Received: 8 April 2018 | Revised: 7 November 2018 | Accepted: 13 November 2018

DOI: 10.1002/wcc.567

WILEY WIREs CLIMATE CHANGE

OPINION

S2S reboot: An argument for greater inclusion of machine learning in subseasonal to seasonal forecasts

Judah Cohen[1,2] | Dim Coumou[3,4] | Jessica Hwang[5] | Lester Mackey[6] | Paulo Orenstein[5] | Sonja Totz[4] | Eli Tziperman[7]

[1]Atmospheric and Environmental Research, Inc., Lexington, Massachusetts

[2]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts

[3]Department of Water and Climate Risk, Institute for Environmental Studies, VU University Amsterdam, Amsterdam, Netherlands

[4]Potsdam Institute for Climate Impact Research, Potsdam, Germany

[5]Stanford University, Stanford, California

[6]Microsoft Research New England, Cambridge, Massachusetts

[7]Department of Earth and Planetary Sciences, School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts

Correspondence
Judah Cohen, Atmospheric and Environmental Research, Inc., Lexington, MA 02421.
Email: jcohen@aer.com

Funding information
Directorate for Geosciences, Grant/Award Number: AGS-1303647, AGS-1622985 and PLR-1504361

Edited by Eduardo Zorita, Domain Editor, and Mike Hulme, Editor-in-Chief

The discipline of seasonal climate prediction began as an exercise in simple statistical techniques. However, today the large government forecast centers almost exclusively rely on complex fully coupled dynamical forecast systems for their subseasonal to seasonal (S2S) predictions while statistical techniques are mostly neglected and those techniques still in use have not been updated in decades. In this Opinion Article, we argue that new statistical techniques mostly developed outside the field of climate science, collectively referred to as machine learning, can be adopted by climate forecasters to increase the accuracy of S2S predictions. We present an example of where unsupervised learning demonstrates higher accuracy in a seasonal prediction than the state-of-the-art dynamical systems. We also summarize some relevant machine learning methods that are most applicable to climate prediction. Finally, we show by comparing real-time dynamical model forecasts with observations from winter 2017/2018 that dynamical model forecasts are almost entirely insensitive to polar vortex (PV) variability and the impact on sensible weather. Instead, statistical forecasts more accurately predicted the resultant sensible weather from a mid-winter PV disruption than the dynamical forecasts. The important implication from the poor dynamical forecasts is that if Arctic change influences mid-latitude weather through PV variability, then the ability of dynamical models to demonstrate the existence of such a pathway is compromised. We conclude by suggesting that S2S prediction will be most beneficial to the public by incorporating mixed or a hybrid of dynamical forecasts and updated statistical techniques such as machine learning.

This article is categorized under:
Climate Models and Modeling > Knowledge Generation with Models

KEYWORDS
climate prediction, machine learning, polar vortex, unsupervised learning

1 | INTRODUCTION

It is estimated that 2.7 trillion dollars of the U.S. economy alone is sensitive to the impacts of weather and climate (National Oceanic and Atmospheric Administration, 2002). Improving our ability to forecast the weather and climate is of interest to all sectors of the economy and government agencies from the local to the national level. In recent years, seasonal climate forecasts have become an important aspect of policy and decision-making, and are utilized in a broad range of socioeconomic applications (Troccoli, 2010).

Traditionally, statistical composites have been used to identify patterns of variability in the atmosphere, an approach on which our understanding and prediction of climate states has been based (e.g., Barnston & Livezey, 1987; Wallace & Gutzler, 1981). However, today the most commonly employed tool in seasonal forecasting at the government supported operational forecast centers is general circulation models or global climate models (GCMs). These highly complex dynamical models represent many of the major processes in the ocean–ice–land–atmosphere climate system.

Skillful seasonal prediction is based on the premise that either statistics such as persistence or multiannual trends or slowly varying boundary forcings can be exploited for producing long-range forecasts that are more accurate than climatology. Slowly varying boundary forcings include ocean temperatures, sea ice, soil moisture, and snow cover (Doblas-Reyes, Garcia-Serrano, Lienert, Biescas, & Rodrigues, 2013).

Long range or statistical forecasts began in the 1950s when scientists first identified large scale atmospheric patterns and recognized relationships between atmospheric variability and ocean temperature anomalies (Namias, 1953; National Academies of Sciences, Engineering, and Medicine, 2016). Seasonal forecasting began with using statistical or empirical forecast techniques. In the 1980s, seasonal prediction was based on lag correlations with observed upper atmosphere geopotential height anomalies (Barnston, Kumar, Goddard, & Hoerling, 2005; Wagner, 1989) or analogs (Barnston et al., 2005; Livezey & Barnston, 1988). Analog techniques typically choose years with similar boundary forcings (e.g., same El Niño/Southern Oscillation or ENSO phase) and use composites of those chosen years as the forecast. Forecasts at the large government centers are issued as probabilities and not as deterministic forecasts (Livezey & Timofeyeva, 2008).

Statistical or empirical forecasts have traditionally identified a relationship between sensible weather, that is, surface temperature and precipitation anomalies with sea surface temperature (SST) anomalies. Most often those relationships are limited to the tropical oceans in general and to the ENSO region in particular, although some attempts have been made to include SSTs from the extratropics as well (Barnett, 1981; Barnston et al., 2005; Barnston & Smith, 1996).

Statistical methods have often relied on linear regression or variations such as canonical correlation analysis (CCA). These methods relate variations in predictor fields to variations in predictand fields (Barnston & Smith, 1996). CCA optimizes the linear combination of predictor data to explain the greatest variance in the predictand array. In CCA, both predictors and the predictands are multidimensional arrays of variables (Barnett & Preisendorfer, 1987).

Starting in the 1980s, seasonal forecasting began to involve dynamical models (National Academies of Sciences, Engineering, and Medicine, 2016; Reeves & Gemmill, 2004) and since the 1990s the seasonal predictions at government forecast centers have transitioned away from statistical techniques and instead focused their efforts and resources on atmosphere–ocean coupled dynamical models (Barnston, Tippett, L'Heureux, Li, & DeWitt, 2012). There are two types of atmosphere–ocean coupling employed at government forecast centers—tier 1 and tier 2. In tier 1 coupled model systems, both the atmosphere and ocean models are fully dynamical models coupled to each other. In tier 2 coupled model systems, a fully dynamical atmosphere model is forced with prescribed SSTs. Model skill for tier 1 coupled systems has been shown to be better than tier 2 (Doblas-Reyes et al., 2013; Kug, Kang, & Choi, 2008).

Initially, tier 1 coupled dynamical forecasts involved an atmosphere model coupled to the Tropical Pacific (Delecluse et al., 1998; Doblas-Reyes et al., 2013). When resources became available, the atmosphere models were coupled to global oceans (Kirtman, Shukla, Huang, Zhu, & Schneider, 1997; Latif, Collins, Pohlmann, & Kennlyside, 1998). However, other boundary forcings such as snow cover, sea ice, and soil moisture were prescribed from climatology (National Academies of Sciences, Engineering, and Medicine, 2016). In today's most advanced dynamical prediction systems, the atmosphere is coupled to dynamical models of the global oceans including sea ice, the land surface including soil moisture and snow cover, although the ocean component remains the most developed (Doblas-Reyes et al., 2013; National Academies of Sciences, Engineering, and Medicine, 2016). Today almost all national and international operational weather centers use coupled dynamical systems to produce seasonal forecasts (National Academies of Sciences, Engineering, and Medicine, 2016).

Despite the notable recent improvement in dynamical seasonal prediction models, most of the inherent skill in these models is derived from the accurate prediction of coupled atmospheric–oceanic phenomena—predominantly related to ENSO (Barnston et al., 1994; Troccoli, 2010; van Oldenborgh, Balmaseda, Ferranti, Stockdale, & Anderson, 2005a, 2005b). Understanding and identifying the global impacts of ENSO was a critical advance in climate prediction on the seasonal time scale. Although important, ENSO explains only a portion of the climate variability and ENSO-based forecasts have been found to produce mixed results at best in the atmospheric prediction of mid- and high-latitude regions far removed from the tropical Pacific Ocean, for example, the North Atlantic sector (Barnston et al., 1999; Cohen & Fletcher, 2007; Spencer & Slingo, 2003). Furthermore, ENSO is weak in boreal summer months and hence does not provide skillful prediction for Northern Hemisphere summers, especially Eurasia. Given that ENSO variability offers only limited atmospheric predictive skill away from the tropics, it has provided a natural constraint on model skill in the extratropics. In the light of limited extratropical predictive skill associated with ENSO, a question remains—what are the prospects for skill improvement in the extratropical latitudes where a large fraction of climate variability still remains unpredictable?

Studies of the comparable skill between ENSO seasonal forecasts made by dynamical and statistical models found that, in general, dynamical models outperformed statistical models (Barnston et al., 2012). This is expected given that the focus of effort and resources has been on dynamical forecast systems while statistical models have been mostly neglected. And while dynamical models are constantly improved and updated, statistical models have basically remained unchanged from the 1990s and even the 1980s (Barnston et al., 2012). An influential National Academies of Science report recommended improvements on subseasonal to seasonal (S2S) predictions that were exclusively directed at dynamical prediction systems including data assimilation, parameterization and multimodel approaches, while statistical models were ignored. The report recommended that "To summarize, investment in research aimed at physical understanding and reducing (dynamical) model errors is seen by this committee as a top priority for improving the skill of S2S predictions." (National Academies of Sciences, Engineering, and Medicine, 2016)

In this *Perspective*, we argue that statistical methods still have an important place in S2S forecasting. Although development of new statistical techniques has not been a focus of the climate community, new statistical techniques have been developed and utilized in other disciplines to manipulate large datasets. We argue that these new statistical techniques, often referred to collectively as "machine learning," improve on traditional statistical techniques with higher forecast accuracy. Recent work shows that these new statistical techniques are often more accurate than the latest generation of dynamical coupled atmosphere–ocean models on S2S timescales. We present two examples where newly developed statistical techniques demonstrate higher skill in hindcasts than both traditional CCA and dynamical models.

In addition, statistical forecasts have the advantage of utilizing continuously acquired knowledge obtained from data analysis of climate variability, which can readily be applied (Doblas-Reyes et al., 2013). In testing the utility of a boundary forcing such as sea ice or snow cover in improving forecast skill, using a statistical model is much easier and faster to employ and implement than a dynamical model. Since statistical models are computationally more expedient than GCMs, they can be used to efficiently search for "windows of predictability (i.e., the ability to predict)," that is, regions, states or timescales that are linked to low frequency processes with better predictive skill. And although dynamical models incorporate boundaries such as ocean variability, snow cover, sea ice, soil moisture and even the stratosphere, model errors could mask potential skill from these possible forcings.

We further argue that even traditional statistical models can be used to uncover weaknesses and guide improvements in dynamical model simulations. Observational analysis has supported that Arctic variability influences mid-latitude weather while most modeling studies have supported mostly no relationship or that only a weak relationship exists (Cohen, Screen, et al., 2014; Overland et al., 2015; Vihma, 2014). Additionally, modeling studies have been used to criticize the observational analysis (Cohen, Screen, et al., 2014). We show, however, that statistical model forecasts using Arctic predictors outperformed dynamical models in predicting winter 2017/2018 surface temperature anomalies and exposing serious shortcomings of dynamical models and an inherent insensitivity to Arctic forcings.

2 | DATASETS

In this study, we used several precursor fields in autumn (September, October, November [SON]), including monthly sea ice concentration (SIC) from the Met Office Hadley Centre at 2.5° latitude X 2.5° longitude (Rayner et al., 2003) and the domain for the SIC is 60°–90°N latitude and 0°W–180°E longitude and for the time period of 1965–2015. We further include monthly snow cover extent (SCE) from the Rutgers Snow Lab with data provided by the National Oceanographic and Atmospheric Administration (NOAA; Robinson, Estilow, & Program, 2012) using the area between 20°–60°N latitude and 0°–180°E longitude. In addition, we include monthly SST from the Met Office Hadley Centre at 1° × 1° (Rayner et al., 2003) over three different regions: the North Atlantic (10°–70°N latitude and −110°W–20°E longitude), the Mediterranean (25°–50°N and 6°W–45°E) and the Pacific region (30°S–30°N and 150°E–70°W). In the Atlantic region SST, we masked portions corresponding to the Pacific Ocean, Mediterranean as well as Hudson Bay. Similarly, we masked the Atlantic region in the selected Pacific Ocean precursor region. Finally, we included as atmospheric precursors the geopotential height at 500 hPa over 20°–80°N and 10°E–180°E and sea level pressure (SLP) over 40°–80°N and 0°–360° from the National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalysis available four times daily at a global resolution at 2.5° × 2.5° resolution (Kalnay et al., 1996). We averaged daily values to create both monthly and seasonal averages. All regions are selected by examining the significance of precursor composites as calculated by a bootstrap method.

For the winter European temperature forecast we used monthly averaged temperature anomalies on a 0.5° × 0.5° grid provided by the NOAA (Fan & van den Dool, 2008). We used the same area as for the precipitation set (25°–75°N latitude and 20°W–45°E longitude). In addition, we included a greater and shifted SST area for the North Atlantic region (5°S–50°N and 90°–6°W) to account for the North Atlantic Oscillation (NAO) pattern, and excluded the Mediterranean region for the temperature anomaly forecast. For Northern Hemisphere surface temperature data we used the NCEP/NCAR reanalysis available four

times daily at a global resolution of ~1.9° (Kalnay et al., 1996). We averaged daily values to create both monthly and seasonal averages.

3 | FORECAST METHODS

In the paper we compare the hindcast skill of three forecast methodologies. The first method is CCA, which is the traditional statistical method employed by forecast centers for long range weather or climate prediction. We also use a new method of statistical prediction based on hierarchical clustering analysis developed by Totz, Tziperman, Coumou, Pfeiffer, and Cohen (2017). The advantage of the clustering-based forecasting method over index-based regression models (e.g., Cohen & Fletcher, 2007) is that it accounts not only for the amplitude of the predictors but also for the geographical distribution using clustering techniques. Its advantage over CCA, which is based on principal component analysis is that, unlike principal components, the clusters do not need to be orthogonal and therefore represent variability patterns more faithfully. Further details on the hierarchical clustering analysis are provided in Supporting Information File S1. A complete description of how hierarchical clustering techniques can be used for forecasting is described in Totz et al. (2017).

In addition, we calculated the ensemble model mean of monthly hindcast data from nine models, which participate in the North American Multi-Model Ensemble (NMME). The NMME system includes coupled models from a number of United States and Canadian modeling centers in an ensemble of opportunity supporting seasonal forecasting experiments (Kirtman et al., 2014). The models used in the NMME are: CMC1-CanCM3, CMC2-CanCM4, NCAR-CESM1, NCEP-CFSv2, COLA-RSMAS-CCSM3, COLA-RSMAS-CCSM4, NASA-GMAO, IRI-ECHAMP4p5-DirectCoupled, IRI-ECHAMP4p5-AnomalyCoupled (Kirtman et al., 2014). The NMME project is managed by NOAA.

We also present a real-time monthly and seasonal forecast based on multilinear regression (Cohen & Fletcher, 2007). The linear regression model is not novel of course but does use mostly Arctic predictors for the winter forecast. The major predictors for the winter are predicted winter ENSO (Niño 3.4 index), September Arctic SIC, October Eurasian SCE and October SLP anomaly in northwestern Asia. No detrending is applied to the data in producing the forecast. Three of the four predictors are high latitude variables with the only tropical predictor being ENSO. Cohen and Fletcher (2007) demonstrated improved hindcast temperature skill over dynamical model hindcast skill using linear regression and Arctic predictors. However, the variance of the model is damped relative to the observations. In those regions where the model has higher skill such as the eastern United States and northern Eurasia, the model variance is 70–80% of the observed values. This model is run operationally prior to each winter and the forecast is posted to the National Science Foundation website (https://www.nsf.gov/news/special_reports/autumnwinter/predicts.jsp). Subsequent studies have supported the use of SIC and SCE as predictors for wintertime temperature anomalies (Furtado, Cohen, Butler, Riddle, & Kumar, 2015; García-Serrano et al., 2016; Gastineau, García-Serrano, & Frankignoul, 2017). Since the model utilizes October data, a preliminary forecast is usually generated the end of October with the forecast updated the first week of November. The forecast for winter (December to February) 2017/2018 was created November 8, 2017.

We show the winter temperature anomaly forecast for the Northern Hemisphere from winter 2017/2018. The statistical forecast is compared to the NMME suite of models. In addition, we also compare to the real-time forecasts from an international suite of models including the CFSv2 (climate forecast system version 2), the European Centre for Medium-range Weather Forecasting (ECMWF) model (Stockdale et al., 2011), the UK Met Office Hadley Center Unified Model (Walters et al., 2017), and the MeteoFrance model (Voldoire et al., 2013). These four models are referred to as the International Multi-Model Ensemble (IMME). In the present work, we evaluated NMME and IMME forecasts from November for the December, January and February period, and from December for the January, February and March period.

4 | ALTERNATIVE STATISTICAL LEARNING APPROACHES

In this Perspective article, we bring examples of statistical techniques applied to S2S prediction using clustering analysis and multilinear regression. However, there are other examples of new techniques of supervised learning often referred to collectively as “machine learning” that may be applied and possibly provide improved forecast skill to operational S2S predictions. These techniques are briefly described in File S1. In the following section, we present hindcast skills from an unsupervised learning technique known as hierarchical clustering.

5 | EUROPEAN WINTER PRECIPITATION HINDCASTS

In a recent paper (Totz et al., 2017), we compared hindcast skill of winter (December–January–February) precipitation for the European and Mediterranean regions for CCA, hierarchical clustering and from the NMME dynamical models. In Totz et al. (2017), the cross validated correlation of the cluster-based hindcasts with observations for the years 1967–2016 were shown. Also shown were the cross-validated correlations of the CCA-based hindcasts with observations for the years 1967–2016. In addition, the cross-validated correlations were shown for hindcasts for the years 1982–2010 from the NMME dynamical models. Finally, to make a direct comparison with the NMME, the cross-validated correlation of the cluster-based hindcasts with observations for the years 1982–2010 were shown as well.

The hierarchical clustering used four fall (three-month average of September, October, and November) predictors for predicting winter precipitation in the European and Mediterranean regions: SCE across Eurasia, SIC in the Arctic and SST in the North Atlantic and Mediterranean regions. The conclusion of that study was that the forecasts based on hierarchical clustering had higher skill than both the more traditional CCA statistical model and the NMME suite of dynamical models. Here we include a similar figure of forecast skill slightly modified from the techniques presented in Totz et al. (2017) (see Figure 1). The hierarchical clustering exhibits improved skill in representing the overall pattern of precipitation anomalies and in the skill score of individual locations throughout the region.

6 | WINTER SURFACE TEMPERATURE HINDCASTS

Totz et al. (2017) only applied hierarchical clustering to precipitation forecasts. We performed a similar analysis for temperature forecasts again for the European sector. We used hierarchical clustering and the Ward method to identify patterns of winter temperature variability and found six dominant clusters (Figure 2). The first cluster shows east–west temperature variability, whereas Clusters 2 and 5 exhibit homogenous warm and cold temperature patterns, respectively. Clusters 3 and 4 exhibit north–south temperature variability that is the classic signature of the NAO. The last cluster exhibits positive temperature anomalies over northern Europe and negative temperature anomalies in southern Europe. We then matched the six clusters with seven fall predictors for predicting winter temperature in the European and Mediterranean regions: SCE across Eurasia, SIC in the Arctic, SST in the North Atlantic region, Mediterranean region and Pacific region as well as geopotential

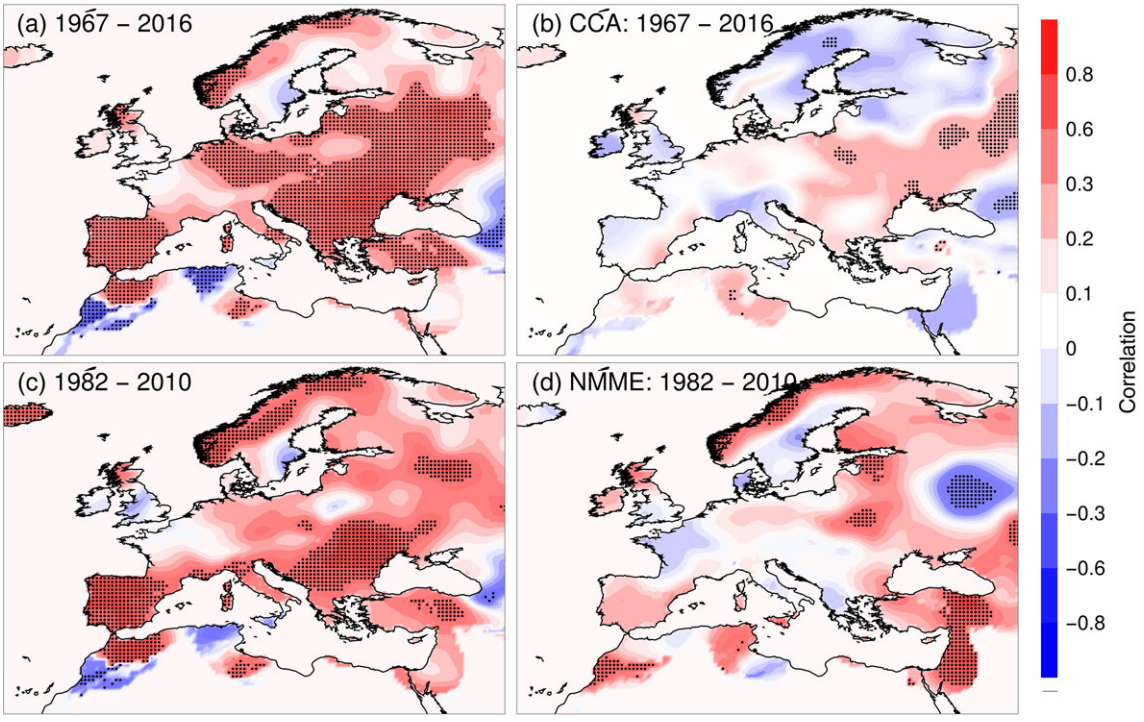


FIGURE 1 Precipitation winter (December, January, and February) forecast skill: (a) cross-validated correlation of the cluster-based hindcasts with observations for the years 1967–2016. (b) Same as (a) but for the hindcast skill using canonical correlation analysis. (c) Same as (a) but for the years 1982–2010. (d) Correlation of the North American Multi-Model Ensemble (NMME) hindcasts for the years 1982–2010 with observations. Significant values ($p < .05$) according to the two-sided Student's t test are shown in hatches. The cluster-based forecast performs better than the NMME models according to the cross-validated correlations. (Reprinted with permission from Totz et al. (2017). Copyright 2017 American Geophysical Union)

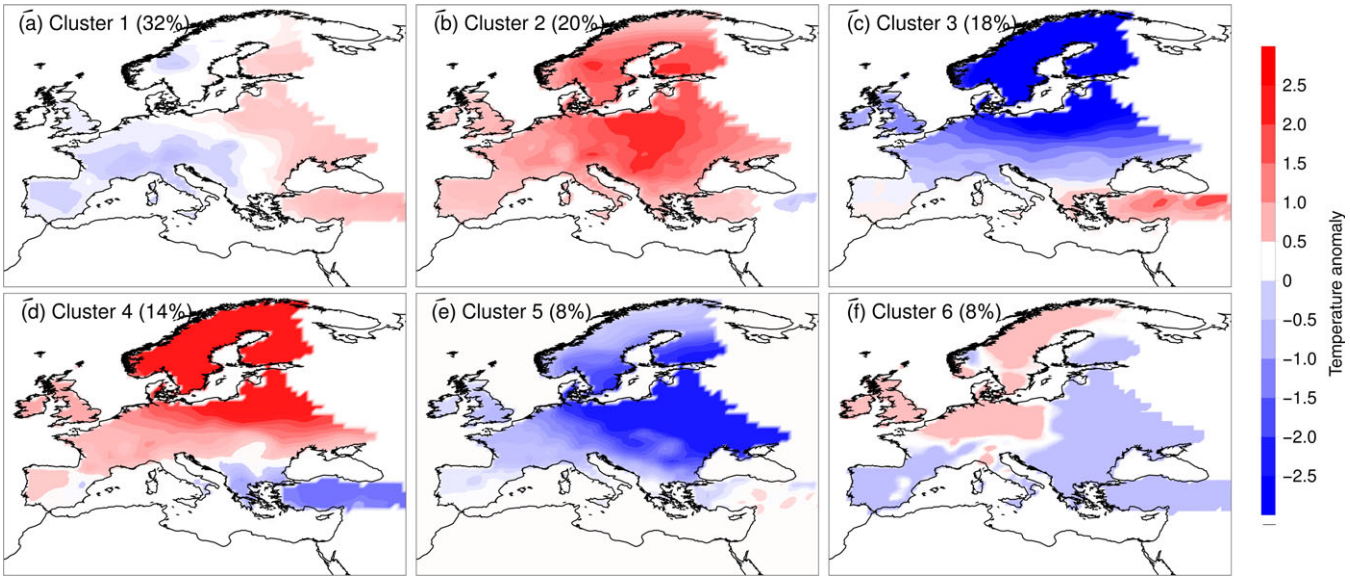


FIGURE 2 Clusters of temperature winter (December, January, and February) anomalies for the European region, ordered by their frequency: Cluster (a) has a frequency of 32%, (b) a frequency of 20%, (c) a frequency of 18%, (d) a frequency of 14%, (e) a frequency of 8%, and (f) a frequency of 8%

height over Eurasia and SLP over the mid- to high-latitudes of the Northern Hemisphere. We implement a few improvements relative to Totz et al. (2017) as follows. First, we remove the mean of each precursor using all data except for the predicted year. Next, we normalize each precursor by its standard deviation. Finally, we combine different precursors into a single vector. In addition, we define a threshold for the calculation of the singular value decomposition-based pseudo-inverse used to calculate the prediction coefficients such that singular values that are smaller than 1% of the largest singular value are set to zero.

All hindcasts were cross-validated, that is, the precursors for the year hindcasted were not included in building the regression coefficients used for the temperature hindcast of that year. We calculated the mean cross-validated skill using all possible configurations of the seven precursors: using all possible seven single precursors, using all possible pairs of precursors chosen out of the possible seven, etc. There is a total of 127 models based on these different precursor choices. Out of these, 16 show a skill higher than 0.1, 12 a skill higher than 0.15, and 12 higher than 0.2. Note that the NMME skill is 0.02. Thus the clustering-based forecast has an improved forecast skill that is robust to model details. We find that the mean skill has the highest value (0.21) using the three predictors: Atlantic SST, Pacific SST, and geopotential height over Europe. These precursors were also used to compare the spatial skill over Europe with CCA and NMME (Figure 3). Our results once again demonstrate that forecasts based on hierarchical clustering had higher skill than both the more traditional CCA statistical model and the NMME suite of dynamical models. The hierarchical clustering exhibits improved skill in representing the overall pattern of surface temperature anomalies and in the skill score of individual locations throughout the region.

We are currently applying hierarchical clustering to temperature forecasts and across North America also in winter. Results are not available at the time of submission of this manuscript but we are preparing analysis for a future publication.

7 | NORTHERN HEMISPHERE WINTER SURFACE TEMPERATURE FORECAST—WINTER 2018

Government operational forecast centers almost exclusively rely on dynamical coupled forecast systems in generating and disseminating seasonal forecast products with some notable exceptions (e.g., the Indian Meteorological Department uses statistical methods for monsoon forecasts). In contrast, AER produces an operational seasonal forecast using a statistical model (Cohen & Fletcher, 2007).

In Figure 4 we show two dynamical, December to February 18, 2017, forecasts from the NMME suite of dynamical models and the IMME suite of dynamical models to compare with the AER seasonal forecast model (the Northern Hemisphere winter forecast was posted on November 27, 2017: https://www.aer.com/siteassets/ao-archives/ao-update-27-nov-17.pdf). In Figure 4 we also show the January to March 2018 forecasts from the NMME suite of dynamical models, the IMME suite of dynamical models and the AER seasonal forecast model.

All the dynamical models predict above normal, to even well above normal, temperatures across the Eurasian continent. The forecasts across the United States were for normal to above normal temperatures but the dynamical models did predict a

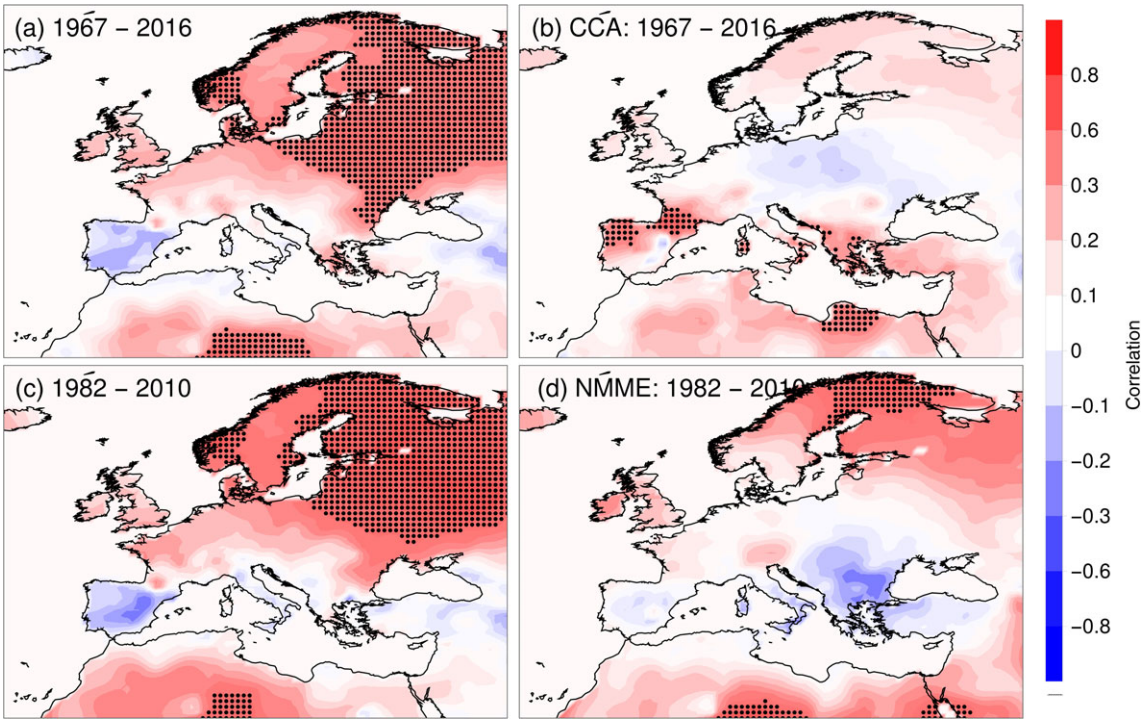


FIGURE 3 Temperature winter (December, January, and February) forecast skill: (a) cross-validated correlation of the cluster-based hindcasts with observations for the years 1967–2016. (b) Cross-validated correlation of the canonical correlation analysis hindcasts with observations for the years 1967–2016. (c) Same (a) but for the years 1982–2010. (d) Correlation of the North American Multi-Model Ensemble (NMME) hindcasts for the years 1982–2010 with observations. Significant values ($p < .05$) according to the two-sided Student's t test are shown in hatches. The cluster-based forecast performs better than the NMME models according to the cross-validated correlations

more regional area of below normal temperatures in southeastern Alaska and northwestern Canada. The cold temperature forecasts in Alaska, Canada, and the Pacific Northwest are likely related to the predicted La Niña for winter 2017/2018. Correlations of ENSO with surface temperatures show a positive correlation in western North America (Figure 5), that is, La Niña favors cold temperatures in southeast Alaska, western Canada, the Pacific northwest and east towards the Canadian Plains.

In contrast to the dynamical model forecasts, the statistical model forecast predicted normal to below normal temperatures for northern Asia, East Asia, and northern Europe. And in North America, the model predicted normal to below normal temperatures for western Canada into the upper midwest of the United States, the Great Lakes and most of the eastern United States. The model also predicted above normal temperatures for the Mediterranean region, the Middle East, southern Asia, the North American Arctic and the southwestern United States. The statistical model based on Arctic predictors correctly predicted cold across large parts of Asia, including east Asia, western Europe, central Canada and the upper midwest of the United States. The predictors that were the biggest contributors to the cold forecast across northern Eurasia were October SLP anomalies and September SIC. The biggest contributors to the cold forecast in the eastern United States were October SLP anomalies and October SCE and in the western United States it was the predicted La Niña. These are regions that were not predicted to be cold across the suite of dynamical models. Cold temperatures were more widespread across the Northern Hemisphere mid-latitudes in the latter half of winter consistent with previous studies, which argued that cold temperatures in the era of Arctic amplification have become more dependent on polar vortex (PV) disruptions that typically occur in mid- to late-winter (Cohen, Barlow, & Saito, 2009; Cohen, Pfeiffer, & Francis, 2018). In addition to the more accurate cold temperature forecast across the continents, the statistical model predicted more amplified warming across the Arctic and closer to the observed compared to the dynamical model forecasts.

The cold forecast by the statistical model but absent in the dynamical models suggests the more accurate cold forecast based on the statistical model is related to Arctic variability and is missing from the dynamical models. The more accurate winter forecast using Arctic variables is supportive of a recent study that showed Northern Hemispheric atmospheric trends are more consistent with Arctic trends than with tropical trends (Cohen, 2016). In this regard, the statistical model can be used to inform or guide improvements in the dynamical models. This argument of the limited ability of dynamical models to simulate Arctic or high latitude processes and coupling with the atmosphere is discussed in greater detail on the subseasonal scale.

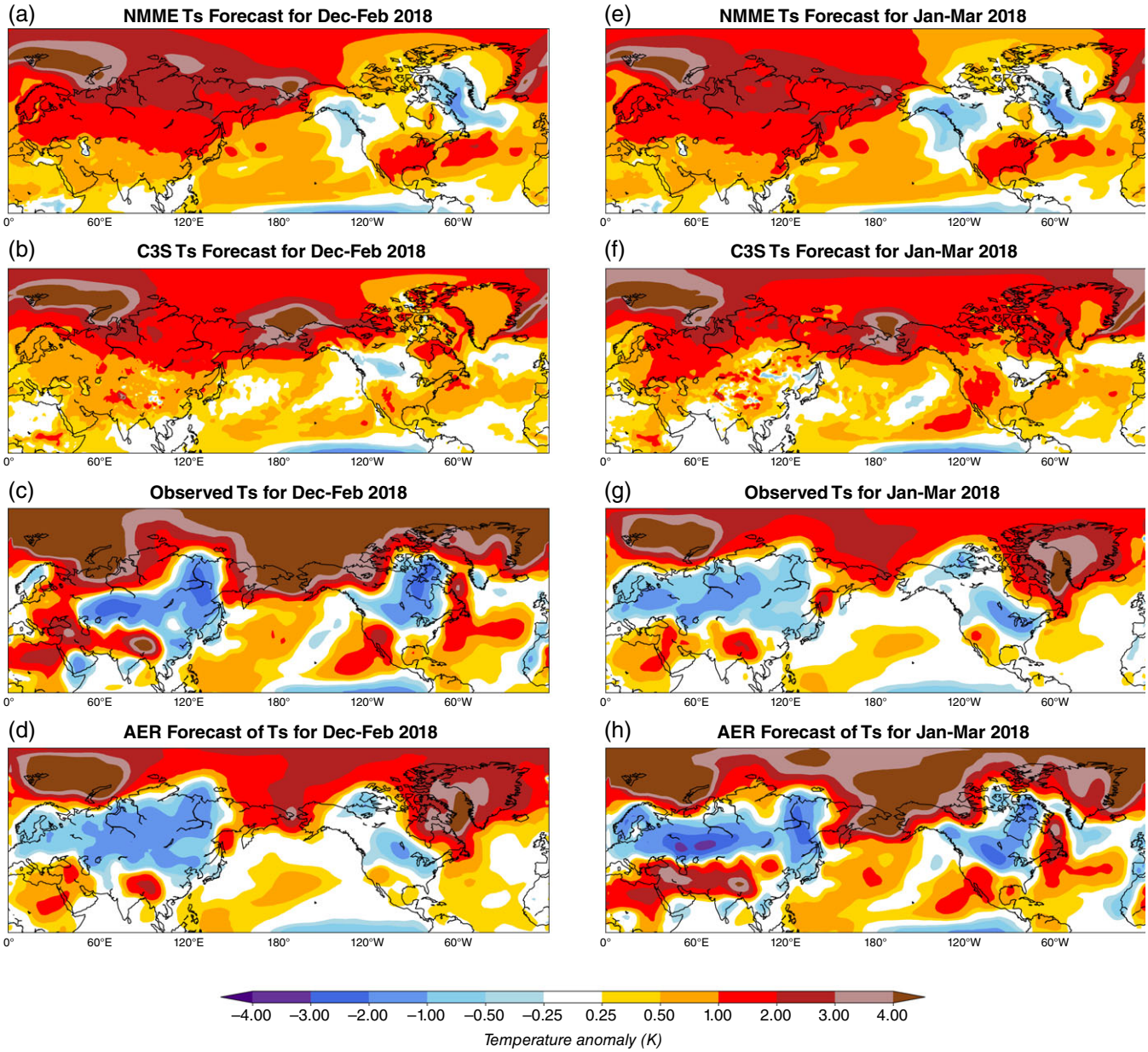


FIGURE 4 Predicted December, January, and February 2017/18 surface temperature anomalies from (a) North American Multi-Model Ensemble (NMME) suite of models, (b) International Multi-Model Ensemble (IMME) suite of models both initialized on November 1, 2017, (c) the observed surface temperature anomalies for December, January, and February 2017/18 and (d) same as (a) but for the AER statistical model initialized on November 8, 2017. Predicted January, February, and March 2018 surface temperature anomalies from the (e) NMME suite of models initialized on December 1, 2017, (f) IMME suite of models both initialized on November 1, 2017, (g) the observed surface temperature anomalies for January, February, and March 2018 and (h) same as (e) but for the AER statistical model initialized on December 1, 2017. Smoothing was applied to the statistical model and observed surface temperature anomalies

8 | POLAR VORTEX DISRUPTION FEBRUARY 2018

In addition to the overall winter seasonal forecast in winter 2017/2018, much attention was given to a significant PV disruption that resulted in a PV split and widespread severe winter weather across the Northern Hemisphere starting in mid-February including record cold and disruptive snowfalls both in Europe and the United States (https://mashable.com/2018/02/15/polar-vortex-split-stratospheric-warming-snow-cold-europe-us/#nMlfNYJ2qmqn). There was also some celebration in the climate community on how well the event was predicted in advance. On February 12, an ongoing PV disruption officially achieved major mid-winter warming (MMW; defined as a reversal of the zonal wind from westerly to easterly at 60°N and 10 hPa) status. The PV split into two pieces and was one of, or possibly the most extreme, PV disruptions observed in the month of February in the satellite era (since 1979) as defined by the magnitude of the easterly wind at 60°N and 10 hPa and the anomalously warm temperatures in the polar stratosphere (averaged between 60–90°N and 10 hPa).

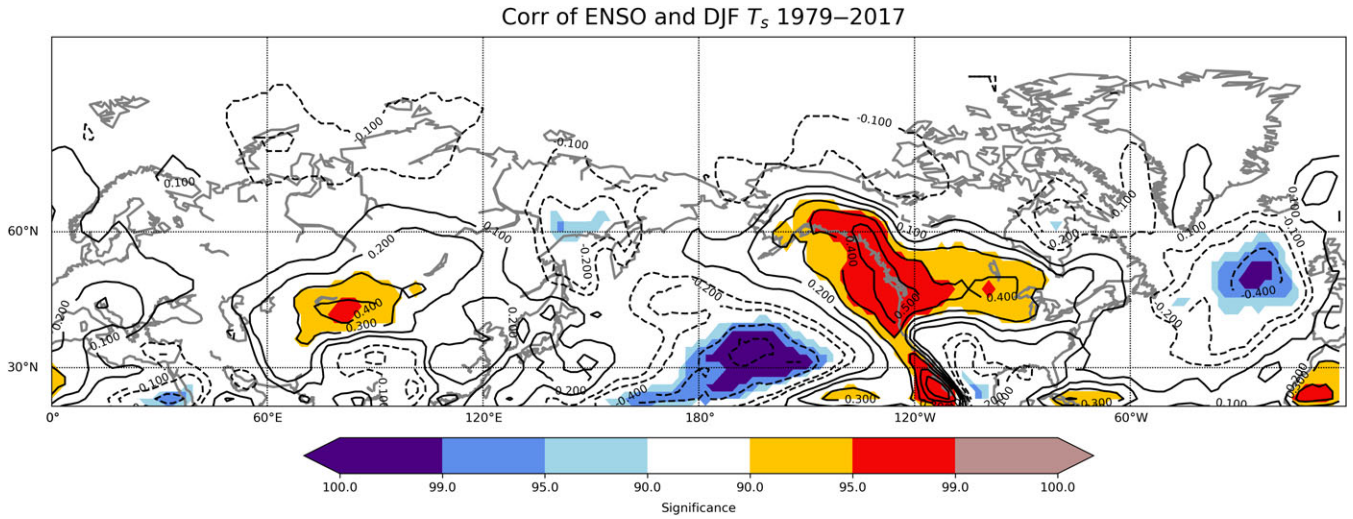


FIGURE 5 Correlation of Niño 3.4 index with surface temperatures across the Northern Hemisphere (contouring). Correlation of 90, 95 and 99% are represented by light, dark and darkest shading, respectively. Red shading represents positive correlations and blue shading represents negative correlations

It has been known for some time that following MMWs, the Northern Hemisphere atmospheric circulation is altered up to 2 months including a shifted storm track, storm frequency and temperature anomalies (Baldwin & Dunkerton, 2001; Kolstad, Breiteig, & Scaife, 2010). It has also been shown that following PV disruptions, the Arctic Oscillation is predominantly in the negative phase (Baldwin & Dunkerton, 2001), which favors increased snowfall across Europe and the northeastern United States (Cohen, Ye, & Jones, 2015). Following the MMW during the second week of February 2018, temperatures turned much colder across Europe along with disruptive snowfalls that continued through late March. In North America, initially temperatures turned colder and storminess increased across the western portion of the continent but with time the cold air and storminess transitioned to the eastern United States. The northeastern United States experienced four nor'easters in relative quick succession, each with heavy snowfall reported (50–100+ cm) during the first 3 weeks of March.

Long lead forecasts of PV disruptions can benefit regional and local municipalities and businesses with advanced warning to prepare for possible inclement winter weather. A forecast based on statistical techniques predicted a significant PV disruption nearly a month in advance. In a weekly blog postdated January 15, 2018, the first author wrote "Our PV forecast model predicts that the PV disruption will peak the second week of February." Based on the expected PV disruption, the first author further wrote "This anticipated stratospheric disruption is likely to differ from the stratospheric PV disruption in late December where the resultant below normal temperatures were focused across North America while much of Eurasia remained relatively mild….I expect the focus of the resultant cold temperatures to be across northern Eurasia, especially Siberia. Temperatures would likely average below normal across much of Siberia and likely elsewhere across northern Eurasia." (The full blog is posted here: https://www.aer.com/siteassets/ao-archives/ao-update-15-jan-18.pdf)

In contrast, dynamical predictions of the PV disruption only came two or more weeks later. A New York Times Op-Ed celebrated as an achievement, the forecast in early February of the PV disruption by dynamical models ("Both the sudden stratospheric warming and the Arctic warming were forecast with outstanding accuracy" from *Hot Times in the Arctic* March 14, 2018). Support for the outstanding forecasts was a discussion from February 3, 2018 (https://www.nymetroweather.com/2018/02/03/sudden-stratospheric-warming-increasingly-likely-mean/). A NOAA blog also argues that the numerical weather prediction models began to simulate the MMW in the late January and early February time frame (Butler, 2018). But the PV would split just a week later and achieve MMW status less than 10 days later. In addition, the greatest poleward heat flux into the polar stratosphere ever observed occurred that week and was well-predicted by the models making a PV disruption all but inevitable just a few days or a week later. Given the fewer degrees of freedom in the stratosphere, the lack of topography and diabatic heating sources and the large wavelengths of the atmospheric circulation, it is questionable whether a correct forecast of an extreme stratospheric event only 1 week to 10 days in advance should be celebrated as an outstanding achievement.

Based on Arctic variability the expectation of a significant PV disruption was noted even months earlier. The first author in a blog dated November 20, 2017 predicted a significant PV disruption in the late winter (the full blog is posted here: https://www.aer.com/siteassets/ao-archives/ao-update-20-nov-17.pdf). This prediction was based on below normal November Arctic SIC, above normal Eurasian October SCE, an easterly quasi-biennial oscillation (QBO) and a slightly negative snow advance index (Cohen & Jones, 2011a). The forecast for the winter surface temperature anomalies were included as well. The reasoning was partially based on two recent studies demonstrating that above normal Eurasian SCE coupled with an easterly

QBO favors PV disruptions later in the winter (Peings, Douville, Colin, Martin, & Magnusdottir, 2017; Tyrrell, Karpechko, & Räisänen, 2018).

The correct forecast of a significant PV disruption in late winter issued both in the fall and in mid-January was based on Arctic or high latitude variability and applied to statistical forecasts. The expectation of a significant PV disruption in November was based on Arctic sea ice and snow cover anomalies. However, the forecast in mid-January of a significant PV disruption the second week of February was mostly based on SLP anomalies across the high latitudes and surface temperatures anomalies across northern Asia. It has been previously shown that a tripole pattern in SLP anomalies with positive SLP anomalies across northwestern Eurasia, centered near the Urals, and negative SLP anomalies both upstream and downstream in the North Atlantic and the North Pacific, respectively, is favorable for disrupting the PV (Cohen, Furtado, et al., 2014; Cohen & Jones, 2011b). Comparison of January 2018 SLP anomalies and the SLP tripole pattern (Cohen, Furtado, et al., 2014) favorable for PV disruptions shows a strong similarity (Figure 6). Furthermore, a statistical model using SLP anomalies to predict the strength of the PV predicted a weakening of the PV that peaks the second week of February, a month in advance. Further support for a significant PV disruption were the cold temperatures across northern Asia during January. Kretschmer et al. (2018) showed that cold temperatures across northern Asia exist prior to significant PV disruptions. This is an example of using statistical techniques and Arctic variability provided a longer lead time to predict a significant PV disruption than dynamical model forecasts.

Besides Arctic variability, tropical variability has been proposed as a forcing/predictor of PV behavior (Butler & Polvani, 2011). In addition to phases of ENSO possibly forcing PV disruptions, the Madden-Julian Oscillation (MJO) has been proposed as a possible forcing/predictor of PV behavior. Specifically, observational analysis showed that there is an increased frequency to PV disruptions between 25 and 36 days following MJO Phase 3 and there is an increased frequency to PV disruptions between 1 and 12 days following MJO Phase 7 (Garfinkel, Feldstein, Waugh, Yoo, & Lee, 2012). Consistent with this study, the MJO was in Phase 3 in mid-January and in Phase 7 in early February. Therefore, it is possible that not only did Arctic/high latitude variability contribute to the PV disruption the second week of February but so did tropical convection. Statistical or empirical techniques using high latitude and tropical variability would have provided up to a month lead forecast of a PV disruption in early to mid-February.

Of course no one lives in the stratosphere and a successful forecast of stratospheric variability is only of importance if it results in an improved forecast of the sensible weather. For society the importance of the PV disruption is that it initiated an increase in severe winter weather, including cold temperatures and disruptive winter storms across the Northern Hemisphere including Europe and the United States for the following 3 months. It has been shown that for PV disruptions of the magnitude observed in February 2018, the temperature response is below normal temperatures widespread across northern Eurasia including Europe but with the largest negative temperature anomalies focused in Siberia (Kretschmer et al., 2018). We show in Figure 7 the monthly temperature anomaly forecasts from the suite of dynamical models for the 2 months when temperature

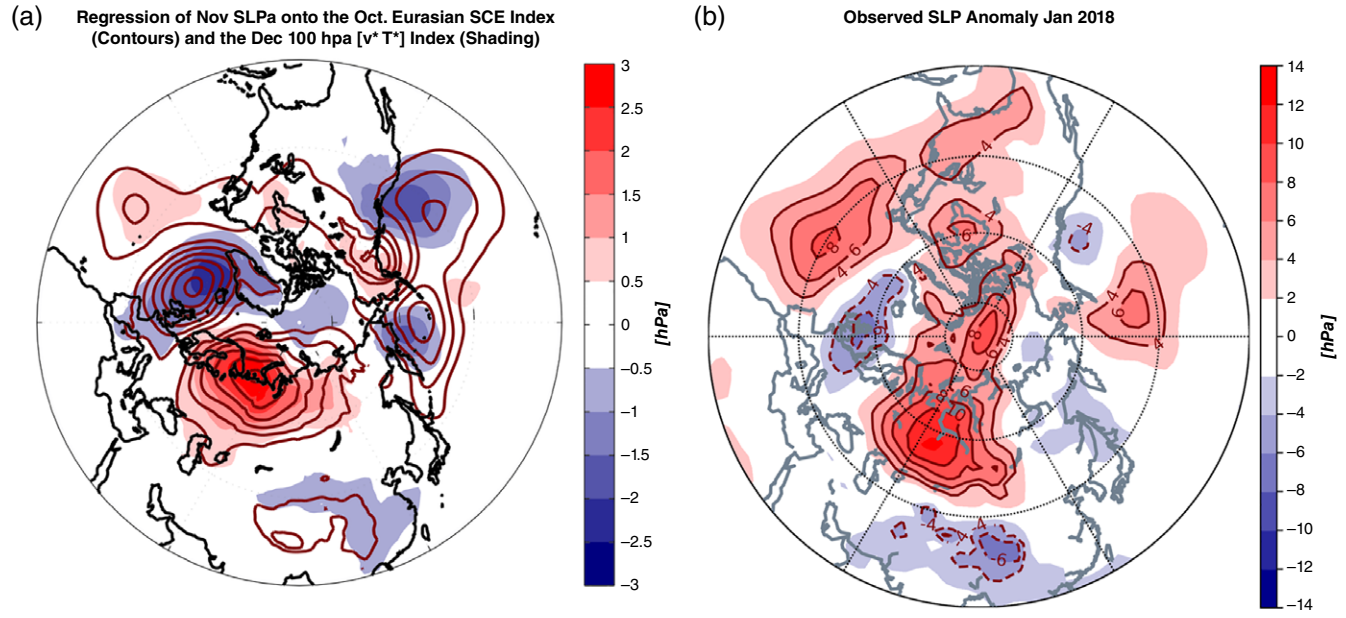


FIGURE 6 (a) Regression of November sea level pressure (SLP) anomalies (hPa) onto October monthly mean Eurasian snow cover extent (contouring) and onto December meridional heat flux anomalies at 100 hPa, averaged between 40°N and 80°N (shading). This figure is the same as fig. 4 from Cohen, Furtado, et al. (2014). (b) Observed mean SLP (contours) and SLP anomalies (shading) for January 2018

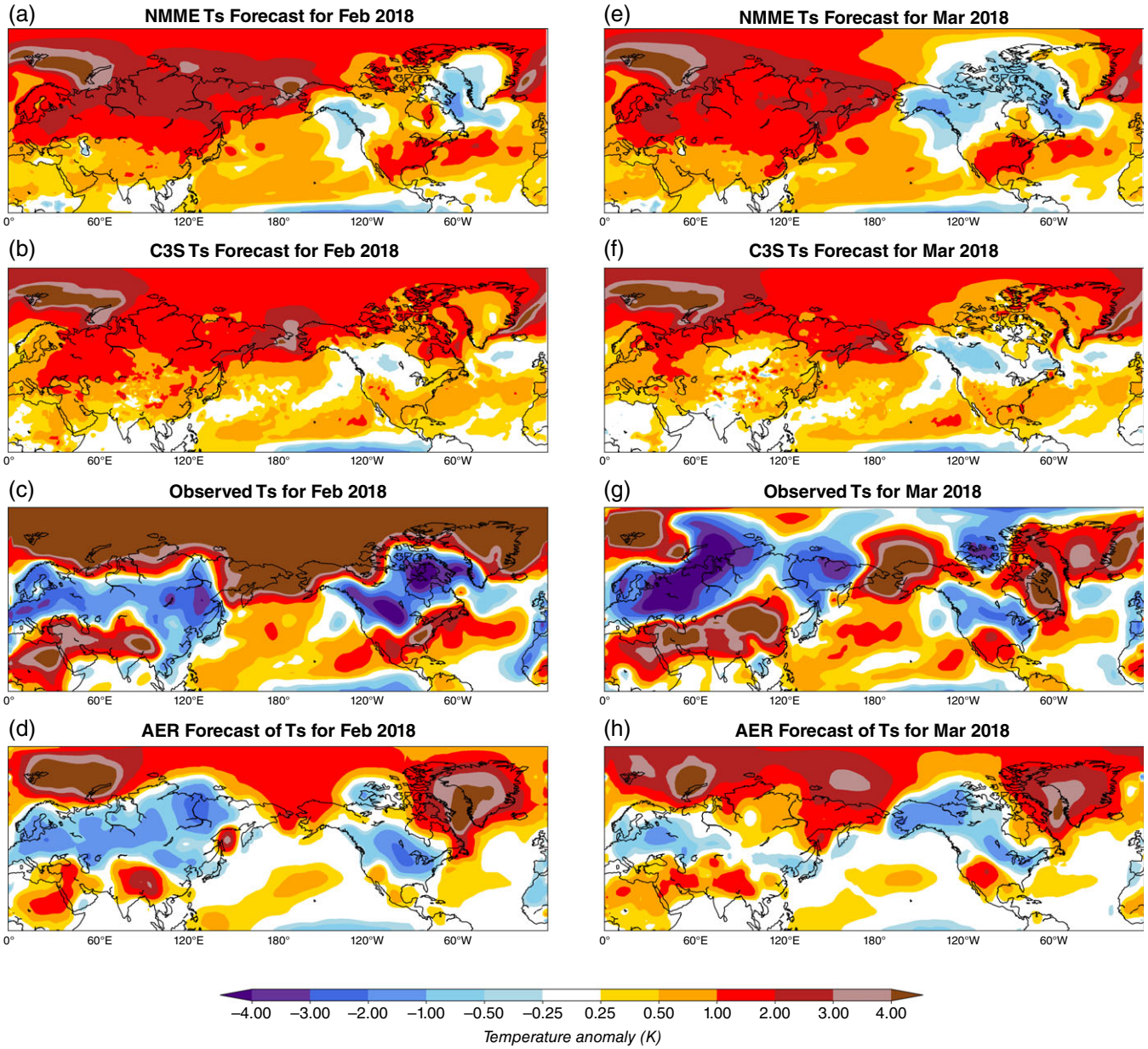


FIGURE 7 Predicted February 2018 surface temperature anomalies from (a) North American Multi-Model Ensemble (NMME) suite of models, (b) International Multi-Model Ensemble (IMME) suite of models both initialized on January 1, 2018, (c) the observed surface temperature anomalies for February 2018, and (d) same as (a) but for the AER statistical model initialized on November 8, 2017. Predicted March 2018 surface temperature anomalies from (e) NMME suite of models initialized on February 1, 2018, (f) IMME suite of models initialized on February 1, 2018, (g) the observed surface temperature anomalies for March 2018, and (h) same as (e) but for the AER statistical model initialized on December 1, 2017. Smoothing was applied to the statistical model and observed surface temperature anomalies

anomalies responded to the PV disruption—February and March. Both the American suite of dynamical models (NMME) and the international suite of models (IMME) predict overall above normal temperatures across all of Eurasia and the United States with the only cold predicted in western Canada, which as discussed with Figure 5 is likely associated with the observed winter La Niña.

We also include the observed temperature anomalies from February to March 2018. In contrast to the model forecasts, the observations show widespread cold temperatures for both months across northern Eurasia including Europe. The dynamical models are initialized with observations a month in advance, therefore, it is unlikely that the models were correctly predicting a PV disruption in early February for the February forecast initialized on January 1. However, for the March forecast initialized on February 1, the models should have correctly simulated a significant PV disruption as discussed above. Yet there is no temperature response to the PV disruption reflected in the March dynamical forecasts, which is nearly identical to the February anomaly temperature forecasts, in sign and in pattern. This is consistent with previous model analysis demonstrating that the circulation anomalies associated with PV disruptions fail to propagate from the stratosphere into the troposphere as seen in the

observations (Furtado et al., 2015). It does appear from comparing dynamical model temperature anomaly forecasts with the observed temperature anomalies, that it is almost irrelevant to the models whether a PV disruption occurs or not. Or, in other words, the dynamical model's tropospheric response is nearly insensitive to PV disruptions and by extension, if Arctic variability influences mid-latitude weather through altering PV behavior, then the models are insensitive to Arctic variability. This lack of sensitivity to Arctic variability needs to be considered as a serious shortcoming of the dynamical models that limits the capability of dynamical models accurately predicting Northern Hemisphere winter temperature variability.

We also include statistical forecasts for the months of February and March generated in the fall. The forecasts include no information about the PV disruption in early February but do include the Arctic variables September Arctic SIC and October Eurasian SCE. Previous studies link both below normal Arctic SIC and above normal October Eurasian SCE to a PV disruption in mid- to late-winter followed by widespread cold temperatures across the Northern Hemisphere continental mid-latitudes (Furtado, Cohen, & Tziperman, 2016; García-Serrano et al., 2016; Gastineau et al., 2017). This expectation of a PV disruption late in the winter was published on the AO blog and on the National Science Foundation website (https://www.nsf.gov/news/special_reports/autumnwinter/predicts.jsp). Although the statistical model forecasts are not perfect, the model did correctly predict below normal temperatures across northern Eurasia for both March and especially February, and supports the idea that the below normal temperatures across northern Eurasia are related to Arctic variability.

Finally, from observational analysis it can be concluded that Arctic variability contributed to the PV disruption of February 2018 or that MJO variability contributed to the PV disruption of February 2018. From the dynamical model forecasts it can be concluded that both Arctic and tropical variability are not significant contributors to mid-latitude variability but one cannot conclude that tropical variability is a significant contributor to mid-latitude variability while Arctic variability is not. This has important implications for the argument that dynamical models support the influence of tropical forcing but not Arctic forcing on mid-latitude weather variability.

9 | CONCLUSION

Over the past two decades, most of the effort and the resources at government-sponsored forecast centers have been dedicated to atmosphere–ocean coupled dynamical models. Current statistical techniques have not been updated since the 1990s and even the 1980s. An influential National Academy of Sciences report on S2S prediction recommended that new resources be dedicated to improving dynamical models but ignored statistical models, contributing to the migration away from statistical models and towards dynamical models. As a consequence, statistical techniques and models are currently mostly ignored at the government-sponsored forecast centers.

We argue that the lack of attention, resources and implementation of statistical techniques is a mistake and deprives the public of immediate and relatively inexpensive improvements to S2S prediction. As we describe above, new statistical techniques, often labeled as machine learning, have been developed that are far more powerful at mining data and recognizing patterns than traditional techniques that can be applied to delivering to the public more skillful forecasts.

As we demonstrated above using seasonal hindcasts, new statistical techniques can provide more skillful forecasts of both precipitation and surface temperature than the current state-of-the-art coupled dynamical systems. The statistical model that we used employed hierarchical clustering and included fall Arctic boundary forcings as predictors. We showed this for Europe and hope to demonstrate shortly for North America as well. We also discussed other novel machine learning techniques developed in other disciplines that may be appropriate to apply to the S2S prediction problem.

Besides demonstrating improved skill for hindcasts, we showed that statistical techniques provided a more accurate winter temperature forecast for winter 2017/2018, better representing the observed “warm Arctic/cold continents” (Cohen, Jones, Furtado, & Tziperman, 2013) pattern than the dynamical models. In addition, statistical or empirical analysis and models can be exploited to guide improvements in dynamical model development. As we discussed above, an extreme PV disruption developed in early February 2018 and achieved MMW status on February 12. Following the PV disruption, cold temperatures and disruptive snowfalls became widespread across northern Eurasia including Europe and across the United States for the next 2 months. The increase in severe winter weather is consistent with observational analysis of the atmospheric response to PV disruptions (Kretschmer et al., 2018).

In contrast to observed temperature anomalies, the dynamical models predicted a relatively mild winter for all months from December to March, both across Eurasia and much of North America. The lone region predicted to experience a cold winter was northwestern North America and is likely related to the predicted and verified La Niña. Little change is seen in the dynamical forecasts between the seasonal mean and individual months. As we showed, there is little change in the hemispheric temperature anomaly dynamical model forecast between February and March temperature anomalies despite that the dynamics of the PV disruption were included in the initialization of the PV disruption for the March forecast initialized in early February but was likely absent in the February forecast initialized in early January. Below normal temperatures became

widespread across the hemisphere following the PV disruption both in February and especially March. Below normal temperatures were already present across northern Asia at the end of January and into early February. Dynamical model forecasts for March were initialized with both the cold temperatures in Asia and the beginnings of the PV disruption. Yet despite the initial conditions, the dynamical models predicted universal warmth across Eurasia for the month of March. This required both incorrectly advecting warm temperatures from the Arctic or the subtropics across the continent, which negated and reversed the cold temperatures already present, and neglecting the atmospheric circulation response to an extreme PV disruption. The possible incorrect advection of the warmth generated in the Arctic from sea ice loss too far south across the continents is consistent with a previous study showing that Arctic warming related to sea ice loss extends further south in the models relative to the observations (Cohen et al., 2013).

Therefore, in winter 2017/2018 the dynamical models predicted a similar temperature anomaly pattern on both climate and synoptic timescales. Climate timescales are dominated by boundary forcings and natural variability. Synoptic timescales are dominated by initial conditions. Yet the dynamical models failed to predict the observed warm Arctic cold continents pattern on longer climate leads in the fall and on shorter synoptic scales in early February for the February and March temperature forecasts. So whether the dynamical models were initialized with the beginnings of the PV disruption or not, they struggled to predict the temperature response to the PV disruption until mid-February (Butler, 2018), when the circulation anomalies associated with the stratospheric PV disruption descended to the troposphere (Figure S1). The consistency between the dynamical model forecasts across timescales does not support that differences between observed and predicted temperatures can be attributed to only natural variability. The dynamical models only correctly predicted the warm Arctic cold continents pattern once they were initialized with a weakened PV in the upper troposphere.

In contrast, the observed temperature anomalies following the PV disruption closely follow the temperature response to such events derived from one of the new statistical techniques discussed above, hierarchical clustering. Also, a statistical model forced with Arctic variability better predicted the late winter hemisphere-wide cold than the dynamical models. In this regard, statistical models can help guide model developers to improve the physical processes in the coupled dynamical models, for example, high latitude processes that involve the PV.

Statistical predictions in general and machine learning techniques in particular are likely to improve subseasonal climate forecasts where there are repeatable patterns in the atmosphere and where there are fairly consistent sequences of events. We provided an example where statistical predictions can help improve temperature predictions following a PV disruption relative to forecasts that are derived solely from dynamical systems. For many PV disruptions there is an identifiable tropospheric wave pattern followed by a weakening of the PV and then cold air outbreaks across the Northern Hemisphere continents but most favored across northern Eurasia. As demonstrated above, dynamical models struggle with simulating this sequence of events. It seems reasonable that weather predictions associated with blocking events and/or periods of amplified known teleconnection patterns would also benefit from statistical techniques and machine learning.

Seasonal prediction had humble beginnings when simple and easy-to-employ statistical techniques such as persistence or composites were used to generate forecasts. During the 1980s, more sophisticated statistical techniques such as CCA were incorporated into seasonal prediction. Then, beginning in the 1990s, forecasts from dynamical models of increasing complexity were introduced into the operational production of S2S forecasts while inclusion of statistical techniques was phased out. We argue that currently the pendulum has swung to the other extreme where S2S forecasts are almost exclusively derived from the coupled dynamical systems while new techniques broadly referred to as "machine learning" that can both improve forecast skill and improve coupled dynamical systems are being ignored. The introduction of new statistical techniques into the operational forecast centers would benefit the public and ultimately the public is best served by hybrid forecasts utilizing both state-of-the-art dynamical models and updated statistical techniques.

ACKNOWLEDGMENTS

We thank two anonymous reviewers for their constructive comments that resulted in significant improvement of the manuscript. J.C. is supported by the National Science Foundation grants AGS-1303647 and PLR-1504361. E.T. is supported by the NSF Climate Dynamics program, grant AGS-1622985, E.T. thanks the Weizmann Institute for its hospitality during parts of this work. GHCN Gridded V2 data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their website at https://www.esrl.noaa.gov/psd/. We thank Karl Pfeiffer for the help in revising some of the figures.

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

REFERENCES

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

Barnett, T. P. (1981). Statistical prediction of North American air temperatures from Pacific predictors. *Monthly Weather Review*, *109*, 1021–1041.

Barnett, T. P., & Preisendorfer, R. W. (1987). Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis. *Monthly Weather Review*, *115*, 1825–1850.

Barnston, A. G., Kumar, A., Goddard, L. M., & Hoerling, M. P. (2005). Improving seasonal prediction practices through attribution of climate variability. *Bulletin of the American Meteorological Society*, *86*, 59–72.

Barnston, A. G., Leetmaa, A., Kousky, V. E., Livezey, R. E., O'Lenic, E. A., van den Dool, H., … Unger, D. A. (1999). NCEP forecasts of the El Niño of 1997–98 and its impacts. *Bulletin of the American Meteorological Society*, *80*, 1829–1852. https://doi.org/10.1175/1520-0477(1999)080<1829:NFOTEN>2.0.CO;2

Barnston, A. G., & Livezey, R. E. (1987). Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Monthly Weather Review*, *115*, 1083–1126.

Barnston, A. G., & Smith, T. M. (1996). Specification and prediction of global surface temperature and precipitation from global SST using CCA. *Journal of Climate*, *9*, 2660–2697.

Barnston, A. G., Tippett, M. K., L'Heureux, M. L., Li, S., & DeWitt, D. G. (2012). Skill of real-time seasonal ENSO model predictions during 2002–11: Is our capability increasing? *Bulletin of the American Meteorological Society*, *93*, 631–651.

Barnston, A. G., van den Dool, H. M., Rodenhuis, D. R., Ropelewski, C. R., Kousky, V. E., O'Lenic, E. A., … Leetmaa, A. (1994). Long-lead seasonal forecasts—Where do we stand? *Bulletin of the American Meteorological Society*, *75*, 2097–2114.

Butler, A. H. (2018). Retrieved from https://www.climate.gov/news-features/blogs/enso/february-and-march-madness-how-winds-miles-above-arctic-may-have-brought#.WuSjok5MN9w.twitter

Butler, A. H., & Polvani, L. M. (2011). El Niño, La Niña, and stratospheric sudden warmings: A reevaluation in light of the observational record. *Geophysical Research Letters*, *38*, L13807. https://doi.org/10.1029/2011GL048084

Cohen, J. (2016). An observational analysis: Tropical relative to Arctic influence on mid-latitude weather in the era of Arctic amplification. *Geophysical Research Letters*, *43*, 5287–5294. https://doi.org/10.1002/2016GL069102

Cohen, J., Barlow, M., & Saito, K. (2009). Decadal fluctuations in planetary wave forcing modulate global warming in late boreal winter. *Journal of Climate*, *22*, 4418–4426.

Cohen, J., & Fletcher, C. (2007). Improved skill for northern hemisphere winter surface temperature predictions based on land–atmosphere fall anomalies. *Journal of Climate*, *20*, 4118–4132.

Cohen, J., Furtado, J., Jones, J., Barlow, M., Whittleston, D., & Entekhabi, D. (2014). Linking Siberian snow cover to precursors of stratospheric variability. *Journal of Climate*, *27*, 5422–5432.

Cohen, J., & Jones, J. (2011a). A new index for more accurate winter predictions. *Geophysical Research Letters*, *38*(21), 1–6. https://doi.org/10.1029/2011GL049626

Cohen, J., & Jones, J. (2011b). Tropospheric precursors and stratospheric warmings. *Journal of Climate*, *24*, 6562–6572.

Cohen, J., Jones, J., Furtado, J. C., & Tziperman, E. (2013). Warm Arctic, cold continents: A common pattern related to Arctic Sea ice melt, snow advance, and extreme winter weather. *Oceanography*, *26*(4), 150–160. https://doi.org/10.5670/oceanog.2013.70

Cohen, J., Pfeiffer, K., & Francis, J. (2018). Warm Arctic episodes linked with increased frequency of extreme winter weather in the United States. *Nature Communications*, *9*, 869. https://doi.org/10.1038/s41467-018-02992-9

Cohen, J., Screen, J. A., Furtado, J. C., Barlow, M., Whittleston, D., Coumou, D., … Jones, J. (2014). Recent Arctic amplification and extreme mid-latitude weather. *Nature Geoscience*, *7*, 627–637. https://doi.org/10.1038/ngeo2234

Cohen, J., Ye, H., & Jones, J. (2015). Trends and variability in rain-on-snow events. *Geophysical Research Letters*, *42*, 7115–7122. https://doi.org/10.1002/2015GL065320

Delecluse, P., Davey, M. K., Kitamura, Y., Philander, S. G. H., Suarez, M., & Bengtsson, L. (1998). Coupled general circulation modeling of the tropical Pacific. *Journal of Geophysical Research: Oceans*, *103*, 14357–14373.

Doblas-Reyes, F. J., Garcia-Serrano, J., Lienert, F., Biescas, F. P., & Rodrigues, L. R. L. (2013). Seasonal climate predictability and forecasting: Status and prospects. *WIREs Climate Change*, *4*, 245–268. https://doi.org/10.1002/wcc.217

Fan, Y., & van den Dool, H. (2008). A global monthly land surface air temperature analysis for 1948-present. *Journal of Geophysical Research*, *113*, D01103. https://doi.org/10.1029/2007JD008470

Furtado, J. C., Cohen, J., & Tziperman, E. (2016). The combined influences of autumnal snow and sea ice on northern hemisphere winters. *Geophysical Research Letters*, *43*, 3478–3485. https://doi.org/10.1002/2016GL068108

Furtado, J. C., Cohen, J. L., Butler, A. H., Riddle, E. E., & Kumar, A. (2015). Eurasian snow cover variability, winter climate, and stratosphere–troposphere coupling in the CMIP5 models. *Climate Dynamics*, *45*, 2591–2605.

García-Serrano, J., Frankignoul, C., King, M. P., Arribas, A., Gao, Y., Guemas, V., … Sanchez-Gomez, E. (2016). Multi-model assessment of linkages between eastern Arctic Sea-ice variability and the Euro-Atlantic atmospheric circulation in current climate. *Climate Dynamics*, *49*, 2407–2429. https://doi.org/10.1007/s00382-016-3454-3

Garfinkel, C. I., Feldstein, S. B., Waugh, D. W., Yoo, C., & Lee, S. (2012). Observed connection between stratospheric sudden warmings and the Madden-Julian Oscillation. *Geophysical Research Letters*, *39*, L18807. https://doi.org/10.1029/.2012GL053144

Gastineau, G., García-Serrano, J., & Frankignoul, C. (2017). The influence of autumnal Eurasian snow cover on climate and its link with Arctic Sea ice cover. *Journal of Climate*, *30*(19), 7599–7619. https://doi.org/10.1175/JCLI-D-16-0623.1

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., … Joseph, D. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, *77*, 437–471.

Kirtman, B. P., Min, D., Infanti, J. M., Kinter, J. L., III, Paolino, D. A., Zhang, Q., … Wood, E. F. (2014). The North American multimodel ensemble: Phase-1 seasonal-to-interannual prediction; Phase-2 toward developing intraseasonal prediction. *Bulletin of the American Meteorological Society*, *95*(4), 585–601. https://doi.org/10.1175/BAMS-D-12-00050.1

Kirtman, B. P., Shukla, J., Huang, B. H., Zhu, Z. X., & Schneider, E. K. (1997). Multiseasonal predictions with a coupled tropical ocean–global atmosphere system. *Monthly Weather Review*, *125*, 789–808.

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

Kretschmer, M., Coumou, D., Agel, L. A., Barlow, M. A., Tziperman, E., & Cohen, J. L. (2018). More-persistent weak stratospheric polar vortex states linked to cold extremes. *Bulletin of the American Meteorological Society*, *99*(1), 49–60. https://doi.org/10.1175/BAMS-D-16-0259.1

Kug, J. S., Kang, I. S., & Choi, D. H. (2008). Seasonal climate predictability with tier-one and tier-two prediction systems. *Climate Dynamics*, *31*, 403–416. https://doi.org/10.1007/s00382-007-0264-7

Latif, M., Collins, M., Pohlmann, H., & Kennlyside, N. (1998). A review of the predictability and prediction of ENSO. *Journal of Geophysical Research*, *103*, 14375–14393.

Livezey, R. E., & Barnston, A. G. (1988). An operational multi-field analog anti-analog prediction system for United States seasonal temperatures. 1. System design and winter experiments. *Journal of Geophysical Research*, *93A*, 10953–10974.

Livezey, R. E., & Timofeyeva, M. M. (2008). The first decade of long-lead U.S. seasonal forecasts. *Bulletin of the American Meteorological Society*, *89*, 843–854.

Namias, J. (1953). Thirty-day forecasting: A review of a ten-year experiment. *Meteorological monograph* (No. 2). American Meteorological Society, Washington, DC.

National Academies of Sciences, Engineering, and Medicine. (2016). *Next generation earth system prediction: Strategies for subseasonal to seasonal forecasts*. Washington, DC: National Academies Press. https://doi.org/10.17226/21873

National Oceanic and Atmospheric Administration. (2002). NOAA economic statistics (p. 26). Office of Policy and Strategic Planning, US Department of Commerce.

Overland, J. E., Francis, J. A., Hall, R., Hanna, E., Kim, S.-J., & Vihma, T. (2015). The melting Arctic and mid-latitude weather patterns: Are they connected? *Journal of Climate*, *28*, 7917–7932. https://doi.org/10.1175/JCLI-D-14-00822.1

Peings, Y., Douville, H., Colin, J., Martin, D. S., & Magnusdottir, G. (2017). Snow—(N)AO teleconnection and its modulation by the quasi-biennial oscillation. *Journal of Climate*, *30*(24), 10211–10235.

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., … Kaplan, A. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, *108*(D14), 4407. https://doi.org/10.1029/2002JD002670

Reeves, R. W., & Gemmill, D. D. (2004). *Climate prediction center: Reflections on 25 years of analysis, diagnosis, and prediction*. Washington, DC: US Government Printing Office.

Robinson, D. A., Estilow, W. T., & Program, N. C. (2012). NOAA climate date record (CDR) of northern hemisphere (NH) snow cover extent (SCE). https://doi.org/10.7289/V5N014G

Spencer, H., & Slingo, J. M. (2003). The simulation of peak and delayed ENSO teleconnections. *Journal of Climate*, *16*, 1757–1774. https://doi.org/10.1175/1520-0442(2003)016<1757:TSOPAD>2.0.CO;2

Stockdale, T. N., Anderson, D. L. T., Balmaseda, M. A., Doblas-Reyes, F. J., Ferranti, L., Mogensen, K., … Vitart, F. (2011). ECMWF seasonal forecast system 3 and its prediction of sea surface temperature. *Climate Dynamics*, *37*, 455–471. https://doi.org/10.1007/s00382-010-0947-3

Totz, S., Tziperman, E., Coumou, D., Pfeiffer, K., & Cohen, J. (2017). Winter precipitation forecast in the European and Mediterranean regions using cluster analysis. *Geophysical Research Letters*, *44*, 12418–12426. https://doi.org/10.1002/2017GL075674

Troccoli, A. (2010). Seasonal climate forecasting. *Meteorological Applications*, *17*, 251–268.

Tyrrell, N. L., Karpechko, A. Y., & Räisänen, P. (2018). The influence of Eurasian snow extent on the northern extratropical stratosphere in a QBO resolving model. *Journal of Geophysical Research: Atmospheres*, *123*(1), 315–328.

van Oldenborgh, G. J., Balmaseda, M. A., Ferranti, L., Stockdale, T. N., & Anderson, D. L. T. (2005a). Did the ECMWF seasonal forecast model outperform statistical ENSO forecast models over the last 15 years? *Journal of Climate*, *18*(16), 3240–3249.

van Oldenborgh, G. J., Balmaseda, M. A., Ferranti, L., Stockdale, T. N., & Anderson, D. L. T. (2005b). Evaluation of atmospheric fields from the ECMWF seasonal forecasts over a 15-year period. *Journal of Climate*, *18*(16), 3250–3269.

Vihma, T. (2014). Effects of Arctic Sea ice decline on weather and climate: A review. *Surveys in Geophysics*, *35*(5), 1175–1214. https://doi.org/10.1007/s10712-014-9284-0

Voldoire, A., Sanchez-Gomez, E., Salas y Mélia, D., Decharme, B., Cassou, C., Sénési, S., … Chauvin, F. (2013). The CNRM-CM5,1 global climate model: Description and basic evaluation. *Climate Dynamics*, *40*(9–10), 2091–2121. https://doi.org/10.1007/s00382-011-1259-y

Wagner, A. J. (1989). Medium- and long-range forecasting. *Weather and Forecasting*, *4*, 413–426.

Wallace, J. M., & Gutzler, D. D. (1981). Teleconnections in the geopotential height field during the northern hemisphere winter. *Monthly Weather Review*, *109*, 784–812.

Walters, D., Boutle, I., Brooks, M., Melvin, T., Stratton, R., Vosper, S., … Xavier, P. (2017). The met Office unified model global atmosphere 6.0/6.1 and JULES global land 6.0/6.1 configurations. *Geoscience Model Development*, *10*, 1487–1520. https://doi.org/10.5194/gmd-10-1487-2017

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Cohen J, Coumou D, Hwang J, et al. S2S reboot: An argument for greater inclusion of machine learning in subseasonal to seasonal forecasts. *WIREs Clim Change*. 2018;e00567. https://doi.org/10.1002/wcc.567

Supplementary-information for the paper: "S2S Reboot: An Argument for Greater Inclusion of Machine Learning in Subseasonal to Seasonal (S2S) Forecasts"

Judah Cohen, Dim Coumou Jessica Hwang, Lester Mackey, Paulo Orenstein, Sonja Totz and Eli Tziperman

1 Cluster-based prediction

1.1 Introduction

The supplemental information contains a description of the clustering-based prediction approach, which is an improved version of that presented in S. Totz, E. Tziperman, D. Coumou, K. Pfeiffer, and J. Cohen, "Winter precipitation forecast in the european and mediterranean regions using cluster analysis." Geophys. Res. Lett., 44 (doi:10.1002/2017GL075674):12,418–12,426, 2017. The corresponding code is available at
https://www.seas.harvard.edu/climate/eli/Downloads/Clustering-based-prediction/European-temperature-2018b/

1.2 Cluster-based prediction methodology

Given the time series of the quantity to be predicted (predictand, e.g., anomaly winter (DJF) precipitation) and precursors (predictors, e.g., autumn (SON) sea ice cover and snow cover extent), we calculate the clusters of the predictand, and then use them to construct the prediction as described below. In order to obtain a cross-validated forecast, we choose one year to be predicted and then use all other years in order to build the prediction model. This is repeated for all years and the skill presented below is the average over all of these prediction calculations. For each predicted year, we

first remove the mean of the precipitation using data from all years except for the predicted year.

Consider a forecast of precipitation anomaly time series at several locations, given by the predictand vector $\mathbf{prcp}(t)$. These precipitation data will be predicted using given precursors, e.g., time series of snow cover extent anomalies at several spatial locations given by the time-dependent vector $\mathbf{sce}(t)$, and time series of sea ice extent at several spatial locations, $\mathbf{sic}(t)$.

We assume that there are $N_{clusters}$ significant precipitation clusters. We use bold upper case variable names to denote clusters and composites, and lower case bold variable names to denote time series data. The prediction procedure requires the winter (DJF) precipitation clusters $\mathbf{PRCP}_i$, $i = 1, \ldots, N_{clusters}$ and the corresponding precursor composites (e.g., sea ice cover and snow cover extent anomalies from the autumn SON mean), $\mathbf{COMPOSITE}_i$. The clusters are calculated using hierarchical clustering of the winter precipitation anomaly data, while the composites for a given cluster i are calculated by averaging the predictors over all times in which the precipitation anomaly is assigned to its cluster i.

We also need a time series of the autumn-mean (averaged over SON) precursor anomaly (predictors) $\mathbf{precursor}_{SON}(t)$, for each spatial location. The time t denotes the year, where the precursors are evaluated during the fall (SON) and the precipitation of that year refers to the following DJF. For example, if the precursors are sea ice and snow cover, the vector of precursors (predictors) time series, and the vector of composites are calculated as follows.

First, we remove the mean of each precursor using all precursors data except the predicted year. Next, we normalize each precursor by the standard deviation. Finally, we combine different precursors into a single vector,

$$\begin{aligned} \mathbf{sic}'_{SON}(t) &= \mathbf{sic}_{SON}(t) - \overline{\mathbf{sic}}_{SON} \\ \widehat{\mathbf{sic}}_{SON}(t) &= \mathbf{sic}'_{SON}(t)/\sigma_{\mathbf{sic}} \\ \mathbf{precursor}_{SON}(t) &= (\widehat{\mathbf{sic}}_{SON}(t), \widehat{\mathbf{sce}}_{SON}(t))^T \end{aligned}$$

The variable $\overline{\mathbf{sic}}_{SON}$ is the time mean of the sea ice concentration using all times except the predicted year. The variable $\sigma_{\mathbf{sic}}$ is the standard deviation over all times and all grid points.

Then, we find the composites of the different autumn predictors by averaging the normalized predictors ($\widehat{\mathbf{sce}}(t)$,$\widehat{\mathbf{sic}}(t)$) over all autumn seasons (SON) for which the following winter precipitation anomaly is assigned to a

given cluster. The predictor's composites of the same cluster are combined into one composite

$$\mathbf{COMPOSITE}_{1,2} \;=\; (\mathbf{SIC}_{1,2}, \mathbf{SCE}_{1,2})^T$$

To obtain the prediction for the precipitation, we first find the projection of the current state of the predictors (snow cover and sea ice) on the $N_{clusters}$ predictor composites corresponding to the precipitation clusters.

Each predictor composite is associated with a precipitation cluster and provides information about the amplitude and spatial structure of winter precipitation expected given the autumn predictor composite. This allows us to calculate the expected precipitation pattern due to the projection of the current state of predictors on each cluster. Finally, we sum the contributions to the precipitation due to all clusters, to obtain the predicted total precipitation anomaly.

Mathematically, this proceeds as follows. To calculate the projection of $\mathbf{precursor}_{SON}(t)$ on the composite $\mathbf{COMPOSITE}_i$, we expand the current precursor state in terms of the precursor composites, to find the expansion coefficients, noting that the composites are not necessarily orthogonal. The expansion takes the form,

$$\mathbf{precursor}_{SON}(t) \approx \sum_{i=1}^{N_{clusters}} a_i(t)\; \mathbf{COMPOSITE}_i.$$

The expansion may only be approximate because the composites are not necessarily a complete set of vectors. To find the expansion coefficients $a_i(t)$, multiply by precursor composite $\mathbf{COMPOSITE}_j$, remembering that they are not necessarily orthogonal,

$$\mathbf{precursor}_{SON}(t) \cdot \mathbf{COMPOSITE}_j \;=\; \sum_{i=1}^{N_{clusters}} a_i(t)\; \mathbf{COMPOSITE}_i \cdot \mathbf{COMPOSITE}_j.$$

Next, we write this as a matrix equation for the unknown vector $\mathbf{a}(t)$ of coefficients $a_i(t)$. Define a matrix, $B_{ij} = \mathbf{COMPOSITE}_i \cdot \mathbf{COMPOSITE}_j$, and the right-hand side $\mathbf{\Gamma}_j(t) = \mathbf{precursor}_{SON}(t) \cdot \mathbf{COMPOSITE}_j$. This leads to the linear equations,

$$B\,\mathbf{a}(t) \;= \mathbf{\Gamma}(t),$$

that may be solved for the coefficients $a_i(t)$ at every time step (year t) in the data. Given that the matrix B may be ill conditioned, there may be

many solutions for $\mathbf{a}(t)$. We choose the one with the smallest norm, using the SVD-based pseudo inverse such that singular values that are smaller than 1% of the largest singular value are set to zero (using python's pinv-function with the threshold set to 0.01).

The final expression for the predicted precipitation anomaly is obtained by summing the contribution of all clusters, each multiplied by the projection of the current state of precursors, $a(i)$,

$$\mathbf{prcp}(t) = \sum_{i=1}^{N_{clusters}} a_i(t)\ \mathbf{PRCP}_i. \tag{1}$$

2 Alternative Statistical Learning Approaches

kNN

Given the vector of features (e.g., lagged measurements, model forecasts, temporal characteristics, and geographic characteristics) associated with a target date and forecast region, a k-nearest neighbor (kNN) method would search for the historical dates and regions ("neighbors") with features most similar to the target. The predicted weather pattern would then be a weighted average of the realized weather patterns associated with all neighbors. Such kNN approaches are especially popular in recommender systems (*Bobadilla et al.*, 2013), where the algorithm is used to recommend items similar to items previously enjoyed by a customer or to recommend items enjoyed by customers similar to target customer. See Chapter 13 of *Hastie et al.* (2001) for more details.

Random forests

A decision tree is a prediction method that hierarchically partitions forecasting targets into homogeneous groups based on associated features (e.g., lagged measurements, location, and model forecasts) and forecasts the average historical weather pattern in each group. A random forest is an ensemble method that aggregates many different trees by averaging their predictions. To make the individual trees more diverse, the method uses only a randomly selected subset of features to create each partition. Random forests (*Breiman*, 2001) and the closely related Bayesian additive regression tree method (*Chipman et al.*, 2010) have led to state-of-the-art performance in a wide variety of prediction tasks including predicting disease progression in Lou Gehrig's disease patients (*Küffner et al.*, 2015) and identifying breast lesions at high risk of cancer (*Bahl et al.*, 2017). For more details and examples, see Chapter 8 of *James et al.* (2013).

Boosted decision trees

Boosting (*Freund & Schapire*, 1997) is a learning method which sequentially combines lower-accuracy prediction rules, like decision trees, into a final higher accuracy ensemble. For boosted decision trees, we train a sequence of decision trees sequentially, each to correct the errors of the last:

start by growing a tree to predict a target variable (e.g., future temperature on a given location), build a second tree to predict the mistakes made by the first tree, and then keep building trees to predict on the errors of the previous one. The aim is to improve prediction performance with the addition of each new tree. Boosting has delivered state-of-the-art performance for a variety of prediction tasks including Higgs Boson classification in high-energy physics and insurance claim classification (*Chen & Guestrin*, 2016). More details can be found in Chapter 8 of *James et al.* (2013).

Gaussian processes

An extremely popular model in spatial statistics, Gaussian process regression, views the response variable (temperature, precipitation, etc.) as a smooth spatial surface. The smoothness of the surface is controlled by the covariance function of the Gaussian process, and spatial trends in the response variable are controlled by the mean function. The mean function can depend on additional features such as lagged measurements or other model forecasts. The result is a regression model that takes into account spatial dependencies. Gaussian processes have been used to forecast wind speed (*Chen et al.*, 2014) and predict forest biomass (*Banerjee et al.*, 2008). For an overview of the methodology, see *Rasmussen & Williams* (2006).

Neural networks

Neural networks are a highly flexible model class for relating a collection of inputs (e.g., lagged measurements or model forecasts at a set of locations) to a collection of outputs (e.g., temperature measurements at a set of locations). The inputs undergo a series of nonlinear transformations in the neural network's hidden layers; as the neural network is trained, the weights associated with these nonlinear transformations are learned in order to minimize prediction error. Neural networks, particularly deep networks with many hidden layers, have dramatically improved performance on a variety of learning tasks, including image recognition and machine translation (see, e.g., *Deng & Yu* (2014)).

Causal effect networks

Causal discovery algorithms allow for interpretation of causal links between variables by determining whether we can say that, statistically, x provides more information about future values of y than past values of y alone. The causal effect network (CEN) aims to detect causal relationships amongst a set of time-series by iteratively testing the partial correlations conditioning on combinations of other time-series at different lags *Kretschmer et al.* (2016). Thus, causal links in the CEN are those for which the linear relationship cannot be explained by the (combined) influence of other included indices or by auto-correlation. CEN is related to Granger-causality but allows for much stronger causal statements beyond, for example, the bi-variate only concept *Kretschmer et al.* (2016).

Classically, one of the major limitations of statistical forecast models has been overfitting, which results in very high correlations to R-squared values on training data but the forecast fails on independent test data. Recent studies have introduced causal discovery algorithms to identify the causal precursors and remove those that arise from spurious correlations. This is an effective way to avoid overfitting problems and has resulted in robust statistical forecasts of polar vortex (PV) strength (*Kretschmer et al.*, 2017) and Indian summer monsoon rainfall (*Di Capua & Coumou*, 2016).

References

Bahl, M., R. Barzilay,A. B. Yedidia, N. J. Locascio, L. Yu, and C. D. Lehman (2017), High-risk breast lesions: A machine learning model to predict pathologic upgrade and reduce unnecessary surgical excision, *Radiology*, *286*(3).doi:10.1148/radiol.2017170549.

Banerjee, S., A. E. Gelfand, A. O. Finley, and H. Sang (2008), Gaussian predictive process models for large spatial data sets, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *70*,825–848.

Bobadilla, J., F. Ortega, A. Hernando, and A. Gutiérrez (2013), Recommender systems survey, *Knowledge-based Systems*, *46*,109–132.

Breiman, L., (2001), Random forests, *Machine Learning*, *45*(1),5–32.

Chen, N., Z. Qian, I. T. Nabney, and X. Meng (2014), Wind power forecasts using Gaussian processes and numerical weather prediction, In *IEEE Transactions on Power Systems*, *29*(2).

Chen, T., and C. Guestrin (2016), Xgboost: A scalable tree boosting system, In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (pp.785–794).ACM.

Chipman, H. A., E. I. George and R. E. McCulloch (2010), BART: Bayesian Additive Regression Trees, *The Annals of Applied Statistics*, *4.1*,266–298.

Deng, L., and D. Yu (2014), Deep learning: methods and applications, *Foundations and Trends in Signal Processing*, *7*(3–4),197–387.

Di Capua, G., and D. Coumou (2016), Changes in meandering of the Northern Hemisphere circulation, *Environ. Res. Lett.*, *11*, 094028, doi:10.1088/1748-9326/11/9/094028.

Freund, Y., and R. E. Schapire (1997), A decision-theoretic generalization of on-line learning and an application to boosting, *Journal of Computer and System Sciences*, *55*(1), 119-–139.

Hastie, T., R. Tibshirani and J. Friedman (2001). The elements of statistical learning, Vol. 1, (pp. 337–387). New York: Springer series in statistics.

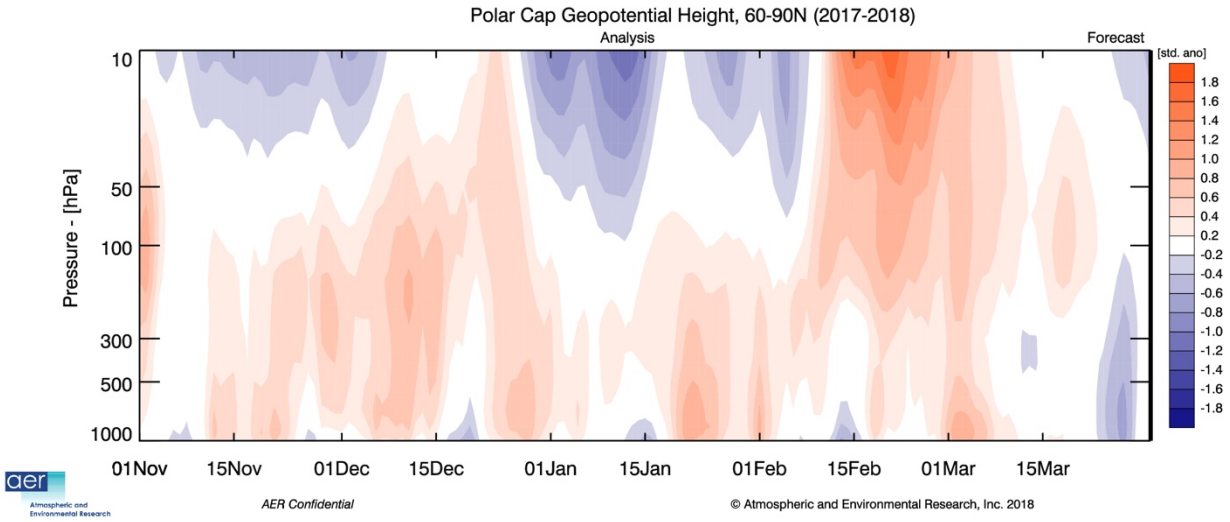
James, G., D. Witten, T. Hastie and R. Tibshirani (2013), An introduction to statistical learning, Vol. 112. New York: Springer.

Kretschmer, M., D. Coumou, J. Donges, and J. Runge, (2016),Using causal effect networks to analyze different Arctic drivers of midlatitude winter circulation, *Journal of Climate*, *29*,4069-–81, doi:10.1175/JCLI-D-15-0654.1.

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

Küffner, R., N. Zach, R. Norel, J. Hawe, D. Schoenfeld, L. Wang,..., and M. Cudkowicz (2015), Crowdsourced analysis of clinical trial data to predict amyotrophic lateral sclerosis progression, *Nature Biotechnology*, *33*(1),51.

Rasmussen, C. E., and C. K. Williams (2006), Gaussian Processes for Machine Learning, The MIT Press, Cambridge, MA.



Supplementary Figure 1. Daily polar cap (area averaged 60-90°N) geopotential height anomalies (PCHs) from November 1, 2017 – March 31, 2018 from the surface through the mid-stratosphere. Stratospheric polar vortex disruption in early February (characterized by transition of PCHs from negative to positive) descends into the troposphere in mid-February.