University of Reading

Early prediction of extreme stratospheric polar vortex states based on causal precursors

Article

Published Version

Kretschmer, M. ORCID: https://orcid.org/0000-0002-2756-9526, Runge, J. ORCID: https://orcid.org/0000-0002-0629-1772 and Coumou, D. ORCID: https://orcid.org/0000-0003-2155-8495 (2017) Early prediction of extreme stratospheric polar vortex states based on causal precursors. Geophysical Research Letters, 44 (16). pp. 8592-8600. ISSN 0094-8276 doi: 10.1002/2017GL074696 Available at https://centaur.reading.ac.uk/92434/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.
Published version at: http://dx.doi.org/10.1002/2017GL074696
To link to this article DOI: http://dx.doi.org/10.1002/2017GL074696

Publisher: American Geophysical Union

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER
10.1002/2017GL074696

Key Points:
- A new scheme to detect causal precursors useful for empirical predictions is introduced
- Subseasonal causal precursors of the stratospheric polar vortex are identified
- Enables skillful predictions of extremely weak and strong vortex states at long lead times

Supporting Information:
- Supporting Information S1

Correspondence to:
M. Kretschmer,
kretschmer@pik-potsdam.de

Citation:
Kretschmer, M., J. Runge, and D. Coumou (2017), Early prediction of extreme stratospheric polar vortex states based on causal precursors, *Geophys. Res. Lett.*, *44*, 8592–8600. doi:10.1002/2017GL074696.

Received 28 JUN 2017
Accepted 25 JUL 2017
Accepted article online 2 AUG 2017
Published online 16 AUG 2017

©2017. American Geophysical Union. All Rights Reserved.

Early prediction of extreme stratospheric polar vortex states based on causal precursors

Marlene Kretschmer[1,2], Jakob Runge[3], and Dim Coumou[1,4]

[1]Potsdam Institute for Climate Impact Research, Earth System Analysis, Potsdam, Germany, [2]Department of Physics, University of Potsdam, Potsdam, Germany, [3]Imperial College, Grantham Institute, London, UK, [4]Department of Water and Climate Risk, Institute for Environmental Studies (IVM), VU University Amsterdam, Amsterdam, Netherlands

Abstract Variability in the stratospheric polar vortex (SPV) can influence the tropospheric circulation and thereby winter weather. Early predictions of extreme SPV states are thus important to improve forecasts of winter weather including cold spells. However, dynamical models are usually restricted in lead time because they poorly capture low-frequency processes. Empirical models often suffer from overfitting problems as the relevant physical processes and time lags are often not well understood. Here we introduce a novel empirical prediction method by uniting a response-guided community detection scheme with a causal discovery algorithm. This way, we objectively identify causal precursors of the SPV at subseasonal lead times and find them to be in good agreement with known physical drivers. A linear regression prediction model based on the causal precursors can explain most SPV variability ($r^2 = 0.58$), and our scheme correctly predicts 58% (46%) of extremely weak SPV states for lead times of 1–15 (16–30) days with false-alarm rates of only approximately 5%. Our method can be applied to any variable relevant for (sub)seasonal weather forecasts and could thus help improving long-lead predictions.

1. Introduction

The extratropical stratosphere in boreal winter is characterized by a westerly circumpolar flow, denoted as the stratospheric polar vortex (SPV). The strength of the SPV can influence tropospheric circulation [*Baldwin et al.*, 2003; *Hitchcock and Simpson*, 2014] although the exact mechanisms remain not well understood [*Kidston et al.*, 2015; *Karpechko et al.*, 2017]. Extremely weak (strong) SPV states often lead to an equatorward (poleward) shift of the tropospheric jet and precede a negative (positive) phase of the Arctic Oscillation (AO) [*Baldwin and Dunkerton*, 1999; *Kolstad et al.*, 2010; *Kidston et al.*, 2015], usually associated with cooler (warmer) near-surface temperatures especially over northern Eurasia. The SPV is further an important source of seasonal predictability for winter weather [*Jia et al.*, 2017] and models initialized at the onset of extreme stratospheric events have better skill [*Kuroda*, 2008; *Sigmond et al.*, 2013; *Scaife et al.*, 2016]. To improve long-range forecasts for winter weather in the midlatitudes, it is thus essential to accurately capture extreme SPV states [*Sigmond et al.*, 2013; *Tripathi et al.*, 2015; *Scaife et al.*, 2016].

Variability in SPV strength depends primarily on the dynamical troposphere-stratosphere coupling via upward propagating planetary waves [*Matsuno*, 1970; *Polvani and Waugh*, 2004; *Shaw et al.*, 2014; *Dunn-Sigouin and Shaw*, 2015]. Subseasonal mechanisms can create vertical wave activity that either constructively or destructively interferes with the climatological wave [*Feldstein and Lee*, 2014]. These mechanisms include tropical teleconnections like El Niño–Southern Oscillation (ENSO) [*Garfinkel and Hartmann*, 2008], the Madden-Julian Oscillation (MJO) [*Garfinkel et al.*, 2014; *Liu et al.*, 2014], and the Quasi-biennial Oscillation (QBO) [*Watson and Gray*, 2014] and also Arctic-related processes like sea ice [*Kim et al.*, 2014] or snow cover variability [*Cohen et al.*, 2014]. Tropospheric processes are thus likely to have predictive information of the SPV at longer lead times, but the exact causal links are not well understood.

Though numerical models capture synoptic processes in the stratosphere well, they usually have low skill in predicting extreme SPV states at lead times beyond ~10 days [*Tripathi et al.*, 2015]. Their skill depends primarily on the models' ability to capture the mean stratospheric flow and upward propagating waves from the troposphere, but both processes suffer from uncertainties in initial conditions [*Mukougawa et al.*, 2005; *Tripathi et al.*, 2015]. In particular, stationary tropospheric waves associated with anomalous vertical flux of wave activity have a biased representation in models [*Woollings et al.*, 2010; *Shaw et al.*, 2014].

Next to numerical models, empirical methods to forecast summer and winter weather conditions have given promising results [*Baldwin et al*., 2003; *Cohen et al*., 2014; *Eden et al*., 2015; *McKinnon et al*., 2016]. The key challenge is, however, to avoid overfitting of the model by finding the physically relevant processes [*Runge et al*., 2015]. Empirical prediction models considering indices like ENSO and the Pacific Decadal Oscillation (PDO) captured SPV variability well over a training period. However, their performance was limited in prediction lead time or dropped drastically when applied to independent test data, indicating overfitting due to the use of noncausal processes [*Blume and Matthes*, 2012; *Minokhin et al*., 2017]. Predictors are often calculated based on principal component analysis or as the spatial mean over some region. The domains over which the indices are defined are thus not calculated directly from the to-be-forecasted variable ("response variable"), likely affecting the predictive skill. To overcome this, *Bello et al*. [2015] introduced *response-guided community detection* to optimize the construction of climate indices. Yet their approach does not require the indices to represent connected geographical regions nor does it account for different time lags or causality.

Here for the first time, we combine the idea of response-guided community detection [*Bello et al*., 2015] with a causal discovery algorithm [*Runge et al*., 2012, 2014, 2015; *Pearl*, 2013] to detect causal precursors of an index of interest. Our approach objectively detects and constructs climate indices from spatially gridded data and avoids overfitting by selecting causal predictors only. We present predictions of the stratospheric polar vortex here, but the method is generic and can be applied to any teleconnection index or meteorological variable of interest.

2. Response-Guided Causal Precursor Detection (RG-CPD)

We introduce a two-step procedure to detect causal precursors of a response variable (see Figure 1 for a schematic presentation). In the first step, the *response-guided community detection* (Figures 1a and 1b), we identify regions of meteorological variables that precede changes in the response index. For each variable, we calculate correlation maps with the considered index for different lags (Figure 1a). Adjacent grid points with significant correlation ($P < 0.005$, according to a two-sided Student's *t* test) of the same sign are grouped into *precursor regions*, and for each region a time series is created by calculating the area-weighted domain mean (Figure 1b). This reduces the data's dimensionality but preserves the spatiotemporal patterns describing the relationship between the lagged precursor variable and the response variable.

In the second step we identify the *causal* precursors of the response variable amongst all precursor regions (Figure 1c). By construction, all precursor time series are at least at one lag significantly correlated with the response variable (left part of Figure 1c) but this does not imply causality: Autocorrelation, common drivers, or indirect links can lead to spurious correlations. We therefore detect all spurious, respectively, noncausal, correlations from our list of precursors by applying a multivariate causal discovery algorithm [*Runge et al*., 2012, 2014; *Pearl*, 2013; *Colombo and Maathuis*, 2014] which is an adaptation of the so called PC algorithm [*Spirtes et al*., 2000; *Runge et al*., 2014] and is explained step by step in *Kretschmer et al*. [2016] (see also supporting information). The method iteratively tests (using partial correlations) if the lagged linear relationship between a precursor index and the response variable can be explained by the influence of other precursors. Precursors which are found to be conditionally dependent with the response variable (at some significance threshold) can be considered its *causal precursors* (right part of Figure 1c) meaning that their relationship with the response variable cannot be explained by the influence of any combination of all other precursor regions.

3. Data

We use gridded ($0.75° \times 0.75°$) ERA-Interim reanalysis data from 1979 to 2015. For our variable of interest, i.e., the SPV, we define an index as the (area-weighted) polar cap mean northward of 60°N of geopotential heights at 10 hPa. High values thus indicate a weak state of the SPV. We create a half-monthly (15 day means, i.e., beyond synoptic timescales) time series and remove long-term means and linear trends for each half-monthly calendar period. We choose this time resolution because we are interested in subseasonal drivers of the SPV. Moreover, especially persistent stratospheric anomalies can influence tropospheric circulation [*Gerber et al*., 2009; *Runde et al*., 2016] and those are thus captured by this choice. Indeed, the 10% weakest

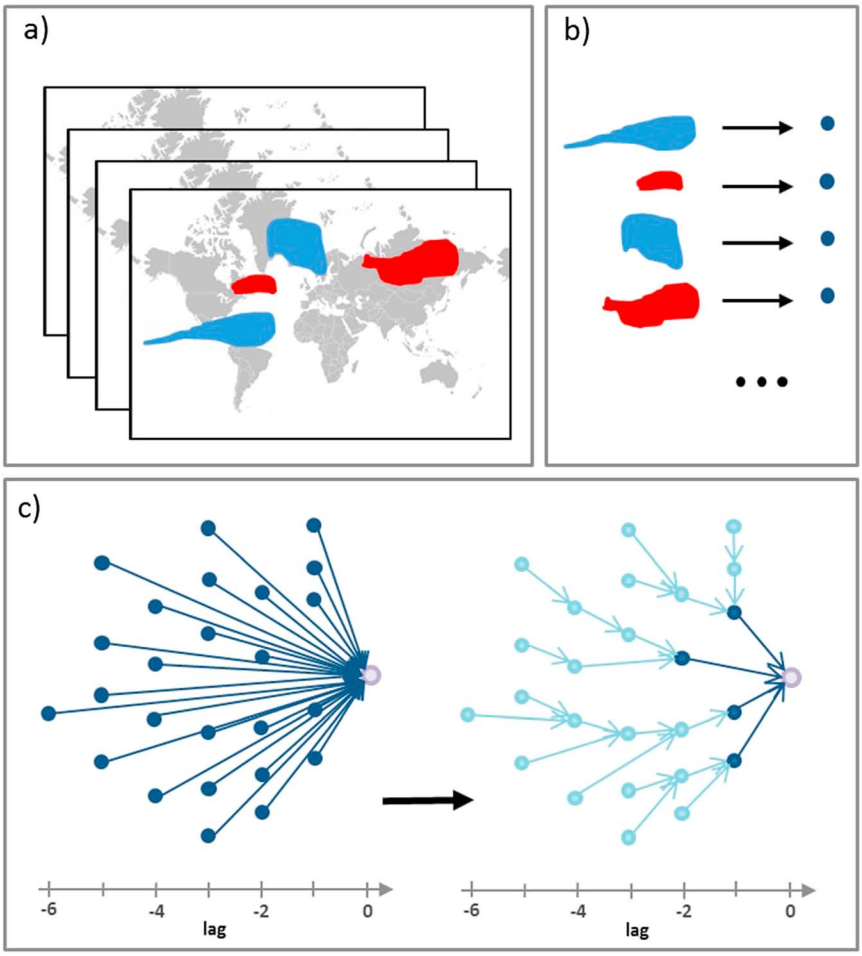

Figure 1. Schematic representation of the response-guided causal precursor detection (RG-CPD) scheme. (a) Detect regions in multi-variate data which correlate positively (red) or negatively (blue) with the response variable at different lead-lag times. (b) Take area-weighted averages of all regions creating time series of precursors. (c) By construction, all time series are (at least) at some lag significantly correlated with the response variable (left). A causality test removes all non-causal links due to common drivers, auto-correlation or indirect links (right).

(strongest) states of our SPV index coincide with negative (positive) AO-like surface temperature anomalies (see Figure S1) and therefore reflect those states of the SPV in which we are ultimately interested.

As potential precursors of the SPV, we allow any location from 20°S to 89°N in the fields of sea surface temperature (SST), sea level pressure (SLP), geopotential heights at 500 hPa (Z_{500}), poleward eddy heat flux at 100 hPa ($v^*T^*_{100}$), and zonal wind velocity at 50 hPa (U_{50}). This choice captures proposed drivers in the tropics (e.g., ENSO or the QBO) and Northern Hemisphere extratropics (e.g., SLP over the Ural Mountains) and thus represents variables and pressure levels that are often used in this context [*Garfinkel and Hartmann*, 2008; *Cohen et al*., 2014; *Kim et al*., 2014]. The regions for which indices are calculated are not selected a priori but are an outcome of the precursor detection step. Only the choice of the variable fields and their spatiotemporal resolution was based on expert judgment. For the SPV we are only interested in the winter (November to March) values, but the lagged precursor fields can contain data from previous seasons up to a lead-lag of 4 months (eight time steps).

4. Results

4.1. Detecting Causal Precursor Regions of the SPV

Using the SPV guided precursor detection approach to all fields for different lags (lag-1 to lag-8), we identify in total 471 precursor regions in the fields of $v^*T^*_{100}$ (72 regions), SLP (40 regions), SST (278 regions), Z_{500}

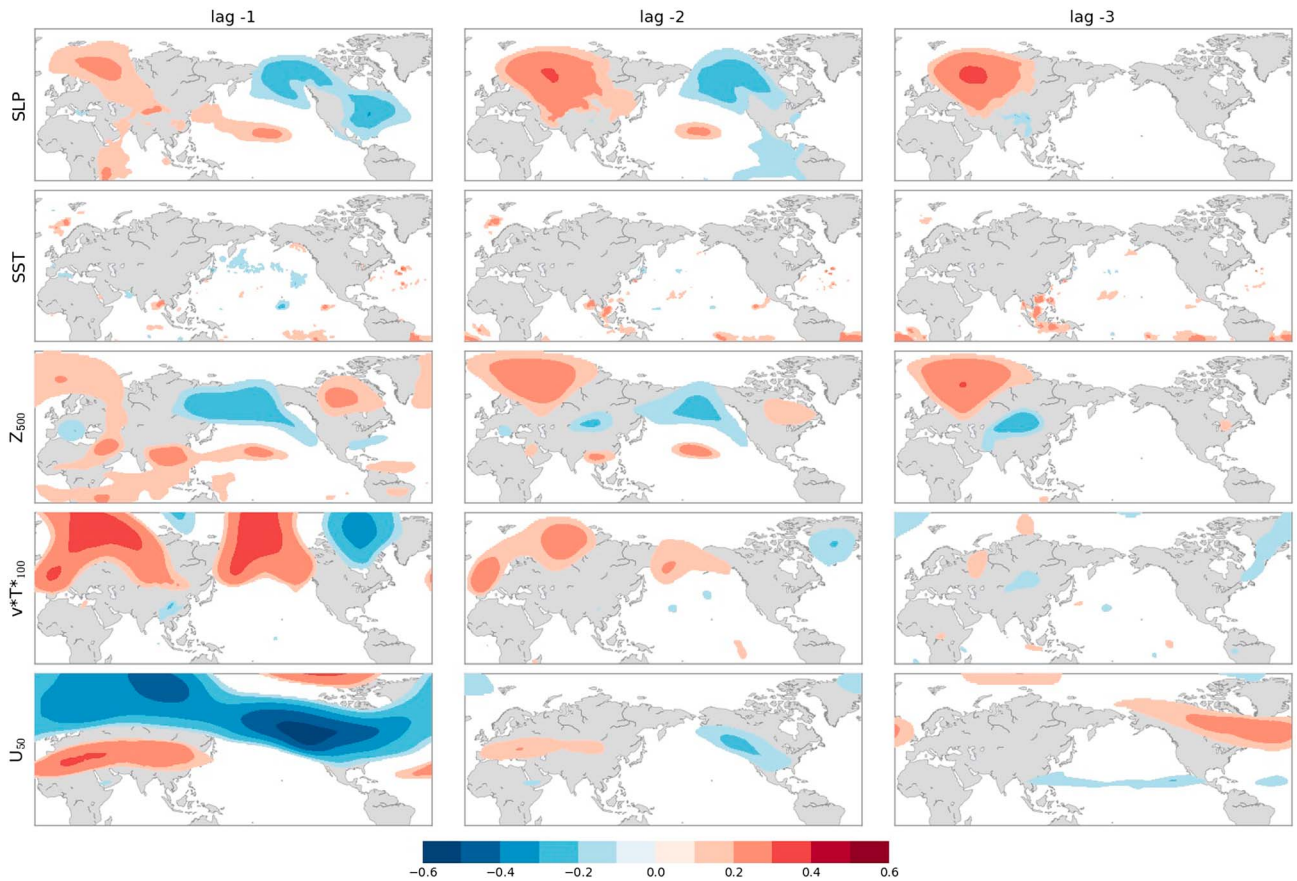

Figure 2. Correlation maps of the different precursor fields at lags from 1–3 with the (half-monthly) winter SPV index. Only significant correlation coefficients ($P < 0.005$, according to a two-sided Student's *t* test) are displayed. Each connected region shaded in blue (negatively correlated) or red (positively correlated) is a precursor region, and a (weighted) domain-mean time series is calculated.

(49 regions), and U_{50} (32 regions) (Figure 2 for lag-1 to lag-3, supporting information). The precursors contain several regions that are linked to proposed physical drivers. For example, we find SLP and Z_{500} over the Ural Mountains to be significantly correlated with the SPV for several lags. This pattern describes a stationary high-pressure system often preceding weak SPV states [*Cohen et al.*, 2014; *Kretschmer et al.*, 2016]. SST precursors in the equatorial Pacific indicate a positive relationship between ENSO-like patterns and weak SPV states [*Baldwin and O'Sullivan*, 1995; *Scaife et al.*, 2017], and we also find a pattern of Z_{500} over the Bering Strait associated with strong SPV states [*Orsolini et al.*, 2009; *Nishii et al.*, 2010]. Moreover, as expected, large precursor regions of $v^*T^*_{100}$, a proxy for vertical wave activity, are discovered at higher latitudes [*Polvani and Waugh*, 2004].

After applying the causal discovery algorithm (with a significance threshold $\alpha = 0.01$), only four causal precursors of the SPV index are identified: The lag-1 autocorrelated SPV index itself as well as three precursor regions (Figure 3a). These include two regions of lag-1 $v^*T^*_{100}$ over Eurasia and the North Pacific and a lag-3 precursor region of SSTs in the Philippine Sea. All detected causal precursors are positively correlated with the SPV (solid contours in Figure 3a).

We perform a series of tests (see supporting information) to assess the robustness of the detection scheme and find almost all detected causal precursors being insensitive to different parameter settings (maximum lag, significance threshold). Furthermore, the causal precursors are robust, when applying the scheme to shorter time intervals. The causal lag-1 precursors are in agreement with modeling studies stressing the importance of the mean stratospheric flow and vertical wave propagation for predicting the SPV [*Tripathi et al.*, 2015]. The relationship of the identified lag-3 SST region (Figure 3a) with the SPV is

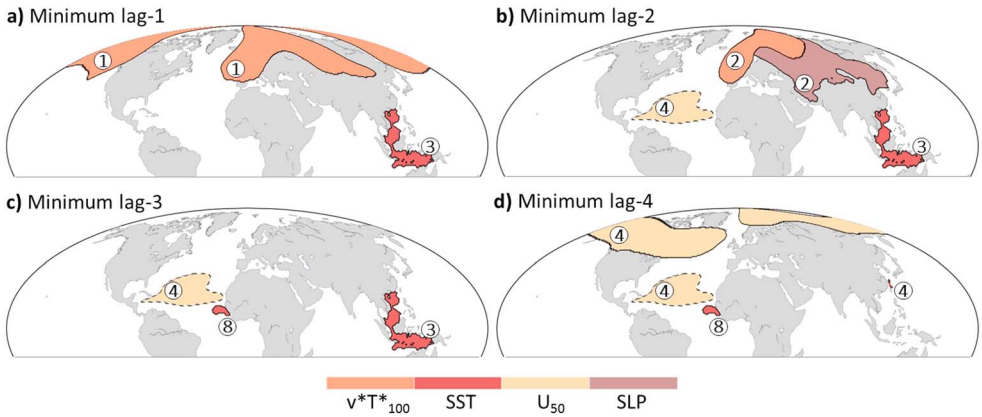

Figure 3. Causal precursor regions of the SPV index for different minimum lag tests. The color denotes the variable of the region, and the contours indicate if the region is positively (solid line) or negatively (dashed line) correlated with the SPV. The numbers next to the regions indicate the lag of the relationship. The individual correlation coefficients of each region with the SPV index are shown in the supporting information.

weakest as compared to the other causal precursors and somewhat less robust (see supporting information). Nevertheless, it reflects a real physical connection as anomalously low outgoing longwave radiation over this region—associated with enhanced deep convection and thus high SSTs—has been linked to different phases of the MJO preceding stratospheric warmings [*Garfinkel et al*., 2014; *Liu et al*., 2014].

In summary, all identified causal precursors are thus in accordance with previous studies and can be understood in a physically consistent way. Yet here their detection was done using an automatized and objective statistical analysis applied to a large multivariate data set, showing the strength of the methodology. In the following, we will focus on the three (most robust) lag-1 causal precursors but we will later discuss the lag-3 SST region again in the context of longer lead-lags.

4.2. Predicting SPV Variability and Extreme States

We perform multivariate linear regression of the lag-1 causal precursors on the SPV index to test their predictive skill. The predicted (based on the lagged regression model) and the observed SPV strongly correlate (r = 0.76, r^2 = 0.58), indicating that approximately 60% of SPV variability can be explained by our model. However, as already mentioned, statistical models often suffer from overfitting and fail when applied to independent test data. Here we intended to overcome this by applying a causal discovery algorithm. To show that our method indeed does not suffer from overfitting problems, we divide the data in a training (1979/1980–2002/2003) and a test (2003/2004–2014/2015) period and train our full method to data from the former only. Thus, we apply the CPD scheme to the training period only and detect the same three lag-1 causal precursors over this shorter time period. Also, the regression model is trained using data from the training period only. We then predict the SPV in the test period based on the causal precursors and regression parameters from the training period. The observed and predicted SPV correlate strongly over the training period (r = 0.79, Figure 4a), and this does not drop notably for the test period (r = 0.73, Figure 4a) proving that our model is not suffering from overfitting. Thus, the causal precursors detected by our scheme represent robust lag-1 predictors of the SPV suitable for forecasting (see supporting information for model parameters).

In the following we focus on hindcasting the SPV. To explore the skill of our regression model in more detail, we perform leave-out cross validation and build the regression model by leaving out *all* time steps of the to-be-forecasted winter to avoid any biased scores due to dependent data. This is sufficient as there is no year-to-year correlation of the SPV ($r = -0.06$). Yet even when leaving out values of more than one winter, we find robust results (see Figure S7). Note that all data processing (calculating anomalies and detrending) is performed excluding the considered winters. The out-of-sample predictions are shown in Figure 4b. The SPV amplitude is captured in magnitude, and the observations mostly lie within the 95% prediction interval. Moreover, the predicted and observed SPV strongly correlate (r = 0.76) indicating good explanatory skill of SPV variability.

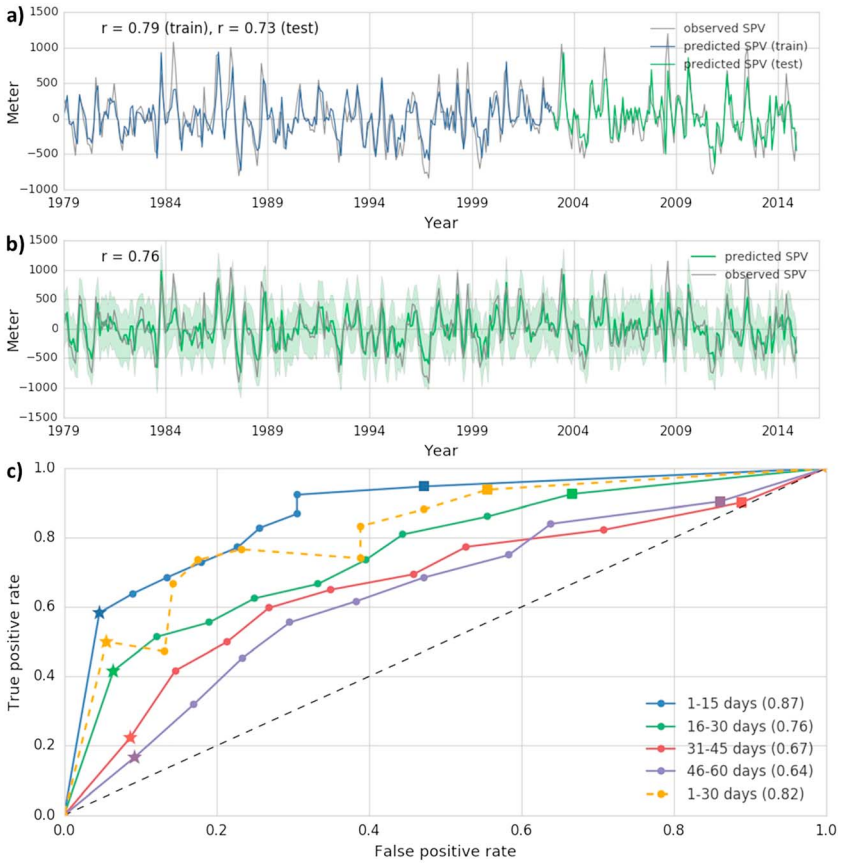

Figure 4. Validation of regression models. (a) Observed (gray line) and predicted SPV over a training (1979/1980 to 2002/2003, blue line) and a test (2003/2004 to 2014/2015, green line) period with causal precursors and regression parameters coming from training model. (b) Observed (gray line) and out-of-sample predictions of leave-out cross-validated SPV (green line) and the 95% prediction interval in green shading. (c) ROC scores for predictions based on different minimum lead-lags on half-monthly (colored solid lines) and monthly (dashed yellow line) time resolutions for percentiles in steps of 10%. The scores for the 10% weakest SPV states are indicated with a star and those of the 90% weakest (complementing the 10% strongest SPV states) with a square.

Being able to predict extreme states of the SPV is most relevant for seasonal winter forecast, as those can influence tropospheric circulation [*Baldwin et al*., 2003]. Therefore, we calculate the receiver operating characteristic (ROC) score of our regression model to test if it can skillfully predict threshold-exceeding events (Figure 4c). More precisely, we calculate the true positive and false positive rates for different percentile thresholds (in steps of 10%) which is a common metric to evaluate the performance of prediction models [*McKinnon et al*., 2016]. The true-positive rate is the number of correctly predicted events, normalized by the number of observed events in the SPV time series. Likewise, the false positive rate is the number of wrongly predicted events, normalized by the number of nonevents. We focus on events exceeding the 90th percentile in 10 hPa geopotential height averaged over the polar cap (i.e., the 10% weakest SPV states) and those below the 10th percentile (i.e., the 10% strongest SPV states). Our model accurately predicted 58% of weak SPV states with a false alarm rate of only ~4% (Figure 4c). This gives an odds ratio of 28.8 (i.e., the ratio of the probability that an event actually occurs when predicted, to the occurrence probability when it is not predicted [*McKinnon et al*., 2016]). Thus, there is a 29-fold increase in the probability of a weak SPV when our model predicts it. Similarly, our model correctly predicts 53% of extremely strong SPV states with a false alarm rate of ~5% giving an odds ratio of 20.2. The ROC score (defined as the area under the ROC curve, Figure 4c) is 0.87, indicating very good predictive skill of our model, way beyond the range of a random classifier with an ROC score of 0.5 by definition. This shows that our approach is not only suitable to model SPV variability well, but it is also capable of predicting extreme states with lead-times of 1–15 days (lag-1).

4.3. Predictions With Longer Lead Times

Troposphere-induced mechanisms influencing the stratospheric flow operate at subseasonal timescales of typically several weeks and might therefore provide longer-range predictive skill [*Polvani and Waugh*, 2004; *Garfinkel et al*., 2014; *Kretschmer et al*., 2016]. Therefore, we now detect causal precursors for the SPV at longer lead times by considering precursors with minimum lags of 2 (16–30 days), 3 (31–45 days), or 4 (46–60 days) time steps only.

For a *minimum* lag-2, Eurasian $v^*T^*_{100}$ at lag-2 as well as the lag-3 SST region are again detected as causal precursors (Figure 3b). Further, lag-2 sea level pressure over Central Eurasia and a lag-4 region of U_{50} over the subtropical Atlantic are identified (Figure 3b). These long-range causal precursors are in agreement with known physical mechanisms: (1) High pressures over the Ural Mountains favor increased vertical wave activity [*Feldstein and Lee*, 2014; *Kretschmer et al*., 2016]. (2) The negative correlation of lag-4 U_{50} region is consistent with the easterly phase of the QBO associated with weak SPV states [*Watson and Gray*, 2014]. For even longer minimum lags, the lag-3 SST region (Figures 3a–3c) and the lag-4 U_{50} region (Figures 3b–3d) as well as two subtropical SST precursors in the Atlantic (Figures 3c and 3d) and Pacific (Figure 3d) are detected.

For these longer lead times, the predicted SPV based on the out-of-sample regression models still correlates significantly ($P < 0.01$) with the observed SPV ($r = 0.56$, minimum lag-2), $r = 0.35$ (minimum lag-3), and $r = 0.29$ (minimum lag-4). We find predictive skill with ROC-scores ranging from 0.76 (minimum lag-2) to 0.65 (minimum lag-4). In particular, the models still correctly predict 42% (16–30 days ahead), 22% (31–45 days ahead), and 17% (46–60 days ahead) of the weakest SPV states resulting in odds ratios of 10.3, 3, and 1.96, respectively (Figure 4c). Similar though somewhat lower skill is achieved for predicting the 10% strongest states with odds ratios of 6.25 (16–30 days ahead), 1.14 (31–45 days ahead), and 1.52 (46–60 days ahead).

Finally, we repeated our analysis using monthly mean data. On this coarser time resolution, we detect again two $v^*T^*_{100}$ regions (as in Figure 3a) and a Central Eurasia SLP region (as in Figure 3b) as causal precursors at lag-1 (1–30 days ahead). On the monthly timescale the predicted and observed SPV correlate significantly ($r = 0.63$) and the regression model has high predictive skill (ROC score = 0.82) with 50% (44%) correctly predicted extremely weak (strong) SPV events and an odds ratio of 17 (12.16). These skill values (Figure 4c) are even better to those using 15 day averaged data for the minimum lag-2 prediction, underlining the robustness of our method.

5. Discussion

Our forecast method provides a generic framework to avoid common problems of empirical studies such as overfitting and finding the relevant time lags and domains with respect to an index of interest. Our scheme proposes an objective way to (1) construct response-guided climate indices and (2) filter out spurious correlations using a multivariate causal discovery algorithm. This way, the most relevant physical processes are identified that can be interpreted as causal precursors in a statistical sense.

From 471 precursor regions, two regions of poleward heat flux (Figure 3a) and the autoregressive SPV index itself were most robustly detected as causal precursors of the SPV at half-monthly lead times, which is consistent with physical understanding [*Matsuno*, 1970; *Polvani and Waugh*, 2004]. By restricting our scheme to longer lead-lags only or using monthly mean data, we showed that tropospheric processes give useful predictive skill up to 2 months, especially SLP over Central Eurasia (Figure 3b) and tropical western Pacific SSTs (Figures 3a–3c) were robustly identified for longer lead times. These regions have been linked to tropical [*Garfinkel et al*., 2014; *Liu et al*., 2014] and Arctic-related [*Cohen et al*., 2014; *Feldstein and Lee*, 2014; *Kretschmer et al*., 2016] processes favoring increased vertical wave activity. The other detected causal precursors are less well documented, but their inclusion, nevertheless, showed useful predictive skill beyond a random classifier. Overall, we achieve higher skill at longer lead times compared to other empirical prediction studies of the SPV with prediction skill limited to approximately 20 days [*Blume and Matthes*, 2012; *Minokhin et al*., 2017]. Though a one-to-one comparison is not possible, our correlation coefficients at lead times form 16–30 days compare mostly favorably to those for operational forecast models [see *Vitart*, 2014, Figure 10].

Our findings give additional knowledge to understand the dynamical troposphere-stratosphere coupling in winter and allow for a causal and thus physical interpretation. Further, given that our method produces

relevant skill at subseasonal timescale, it could have applications as a stand-alone tool in forecasting extreme SPV states. A possibly more powerful approach would combine our empirical forecast method with those of dynamical models to improve subseasonal probabilistic forecasts [*Weisheimer and Palmer*, 2014; *Jia et al.*, 2017]. *Ferrone et al.* [2017] showed that postprocessing of an ensemble of subseasonal forecasts can enhance their probabilistic skill. Predictors resulting from slowly varying components of the climate system provide information useful for statistical postprocessing of model ensembles. This can correct for model biases, which are often substantial for low-frequency variability. Physically meaningful precursors are needed to define proper statistical relationships, to avoid overfitting issues, and here our causal precursor detection scheme could proof useful.

Acknowledgments
We thank ECMWF for making the ERA-Interim data available. The work was supported by the German Federal Ministry of Education and Research, grant 01LN1304A (M.K. and D.C.). J.R. is funded by the James S. McDonnell Foundation. Code for the causal discovery method is freely available in the *Tigramite* Python software package https://github.com/jakobrunge/tigramite.

References

Baldwin, M. P., and T. J. Dunkerton (1999), Propagation of the Arctic oscillation from the stratosphere to the troposphere, *J. Geophys. Res.*, *104*, 30,937–30,946, doi:10.1029/1999JD900445.

Baldwin, M. P., and D. O'Sullivan (1995), Stratospheric effects of ENSO-related tropospheric circulation anomalies, *J. Clim.*, *8*(4), 649–667, doi:10.1175/1520-0442(1995)008<0649:SEOERT>2.0.CO;2.

Baldwin, M. P., D. B. Stephenson, D. W. J. Thompson, T. J. Dunkerton, A. J. Charlton, and A. O'Neill (2003), Stratospheric memory and skill of extended-range weather forecasts, *Science*, *301*(5633), 636–640, doi:10.1126/science.1087143.

Bello, G. A., et al. (2015), Response-guided community detection: Application to climate index discovery, in *Machine Learning and Knowledge Discovery in Databases*, vol. 9285, edited by A. Appice et al., ECML PKDD 2015. Lecture Notes in Computer Science, Springer, Cham.

Blume, C., and K. Matthes (2012), Understanding and forecasting polar stratospheric variability with statistical models, *Atmos. Chem. Phys.*, *12*(13), 5691–5701, doi:10.5194/acp-12-5691-2012.

Cohen, J., J. C. Furtado, J. Jones, M. Barlow, D. Whittleston, and D. Entekhabi (2014), Linking Siberian snow cover to precursors of stratospheric variability, *J. Clim.*, *27*(14), 5422–5432, doi:10.1175/JCLI-D-13-00779.1.

Colombo, D., and M. H. Maathuis (2014), Order-independent constraint-based causal structure learning, *J. Mach. Learn. Res.*, *15*, 3921–3962.

Dunn-Sigouin, E., and T. A. Shaw (2015), Comparing and contrasting extreme stratospheric events, including their coupling to the tropospheric circulation, *J. Geophys. Res. Atmos.*, *120*, 1374–1390, doi:10.1002/2014JD022116.

Eden, J. M., G. J. van Oldenborgh, E. Hawkins, and E. B. Suckling (2015), A global empirical system for probabilistic seasonal climate prediction, *Geosci. Model Dev.*, *8*(12), 3947–3973, doi:10.5194/gmd-8-3947-2015.

Feldstein, S. B., and S. Lee (2014), Intraseasonal and Interdecadal jet shifts in the northern hemisphere: The role of warm pool tropical convection and sea ice, *J. Clim.*, *27*(17), 6497–6518, doi:10.1175/JCLI-D-14-00057.1.

Ferrone, A., D. Mastrangelo, and P. Malguzzi (2017), Multimodel probabilistic prediction of 2 m temperature anomalies on the monthly timescale, *Adv. Sci. Res.*, *14*, 123–129, doi:10.5194/asr-14-123-2017.

Garfinkel, C. I., and D. L. Hartmann (2008), Different ENSO teleconnections and their effects on the stratospheric polar vortex, *J. Geophys. Res.*, *113*, D18114, doi:10.1029/2008JD009920.

Garfinkel, C. I., J. J. Benedict, and E. D. Maloney (2014), Impact of the MJO on the boreal winter extratropical circulation, *Geophys. Res. Lett.*, *41*, 6055–6062, doi:10.1002/2014GL061094.

Gerber, E. P., C. Orbe, and L. M. Polvani (2009), Stratospheric influence on the tropospheric circulation revealed by idealized ensemble forecasts, *Geophys. Res. Lett.*, *36*, L24801, doi:10.1029/2009GL040913.

Hitchcock, P., and I. R. Simpson (2014), The downward influence of stratospheric sudden warmings*, *J. Atmos. Sci.*, *71*(10), 3856–3876, doi:10.1175/JAS-D-14-0012.1.

Jia, L., et al. (2017), Seasonal prediction skill of northern extratropical surface temperature driven by the stratosphere, *J. Clim.*, *30*, 4463–4475, doi:10.1175/JCLI-D-16-0475.1.

Karpechko, A. Y., P. Hitchcock, D. H. W. Peters, and A. Schneidereit (2017), Predictability of downward propagation of major sudden stratospheric warmings, *Q. J. R. Meteorol. Soc.*, *143*, 1459–1470, doi:10.1002/qj.3017.

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

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

Kolstad, E. W., T. Breiteig, and A. A. Scaife (2010), The association between stratospheric weak polar vortex events and cold air outbreaks in the northern hemisphere, *Q. J. R. Meteorol. Soc.*, *136*(649), 886–893, doi:10.1002/qj.620.

Kretschmer, M., D. Coumou, J. F. Donges, and J. Runge (2016), Using causal effect networks to analyze different Arctic drivers of mid-latitude winter circulation, *J. Clim.*, *29*, 4069–4081, doi:10.1175/JCLI-D-15-0654.1.

Kuroda, Y. (2008), Role of the stratosphere on the predictability of medium-range weather forecast: A case study of winter 2003–2004, *Geophys. Res. Lett.*, *35*, L19701, doi:10.1029/2008GL034902.

Liu, C., B. Tian, K.-F. Li, G. L. Manney, N. J. Livesey, Y. L. Yung, and D. E. Waliser (2014), Northern hemisphere mid-winter vortex-displacement and vortex-split stratospheric sudden warmings: Influence of the madden-Julian oscillation and quasi-biennial oscillation, *J. Geophys. Res. Atmos.*, *119*, 12,599–12,620, doi:10.1002/2014JD021876.

Matsuno, T. (1970), Vertical propagation of stationary planetary waves in the winter northern hemisphere, *J. Atmos. Sci.*, *27*(6), 871–883, doi:10.1175/1520-0469(1970)027<0871:VPOSPW>2.0.CO;2.

McKinnon, K. A., A. Rhines, M. P. Tingley, and P. Huybers (2016), Long-lead predictions of eastern United States hot days from Pacific sea surface temperatures, *Nat. Geosci.*, *9*(5), 389–394, doi:10.1038/ngeo2687.

Minokhin, I., C. G. Fletcher, and A. Brenning (2017), Forecasting northern polar stratospheric variability with competing statistical learning models, *Q. J. R. Meteorol. Soc.*, *143*, 1816–1827, doi:10.1002/qj.3043.

Mukougawa, H., H. Sakai, and T. Hirooka (2005), High sensitivity to the initial condition for the prediction of stratospheric sudden warming, *Geophys. Res. Lett.*, *32*, L17806, doi:10.1029/2005GL022909.

Nishii, K., H. Nakamura, and Y. J. Orsolini (2010), Cooling of the wintertime Arctic stratosphere induced by the western Pacific teleconnection pattern, *Geophys. Res. Lett.*, *37*, L13805, doi:10.1029/2010GL043551.

Orsolini, Y. J., A. Y. Karpechko, and G. Nikulin (2009), Variability of the northern hemisphere polar stratospheric cloud potential: The role of North Pacific disturbances, *Q. J. R. Meteorol. Soc.*, *135*(641), 1020–1029, doi:10.1002/qj.409.

Pearl, J. (2013), Linear models: A useful "microscope" for causal analysis, *J. Causal Inference*, *1*(1), 155–170, doi:10.1515/jci-2013-0003.

Polvani, L. M., and D. W. Waugh (2004), Upward wave activity flux as a precursor to extreme stratospheric events and subsequent anomalous surface weather regimes, *J. Clim.*, *17*(18), 3548–3554, doi:10.1175/1520-0442(2004)017<3548:UWAFAA>2.0.CO;2.

Runde, T., M. Dameris, H. Garny, and D. E. Kinnison (2016), Classification of stratospheric extreme events according to their downward propagation to the troposphere, *Geophys. Res. Lett.*, *43*, 6665–6672, doi:10.1002/2016GL069569.

Runge, J., J. Heitzig, N. Marwan, and J. Kurths (2012), Quantifying causal coupling strength: A lag-specific measure for multivariate time series related to transfer entropy, *Phys. Rev. E*, *86*(6), 61121, doi:10.1103/PhysRevE.86.061121.

Runge, J., V. Petoukhov, and J. Kurths (2014), Quantifying the strength and delay of climatic interactions: The ambiguities of cross correlation and a novel measure based on graphical models, *J. Clim.*, *27*(2), 720–739, doi:10.1175/JCLI-D-13-00159.1.

Runge, J., R. V. Donner, and J. Kurths (2015), Optimal model-free prediction from multivariate time series, *Phys. Rev. E*, *91*(5), 052909, doi:10.1103/PhysRevE.91.052909.

Scaife, A. A., et al. (2016), Seasonal winter forecasts and the stratosphere, *Atmos. Sci. Lett.*, *17*(1), 51–56, doi:10.1002/asl.598.

Scaife, A. A., et al. (2017), Tropical rainfall, Rossby waves and regional winter climate predictions, *Q. J. R. Meteorol. Soc.*, *143*(702), 1–11, doi:10.1002/qj.2910.

Shaw, T. A., J. Perlwitz, and O. Weiner (2014), Troposphere-stratosphere coupling: Links to North Atlantic weather and climate, including their representation in CMIP5 models, *J. Geophys. Res. Atmos.*, *119*, 5864–5880, doi:10.1002/2013JD021191.

Sigmond, M., J. F. Scinocca, V. V. Kharin, and T. G. Shepherd (2013), Enhanced seasonal forecast skill following stratospheric sudden warmings, *Nat. Geosci.*, *6*(2), 98–102, doi:10.1038/ngeo1698.

Spirtes, P., C. Glymour, and R. Scheines (2000), *Causation, Prediction, and Search*, vol. 81, MIT Press, Cambridge.

Tripathi, O. P., et al. (2015), The predictability of the extratropical stratosphere on monthly time-scales and its impact on the skill of tropospheric forecasts, *Q. J. R. Meteorol. Soc.*, *141*(689), 987–1003, doi:10.1002/qj.2432.

Vitart, F. (2014), Evolution of ECMWF sub-seasonal forecast skill scores, *Q. J. R. Meteorol. Soc.*, *140*(683), 1889–1899, doi:10.1002/qj.2256.

Watson, P. A. G., and L. J. Gray (2014), How does the quasi-biennial oscillation affect the stratospheric polar vortex?, *J. Atmos. Sci.*, *71*(1), 391–409, doi:10.1175/JAS-D-13-096.1.

Weisheimer, A., and T. N. Palmer (2014), On the reliability of seasonal climate forecasts, *J. R. Soc. Interface*, *11*(96), doi:10.1098/rsif.2013.1162.

Woollings, T., A. Charlton-Perez, S. Ineson, A. G. Marshall, and G. Masato (2010), Associations between stratospheric variability and tropospheric blocking, *J. Geophys. Res.*, *115*, D06108, doi:10.1029/2009JD012742.