A minimal model to diagnose the contribution of the stratosphere to tropospheric forecast skill

Charlton-Perez, A. J. ORCID: https://orcid.org/0000-0001-8179-6220, Bröcker, J. ORCID: https://orcid.org/0000-0002-0864-6530, Karpechko, A. Y. ORCID: https://orcid.org/0000-0003-0902-0414, Lee, S. H. ORCID: https://orcid.org/0000-0003-0986-0093, Sigmond, M. ORCID: https://orcid.org/0000-0003-2191-9756 and Simpson, I. R. ORCID: https://orcid.org/0000-0002-2915-1377 (2021) A minimal model to diagnose the contribution of the stratosphere to tropospheric forecast skill. Journal of Geophysical Research: Atmospheres, 126 (24). ISSN 2169-8996 doi: 10.1029/2021JD035504 Available at https://centaur.reading.ac.uk/102160/

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

AGU ADVANCING EARTH AND SPACE SCIENCE

JGR Atmospheres

RESEARCH ARTICLE

Key Points:
- We propose a model that demonstrates how forecast skill present in the lowermost stratosphere contributes to tropospheric forecast skill
- The model can explain the greater correlation skill in the troposphere for forecasts during weak or strong vortex events
- The model shows how tropospheric skill arising from the stratosphere can sometimes be confounded by uncorrelated tropospheric signals

Correspondence to:
A. J. Charlton-Perez,
a.j.charlton-perez@reading.ac.uk

A Minimal Model to Diagnose the Contribution of the Stratosphere to Tropospheric Forecast Skill

Andrew J. Charlton-Perez[1], Jochen Bröcker[1,2], Alexey Yu. Karpechko[3], Simon H. Lee[1], Michael Sigmond[4], and Isla R. Simpson[5]

[1]Department of Meteorology, University of Reading, Reading, UK, [2]Department of Mathematics and Statistics, University of Reading, Reading, UK, [3]Finnish Meteorological Institute, Helsinki, Finland, [4]Canadian Centre for Climate Modelling and Analysis, Victoria, BC, Canada, [5]National Center for Atmospheric Research, Boulder, CO, USA

Abstract Many recent studies have confirmed that variability in the stratosphere is a significant source of surface sub-seasonal prediction skill during Northern Hemisphere winter. It may be beneficial, therefore, to think about times in which there might be windows-of-opportunity for skillfull sub-seasonal predictions based on the initial or predicted state of the stratosphere. In this study, we propose a simple, minimal model that can be used to understand the impact of the stratosphere on tropospheric predictability. Our model purposefully excludes state dependent predictability in either the stratosphere or troposphere or in the coupling between the two. Model parameters are set up to broadly represent current sub-seasonal prediction systems by comparison with four dynamical models from the Sub-Seasonal to Seasonal Prediction Project database. The model can reproduce the increases in correlation skill in sub-sets of forecasts for weak and strong lower stratospheric polar vortex states over neutral states despite the lack of dependence of coupling or predictability on the stratospheric state. We demonstrate why different forecast skill diagnostics can give a very different impression of the relative skill in the three sub-sets. Forecasts with large stratospheric signals and low amounts of noise are demonstrated to also be windows-of-opportunity for skillfull tropospheric forecasts, but we show that these windows can be obscured by the presence of unrelated tropospheric signals.

Plain Language Summary For successful forecasts of surface winter conditions between two weeks and one season ahead, the stratosphere has been shown to be a key source of information. Despite many studies examining how well the stratosphere can be predicted in computer-based forecasting systems, there remains a lack of understanding of which surface forecasts the stratosphere is most important for. This study is an attempt to step back from examining the role of the stratosphere in any particular forecasting system and instead to determine a simple framework that can be used to understand when and how the stratosphere is important. Using our framework we can construct a series of simple experiments that help to understand how important the stratosphere is in the longer range forecasting problem. Our experiments show that forecasts made during periods in which the Arctic stratospheric winds are unusually strong or weak have greater skill, but this does not depend on how unusually weak or strong the stratospheric winds are. The results are particularly important for thinking about the times in which longer range forecasts might be more skillfull than on average, so called windows-of-opportunity, and how these depend on the stratosphere.

1. Introduction and Motivation

On sub-seasonal and seasonal timescales, coupling between the stratosphere and troposphere is a significant part of the available tropospheric forecast skill (e.g., Domeisen, Butler, et al., 2020; Scaife et al., 2016). The contrasting impact of stratospheric variability during recent winter seasons with similar stratospheric disturbances (Knight et al., 2021) has, however, brought into sharp relief the lack of a full quantitative understanding of the stratospheric contribution to tropospheric prediction skill. The Sudden Stratospheric Warming (SSW) which occurred in February 2018 has been clearly linked to enhanced sub-seasonal predictive skill of cold conditions in Europe during late-winter and spring (Karpechko et al., 2018; Kautz et al., 2020). In contrast, the SSW which occurred in January 2019 is not thought to have had a major surface impact or have contributed to enhanced skill (Rao et al., 2020). Early examination of the January 2021 SSW (Lee, 2021) suggest this event was strongly coupled to the surface Northern Annular Mode (NAM). Some authors have also proposed links to weather impacts in both Texas and Greece (Wright et al., 2021). The exceptionally strong and predictable polar vortex during the 2019/20 season also seems to have enhanced tropospheric seasonal forecast skill during late winter

and spring (Lee et al., 2020). Explanations for these differences often focus on several different aspects of the stratosphere-troposphere coupling. On the one hand, there is evidence that not all SSWs produce the necessary lower stratospheric signals associated with strong coupling to the surface (Karpechko et al., 2017). Related work notes a difference in the tropospheric response to the morphology of SSW, vortex displacement or vortex split, with enhanced coupling following splitting events (White et al., 2020). Other recent work suggests tropospheric drivers of sub-seasonal skill, unconnected to the stratosphere, significantly influence the sign and predictability of the tropospheric response (e.g., Afargan-Gerstman & Domeisen, 2020; Knight et al., 2021). It has also been proposed that coupling between the stratosphere and troposphere is strongly dependent on the tropospheric state at the time of the stratospheric perturbation (Charlton-Perez et al., 2018; Domeisen, Grams, & Papritz, 2020; Maycock et al., 2020). These ideas relate, more generally, to the concept of intermittent "windows-of-opportunity" for sub-seasonal prediction (Mariotti et al., 2020). In order to critically assess and quantify the importance of these ideas, it would be helpful to have a simple "toy" or "null" model of how predictability in the stratosphere and troposphere is connected. By this, we mean a simple stochastic model that captures the relationships between the predictable signal in the stratosphere and troposphere with as few parameters as possible. A toy model with a minimal set of parameters is important, because it provides a quantitative lens through which to examine explanations of why some stratospheric perturbations appear to have a larger impact on tropospheric predictability than others. Put another way, are the differences in the impact on tropospheric skill of recent SSW and strong vortex events just an effect of random variability in stratosphere-troposphere coupling or do they reflect more complex dynamics? To the best of our knowledge, no such null model exists. When considering tropospheric predictability more generally, the simple models of, for example, Weigel et al. (2008) and Siegert et al. (2016) have been widely used for similar purposes. We choose to call this model a "minimal" model since it contains what we believe is the simplest description of the coupled distribution of forecasts and observations in the stratosphere and troposphere. Having proposed a minimal model of the stratospheric contribution to tropospheric predictability, in the second half of the paper we perform a number of simple thought experiments to show the extent to which many of the characteristic properties of stratosphere-troposphere coupling can be reproduced by this simple model. Given the limited size of hindcast datasets that can be exploited to understand the properties of stratosphere-troposphere coupling, we also hope that this minimal model can be used to test and refine diagnostic tools for examining coupling in operational prediction systems.

2. Model Design

We start from the signal-noise model for ensemble forecasts developed by Charlton-Perez et al. (2019) from Siegert et al. (2016):

$$Y(t) = \mu_y + \beta_y S(t) + \varepsilon O(t) \tag{1}$$

$$X_k(t) = \mu_x + \beta_x S(t) + \eta P_k(t) \qquad \text{for } k = 1, \ldots, K \tag{2}$$

In this model, $Y(t)$ is the observed time series of the parameter of interest for forecasts made at different times, t. $X_k(t)$ are the matching ensemble forecasts; $S(t), O(t), P_1(t), \ldots, P_K(t)$ are independent standard normal random variables that are also independent over time (i.e., for different t); μ_y and μ_x are the climatological means. Key to the model is the shared "signal" term that is identical in the forecasts and observations, $S(t)$. The noise terms, $O(t), P_1, \ldots, P_K(t)$ are uncorrelated with $S(t)$ and with each other. The two parameters β_y and β_x scale the signal term, allowing for under or over confidence in the forecasts. The ε and η terms similarly scale the noise components. This model can be further simplified by considering the case where the forecast and observations are normalised climate indices (e.g., the North Atlantic Oscillation index, NAO) with mean of zero and variance of one. In this case μ_y and μ_x are zero and can be removed. If the variance of $Y(t)$ and $X_k(t)$ are known to be one, then the amplitude of the signal and noise terms are related (since the variance of the sum of uncorrelated variables is the sum of the variances of each variable):

$$\beta_y^2 = 1 - \varepsilon^2, \tag{3}$$

$$\beta_x^2 = 1 - \eta^2. \tag{4}$$

To extend this model to consider coupling between the stratosphere and troposphere we adopt a series of principles and assumptions:

1. The stratospheric observations and forecasts should have identical structure to the model above.
2. The predictable signal in the troposphere should have two uncorrelated components: a predictable signal linked linearly to the stratospheric state and one which captures skill from purely tropospheric processes, such as the Madden-Julian Oscillation (MJO) or from other Earth System processes like sea-ice, sea-surface temperature or soil moisture.
3. Coupling between the stratosphere and troposphere in the model forecasts is linked to the full stratospheric state in each ensemble member, not solely the predictable component.
4. The tropospheric and stratospheric variables should represent normalised climate indices. We consider these to be the NAM index in the lower stratosphere (for example 100 hPa) where forecast skill is high (Son et al., 2020) and the NAO index, but any normalised climate index would be equally appropriate.

By adopting these design choices, the model does not seek to explicitly consider upward coupling that is, the role of the tropospheric state in determining the predictable signal in the stratosphere. Anomalous stratospheric states have been shown to be a strong function of the integrated lower stratospheric meridional heatflux (Hinssen & Ambaum, 2010; Polvani & Waugh, 2004). The model does not seek to capture this process. Since on sub-seasonal and seasonal timescales, the lower stratosphere is the most predictable part of the extra-tropical atmosphere and predicting the tropospheric state is the ultimate goal of any forecasting system it is natural to design the model in this way. Quoting a reviewer of the paper, in a sentiment we agree with, this means we "do not seek to produce a model which can be judged "wrong" due to a lack of realism rather one which is relevant and useful." Other authors may wish to produce a model that includes other interesting effects such as upward coupling, interactions between the mean state and wave part of the flow or state dependence of stratosphere-troposphere coupling. Our choice here is to sacrifice model complexity in order to make the model easy to understand and conduct experiments with.

The proposed model is then as follows:

$$Y_S(t) = \beta_y S(t) + \varepsilon O(t), \tag{5}$$

$$X_{Sk}(t) = \beta_x S(t) + \eta P_k(t) \qquad \text{for} \quad k = 1, \dots, K, \tag{6}$$

$$Y_T(t) = C_y Y_S(t) + \alpha_y T(t) + \lambda Q(t), \tag{7}$$

$$X_{Tk}(t) = C_x X_{Sk}(t) + \alpha_x T(t) + \xi R_k(t) \qquad \text{for} \quad k = 1, \dots, K, \tag{8}$$

where an added subscript S means stratosphere and T means troposphere. The predictable signal in the troposphere, which is common to the forecasts and observations but uncorrelated with the stratospheric signal $S(t)$ is denoted $T(t)$. In the real world, some part of the tropospheric and stratospheric signal may be correlated with a common forcing, for example, from the tropics (Barnes et al., 2019), and so this is an additional way in which the model is minimal rather than realistic. α_y and α_x are the amplitude of $T(t)$ in the observations and model respectively. $Q(t), R_1(t), \dots, R_K(t)$ $Q(t)$ are the noise terms in the troposphere. They are uncorrelated with each other and with the noise terms in the stratosphere. λ and ξ are the amplitude of the tropospheric noise terms. By substitution, the expressions for the tropospheric observations and forecast can then be expanded as:

$$Y_T(t) = C_y \beta_y S(t) + C_y \varepsilon O(t) + \alpha_y T(t) + \lambda Q(t),$$

$$X_{Tk}(t) = C_x \beta_x S(t) + C_x \eta P_k(t) + \alpha_x T(t) + \xi R_k(t) \qquad \text{for} \quad k = 1, \dots, K.$$

the variance, covariance and correlation structure of the model, along with expression for the signal-to-noise ratio, are shown in the Supporting Information S1. Since the forecasts and observations are normalised, with variance of one, the correlation and covariance between individual ensemble members and the observations are equal.

$$\rho(X_{Sk}, Y_S) = \beta_x \beta_y \tag{9}$$

$$\rho(Y_S, Y_T) = C_y \tag{10}$$

$$\rho(X_{Sk}, X_{Tk}) = C_x \tag{11}$$

$$\rho(X_{Tk}, Y_T) = C_y \beta_y C_x \beta_x + \alpha_y \alpha_x \tag{12}$$

Table 1
Week 3 Correlation Structure Measured From Samples of Sub-Seasonal Forecasts

Center (model version)	Number of forecasts, ensemble size	Stratosphere-troposphere correlation (observations)	Stratosphere-troposphere correlation (forecasts)	Forecast-observation correlation (stratosphere)	Forecast-observation correlation (troposphere)	Ensemble mean correlation (stratosphere)	Ensemble mean correlation (troposphere)
	N, K	$\rho(Y_S, Y_T)$	$\rho(X_{Sk}, X_{Tk})$	$\rho(X_{Sk}, Y_S)$	$\rho(X_{Tk}, Y_T)$	$\rho(\overline{X_S}, Y_S)$	$\rho(\overline{X_T}, Y_T)$
ECMWF (cy45r1)	700, 11	0.44	0.46	0.63	0.24	0.77	0.43
UKMO (GloSea5)	384, 7	0.43	0.44	0.67	0.24	0.75	0.36
NCEP (CFS v2)	467, 12	0.45	0.41	0.61	0.26	0.73	0.44
Meteo-France (cnrm-cm 6.0)	352, 15	0.43	0.49	0.60	0.26	0.72	0.42
Assumed representative system		0.45	0.45	0.6	0.25		

Note. Here, Week 3 means the average of days 14–20 of the forecast, and correlations are calculated by first taking the mean value of the index over these days. $\rho(X_{Sk}, X_{Tk})$, $\rho(X_{Sk}, Y_S)$ and $\rho(X_{Tk}, Y_T)$ are the ensemble mean of the correlation in each individual ensemble member. N is the number of forecast initialisations considered, K is the number of ensemble members. The final line shows summary values assumed in our minimal model calculations which should represent a typical sub-seasonal forecasting system. No values are assumed for the ensemble size or number of forecasts since this variable is explored in later calculations, also changing the ensemble mean correlation.

For the stratosphere, the Pearson's correlation, ρ, between each ensemble member and the observations (Equation 9) is simply the product of the two signal amplitudes, $\beta_y \beta_x$. The correlation between the stratospheric and tropospheric index in the observations (Equation 10) is C_y, and for each ensemble member (Equation 11) is C_x. The correlation between each ensemble member and the observations in the troposphere (Equation 12) has a component which depends on both the strength of the coupling between the model and observations and the size of the stratospheric signal ($C_y \beta_y C_x \beta_x$) and a component that depends on the size of the tropospheric signal ($\alpha_y \alpha_x$).

3. Model Parameters

To complete the minimal model, we need to set values of the parameters in Equations 5–8. As in Siegert et al. (2016), a Bayesian approach could be used to fit the model to a set of model hindcasts to determine these parameters. In this study, since the aim is to describe and analyze the simple model we do not take this approach but it is a clear extension. Instead, as a simple starting point, we can examine the correlation structure of four sets of example hindcasts available from the sub-seasonal to seasonal project database (S2S, Vitart et al., 2017). As a representative stratospheric index we chose the NAM at 100 hPa derived using the zonal mean principal component method of Baldwin and Thompson (2009). As a representative tropospheric index we chose the NAO, here defined as the mean sea-level pressure difference between a 2.5° × 2.5° grid box centered at 65N and 20W and one centered at 37.5N and 25W, that is, over Iceland and the Azores respectively. In both cases, these indices are chosen to be illustrative only and different choices would result in slightly different numerical values for the calculations. All hindcasts in the database initialized between November and February for the particular model version in question are considered. No attempt is made to standardize the period over which the forecasts are made. For the NCEP CFS model, a lagged ensemble is created by combining forecasts initialized over three consecutive days. The lead time dependent bias is removed, prior to analysis. For the week 3 forecast, Table 1 shows the correlation structure.

Broadly, the correlation structure in all four models is similar for these climate indices and this lead time:

1. The correlation between the stratosphere and troposphere in the observations, $\rho(Y_S, Y_T)$, is approximately 0.45.
2. The correlation between the stratosphere and troposphere in the models, $\rho(X_{Sk}, X_{Tk})$ is also approximately 0.45.
3. The correlation between the observed NAM and an individual ensemble member, $\rho(X_{Sk}, Y_S)$, is approximately 0.6.
4. The correlation between the observed NAO and an individual ensemble member, $\rho(X_{Tk}, Y_T)$, is approximately 0.25.

Using these four representative correlations, we can derive example parameters for the minimal model. For the remainder of the manuscript, all calculations and estimates assume these values, and no further reference to the four sets of sub-seasonal forecasts is made. We first assume that the amplitude of the signal in the observations and model is the same in the stratosphere ($\beta_y = \beta_x$, a perfect model assumption which is relaxed later). We can use the representative correlation $\rho(X_{Sk}, Y_S) = 0.6$ above to calculate values for the stratospheric parameters from Equation 9 and Equations 3 and 4. In this case, $\beta_y = \beta_x = 0.77$, $\epsilon = \eta = 0.63$. This is equivalent to an identical signal-to-noise ratio in the model and observations of $SNR_{Sy} = SNR_{Sx} = 1.22$ (see the Supporting Information S1). Note that we have assumed positive values for β_y and β_x, although they could produce the same positive correlation if both values were negative. If we assume that the model has these parameter values, we can calculate a representative correlation between the ensemble mean forecast and the observations by assuming a value for the ensemble size, K (for details of the calculation of the ensemble mean correlation see the Supporting Information S1). For an example ensemble size of 51 members (which corresponds to the size of the operational ECMWF ensemble forecast), the correlation between the ensemble mean forecast and observations ($\rho(\overline{X_S}, Y_S)$) is 0.77. Making a further assumption that the amplitude of the uncoupled part of the tropospheric signal is the same in the model and observations (i.e., that α_y and α_x are the same) and using Equation 12, we can derive the following parameters for the tropospheric part of the minimal model. The size of the tropospheric signal, $\alpha_y = \alpha_x = 0.36$ and the residual noise terms $\lambda = \xi = 0.82$. These values give an overall signal-to-noise ratio in the troposphere, $SNR_{Ty} = SNR_{Tx} = 0.58$ and for an assumed ensemble size of 51 members the correlation between the ensemble mean forecast and observations ($\rho(\overline{X_T}, Y_T)$) is 0.49. In the next sections, these model parameters are further perturbed to explore the importance of the stratosphere-troposphere coupling for sub-seasonal predictability. Later in the paper, the minimal model with the same parameters is used to generate synthetic forecast sets with 1 million forecast initialisations.

4. Thought Experiment 1: Impact of Stratospheric Skill Improvement and Increased Ensemble Size

A common question posed about the importance of the stratosphere for sub-seasonal prediction is understanding the trade off between spending additional computational resources to improve stratospheric skill (for example by increasing the complexity of the gravity wave drag parameterization or enhancing model vertical resolution) versus using the same resources to increase the ensemble size. While it is difficult to quantify the impact of any given model improvement on forecast skill, the minimal model does give some insight into this question. Assuming that improving skill in the model stratosphere does not affect tropospheric predictability from other sources (i.e., the α terms remain constant) or the coupling between the stratosphere and troposphere (i.e., the C terms remain constant), the impact of increased stratospheric skill ($\delta\rho$) is proportional to $\rho(X_{Sk}, Y_S)$ multiplied by $C_x C_y = 0.45^2 = 0.20$ for parameters representative of the four sub-seasonal forecast models above. So (from Equations 9 and 12),

$$\delta\rho(X_{Tk}, Y_T) \propto 0.2\delta\rho(X_{Sk}, Y_S)$$

The model can be used to anticipate the impact of this increased skill on the ensemble mean correlation skill as shown in the Supporting Information S1. Assuming all other parameters are fixed, for an ensemble size of 51 members, modifying $\rho(X_{Sk}, Y_S)$ between 0 (no skill) and 1 (perfect skill) results in $\rho(\overline{X_T}, Y_T)$ increasing from 0.34 to 0.56. Further calculations of the impact of changes in stratospheric skill are shown in Figure 1a, with the case with 51 ensemble members shown in the solid line. The close to linear increase of $\rho(\overline{X_T}, Y_T)$ for the range of values around $\rho(X_{Sk}, Y_S) = 0.6$ typical of most S2S modeling systems does not depend strongly on the size of the model ensemble (see for example the dashed and dotted lines showing the cases with 101 and 11 ensemble members respectively). The contrasting impact of increasing ensemble size on $\rho(\overline{X_T}, Y_T)$ is shown in Figure 1b. The impact of the increasing ensemble size is relatively large as the ensemble size increases between 11 and 50 members but begins to saturate for much larger sizes. This effect is also not strongly dependent on the skill of the stratospheric forecast, with a similar dependence for $\rho(\overline{X_T}, Y_T)$ when $\rho(X_{Sk}, Y_S)$ is equal to 0.4 (dotted line) or 0.8 (dashed line). Note that the quantification here is only relevant for forecasting systems with similar signal-to-noise properties as the minimal model in this configuration. Further experiments could use the same framework to explore how to target model investment in under or over confident forecasting systems, which we will consider next.
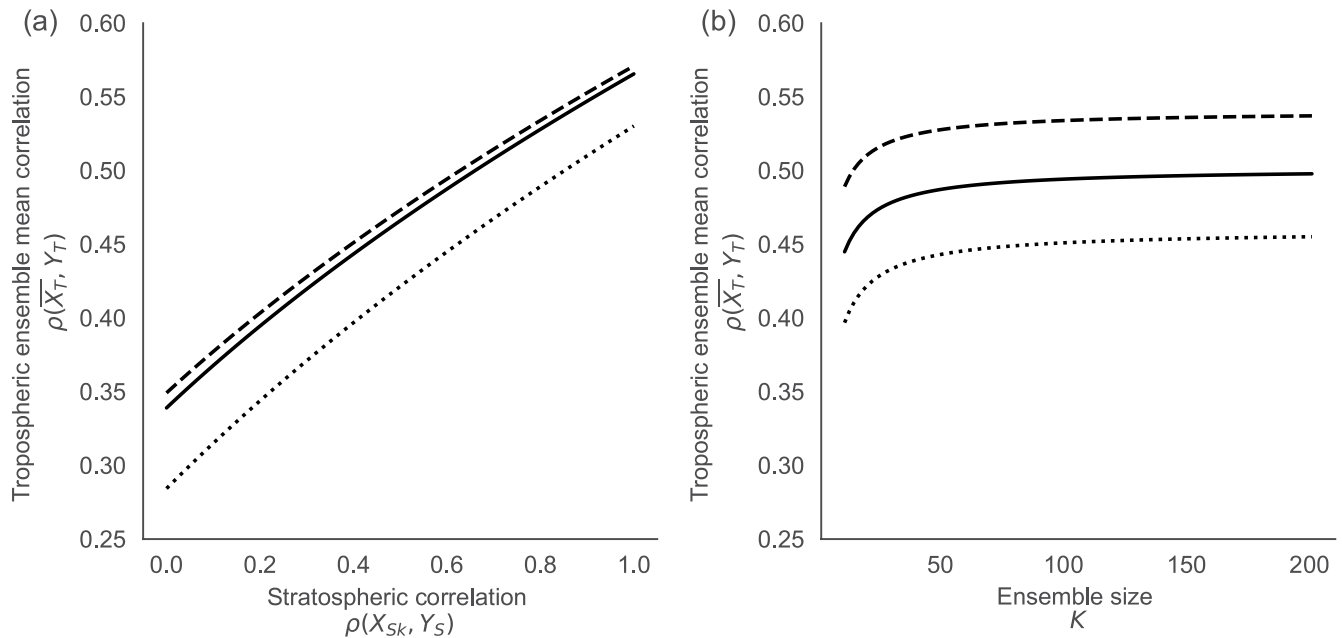
Figure 1. Impact of improving model skill in the stratosphere and increasing ensemble size on tropospheric ensemble mean skill. (a) Shows the resulting tropospheric ensemble mean skill plotted against the imposed correlation between an individual ensemble member and the observations in the stratosphere for an ensemble with 11 ensemble members (dotted line), 51 ensemble members (solid line) and 101 ensemble members (dashed line). (b) Shows the impact of changing the ensemble size (K) on tropospheric ensemble mean skill where correlation between an individual ensemble member and the observations in the stratosphere ($\rho(X_{Sk}, Y_S)$) is 0.4 (dotted line), 0.6 (solid line) and 0.8 (dashed line). All other parameters follow the base model.

5. Thought Experiment 2: Differences in Signal-To-Noise Ratio Between Forecasts and Observations

Given that most forecasting systems are not explicitly designed to perturb the stratosphere (or target error growth there) then one could imagine a case whereby the stratospheric forecast is over confident. Equally there has been much discussion in the literature of under confidence of forecast systems on seasonal and longer timescales, particularly in the North Atlantic and for the NAO (Eade et al., 2014). Using the minimal model, there are a number of different ways to explore the relationships between correlation and signal-to-noise ratio in this simple system, diverging from the perfect ensemble approach in Thought Experiment 1. One approach would be to fix the value of the signal-to-noise ratio in the observations and perturb the signal-to-noise ratio of the forecast ensemble members (not shown). Another approach, is to fix the value of the correlation between the forecasts and observations in the stratosphere and examine the different values of signal-to-noise ratio and tropospheric correlation which are consistent with these values. This set of experiments is reported here, since we assume that the correlation between forecasts and observations in the stratosphere ($\rho(X_{Sk}, Y_S)$) is mostly robustly known (note that in the four systems above, there is little variation in this parameter). By assuming the correlation is known, this means that different cases can only be generated by varying the SNR in both the observations and forecasts, so in the extreme cases the SNR in the observations is either much greater or much smaller than is assumed in the base case in thought experiment 1.

In the base case above, we assume that the size of the signal in the model stratosphere is equal to that of the observations, that is, $\beta_x = \beta_y = \beta_S = \sqrt{\rho(X_{Sk}, Y_S)}$.

We can relax this assumption, without changing $\rho(X_{Sk}, Y_S)$ by introducing a parameter, c, which represents the fractional change in the size of β_S for the base case above that is, $\beta_y = c\beta_S$ and $\beta_x = \frac{1}{c}\beta_S$.

The impact of model under or overconfidence on the skill of tropospheric forecasts, for an ensemble size of 51 members, is shown in Figure 2. For clarity, when we refer to model under confidence, this means that the signal-to-noise ratio of the model is smaller than the signal-noise ratio of the observations and vice versa for model over confidence. As might be expected, the tropospheric signal-to-noise ratio somewhat follows the signal-to-noise
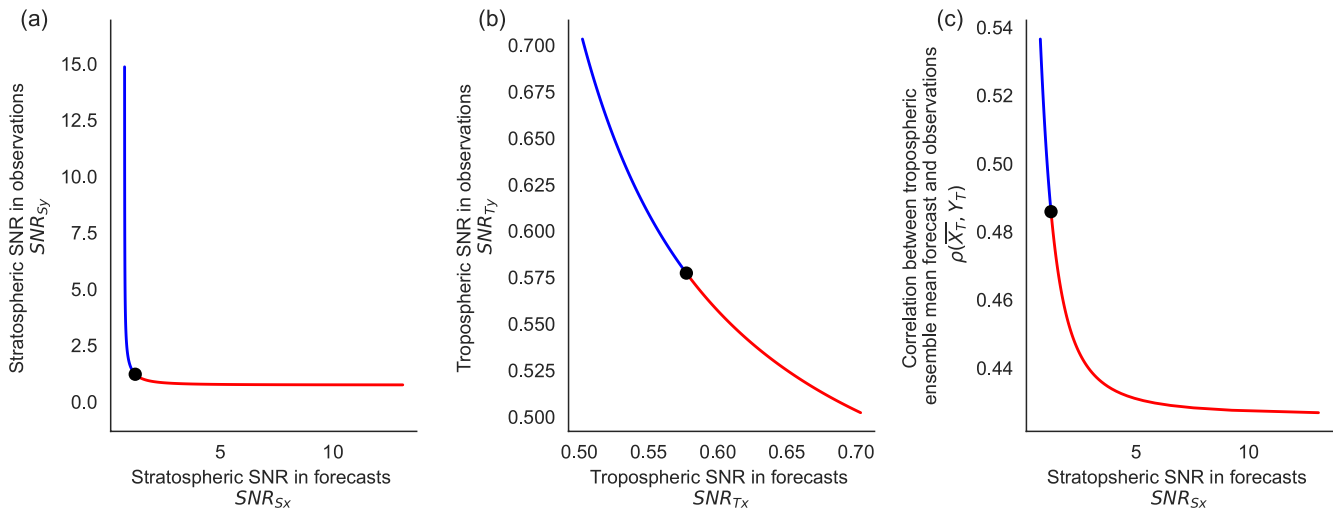
Figure 2. Impact of stratospheric signal-to-noise ratio biases on the tropospheric forecast (for an ensemble size of 51 members). In all plots, blue lines show cases where the model is under confident and red lines where the model is over confident. The black dot shows the base solution where the size of the signal in model and observations is equal. (a) Shows the possible range of signal-to-noise ratio for the model (x-axis) and observations (y-axis) when the stratospheric model skill is fixed as the same value as the base model. (b) Shows the same quantities in the troposphere. (c) Shows the tropospheric ensemble mean skill plotted against the signal-to-noise ratio in the model stratosphere.

ratio in the stratosphere. Where the model is under confident in the stratosphere (blue part of the lines) it is also under confident (although to a lesser degree) in the troposphere. The relative size of the different terms mean that for the range in the center of the plot, the impact in the troposphere is modest. Figure 2c demonstrates the impact of stratospheric under or over confidence on the correlation skill of the tropospheric ensemble mean. Compared to the case in which the amplitude of the stratospheric signal-to-noise ratio is correct in the model (black dot), $\rho(\overline{X_T}, Y_T)$ is reduced for an ensemble with an over confident stratosphere and increased for an under confident stratosphere. Due to the fact that:

$$\rho(\overline{X_T}, Y_T) = \frac{C_x C_y \beta_x \beta_y + \alpha_x \alpha_y}{\sqrt{1 - \frac{K-1}{K}\left(\xi^2 + C_x^2 \eta^2\right)}},$$

and since in this experiment, all parameters apart from η, β_x and β_y are fixed and the product $\beta_x \beta_y$ is fixed, this dependence is related to the value of η. In the over confident case, η is small, increasing the denominator and therefore decreasing the correlation in the expression above. In the limiting case, where stratospheric forecasts have no noise (but there is still an unpredictable, noise component in the observations), $\eta = 0$ and $\rho(\overline{X_T}, Y_T) = \frac{\rho(X_{Tk}, Y_T)}{\sqrt{1 - \frac{K-1}{K}(\xi^2)}} = 0.41$.

6. Thought Experiment 3: Perfect Stratospheric Forecasts

An increasingly common method used to interrogate the stratosphere-troposphere coupling in models is to add artificial physics to the model to nudge the state in the model stratosphere towards the observed state (e.g., Douville, 2009; Hitchcock & Simpson, 2014). The minimal model can be used to think about what these experiments might reveal. Although the minimal model is, of course, much simpler than the real world it can be used to think about limits to the gain in correlation skill that nudging experiments could yield.

This thought experiment starts from a hypothetical case of a perfectly predictable stratosphere in which there is no noise. In this case, $X_{Sk} = Y_S$ and so $\rho(X_{Sk}, Y_S) = 1$. The only way this can be achieved in the minimal model is if $\beta_x, \beta_y = 1$ and $\varepsilon, \eta = 0$. This then implies that $\rho(X_{Tk}, Y_T) = C_x C_y + \alpha_x \alpha_y$.

It follows that for the same set of parameters used in Thought Experiment 1, $\rho(X_{Tk}, Y_T) = 0.33$ and $\rho(\overline{X_T}, Y_T) = 0.56$ with 51 ensemble members, for a perfect, deterministic stratospheric forecast. In practice, stratospheric nudging could never achieve this perfect forecast state, partly because the strength of nudging required would lead to

numerical instability, but it predicts an upper limit for tropospheric correlation that might be achieved in a series of nudging experiments. This value should be compared to the tropospheric ensemble mean correlation derived from the model with standard parameters, for an identical ensemble size of 51 ensemble members this value is 0.49 as stated in Section 3. In other words, the minimal model predicts a maximum increase of 0.07 in correlation skill for the week 3 sub-seasonal forecast for a set of forecasts with strong stratospheric nudging, compared to free running control experiments.

7. Synthetic Experiment 1: Skill for Weak, Neutral and Strong Vortex Cases in the Stratosphere

A common method to explore the impact of the stratosphere on prediction skill is to separate forecasts into categories where the initial stratospheric state has a weak, strong or neutral vortex (as in Domeisen, Butler, et al., 2020; Sigmond et al., 2013; Son et al., 2020; Tripathi et al., 2015). One approach to quantify the differences in skill between the forecasts is to use the correlation skill score:

$$CSS = \frac{\frac{1}{M} \sum_{m=1}^{M} Y \cdot \overline{X}}{\sqrt{\frac{1}{M} \sum_{m=1}^{M} Y^2 \cdot \frac{1}{M} \sum_{m=1}^{M} \overline{X}^2}}. \tag{13}$$

In this and subsequent expressions, M indicates the number of forecast initialisations in each subset. An alternative is to measure the correlation between the observations and ensemble mean forecasts within each sub-selected ensemble, which we call here the sub-set correlation (SSC)

$$SSC = \frac{\frac{1}{M} \sum_{m=1}^{M} (Y - [Y]) \cdot (\overline{X} - [\overline{X}])}{\sqrt{\frac{1}{M} \sum_{m=1}^{M} (Y - [Y])^2 \cdot \frac{1}{M} \sum_{m=1}^{M} (\overline{X} - [\overline{X}])^2}} \tag{14}$$

Here, $[Y]$, $[\overline{X}]$ are, respectively, the mean of the observations and ensemble-mean forecasts within each sub-set. A further measure used to quantify the differences in skill of the different sub-sets of forecasts (Domeisen, Butler, et al., 2020) is the Root Mean Square Error.

$$RMSE = \sqrt{\frac{1}{M} \sum_{m=1}^{M} (Y - \overline{X})^2} \tag{15}$$

To simulate these calculations, and explore their relationship with the predictable signal, we can sub-set synthetic forecasts by the observed stratospheric state (Y_s). In the studies referenced above, sub-setting is normally performed on the observed state at the start of the forecast. Since the minimal model does not simulate the time development of the forecast, this experiment assumes that the state during week 3 is well correlated with the initial state, which is a reasonable assumption given the long autocorrelation timescale in the lower stratosphere. As an illustrative example, we define weak and strong cases as being below the 20th percentile or above the 80th percentile of the index, and generate one million synthetic forecasts using simple random draws for the signal and noise terms in Equations 5–8.

The minimal model reproduces the behavior seen in real forecast ensembles. The CSS for the weak and strong sub-sets is significantly larger than the neutral sub-set in both the stratosphere and troposphere. Note also that the CSS and SSC are equal in the neutral sub-set but that the CSS is substantially higher than the SSC in the weak and strong subsets. This difference doesn't reflect greater correlation within each sub-set. Rather it represents a shift of the PDF of the signal term. Since the signal term is common to the observations and forecasts, this results in a larger CSS for the weak and strong cases. Put another way, the larger CSS in the weak and strong sub-sets reflects their larger signal-to-noise ratio. Even more simply put, forecasts made during weak and strong stratospheric states can have greater CSS than forecasts made during neutral states even when the correlation between the stratosphere and troposphere is identical for all the sub-sets as is explicitly the case in the minimal model.

These results shed new light on the results of Sigmond et al. (2013), who compared the forecast skill of a group of forecasts initialized during SSW events and a control case with no SSWs. The CSS reported in that study for
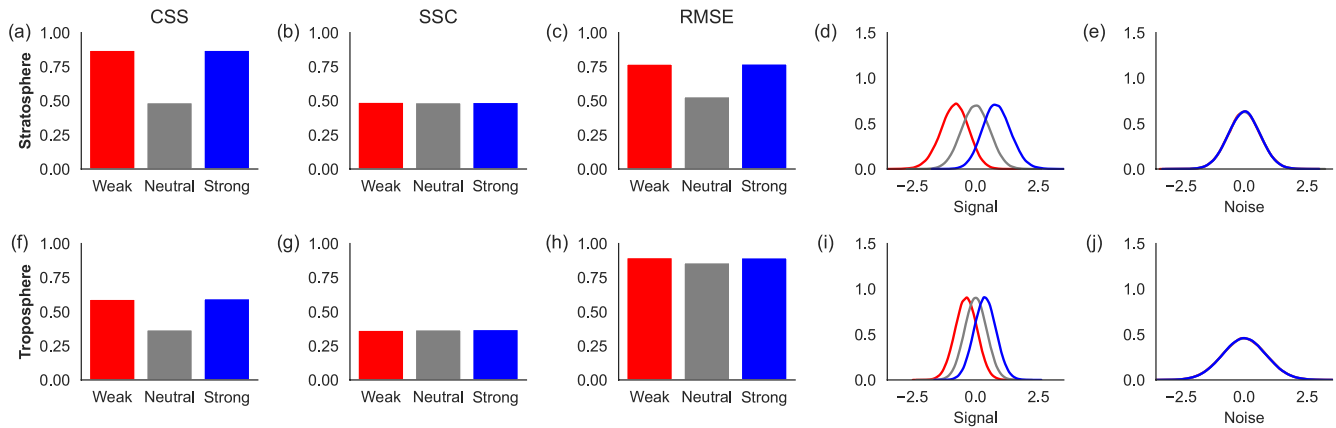
Figure 3. Upper row shows diagnostics for the stratospheric forecast and lower row shows diagnostics for the tropospheric forecasts. (a) & (f) show the CSS for the three observed sub-sets (with weak, neutral and strong conditions in the stratosphere). (b) & (g) show the SSC and (c) & (h) show the RMSE. (d) & (i) show Kernel Density Estimates (KDE) of the signal and (e) & (j) show noise terms for the three sub-sets. In the troposphere, the signal term includes signal from intrinsic tropospheric processes and due to coupling with the stratosphere.

the NAM at 1,000 hPa is 0.55 for the SSW case and −0.01 for the control case, a difference that is statistically significant (p < 0.01). We reanalyzed the Sigmond et al. (2013) results and found that the SSC is 0.23 for the SSW case and −0.02 for the control case, a difference that is not statistically significant. This implies that the difference in CSS reported in that study is mainly due to a shift of the PDF of the signal term, and is not the result of a (significantly) greater correlation within the SSW sub-set compared to the control sub-set.

In contrast, the RMSE in the weak and strong sub-sets is much larger than the neutral sub-set for the stratospheric forecasts. The difference in RMSE between weak/strong and neutral sub-sets is replicated in the troposphere, although the relative size of the difference between the sub-sets is smaller. Since the signal term is common to the observations and forecasts, the RMSE for the stratosphere depends only on the properties of the noise terms, $\varepsilon O(t)$ and $\frac{\eta}{K} \sum_{k=1}^{K} P_k(t)$. Sub-setting the forecasts by the magnitude of Y_S means that the distribution of $\varepsilon O(t)$ is biased towards negative or positive values in the weak or strong sub-sets (see Supporting Information S1). There is no corresponding bias in $\sum_{k=1}^{K} P_k(t)$. As the RMSE is a property of the distribution of $Y_S - \overline{X_S}$ and the distribution is dominated by the distribution of $\varepsilon O(t)$ this leads to the difference in RMSE demonstrated in Figure 3.

Another approach to assess skill in different states is to make the sub-sets on the basis of the ensemble mean forecast rather than the observed state (Figure 4). Sub-setting on this basis, produces a very similar result to the one in Figure 3 for CSS and SSC, but with a clean separation of the three states by the size of the signal term through the effective elimination of the noise term when taking the ensemble average. Since there is no bias in $\varepsilon O(t)$ in the three sub-sets introduced by this method, the RMSE is identical for the three sub-sets and very close to ε.
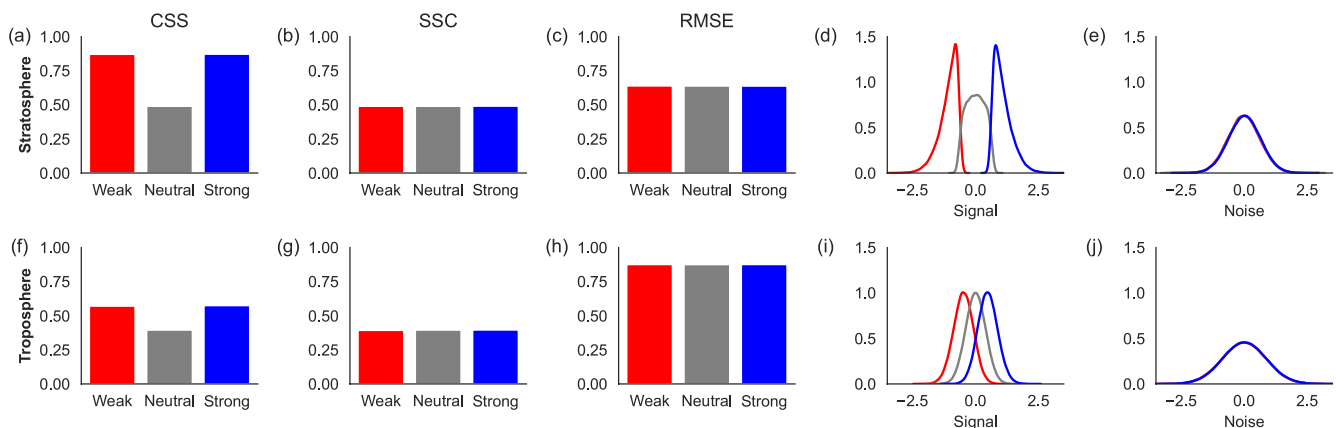

Figure 4. As Figure 3 but for sub-setting based on the forecast ensemble mean state in the stratosphere.
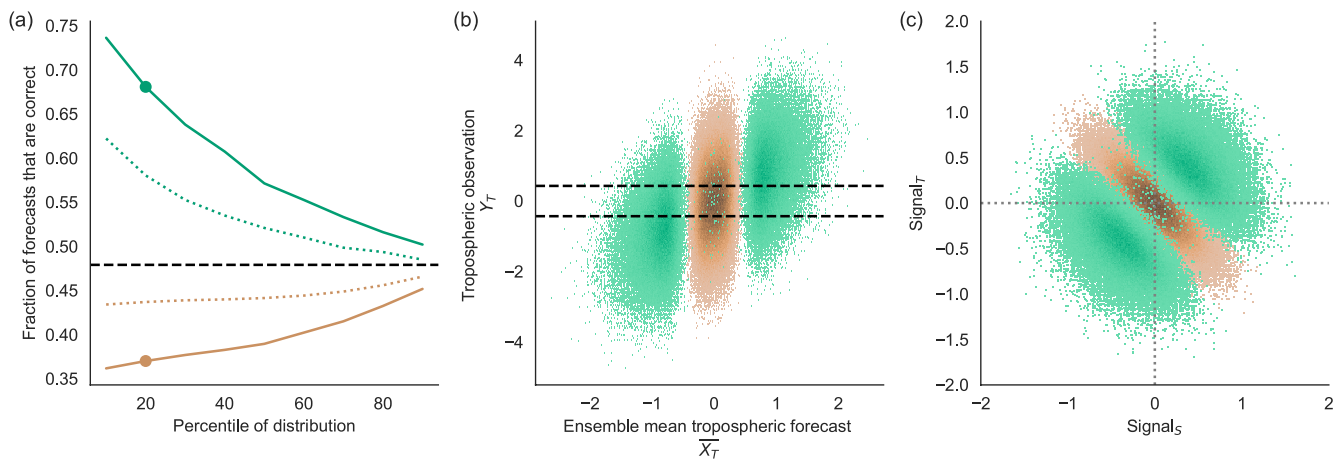
Figure 5. (a) Shows the fraction of tropospheric forecasts which are correct for two categories of forecast, those classified as highly (green) opoorly (brown) predictable. Results are presented as a function of the percentile used to define each class. For example, the value plotted at 20% on the green line is for all forecasts with modal occupation frequency greater than the highest 20% of forecasts and the value plotted at 20% on the brown line is for all forecasts with modal occupation frequency smaller than the lowest 20% of forecasts. The dashed lines show the same calculation, but with forecasts sub-set based on the modal category of the stratospheric forecast. (b) Shows 2D histograms of Y_T and $\overline{X_T}$ for the two sub-sets based on the tropospheric forecasts for the 20% case (dots in the left panel). Dashed black lines show the values of the observed state that define the weak, neutral and strong forecast categories. (c) Shows histograms of the part of the tropospheric signal due to the stratosphere (on the *x*-axis) and intrinsic to the troposphere (on the *y*-axis).

When analyzing real hindcast data, there are arguments for using either of these methods to compare skill in different sub-sets. When comparing forecasts sub-set based on the initial state in the stratosphere, this analysis shows that caution is needed when interpreting simple measures of forecast skill, something well-known in the forecast verification community but perhaps not so widely appreciated when considering atmospheric processes, such as stratosphere-troposphere coupling. The arguments presented in this section are not unique to coupling between the stratosphere and troposphere. Our model could also be applied to other cases in which a predictable component is weakly coupled to the extra-tropical troposphere. Examples might include coupling of the MJO and El Niño Southern Oscillation to the North Atlantic.

8. Synthetic Experiment 2: Windows of Opportunity

There has been a lot of recent interest in understanding when there might be "windows of opportunity" (Mariotti et al., 2020) for sub-seasonal forecasts that are more skillfull than on average based on the presence of a particular dynamical forcing. The minimal model can be used to explore to what extent the stratospheric signal can provide windows of opportunity when this stratosphere-troposphere coupling is linear and independent of the tropospheric state.

Since there is no widely agreed diagnostic of a window of opportunity, here a simple diagnostic from Ziehmann (2001) is used. First observations and forecasts are assigned to categorical bins (in this case based on terciles of the observed tropospheric state). The forecast "state" for each forecast is the modal category (i.e., the one with most forecasts, with random assignment for rare bimodal cases). A forecast is counted as successful when the forecast and observed states are the same, with skill measured simply as the fraction of forecasts for which this is true. Windows of opportunity can be explored by sub-setting the forecasts based on the occupation frequency of the modal category. A more confident forecast is one in which the number of forecasts in the modal category is large, a less confident forecast where the number of forecasts in the modal category is small. Figure 5a shows the result of this calculation. The dashed black line shows the average fraction of successful forecasts. The green solid line shows the success rate of forecasts with high confidence, as a function of the percentile of the number of members in the modal category used to define the "high confidence" sub-set. Forecasts with high confidence are much more likely to be successful than an average forecast. Conversely, forecasts with low confidence, shown in the brown line, are much less likely to be successful. In other words, the forecast spread, as quantified here by the number of members in the modal category, is a good predictor of forecast windows of opportunity.

How much are the windows of opportunity due to the signal present in the stratosphere? One way to quantify this effect is to repeat the calculation but sub-set the forecasts based on the size of the modal category in the stratospheric forecast. This is shown in the dashed lines in Figure 5a. While not as good a predictor of skill as the size of the modal category of the tropospheric forecast, this diagnostic can also be used to identify windows of opportunity. The set of forecasts in the high and low confidence sub-sets for the case shown by the dots in Figure 5a are shown in Figure 5b. Forecasts with high confidence (green) are those in which the ensemble mean is generally large and positive or negative. Since the signal and noise terms are uncorrelated by construction, when the signal term is large, the likelihood of more members of the ensemble being in the same category as the ensemble mean forecast and the observations is increased. Confident forecasts generally have a large signal resulting from the stratosphere with a large tropospheric signal of the same sign, as shown in Figure 5c. Forecasts with low confidence include both those with little signal from either process, and cases where there is an opposing signal from the stratosphere and troposphere (brown points). Forecasts in which there is a large stratospheric signal are therefore windows of opportunity for skillfull tropospheric sub-seasonal predictions. Sometimes, as seen for example, in the contrasting forecasts of the 2018 and 2019 SSW events, opposing stratospheric and tropospheric signals might mask this predictability.

9. Conclusions

In this study, we have attempted to define and investigate a minimal model which describes how skillfull forecasts in the stratosphere contribute to forecast skill in the troposphere. The model is developed from the earlier toy model of Siegert et al. (2016). The key addition to the model to allow the link between the stratosphere and troposphere to be examined is a term coupling the observed and forecast indices in the troposphere with those in the stratosphere. This coupling is independent of the state in the stratosphere or troposphere (i.e., it doesn't depend on the value of Y_s or X_{Sk}), and should be thought of as the simplest possible representation of stratosphere-troposphere coupling. There is no reason to think that this model rules out the need for more complex explanations of the contribution of the stratosphere to tropospheric forecast skill, but it should be regarded as a minimum standard that more complex explanations should be judged against. The simplicity of the model, and its similarity to other simple models of forecasts and observations mean that the conclusions about conditional skill and windows of opportunity that we highlight below are likely to be familiar to a reader who is well versed in thinking about ensemble forecasts and skill scores. Indeed, many of the conclusions below could be derived from a model like that of Weigel et al. (2008) or Siegert et al. (2016) in which only the relationship between a set of forecasts and observations are considered. A key message of this paper is that this body of understanding and literature is also relevant to the case of stratosphere-troposphere and other types of coupling if that coupling is independent of the underlying state of the system. For the broader community involved in analyzing and understanding weak coupling between a relatively well predictable component of the Earth system like the stratosphere and a more weakly predictable component like the extra-tropical troposphere we hope that the model provides a useful interpretive framework with which to consider sets of real sub-seasonal forecasts.

The model reproduces a number of features of the observed properties of real sub-seasonal and seasonal prediction systems, particularly when considering sub-sets of forecasts during weak, neutral and strong lower stratospheric NAM states. The increased CSS for these sub-sets reflects the greater signal-to-noise ratio in these cases. By construction, and as demonstrated by the calculations of SSC, correlation between forecast and observed states is identical within each sub-set. The analysis of the minimal model also demonstrates that care should be taken when constructing sub-sets of forecasts. Choosing to sub-set based on the observed stratospheric state can lead to biases in the RMSE because this method inherently chooses cases with larger than average noise. An alternative approach for analysis of sub-seasonal forecasts could be to sub-set based on the ensemble mean forecast in the stratosphere since this better isolates cases with a large predictable signal.

In a similar vein, windows-of-opportunity for skillfull tropospheric forecasts can be identified by considering the spread of forecasts in the stratosphere. Results from the simple model suggest that focusing detailed dynamical analysis on stratospheric forecasts with high confidence could be a way to identify windows-of-opportunity for skillfull sub-seasonal and seasonal forecasts. Often, confidence in stratospheric forecasts is largest once the signal of vortex disturbances is present in the upper and middle stratosphere. For the parameter choices used in this study, the similar size of the tropospheric signal derived from coupling to the stratosphere and from other unre-

lated tropospheric processes mean that there can often be confounding between the two signals. If this model is a good representation of real forecasting systems, this means that on the sub-seasonal timescale, the development of methods to disaggregate these signals could be an important forecast post-processing tool.

In the future, we aim to use Bayesian methods to fit this model to sub-seasonal and seasonal hindcast datasets in order to compare and contrast different prediction systems. The synthetic forecast datasets that can be easily generated by the minimal model could be used to test other ideas about the predictability associated with stratosphere-troposphere coupling. Obvious examples are tests of the windows-of-opportunity for tropospheric forecasts produced by models that have lower signal-to-noise ratio than the real world and comparison of a broader range of skill metrics to test their performance in detecting model skill derived from stratosphere-troposphere coupling.

Data Availability Statement

Sub-seasonal forecast data used in the study was obtained from the sub-seasonal to seasonal (S2S) archive which can be found at: https://apps.ecmwf.int/datasets/.

References

Afargan-Gerstman, H., & Domeisen, D. I. V. (2020). Pacific modulation of the North Atlantic storm track response to sudden stratospheric warming events. *Geophysical Research Letters*, *47*(2). https://doi.org/10.1029/2019gl085007

Baldwin, M. P., & Thompson, D. W. (2009). A critical comparison of stratosphere–troposphere coupling indices. *Quarterly Journal of the Royal Meteorological Society*, *135*(644), 1661–1672. https://doi.org/10.1002/qj.479

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

Charlton-Perez, A. J., Bröcker, J., Stockdale, T. N., & Johnson, S. (2019). When and where do ECMWF seasonal forecast systems exhibit anomalously low signal-to-noise ratio? *Quarterly Journal of the Royal Meteorological Society*, *145*(725), 3466–3478.

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.

Domeisen, D. I., Butler, A. H., Charlton-Perez, A. J., Ayarzagüena, B., Baldwin, M. P., et al. (2020). The role of the stratosphere in sub-seasonal to seasonal prediction: 2. Predictability arising from stratosphere-troposphere coupling. *Journal of Geophysical Research*, *125*(2), e2019JD030923. https://doi.org/10.1029/2019jd030923

Domeisen, D. I. V., Grams, C. M., & Papritz, L. (2020). 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

Douville, H. (2009). Stratospheric polar vortex influence on Northern Hemisphere winter climate variability. *Geophysical Research Letters*, *36*(18), L18703. https://doi.org/10.1029/2009gl039334

Eade, R., Smith, D., Scaife, A., Wallace, E., Dunstone, N., Hermanson, L., & Robinson, N. (2014). Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophysical Research Letters*, *41*(15), 5620–5628. https://doi.org/10.1002/2014gl061146

Hinssen, Y. B. L., & Ambaum, M. H. P. (2010). Relation between the 100-hPa heat flux and stratospheric potential vorticity. *Journal of the Atmospheric Sciences*, *67*(12), 4017–4027. https://doi.org/10.1175/2010jas3569.1

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

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. https://doi.org/10.1029/2018gl081091

Karpechko, A. Y., Hitchcock, P., Peters, D. H., & 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

Knight, J., Scaife, A., Bett, P. E., Collier, T., Dunstone, N., Gordon, M., et al. (2021). Predictability of European winters 2017/2018 and 2018/2019: Contrasting influences from the tropics and stratosphere. *Atmospheric Science Letters*, *22*(1), e1009. https://doi.org/10.1002/asl.1009

Lee, S. H. (2021). The January 2021 sudden stratospheric warming. *Weather*, *76*(4), 135–136. https://doi.org/10.1002/wea.3966

Lee, S. H., Lawrence, Z. D., Butler, A. H., & Karpechko, A. Y. (2020). Seasonal forecasts of the exceptional northern Hemisphere winter of 2020. *Geophysical Research Letters*, *47*(21), e2020GL090328. https://doi.org/10.1029/2020gl090328

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

Maycock, A. C., Masukwedza, G. I., Hitchcock, P., & Simpson, I. R. (2020). A regime perspective on the North Atlantic eddy-driven jet response to sudden stratospheric warmings. *Journal of Climate*, *33*(9), 3901–3917. https://doi.org/10.1175/jcli-d-19-0702.1

Polvani, L. M., & Waugh, D. W. (2004). Upward wave activity flux as a precursor to extreme stratospheric events and subsequent anomalous surface weather regimes. *Journal of Climate*, *17*(18), 3548–3554. https://doi.org/10.1175/1520-0442(2004)017<3548:uwafaa>2.0.co;2

Rao, J., Garfinkel, C. I., & White, I. P. (2020). Predicting the downward and surface influence of the February 2018 and January 2019 sudden stratospheric warming events in subseasonal to seasonal (S2S) models. *Journal of Geophysical Research*, *125*(2), e2019JD031919. https://doi.org/10.1029/2019jd031919

Scaife, A., Karpechko, A. Y., Baldwin, M., Brookshaw, A., Butler, A., Eade, R., et al. (2016). Seasonal winter forecasts and the stratosphere. *Atmospheric Science Letters*, *17*(1), 51–56. https://doi.org/10.1002/asl.598

Siegert, S., Stephenson, D. B., Sansom, P. G., Scaife, A. A., Eade, R., & Arribas, A. (2016). A Bayesian framework for verification and re-calibration 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., Kharin, V., & Shepherd, T. (2013). Enhanced seasonal forecast skill following stratospheric sudden warmings. *Nature Geoscience*, *6*(2), 98–102. https://doi.org/10.1038/ngeo1698

Son, S.-W., Kim, H., Song, K., Kim, S.-W., Martineau, P., Hyun, Y.-K., & Kim, Y. (2020). Extratropical prediction skill of the Subseasonal-to-Seasonal (S2S) prediction models. *Journal of Geophysical Research*, *125*(4), e2019JD031273. https://doi.org/10.1029/2019jd031273

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

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

Weigel, A. P., Liniger, M., & Appenzeller, C. (2008). Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts? *Quarterly Journal of the Royal Meteorological Society*, *134*(630), 241–260. https://doi.org/10.1002/qj.210

White, I. P., Garfinkel, C. I., Cohen, J., Jucker, M., & Rao, J. (2020). The impact of split and displacement sudden stratospheric warmings on the troposphere. *Journal of Geophysical Research*, *126*, e2020JD033989

Wright, C. J., Hall, R. J., Banyard, T. P., Hindley, N. P., Mitchell, D. M., & Seviour, W. J. (2021). Dynamical and surface impacts of the January 2021 sudden stratospheric warming in novel Aeolus wind observations, MLS and ERA5. *Weather and Climate Dynamics Discussions*, *1–39*.

Ziehmann, C. (2001). Skill prediction of local weather forecasts based on the ECMWF ensemble. *Nonlinear Processes in Geophysics*, *8*(6), 419–428. https://doi.org/10.5194/npg-8-419-2001