*, *71*(11), 4028–4054. https://doi.org/10.1175/JAS-D-14-0026.1

Allan, R. P., & Soden, B. J. (2008). Atmospheric Warming and the Amplification of Precipitation Extremes. *Science*, *321*(5895), 1481–1484. https://doi.org/10.1126/science.1160787

Ambaum, M. H. P., & Hoskins, B. J. (2002). The NAO Troposphere–Stratosphere Connection. *Journal of Climate*, *15*(14), 1969–1978. https://doi.org/10.1175/1520-0442(2002)015<1969:TNTSC>2.0.CO;2

Ambaum, M. H. P., Hoskins, B. J., & Stephenson, D. B. (2001). Arctic Oscillation or North Atlantic Oscillation? *Journal of Climate*, *14*(16), 3495–3507. https://doi.org/10.1175/1520-0442(2001)014<3495:AOONAO>2.0.CO;2

Andrews, D. G., Leovy, C. B., & Holton, J. R. (1987). *Middle Atmosphere Dynamics* (1st editio). Academic Press. ISBN: 9780120585762.

Anstey, J. A., Simpson, I. R., Richter, J. H., Naoe, H., Taguchi, M., Serva, F., Gray, L. J., Butchart, N., Hamilton, K., Osprey, S., Bellprat, O., Braesicke, P., Bushell, A. C., Cagnazzo, C., Chen, C.-C., Chun, H.-Y., Garcia, R. R., Holt, L., Kawatani, Y., … Yukimoto, S. (2022). Teleconnections of the Quasi-Biennial Oscillation in a multi-model ensemble of QBO-resolving models. *Quarterly Journal of the Royal Meteorological Society*, *148*(744), 1568–1592. https://doi.org/10.1002/qj.4048

Augier, P., & Lindborg, E. (2013). A New Formulation of the Spectral Energy Budget of the Atmosphere, with Application to Two High-Resolution General Circulation Models. *Journal of the Atmospheric Sciences*, *70*(7), 2293–2308. https://doi.org/10.1175/JAS-D-12-0281.1

Ayarzagüena, B., Barriopedro, D., Garrido-Perez, J. M., Abalos, M., de la Cámara, A., García-Herrera, R., Calvo, N., & Ordóñez, C. (2018). Stratospheric Connection to the Abrupt End of the 2016/2017 Iberian Drought. *Geophysical Research Letters*, *45*(22). https://doi.org/10.1029/2018GL079802

Ayarzagüena, B., Palmeiro, F. M., Barriopedro, D., Calvo, N., Langematz, U., & Shibata, K. (2019). On the representation of major stratospheric warmings in reanalyses. *Atmospheric Chemistry and Physics*, *19*(14), 9469–9484. https://doi.org/10.5194/acp-19-9469-2019

Baldwin, M. P., Gray, L. J., Dunkerton, T. J., Hamilton, K., Haynes, P. H., Randel, W. J., Holton, J. R., Alexander, M. J., Hirota, I., Horinouchi, T., Jones, D. B. A., Kinnersley, J. S., Marquardt, C., Sato, K., & Takahashi, M. (2001). The quasi-biennial oscillation. *Reviews of Geophysics*, *39*(2), 179–229. https://doi.org/10.1029/1999RG000073

Baldwin, M., Stephenson, D. B., Thompson, D. W., Dunkerton, T., Charlton, A., & O'Neill, A. (2003). Stratospheric Memory and Skill of Extended-Range Weather Forecasts. *Science*, *301*(5633), 636–640. https://doi.org/10.1126/science.1087143

Baldwin, M. P., Ayarzagüena, B., Birner, T., Butchart, N., Butler, A. H., Charlton-Perez, A. J., Domeisen, D. I. V., Garfinkel, C. I., Garny, H., Gerber, E. P., Hegglin, M. I., Langematz, U., & Pedatella, N. M. (2021). Sudden Stratospheric Warmings. *Reviews of Geophysics*, *59*(1), 1–37. https://doi.org/10.1029/2020RG000708

Baldwin, M. P., & Dunkerton, T. J. (1999). Propagation of the Arctic Oscillation from the stratosphere to the troposphere. *Journal of Geophysical Research: Atmospheres*, *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

Baldwin, M. P., & Holton, J. R. (1988). Climatology of the Stratospheric Polar Vortex and Planetary Wave Breaking. *Journal of the Atmospheric Sciences*, *45*(7), 1123–1142. ISBN: 9781626239777 https://doi.org/10.1175/1520-0469(1988)045<1123:COTSPV>2.0.CO;2

Bancalá, S., Krüger, K., & Giorgetta, M. (2012). The preconditioning of major sudden stratospheric warmings. *Journal of Geophysical Research: Atmospheres*, *117*(D4). https://doi.org/10.1029/2011JD016769

Barnes, E. A., Samarasinghe, S. M., Ebert-Uphoff, I., & Furtado, J. C. (2019). Tropospheric and Stratospheric Causal Pathways Between the MJO and NAO. *Journal of Geophysical Research: Atmospheres*, *124*(16), 9356–9371. https://doi.org/10.1029/2019JD031024

Bauer, P. (2022). *The Role of Ural Blockings in Stratosphere-Troposphere Coupling Based on Extended-Range Ensemble Forecasts* [Bachelor's Thesis]. Ludwig-Maximilians-University Munich.

Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, *525*(7567), 47–55. https://doi.org/10.1038/nature14956

Baumgart, M., Ghinassi, P., Wirth, V., Selz, T., Craig, G. C., & Riemer, M. (2019). Quantitative View on the Processes Governing the Upscale Error Growth up to the Planetary Scale Using a Stochastic Convection Scheme. *Monthly Weather Review*, *147*(5), 1713–1731. https://doi.org/10.1175/MWR-D-18-0292.1

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

Bett, P. E., Scaife, A. A., Hardiman, S. C., Thornton, H. E., Shen, X., Wang, L., & Pang, B. (2023). Using large ensembles to quantify the impact of sudden stratospheric warmings and their precursors on the North Atlantic Oscillation. *Weather and Climate Dynamics*, *4*(1), 213–228. https://doi.org/10.5194/wcd-4-213-2023

Birner, T., & Albers, J. R. (2017). Sudden stratospheric warmings and anomalous upward wave activity flux. *Scientific Online Letters on the Atmosphere*, *13*(A), 8–12. https://doi.org/10.2151/sola.13A-002

Bjerknes, V. (1904). Das Problem der Wettervorhersage, betrachtet vom Standpunkte der Mechanik und der Physik. *Meteorol. Z.*, *21*, 1–7.

Boehm, C. L., & Thompson, D. W. J. (2023). The Key Role of Cloud–Climate Coupling in Extratropical Sea Surface Temperature Variability. *Journal of Climate*, *36*(8), 2753–2762. https://doi.org/10.1175/JCLI-D-22-0362.1

Brönnimann, S. (2007). Impact of El Niño–Southern Oscillation on European climate. *Reviews of Geophysics*, *45*(3). https://doi.org/10.1029/2006RG000199

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

Büeler, D., Sprenger, M., & Wernli, H. (2024). Northern Hemisphere extratropical cyclone biases in ECMWF subseasonal forecasts. *Quarterly Journal of the Royal Meteorological Society*, *150*(759), 1096–1123. https://doi.org/10.1002/qj.4638

Buizza, R., Leutbecher, M., & Thorpe, A. (2015). Living with the butterfly effect : a seamless view of predictability. *ECMWF Newsletter*, *145*, 18–23.

Butchart, N. (2014). The Brewer-Dobson circulation. *Reviews of Geophysics*, *52*(2), 157–184. https://doi.org/10.1002/2013RG000448

Butler, A., Charlton-Perez, A., Domeisen, D. I., Garfinkel, C., Gerber, E. P., Hitchcock, P., Karpechko, A. Y., Maycock, A. C., Sigmond, M., Simpson, I., & Son, S.-W. (2019). Sub-seasonal Predictability and the Stratosphere. In *Sub-seasonal to seasonal prediction* (pp. 223–241). Elsevier. ISBN: 9780128117149 https://doi.org/10.1016/B978-0-12-811714-9.00011-5

Butler, A. H., & Domeisen, D. I. V. (2021). The wave geometry of final stratospheric warming events. *Weather and Climate Dynamics*, *2*(2), 453–474. https://doi.org/10.5194/wcd-2-453-2021

Butler, A. H., Seidel, D. J., Hardiman, S. C., Butchart, N., Birner, T., & Match, A. (2015). Defining Sudden Stratospheric Warmings. *Bulletin of the American Meteorological Society*, *96*(11), 1913–1928. https://doi.org/10.1175/BAMS-D-13-00173.1

Butler, A. H., Sjoberg, J. P., Seidel, D. J., & Rosenlof, K. H. (2017). A sudden stratospheric warming compendium. *Earth System Science Data*, *9*(1), 63–76. https://doi.org/10.5194/essd-9-63-2017

Cassou, C. (2008). Intraseasonal interaction between the Madden–Julian Oscillation and the North Atlantic Oscillation. *Nature*, *455*(7212), 523–527. https://doi.org/10.1038/nature07286

Chang, E. K. M. (2005). The Impact of Wave Packets Propagating across Asia on Pacific Cyclone Development. *Monthly Weather Review*, *133*(7), 1998–2015. https://doi.org/10.1175/MWR2953.1

Charlton, A. J., & Polvani, L. M. (2007). A New Look at Stratospheric Sudden Warmings. Part I: Climatology and Modeling Benchmarks. *Journal of Climate*, *20*(3), 449–469. https://doi.org/10.1175/JCLI3996.1

Charlton-Perez, A. J., Polvani, L. M., Austin, J., & Li, F. (2008). The frequency and dynamics of stratospheric sudden warmings in the 21st century. *Journal of Geophysical Research: Atmospheres*, *113*(D16). https://doi.org/10.1029/2007JD009571

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

Charney, J. G., & Drazin, P. G. (1961). Propagation of planetary-scale disturbances from the lower into the upper atmosphere. *Journal of Geophysical Research (1896-1977)*, *66*(1), 83–109. https://doi.org/10.1029/JZ066i001p00083

Charney, J. G., Fjörtoft, R., & Von Neumann, J. (1950). Numerical Integration of the Barotropic Vorticity Equation. *Tellus A: Dynamic Meteorology and Oceanography*, *2*(4), 237–254. https://doi.org/10.3402/tellusa.v2i4.8607

Chen, L., Zhong, X., Zhang, F., Cheng, Y., Xu, Y., Qi, Y., & Li, H. (2023). FuXi: a cascade machine learning forecasting system for 15-day global weather forecast. *npj Climate and Atmospheric Science*, *6*(1), 190. https://doi.org/10.1038/s41612-023-00512-1

Cohen, J., Furtado, J. C., Jones, J., Barlow, M., Whittleston, D., & Entekhabi, D. (2014). Linking Siberian Snow Cover to Precursors of Stratospheric Variability. *Journal of Climate*, *27*(14), 5422–5432. https://doi.org/10.1175/JCLI-D-13-00779.1

Cover, T. M., & Thomas, J. A. (2005, September). Entropy, Relative Entropy, and Mutual Information. In *Elements of information theory* (pp. 13–55). Wiley. ISBN: 9780471748823 https://doi.org/10.1002/047174882X.ch2

Craig, G. C., Puh, M., Keil, C., Tempest, K., Necker, T., Ruiz, J., Weissmann, M., & Miyoshi, T. (2022). Distributions and convergence of forecast variables in a 1,000-member convection-permitting ensemble. *Quarterly Journal of the Royal Meteorological Society*, *148*(746), 2325–2343. https://doi.org/10.1002/qj.4305

Dask Development Team. (2016). *Dask: Library for dynamic task scheduling*. http://dask.pydata.org

Davis, R. E. (1976). Predictability of Sea Surface Temperature and Sea Level Pressure Anomalies over the North Pacific Ocean. *Journal of Physical Oceanography*, *6*(3), 249–266. https://doi.org/10.1175/1520-0485(1976)006<0249:POSSTA>2.0.CO;2

Davis, R. E. (1978). Predictability of Sea Level Pressure Anomalies Over the North Pacific Ocean. *Journal of Physical Oceanography*, *8*(2), 233–246. https://doi.org/10.1175/1520-0485(1978)008<0233:POSLPA>2.0.CO;2

de la Cámara, A., Albers, J. R., Birner, T., Garcia, R. R., Hitchcock, P., Kinnison, D. E., & Smith, A. K. (2017). Sensitivity of Sudden Stratospheric Warmings to Previous Stratospheric Conditions. *Journal of the Atmospheric Sciences*, *74*(9), 2857–2877. https://doi.org/10.1175/JAS-D-17-0136.1

de la Cámara, A., Birner, T., & Albers, J. R. (2019). Are Sudden Stratospheric Warmings Preceded by Anomalous Tropospheric Wave Activity? *Journal of Climate*, *32*(21), 7173–7189. https://doi.org/10.1175/JCLI-D-19-0269.1

Deser, C., Simpson, I. R., Phillips, A. S., & McKinnon, K. A. (2018). How Well Do We Know ENSO's Climate Impacts over North America, and How Do We Evaluate Models Accordingly? *Journal of Climate*, *31*(13), 4991–5014. https://doi.org/10.1175/JCLI-D-17-0783.1

Domeisen, D. I. V., & Butler, A. H. (2020). Stratospheric drivers of extreme events at the Earth's surface. *Communications Earth & Environment*, *1*(1), 59. https://doi.org/10.1038/s43247-020-00060-z

Domeisen, D. I. V., Butler, A. H., Charlton-Perez, A. J., Ayarzagüena, B., Baldwin, M. P., Dunn-Sigouin, E., Furtado, J. C., Garfinkel, C. I., Hitchcock, P., Karpechko, A. Y., Kim, H., Knight, J., Lang, A. L., Lim, E.-P., Marshall, A., Roff, G., Schwartz, C., Simpson, I. R., Son, S.-W., & Taguchi, M. (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

Domeisen, D. I. V., Sun, L., & Chen, G. (2013). The role of synoptic eddies in the tropospheric response to stratospheric variability. *Geophysical Research Letters*, *40*(18), 4933–4937. https://doi.org/10.1002/grl.50943

Domeisen, D. I., Butler, A. H., Charlton-Perez, A. J., Ayarzagüena, B., Baldwin, M. P., Dunn-Sigouin, E., Furtado, J. C., Garfinkel, C. I., Hitchcock, P., Karpechko, A. Y., Kim, H., Knight, J., Lang, A. L., Lim, E.-P., Marshall, A., Roff, G., Schwartz, C., Simpson, I. R., Son, S.-W., & Taguchi, M. (2020c). The Role of the Stratosphere in Subseasonal to Seasonal Prediction: 1. Predictability of the Stratosphere. *Journal of Geophysical Research: Atmospheres*, *125*(2), 1–17. https://doi.org/10.1029/2019JD030920

Domeisen, D. I., Garfinkel, C. I., & Butler, A. H. (2019). The Teleconnection of El Niño Southern Oscillation to the Stratosphere. *Reviews of Geophysics*, *57*(1), 5–47. https://doi.org/10.1029/2018RG000596

Epstein, E. S. (1969). Stochastic dynamic prediction. *Tellus A: Dynamic Meteorology and Oceanography*, *21*(6), 739. https://doi.org/10.3402/tellusa.v21i6.10143

Falkena, S. K., de Wiljes, J., Weisheimer, A., & Shepherd, T. G. (2022). Detection of interannual ensemble forecast signals over the North Atlantic and Europe using atmospheric circulation regimes. *Quarterly Journal of the Royal Meteorological Society*, *148*(742), 434–453. https://doi.org/10.1002/qj.4213

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

Finkel, J., Gerber, E. P., Abbot, D. S., & Weare, J. (2023a). Revealing the Statistics of Extreme Events Hidden in Short Weather Forecast Data. *AGU Advances*, *4*(2), 1–16. https://doi.org/10.1029/2023AV000881

Finkel, J., Webber, R. J., Gerber, E. P., Abbot, D. S., & Weare, J. (2023b). Data-Driven Transition Path Analysis Yields a Statistical Understanding of Sudden Stratospheric Warming Events in an Idealized Model. *Journal of the Atmospheric Sciences*, *80*(2), 519–534. https://doi.org/10.1175/JAS-D-21-0213.1

Fischer, E. M., & Schär, C. (2009). Future changes in daily summer temperature variability: driving processes and role for temperature extremes. *Climate Dynamics*, *33*(7-8), 917–935. https://doi.org/10.1007/s00382-008-0473-8

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

Fujiwara, M., Manney, G. L., Gray, L. J., & Wright, J. S. (2022, January). *SPARC Reanalysis Intercomparison Project (S-RIP) Final Report* (M. Heckl, S. Zechlau, & B. Ziegele, Eds.; tech. rep.). https://doi.org/10.17874/800dee57d13
10th assessment report of the SPARC project, published by the International Project Office at DLR-IPA. also: WCRP Report 6/2021.

Garfinkel, C. I., & Hartmann, D. L. (2007). Effects of the El Niño–Southern Oscillation and the Quasi-Biennial Oscillation on polar temperatures in the stratosphere. *Journal of Geophysical Research: Atmospheres*, *112*(D19), 1–13. https://doi.org/10.1029/2007JD008481

Garfinkel, C. I., Benedict, J. J., & Maloney, E. D. (2014). Impact of the MJO on the boreal winter extratropical circulation. *Geophysical Research Letters*, *41*(16), 6055–6062. https://doi.org/10.1002/2014GL061094

Garfinkel, C. I., Chen, W., Li, Y., Schwartz, C., Yadav, P., & Domeisen, D. (2022). The Winter North Pacific Teleconnection in Response to ENSO and the MJO in Operational Subseasonal Forecasting Models Is Too Weak. *Journal of Climate*, *35*(24), 8013–8030. https://doi.org/10.1175/JCLI-D-22-0179.1

Garfinkel, C. I., Hartmann, D. L., & Sassi, F. (2010). Tropospheric Precursors of Anomalous Northern Hemisphere Stratospheric Polar Vortices. *Journal of Climate*, *23*(12), 3282–3299. https://doi.org/10.1175/2010JCLI3010.1

Garfinkel, C. I., Lawrence, Z. D., Butler, A. H., Dunn-sigouin, E., Statnaia, I., Karpechko, A. Y., Koren, G., Abalos, M., Ayarzagüena, B., & Barriopedro, D. (2024). A process-based evaluation of biases in extratropical stratosphere-troposphere coupling in subseasonal forecast systems. *Weather and Climate Dynamics Discussions*, *preprint*(June). https://doi.org/10.5194/egusphere-2024-1762

Garfinkel, C. I., Shaw, T. A., Hartmann, D. L., & Waugh, D. W. (2012). 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

Geisler, J. E. (1974). A numerical model of the sudden stratospheric warming mechanism. *Journal of Geophysical Research*, *79*(33), 4989–4999. https://doi.org/10.1029/JC079i033p04989

Gómara, I., Pinto, J. G., Woollings, T., Masato, G., Zurita-Gotor, P., & Rodríguez-Fonseca, B. (2014). Rossby wave-breaking analysis of explosive cyclones in the Euro-Atlantic sector. *Quarterly Journal of the Royal Meteorological Society*, *140*(680), 738–753. https://doi.org/10.1002/qj.2190

González-Alemán, J. J., Grams, C. M., Ayarzagüena, B., Zurita-Gotor, P., Domeisen, D. I. V., Gómara, I., Rodríguez-Fonseca, B., & Vitart, F. (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., Mueller, S., Chen, X., Chen, T.-C., Pinto, J. G., & Grams, C. M. (2023a). The Linkage of Serial Cyclone Clustering in Western Europe and Weather Regimes in the North Atlantic-European Region in Boreal Winter. *Geophysical Research Letters*, *50*(2), 1–10. https://doi.org/10.1029/2022GL101900

Hauser, S., Teubler, F., Riemer, M., Knippertz, P., & Grams, C. M. (2023b). 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., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hitchcock, P., & Simpson, I. R. (2014). The Downward Influence of Stratospheric Sudden Warmings*. *Journal of the Atmospheric Sciences*, *71*(10), 3856–3876. https://doi.org/10.1175/JAS-D-14-0012.1

Hitchcock, P., & Simpson, I. R. (2016). Quantifying eddy feedbacks and forcings in the tropospheric response to stratospheric sudden warmings. *Journal of the Atmospheric Sciences*, *73*(9), 3641–3657. https://doi.org/10.1175/JAS-D-16-0056.1

Holton, J. R., & Lindzen, R. S. (1972). An Updated Theory for the Quasi-Biennial Cycle of the Tropical Stratosphere. *Journal of the Atmospheric Sciences*, *29*(6), 1076–1080. https://doi.org/10.1175/1520-0469(1972)029<1076:AUTFTQ>2.0.CO;2

Holton, J. R., & Mass, C. (1976). Stratospheric Vacillation Cycles. *Journal of the Atmospheric Sciences*, *33*(11), 2218–2225. https://doi.org/10.1175/1520-0469(1976)033<2218:SVC>2.0.CO;2

Holton, J. R., & Tan, H.-C. (1980). The Influence of the Equatorial Quasi-Biennial Oscillation on the Global Circulation at 50 mb. *Journal of the Atmospheric Sciences*, *37*(10), 2200–2208. https://doi.org/10.1175/1520-0469(1980)037<2200:TIOTEQ>2.0.CO;2

Horel, J. D., & Wallace, J. M. (1981). Planetary-Scale Atmospheric Phenomena Associated with the Southern Oscillation. *Monthly Weather Review*, *109*(4), 813–829. https://doi.org/10.1175/1520-0493(1981)109<0813:PSAPAW>2.0.CO;2

Hoyer, S., & Hamman, J. (2017). xarray: N-D labeled Arrays and Datasets in Python. *Journal of Open Research Software*, *5*(1), 10. https://doi.org/10.5334/jors.148

Huang, J., Hitchcock, P., Maycock, A. C., McKenna, C. M., & Tian, W. (2021). Northern hemisphere cold air outbreaks are more likely to be severe during weak polar vortex conditions. *Communications Earth & Environment*, *2*(1), 147. https://doi.org/10.1038/s43247-021-00215-6

Ineson, S., Dunstone, N. J., Scaife, A. A., Andrews, M. B., Lockwood, J. F., & Pang, B. (2023). Statistics of sudden stratospheric warmings using a large model ensemble. *Atmospheric Science Letters*, (November), 1–10. https://doi.org/10.1002/asl.1202

Jucker, M., Fueglistaler, S., & Vallis, G. K. (2014). Stratospheric sudden warmings in an idealized GCM. *Journal of Geophysical Research: Atmospheres*, *119*(19). https://doi.org/10.1002/2014JD022170

Jucker, M., & Reichler, T. (2018). Dynamical Precursors for Statistical Prediction of Stratospheric Sudden Warming Events. *Geophysical Research Letters*, *45*(23), 13, 124–13, 132. https://doi.org/10.1029/2018GL080691

Karpechko, A. Y. (2018). Predictability of Sudden Stratospheric Warmings in the ECMWF Extended-Range Forecast System. *Monthly Weather Review*, *146*(4), 1063–1075. https://doi.org/10.1175/MWR-D-17-0317.1

Karpechko, A. Y., Afargan-Gerstman, H., Butler, A. H., Domeisen, D. I. V., Kretschmer, M., Lawrence, Z., Manzini, E., Sigmond, M., Simpson, I. R., & Wu, Z. (2022). Northern Hemisphere Stratosphere-Troposphere Circulation Change in CMIP6 Models: 1. Inter-Model Spread and Scenario Sensitivity. *Journal of Geophysical Research: Atmospheres*, *127*(18). https://doi.org/10.1029/2022JD036992

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), 13, 538–13, 546. https://doi.org/10.1029/2018GL081091

Karpechko, A. Y., Hitchcock, P., Peters, D. H. W., & Schneidereit, A. (2017). Predictability of downward propagation of major sudden stratospheric warmings. *Quarterly Journal of the Royal Meteorological Society*, *143*(704), 1459–1470. https://doi.org/10.1002/qj.3017

Kautz, L.-A., 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

Kelder, T., Müller, M., Slater, L. J., Marjoribanks, T. I., Wilby, R. L., Prudhomme, C., Bohlinger, P., Ferranti, L., & Nipen, T. (2020). Using UNSEEN trends to detect decadal changes in 100-year precipitation extremes. *npj Climate and Atmospheric Science*, *3*(1), 47. https://doi.org/10.1038/s41612-020-00149-4

Kidston, J., Scaife, A. A., Hardiman, S. C., Mitchell, D. M., Butchart, N., Baldwin, M. P., & Gray, L. J. (2015). Stratospheric influence on tropospheric jet streams, storm tracks and surface weather. *Nature Geoscience*, *8*(6), 433–440. https://doi.org/10.1038/ngeo2424

Kim, B.-M., Son, S.-W., Min, S.-K., Jeong, J.-H., Kim, S.-J., Zhang, X., Shim, T., & Yoon, J.-H. (2014). Weakening of the stratospheric polar vortex by Arctic sea-ice loss. *Nature Communications*, *5*(1), 4646. https://doi.org/10.1038/ncomms5646

Köhler, R. H., Jaiser, R., & Handorf, D. (2023). How do different pathways connect the stratospheric polar vortex to its tropospheric precursors? *Weather and Climate Dynamics*, *4*(4), 1071–1086. https://doi.org/10.5194/wcd-4-1071-2023

Kolstad, E. W., Breiteig, T., & Scaife, A. A. (2010). The association between stratospheric weak polar vortex events and cold air outbreaks in the Northern Hemisphere. *Quarterly Journal of the Royal Meteorological Society*, *136*(649), 886–893. https://doi.org/10.1002/qj.620

Kretschmer, M., Coumou, D., Donges, J. F., & Runge, J. (2016). Using Causal Effect Networks to Analyze Different Arctic Drivers of Midlatitude Winter Circulation. *Journal of Climate*, *29*(11), 4069–4081. https://doi.org/10.1175/JCLI-D-15-0654.1

Kretschmer, M., Runge, J., & Coumou, D. (2017). Early prediction of extreme stratospheric polar vortex states based on causal precursors. *Geophysical Research Letters*, *44*(16), 8592–8600. https://doi.org/10.1002/2017GL074696

Lang, S., Alexe, M., Chantry, M., Dramsch, J., Pinault, F., Raoult, B., Clare, M. C. A., Lessig, C., Maier-Gerber, M., Magnusson, L., Bouallègue, Z. B., Nemesio, A. P., Dueben, P. D., Brown, A., Pappenberger, F., & Rabier, F. (2024a). AIFS - ECMWF's data-driven forecasting system. http://arxiv.org/abs/2406.01465

Lang, S., Chantry, M., & The AIFS Team. (2024b). Enter the ensembles. https://www.ecmwf.int/en/about/media-centre/aifs-blog/2024/enter-ensembles

Lawrence, Z. D., Abalos, M., Ayarzagüena, B., Barriopedro, D., Butler, A. H., Calvo, N., de la Cámara, A., Charlton-Perez, A., Domeisen, D. I. V., Dunn-Sigouin, E., García-Serrano, J., Garfinkel, C. I., Hindley, N. P., Jia, L., Jucker, M., Karpechko, A. Y., Kim, H., Lang, A. L., Lee, S. H., … Wu, R. W.-Y. (2022). Quantifying stratospheric biases and identifying their potential sources in subseasonal forecast systems. *Weather and Climate Dynamics*, *3*(3), 977–1001. https://doi.org/10.5194/wcd-3-977-2022

Lawrence, Z. D., Perlwitz, J., Butler, A. H., Manney, G. L., Newman, P. A., Lee, S. H., & Nash, E. R. (2020). The Remarkably Strong Arctic Stratospheric Polar Vortex of Winter 2020: Links to Record-Breaking Arctic Oscillation and Ozone Loss. *Journal of Geophysical Research: Atmospheres*, *125*(22), 1–21. https://doi.org/10.1029/2020JD033271

Lee, R. W., Woolnough, S. J., Charlton-Perez, A. J., & Vitart, F. (2019a). ENSO Modulation of MJO Teleconnections to the North Atlantic and Europe. *Geophysical Research Letters*, *46*(22), 13535–13545. https://doi.org/10.1029/2019GL084683

Lee, S. H., Furtado, J. C., & Charlton-Perez, A. J. (2019b). Wintertime North American Weather Regimes and the Arctic Stratospheric Polar Vortex. *Geophysical Research Letters*, *46*(24), 14892–14900. https://doi.org/10.1029/2019GL085592

Lehtonen, I., & Karpechko, A. Y. (2016). Observed and modeled tropospheric cold anomalies associated with sudden stratospheric warmings. *Journal of Geophysical Research: Atmospheres*, *121*(4), 1591–1610. https://doi.org/10.1002/2015JD023860

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

L'Heureux, M. L., & Higgins, R. W. (2008). Boreal Winter Links between the Madden–Julian Oscillation and the Arctic Oscillation. *Journal of Climate*, *21*(12), 3040–3050. https://doi.org/10.1175/2007JCLI1955.1

Li, H., Chen, L., & Zhong, X. (2024a). A machine learning model that outperforms conventional global subseasonal forecast models. *PREPRINT (Version 1) available at Research Square*. ISBN: 0000000167062 https://doi.org/10.21203/rs.3.rs-3776375/v1

Li, L., Carver, R., Lopez-Gomez, I., Sha, F., & Anderson, J. (2024b). Generative emulation of weather forecast ensembles with diffusion models. *Science Advances*, *10*(13). https://doi.org/10.1126/sciadv.adk4489

Lilly, D. K. (1972). Numerical simulation studies of two-Dimensional turbulence: II. Stability and predictability studies. *Geophysical Fluid Dynamics*, *4*(1), 1–28. https://doi.org/10.1080/03091927208236087

Limpasuvan, V., & Hartmann, D. L. (2000). Wave-Maintained Annular Modes of Climate Variability*. *Journal of Climate*, *13*(24), 4414–4429. https://doi.org/10.1175/1520-0442(2000)013<4414:WMAMOC>2.0.CO;2

Lindgren, E. A., Sheshadri, A., & Plumb, R. A. (2018). Sudden Stratospheric Warming Formation in an Idealized General Circulation Model Using Three Types of Tropospheric Forcing. *Journal of Geophysical Research: Atmospheres*, *123*(18). https://doi.org/10.1029/2018JD028537

Lindzen, R. S., & Holton, J. R. (1968). A Theory of the Quasi-Biennial Oscillation. *Journal of the Atmospheric Sciences*, *25*(6), 1095–1107. https://doi.org/10.1175/1520-0469(1968)025<1095:ATOTQB>2.0.CO;2

Lorenz, E. N. (1969). The predictability of a flow which possesses many scales of motion. *Tellus A: Dynamic Meteorology and Oceanography*, *21*(3), 289. https://doi.org/10.3402/tellusa.v21i3.10086

Lorenz, E. N. (1972). Predictability: Does the Flap of a Butterfly's wings in Brazil Set Off a Tornado in Texas? *American Association for the Advancement of Science*. http://eaps4.mit.edu/research/Lorenz/Butterfly%7B%5C_%7D1972.pdf

Madden, R. A., & Julian, P. R. (1971). Detection of a 40–50 Day Oscillation in the Zonal Wind in the Tropical Pacific. *Journal of the Atmospheric Sciences*, *28*(5), 702–708. https://doi.org/10.1175/1520-0469(1971)028<0702:DOADOI>2.0.CO;2

Mariotti, A., Baggett, C., Barnes, E. A., Becker, E., Butler, A., Collins, D. C., Dirmeyer, P. A., Ferranti, L., Johnson, N. C., Jones, J., Kirtman, B. P., Lang, A. L., Molod, A., Newman, M., Robertson, A. W., Schubert, S., Waliser, D. E., & Albers, J. (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

Martius, O., Polvani, L. M., & Davies, H. C. (2009). Blocking precursors to stratospheric sudden warming events. *Geophysical Research Letters*, *36*(14), 1–5. https://doi.org/10.1029/2009GL038776

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

Matsuno, T. (1971). A Dynamical Model of the Stratospheric Sudden Warming. *Journal of the Atmospheric Sciences*, *28*(8), 1479–1494. https://doi.org/10.1175/1520-0469(1971)028<1479:ADMOTS>2.0.CO;2

Matthewman, N. J., & Esler, J. G. (2011). Stratospheric Sudden Warmings as Self-Tuning Resonances. Part I: Vortex Splitting Events. *Journal of the Atmospheric Sciences*, *68*(11), 2481–2504. https://doi.org/10.1175/JAS-D-11-07.1

McIntyre, M. E. (1982). How Well do we Understand the Dynamics of Stratospheric Warmings? *Journal of the Meteorological Society of Japan. Ser. II*, *60*(1), 37–65. https://doi.org/10.2151/jmsj1965.60.1_37

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

Molteni, F., Buizza, R., Palmer, T., & Petroliags, T. (1996). The ECMWF Ensemble Prediction System: Methodology and validation. *Quarterly Journal of the Royal Meteorological Society*, *122*(529), 73–119. https://doi.org/10.1256/smsqj.52904

Mori, M., & Watanabe, M. (2008). The Growth and Triggering Mechanisms of the PNA: A MJO-PNA Coherence. *Journal of the Meteorological Society of Japan. Ser. II*, *86*(1), 213–236. https://doi.org/10.2151/jmsj.86.213

National Academies of Sciences Engineering and Medicine. (2016, July). *Attribution of Extreme Weather Events in the Context of Climate Change*. National Academies Press. ISBN: 978-0-309-38094-2 https://doi.org/10.17226/21852

Oehrlein, J., Polvani, L. M., Sun, L., & Deser, C. (2021). How Well Do We Know the Surface Impact of Sudden Stratospheric Warmings? *Geophysical Research Letters*, *48*(22), 1–11. https://doi.org/10.1029/2021GL095493

Osman, M., Beerli, R., Büeler, D., & Grams, C. M. (2023). Multi-model assessment of subseasonal 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. N. (1981). Aspects of stratospheric sudden warmings studied from a transformed Eulerian-mean viewpoint. *Journal of Geophysical Research: Oceans*, *86*(C10), 9679–9687. https://doi.org/https://doi.org/10.1029/JC086iC10p09679

Palmer, T. (2019). The ECMWF ensemble prediction system: Looking back (more than) 25 years and projecting forward 25 years. *Quarterly Journal of the Royal Meteorological Society*, *145*(S1), 12–24. https://doi.org/10.1002/qj.3383

Palmer, T., Molteni, F., Mureau, R., Buizza, R., Chapalet, P., & Tribbia, J. (1992). Ensemble Prediction. *ECMWF Technical Memorandum*, *188*(July), 2546–2557. https://www.ecmwf.int/en/elibrary/75930-ensemble-prediction

Palmer, T. N., Buizza, R., Hagedorn, R., Lawrence, A., Leutbecher, M., & Smith, L. A. (2006). Ensemble prediction: a pedagogical perspective. *ECMWF Newsletter*, *106*(106), 10–17. https://doi.org/10.21957/ab129056ew

Pearl, J. (2009, September). *Causality* (2nd ed.). Cambridge University Press. ISBN: 9780511803161 https://doi.org/10.1017/CBO9780511803161

Peto, R. (2000). Smoking, smoking cessation, and lung cancer in the UK since 1950: combination of national statistics with two case-control studies. *BMJ*, *321*(7257), 323–329. https://doi.org/10.1136/bmj.321.7257.323

Pinto, J. G., Zacharias, S., Fink, A. H., Leckebusch, G. C., & Ulbrich, U. (2009). Factors contributing to the development of extreme North Atlantic cyclones and their relationship with the NAO. *Climate Dynamics*, *32*(5), 711–737. https://doi.org/10.1007/s00382-008-0396-4

Plumb, R. A. (1981). Instability of the Distorted Polar Night Vortex: A Theory of Stratospheric Warmings. *Journal of the Atmospheric Sciences*, *38*(11), 2514–2531. https://doi.org/10.1175/1520-0469(1981)038<2514:IOTDPN>2.0.CO;2

Polichtchouk, I., van Niekerk, A., & Wedi, N. (2023). Resolved Gravity Waves in the Extratropical Stratosphere: Effect of Horizontal Resolution Increase from O(10) to O(1) km. *Journal of the Atmospheric Sciences*, *80*(2), 473–486. https://doi.org/10.1175/JAS-D-22-0138.1

Polvani, L. M., Sun, L., Butler, A. H., Richter, J. H., & Deser, C. (2017). Distinguishing stratospheric sudden warmings from ENSO as key drivers of wintertime climate variability over the North Atlantic and Eurasia. *Journal of Climate*, *30*(6), 1959–1969. https://doi.org/10.1175/JCLI-D-16-0277.1

Price, I., Sanchez-Gonzalez, A., Alet, F., Andersson, T. R., El-Kadi, A., Masters, D., Ewalds, T., Stott, J., Mohamed, S., Battaglia, P., Lam, R., & Willson, M. (2023). GenCast: Diffusion-based ensemble forecasting for medium-range weather. (1050). http://arxiv.org/abs/2312.15796

Richardson, L. F. (1922). *Weather Prediction by Numerical Process.* Cambridge University Press.

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

Robertson, A. W., & Vitart, F. (2019). *Sub-Seasonal to Seasonal Prediction.* Elsevier. ISBN: 9780128117149 https://doi.org/10.1016/C2016-0-01594-2

Rodwell, M. J., Richardson, D. S., Parsons, D. B., & Wernli, H. (2018). Flow-Dependent Reliability: A Path to More Skillful Ensemble Forecasts. *Bulletin of the American Meteorological Society*, *99*(5), 1015–1026. https://doi.org/10.1175/BAMS-D-17-0027.1

Rotunno, R., & Snyder, C. (2008). A Generalization of Lorenz's Model for the Predictability of Flows with Many Scales of Motion. *Journal of the Atmospheric Sciences*, *65*(3), 1063–1076. https://doi.org/10.1175/2007JAS2449.1

Rupp, P., Spaeth, J., Afargan-Gerstman, H., Büeler, D., Sprenger, M., & Birner, T. (2024). The impact of synoptic storm likelihood on European subseasonal forecast uncertainty and their modulation by the stratosphere. *EGUsphere [preprint]*, 1–19. https://doi.org/10.5194/egusphere-2024-1423

Rupp, P., & Birner, T. (2021). Tropospheric eddy feedback to different stratospheric conditions in idealised baroclinic life cycles. *Weather and Climate Dynamics*, *2*(1), 111–128. https://doi.org/10.5194/wcd-2-111-2021

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). https://doi.org/10.1029/2021JD035667

Rupp, P., Spaeth, J., Garny, H., & Birner, T. (2023). Enhanced Polar Vortex Predictability Following Sudden Stratospheric Warming Events. *Geophysical Research Letters*, *50*(17). https://doi.org/10.1029/2023GL104057

Scaife, A. A., Comer, R. E., Dunstone, N. J., Knight, J. R., Smith, D. M., MacLachlan, C., Martin, N., Peterson, K. A., Rowlands, D., Carroll, E. B., Belcher, S., & Slingo, J. (2017). Tropical rainfall, Rossby waves and regional winter climate predictions. *Quarterly Journal of the Royal Meteorological Society*, *143*(702), 1–11. https://doi.org/10.1002/qj.2910

Scaife, A. A., & Smith, D. (2018). A signal-to-noise paradox in climate science. *npj Climate and Atmospheric Science*, *1*(1), 28. https://doi.org/10.1038/s41612-018-0038-4

Schär, C., Vidale, P. L., Lüthi, D., Frei, C., Häberli, C., Liniger, M. A., & Appenzeller, C. (2004). The role of increasing temperature variability in European summer heatwaves. *Nature*, *427*(6972), 332–336. https://doi.org/10.1038/nature02300

Scherhag, R. (1952). Die explosionsartigen Stratosphärenerwärmungen des Spätwinters 1951/52. *Berichte des Deutschen Wetterdienstes in der US-Zone*, *6*(38), 51–63.

Schneider, T., Bischoff, T., & Płotka, H. (2015). Physics of Changes in Synoptic Midlatitude Temperature Variability. *Journal of Climate*, *28*(6), 2312–2331. https://doi.org/10.1175/JCLI-D-14-00632.1

Schwartz, C., & Garfinkel, C. I. (2020). Troposphere-Stratosphere Coupling in Subseasonal-to-Seasonal Models and Its Importance for a Realistic Extratropical Response to the Madden-Julian Oscillation. *Journal of Geophysical Research: Atmospheres*, *125*(10). https://doi.org/10.1029/2019JD032043

Selz, T. (2019). Estimating the Intrinsic Limit of Predictability Using a Stochastic Convection Scheme. *Journal of the Atmospheric Sciences*, *76*(3), 757–765. https://doi.org/10.1175/JAS-D-17-0373.1

Selz, T., & Craig, G. C. (2015). Upscale Error Growth in a High-Resolution Simulation of a Summertime Weather Event over Europe*. *Monthly Weather Review*, *143*(3), 813–827. https://doi.org/10.1175/MWR-D-14-00140.1

Selz, T., Riemer, M., & Craig, G. C. (2022). The Transition from Practical to Intrinsic Predictability of Midlatitude Weather. *Journal of the Atmospheric Sciences*, *79*(8), 2013–2030. https://doi.org/10.1175/JAS-D-21-0271.1

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shaw, T. A., & Miyawaki, O. (2024). Fast upper-level jet stream winds get faster under climate change. *Nature Climate Change*, *14*(1), 61–67. https://doi.org/10.1038/s41558-023-01884-1

Siegert, S., Stephenson, D. B., Sansom, P. G., Scaife, A. A., Eade, R., & Arribas, A. (2016). A Bayesian Framework for Verification and Recalibration of Ensemble Forecasts: How Uncertain is NAO Predictability? *Journal of Climate*, *29*(3), 995–1012. https://doi.org/10.1175/JCLI-D-15-0196.1

Sigmond, M., Scinocca, J. F., Kharin, V. V., & Shepherd, T. G. (2013). Enhanced seasonal forecast skill following stratospheric sudden warmings. *Nature Geoscience*, *6*(2), 98–102. https://doi.org/10.1038/ngeo1698

Sirbescu, M. (2023). *Troposphärische Signale vor und nach plötzlichen Stratosphärenerwärmungen in Langzeit-Ensemble-Vorhersagen* [Master's Thesis]. Ludwig-Maximilians-University Munich.

Smith, K. L., & Scott, R. K. (2016). The role of planetary waves in the tropospheric jet response to stratospheric cooling. *Geophysical Research Letters*, *43*(6), 2904–2911. https://doi.org/10.1002/2016GL067849

Spaeth, J. (2020). *Statistical Analyses of Stratosphere-Troposphere Coupling in Extended-Range Ensemble Forecasts* [Master's Thesis]. Ludwig-Maximilians-Universität Munich. https://www.meteo.physik.uni-muenchen.de/DokuWiki/lib/exe/fetch.php?media=intern:abschlussarbeiten:2020:ma2020%7B%5C_%7Dspaeth%7B%5C_%7Djonas.pdf

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. (2024a). 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

Spaeth, J., Rupp, P., Osman, M., Grams, C. M., & Birner, T. (2024b). Flow-Dependence of Ensemble Spread of Subseasonal Forecasts Explored via North Atlantic-European Weather Regimes. *Geophysical Research Letters*, *51*(14). https://doi.org/10.1029/2024GL109733

Sprenger, M., Fragkoulidis, G., Binder, H., Croci-Maspoli, M., Graf, P., Grams, C. M., Knippertz, P., Madonna, E., Schemm, S., Škerlak, B., & Wernli, H. (2017). Global Climatologies of Eulerian and Lagrangian Flow Features based on ERA-Interim. *Bulletin of the American Meteorological Society*, *98*(8), 1739–1748. https://doi.org/10.1175/BAMS-D-15-00299.1

Stan, C., Zheng, C., Chang, E. K.-m., Domeisen, D. I. V., Garfinkel, C. I., Jenney, A. M., Kim, H., Lim, Y.-k., Lin, H., Robertson, A., Schwartz, C., Vitart, F., Wang, J., & Yadav, P. (2022). Advances in the Prediction of MJO Teleconnections in the S2S Forecast Systems. *Bulletin of the American Meteorological Society*, *103*(6), E1426–E1447. https://doi.org/10.1175/BAMS-D-21-0130.1

Tamarin-Brodsky, T., & Harnik, N. (2023). The intrinsic relationship between cyclones, anticyclones, and Rossby Wave Breakings in the North-Atlantic. *EGUsphere*, *preprint*, 1–30. https://doi.org/https://doi.org/10.5194/egusphere-2023-534
preprint.

Tamarin-Brodsky, T., Hodges, K., Hoskins, B. J., & Shepherd, T. G. (2019). A Dynamical Perspective on Atmospheric Temperature Variability and Its Response to Climate Change. *Journal of Climate*, *32*(6), 1707–1724. https://doi.org/10.1175/JCLI-D-18-0462.1

Taschetto, A. S., Ummenhofer, C. C., Stuecker, M. F., Dommenget, D., Ashok, K., Rodrigues, R. R., & Yeh, S.-W. (2020, November). ENSO Atmospheric Teleconnections. In *Geophysical monograph series* (pp. 309–335). https://doi.org/10.1002/9781119548164.ch14

Thompson, D. W. J., Baldwin, M. P., & Wallace, J. M. (2002). Stratospheric Connection to Northern Hemisphere Wintertime Weather: Implications for Prediction. *Journal of Climate*, *15*(12), 1421–1428. https://doi.org/10.1175/1520-0442(2002)015<1421:SCTNHW>2.0.CO;2

Thompson, D. W. J., & Wallace, J. M. (1998). The Arctic oscillation signature in the wintertime geopotential height and temperature fields. *Geophysical Research Letters*, *25*(9), 1297–1300. https://doi.org/10.1029/98GL00950

Thompson, V., Dunstone, N. J., Scaife, A. A., Smith, D. M., Slingo, J. M., Brown, S., & Belcher, S. E. (2017). High risk of unprecedented UK rainfall in the current climate. *Nature Communications*, *8*(1), 107. https://doi.org/10.1038/s41467-017-00275-3

Timmermann, A., An, S.-I., Kug, J.-S., Jin, F.-F., Cai, W., Capotondi, A., Cobb, K. M., Lengaigne, M., McPhaden, M. J., Stuecker, M. F., Stein, K., Wittenberg, A. T., Yun, K.-S., Bayr, T., Chen, H.-C., Chikamoto, Y., Dewitte, B., Dommenget, D., Grothe, P., ... Zhang, X. (2018). El Niño–Southern Oscillation complexity. *Nature*, *559*(7715), 535–545. https://doi.org/10.1038/s41586-018-0252-6

Tomassini, L., Gerber, E. P., Baldwin, M. P., Bunzel, F., & Giorgetta, M. (2012). The role of stratosphere-troposphere coupling in the occurrence of extreme winter cold spells over northern Europe. *Journal of Advances in Modeling Earth Systems*, *4*(4). https://doi.org/10.1029/2012MS000177

Toth, Z., & Buizza, R. (2019). Weather Forecasting: What Sets the Forecast Skill Horizon? In A. W. Robertson & F. Vitart (Eds.), *Sub-seasonal to seasonal prediction* (pp. 17–45). Elsevier. ISBN: 978-0-12-811714-9 https://doi.org/10.1016/B978-0-12-811714-9.00002-4

Trenberth, K. E. (1997). The Definition of El Niño. *Bulletin of the American Meteorological Society*, *78*(12), 2771–2777. https://doi.org/10.1175/1520-0477(1997)078<2771:TDOENO>2.0.CO;2

Tripathi, O. P., Charlton-Perez, A., Sigmond, M., & Vitart, F. (2015). Enhanced long-range forecast skill in boreal winter following stratospheric strong vortex conditions. *Environmental Research Letters*, *10*(10), 104007. https://doi.org/10.1088/1748-9326/10/10/104007

Vallis, G. K. (2017, June). *Atmospheric and Oceanic Fluid Dynamics*. Cambridge University Press. ISBN: 9781107065505 https://doi.org/10.1017/9781107588417

van den Brink, H. W., Können, G. P., Opsteegh, J. D., van Oldenborgh, G. J., & Burgers, G. (2004). Improving 10 4 -year surge level estimates using data of the ECMWF seasonal prediction system. *Geophysical Research Letters*, *31*(17), 1–4. https://doi.org/10.1029/2004GL020610

Vitart, F. (2017). Madden—Julian Oscillation prediction and teleconnections in the S2S database. *Quarterly Journal of the Royal Meteorological Society*, *143*(706), 2210–2220. https://doi.org/10.1002/qj.3079

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.-S., Kumar, A., Lin, H., Liu, G., Liu, X., Malguzzi, P., Mallas, I., ... Zhang, L. (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

Weisheimer, A., Baker, L. H., Bröcker, J., Garfinkel, C. I., Hardiman, S. C., Hodson, D. L. R., Palmer, T. N., Robson, J. I., Scaife, A. A., Screen, J. A., Shepherd, T. G., Smith, D. M., & Sutton, R. T. (2024). The Signal-to-Noise Paradox in Climate Forecasts: Revisiting Our Understanding and Identifying Future Priorities. *Bulletin of the American Meteorological Society*, *105*(3), E651–E659. https://doi.org/10.1175/BAMS-D-24-0019.1

Wernli, H., & Schwierz, C. (2006). Surface Cyclones in the ERA-40 Dataset (1958–2001). Part I: Novel Identification Method and Global Climatology. *Journal of the Atmospheric Sciences*, *63*(10), 2486–2507. https://doi.org/10.1175/JAS3766.1

Wheeler, M. C., & Hendon, H. H. (2004). An All-Season Real-Time Multivariate MJO Index: Development of an Index for Monitoring and Prediction. *Monthly Weather Review*, *132*(8), 1917–1932. https://doi.org/10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2

White, I. P., Garfinkel, C. I., Gerber, E. P., Jucker, M., Hitchcock, P., & Rao, J. (2020). The Generic Nature of the Tropospheric Response to Sudden Stratospheric Warmings. *Journal of Climate*, *33*(13), 5589–5610. https://doi.org/10.1175/JCLI-D-19-0697.1

Wicker, W., Polichtchouk, I., & Domeisen, D. I. V. (2023). Increased vertical resolution in the stratosphere reveals role of gravity waves after sudden stratospheric warmings. *Weather and Climate Dynamics*, *4*(1), 81–93. https://doi.org/10.5194/wcd-4-81-2023

Wilks, D. S. (2005). *Statistical Methods in the Atmospheric Sciences* (2nd Editio). Academic Press. ISBN: 9780127519661.

Williams, N. C., Scaife, A. A., & Screen, J. A. (2023). Underpredicted ENSO Teleconnections in Seasonal Forecasts. *Geophysical Research Letters*, *50*(5), 1–9. https://doi.org/10.1029/2022GL101689

Williams, P. D., Alexander, M. J., Barnes, E. A., Butler, A. H., Davies, H. C., Garfinkel, C. I., Kushnir, Y., Lane, T. P., Lundquist, J. K., Martius, O., Maue, R. N., Peltier, W. R., Sato, K., Scaife, A. A., & Zhang, C. (2017). A Census of Atmospheric Variability From Seconds to Decades. *Geophysical Research Letters*, *44*(21). https://doi.org/10.1002/2017GL075483

Wirth, V., & Eichhorn, J. (2014). Long-lived Rossby wave trains as precursors to strong winter cyclones over Europe. *Quarterly Journal of the Royal Meteorological Society*, *140*(680), 729–737. https://doi.org/10.1002/qj.2191

Wirth, V., Riemer, M., Chang, E. K. M., & Martius, O. (2018). Rossby Wave Packets on the Midlatitude Waveguide—A Review. *Monthly Weather Review*, *146*(7), 1965–2001. https://doi.org/10.1175/MWR-D-16-0483.1

Woolnough, S. J. (2019). The Madden-Julian Oscillation. In *Sub-seasonal to seasonal prediction* (pp. 93–117). Elsevier. ISBN: 9780128117149 https://doi.org/10.1016/B978-0-12-811714-9.00005-X

Yoo, C., & Son, S.-W. (2016). Modulation of the boreal wintertime Madden-Julian oscillation by the stratospheric quasi-biennial oscillation. *Geophysical Research Letters*, *43*(3), 1392–1398. https://doi.org/10.1002/2016GL067762

Zhang, C. (2005). Madden-Julian Oscillation. *Reviews of Geophysics*, *43*(2). https://doi.org/10.1029/2004RG000158

Zhang, F., Bei, N., Rotunno, R., Snyder, C., & Epifanio, C. C. (2007). Mesoscale Predictability of Moist Baroclinic Waves: Convection-Permitting Experiments and Multistage Error Growth Dynamics. *Journal of the Atmospheric Sciences*, *64*(10), 3579–3594. https://doi.org/10.1175/JAS4028.1

Zhang, F., Sun, Y. Q., Magnusson, L., Buizza, R., Lin, S.-J., Chen, J.-H., & Emanuel, K. (2019). What Is the Predictability Limit of Midlatitude Weather? *Journal of the Atmospheric Sciences*, *76*(4), 1077–1091. https://doi.org/10.1175/JAS-D-18-0269.1

Zhong, X., Chen, L., Li, H., Feng, J., & Lu, B. (2024). FuXi-ENS: A machine learning model for medium-range ensemble weather forecasting, 1–24. http://arxiv.org/abs/2405.05925

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

Acknowledgements

Firstly, I would like to thank Thomas Birner, who provided a foundation not only on a scientific level but also on a personal one, allowing me to feel truly welcomed and valued in the group as well as in the scientific community. I found it exciting to learn about the atmosphere, statistics, scientific writing, outreach, and much more. His supervision made doing science enjoyable, providing me with both freedom and consistent, helpful feedback.

Moreover, I am grateful to everyone in the group over the past years. I enjoyed our meetings, excursions and the generally pleasant atmosphere we shared. In particular, I would like to thank Hella Garny for the insightful discussions, for clearly explaining complex concepts, and for co-evaluating this thesis. I appreciate the trust of Paul Bauer and Marco Sirbescu, whose theses I co-supervised, making for a pleasant and rewarding collaboration that produced results I was able to further develop in this thesis.

Thanks to Philip Rupp, with whom it was fun to share an office, go for beers at conferences, and collaborate. I believe we had some cool ideas, which might suffice to add another four or five chapters to this thesis. Furthermore, thanks to Aman, Takumi, Frederik, Yvonne, Flo and Philip, it really was a pleasure to share an office with you! Thanks to Matjaž, Sören, Hasanain, and many more, who accompanied me during my PhD journey.

I would like to thank Inna Polichtchouk for giving me the opportunity to visit ECMWF, present my results there and engage with many experts who provided valuable feedback.

During this PhD project, I have met various scientific experts, in Munich, at meetings, and at conferences. Although I cannot list them all, I would kindly like to appreciate their significant input, particularly Prof. Edwin Gerber for his insightful comments on the UNSEEN approach as well as on the physical link between ensemble mean and spread.

I would also like to thank Marisol Osman, Christian Grams and Hilla Afargan-Gerstman for the seamless and productive collaborations.

I'd like to acknowledge the *Waves to Weather* project and its community, particularly George Craig, who led the project and whose knowledge and enthusiasm have been a true inspiration for young scientists, and Audine Laurian, who organized excellent meetings. I enjoyed being part of *Waves to Weather*, met new friends and benefitted from the scientific network.

Finally, I am deeply grateful to my friends, with whom I shared wonderful days in Munich, in the mountains, at the Bavarian lakes, in Sardinia, and beyond; to my sister, for always being supportive and positive, and also for some delightfully nerdy statistics discussions; and, of course, to my parents, for always supporting me in every way they could. I'd also like to acknowledge the patience of all of you when my answers to questions like "What's the weather going to be like? Will it be windy tomorrow?" often began with, "Well, it depends..." Last but not least, thank you, Franzi, for being by my side throughout my PhD journey, for always offering a listening ear, and for your love and support ♡.

Danke!