



# An Earth-System-Oriented View of the S2S Predictability of North American Weather Regimes

Jhayron S. Pérez-Carrasquilla<sup>a</sup> and Maria J. Molina<sup>a</sup>

<sup>a</sup> *Department of Atmospheric and Oceanic Science, University of Maryland, College Park,  
Maryland, USA*

*Corresponding author:* Jhayron S. Pérez-Carrasquilla, [jhayron@umd.edu](mailto:jhayron@umd.edu)

**Early Online Release:** This preliminary version has been accepted for publication in *Artificial Intelligence for the Earth Systems*, may be fully cited, and has been assigned DOI 10.1175/AIES-D-24-0075.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

© 2025 American Meteorological Society. This is an Author Accepted Manuscript distributed under the terms of the default AMS reuse license. For information regarding reuse and general copyright information, consult the AMS Copyright Policy ([www.ametsoc.org/PUBSReuseLicenses](http://www.ametsoc.org/PUBSReuseLicenses)).

**ABSTRACT:** It is widely agreed that subseasonal-to-seasonal (S2S) predictability arises from the atmospheric initial state during early lead times and from the land and ocean during long lead times. We test this hypothesis for the large-scale mid-latitude atmosphere by training numerous XGBoost models to predict weather regimes (WRs) over North America at 1-to-8-week lead times. Each model uses a different predictor from one Earth system component (atmosphere, ocean, or land) sourced from reanalysis. According to the models, the atmosphere provides more predictability during the first two forecast weeks, and the three components performed similarly afterward. However, the skill and sources of predictability are highly dependent on the season and target WR. Our results show greater WR predictability in fall and winter, particularly for the Pacific Trough and Pacific Ridge regimes, driven primarily by the ocean (e.g., El Niño-Southern Oscillation and sea ice). For the Pacific Ridge in winter, the stratosphere also contributes significantly to predictability across most S2S lead times. Additionally, the initial large-scale tropospheric structure (encompassing the tropics and extra-tropics, e.g., Madden-Julian Oscillation) and soil conditions play a relevant role—most notably for the Greenland High regime in winter. This study highlights previously identified sources of predictability for the large-scale atmosphere and gives insight into new sources for future study. Given how closely linked WRs are to surface precipitation and temperature anomalies, storm tracks, and extreme events, the study results contribute to improving S2S prediction of surface weather.

**SIGNIFICANCE STATEMENT:** Recurrent and persistent large-scale atmospheric patterns are referred to as weather regimes. Algorithms can be trained to predict weather regimes over North America up to eight weeks into the future. Contributions to prediction skill can originate from different parts of the Earth system, such as stratospheric wind anomalies, sub-surface ocean heat content, and soil water content. In this study, we found that contributions to the predictability of weather regimes can vary in importance depending on the lead time, season, and predicted weather regime. These results show that predictability contributions can vary more than previously documented and deepen our understanding of sources of predictability, which can help improve forecasts extending beyond two weeks for the benefit of society and decision-makers.

## 1. Introduction

The subseasonal-to-seasonal timescale (S2S; two weeks to two months) is often called a “predictability desert” (Vitart et al. 2012). Despite research advances (Becker et al. 2022; Sengupta et al. 2022), S2S forecasting remains challenging due to the chaotic nature of the atmosphere, imprecise knowledge of initial conditions, and multiple interactions and feedbacks among Earth system components (Lorenz 1963; White et al. 2017; Pegion et al. 2019; Merryfield et al. 2020). For example, atmospheric features evolve more quickly than the land and ocean, but they partly determine future land and ocean conditions. The resulting oceanic and land states also affect the subsequent atmospheric state via vertical fluxes of heat and moisture (Teng et al. 2019). Recently, machine learning (ML) has shown potential for S2S forecasting given its ability to capture nonlinear and complex relationships (e.g., Molina et al. 2023b; Chen et al. 2024). Here, we leverage ML to explore Earth system component contributions to S2S predictability (e.g., Mayer and Barnes 2021; Molina et al. 2023a).

Aiming to make the forecasting problem more general while maintaining its relevance, we focus on the predictability of the large-scale atmosphere instead of focusing on specific surface weather anomalies. K-means clustering (i.e., an unsupervised ML method) has been used to identify large-scale weather or circulation regimes (WRs), which are planetary-scale flow configurations that are recurrent, persistent, and quasi-stationary (Michelangeli et al. 1995; Reinhold and Pierrehumbert 1982; Straus et al. 2007; Molina et al. 2023a; Lee et al. 2023; Lee and Messori 2024). These regimes persist longer and have a slower evolution than small-scale atmospheric events, some of

which may be stochastic (e.g., thunderstorms), which makes them useful for S2S atmospheric prediction (Jennrich et al. 2024). WRs have an imprint on surface precipitation and temperature anomalies by creating persistent dynamic or thermodynamic conditions that lead to subsidence or convection, and by guiding the tracks of extratropical cyclones (e.g., Vigaud et al. 2018). WRs also modulate the probability of extreme events (Lee et al. 2023; Jennrich et al. 2024). For the boreal winter, Jennrich et al. (2024) demonstrated that, in some cases, precipitation and temperature forecasts could improve by using WR forecasts and their respective composited surface anomalies instead of the raw output fields from the models, and Molina et al. (2023a) found that upstream outgoing longwave radiation (OLR) and sea surface temperatures (SSTs) help skillfully predict WRs several weeks in advance. Finally, Lee et al. (2023) redefined North American regimes, allowing them to be used for the whole year instead of just for the cold season.

Beyond helping identify persistent and recurrent patterns, ML can be used to forecast WRs. Many recent developments, mainly using neural networks, have demonstrated that the performance of ML is comparable to state-of-the-art numerical models at S2S lead times (Weyn et al. 2021, 2019; Bi et al. 2023; Lang et al. 2024; Chen et al. 2024). However, neural networks can have a large number of hyperparameters, be prone to overfitting, and take substantial computing to train (e.g., Schmitt 2022). On the other hand, tree-based algorithms can still decipher nonlinear patterns among training data (Herman and Schumacher 2018), and have the capacity to achieve similar performance as neural networks at S2S lead times at comparably lower computational costs during training (e.g., Vitart et al. 2022). Tree-based models include Extreme Gradient Boosting (XGBoost; Friedman 2001), which has been used frequently during recent years to forecast a wide range of phenomena (Molina et al. 2023b), such as lightning (Mostajabi et al. 2019), daily precipitation (Dong et al. 2023), and seasonal temperature (Qian et al. 2020). XGBoost presents several advantages over other tree-based methods because it can better handle imbalanced datasets, has less likelihood of overfitting, and exhibits better detection of non-linear patterns (Fatima et al. 2023). These advantages are mainly due to a more flexible and adaptable architecture and the dynamic adjustment of hyperparameters during training.

Given the intertwined temporal evolution of the atmosphere, land, and ocean, all three components are relevant for predictability at the S2S timescale (Alexander 1992; Koster et al. 2010, 2011; Guo et al. 2012; Hartmann 2015). It has been largely understood that predictability arises

from the atmospheric initial state at lead times of less than two weeks, in contrast to the land and ocean which provide predictability during longer lead times (Meehl et al. 2021). However, Richter et al. (2024) demonstrated that contributions to predictability from Earth system components may be more complicated, with predictability from the atmosphere playing a greater role during S2S lead times. The initial conditions of the atmosphere, land, and ocean sometimes provide “forecasts of opportunity” (Mariotti et al. 2020), some of which can manifest during certain states of the Madden-Julian Oscillation (MJO; Robertson and Vitart 2018), sudden stratospheric warmings (Kidston et al. 2015), and the El Niño-Southern Oscillation (ENSO; DelSole et al. 2017). Considering the theoretical and technical advances mentioned above, this study aims to quantify relative initial state contributions to WR predictability from various Earth system components at S2S lead times using the XGBoost algorithm (Chen and Guestrin 2016).

## 2. Data and Methods

### *a. Weather Regimes Computation and Output of the Models*

Hourly geopotential height at 500-hPa (Z500) on an approximately 31-km grid was sourced from the ECMWF fifth generation reanalysis (ERA5; Hersbach et al. 2020) to compute WR classes. Our study was constrained to 1981-2020 due to the limited availability of ocean data below the sea surface from a different product. Following the Lee et al. (2023) framework, we computed year-round daily WRs. To do this, we first computed the daily average for each grid cell over the region shown in Figure 1 (20°N to 80°N and 180°W to 30°W). Then, we calculated daily anomalies by subtracting the respective grid cell’s multi-year daily average from 1981 to 2020. We applied a 10-day low-pass Fourier filter (across the time dimension) on the anomalies to emphasize synoptic-scale variability. Subsequently, we detrended the data by subtracting the daily area-averaged (cosine-latitude weighted) linear trend of the Z500 field over the domain (5.73 m/decade), and we divided the data by the multi-year daily standard deviation averaged over the region to obtain standardized anomalies. Both the annual cycle of the mean and the standard deviation were smoothed using a 60-day window. Although considering the full period for computing the climatology is not the standard procedure when training ML models, we emphasize that our objective is to gain insight into the sources of predictability and not to develop a realistic prediction system. Thus, this represents a minimal study limitation.

After obtaining the detrended standardized anomalies at a daily temporal resolution, the methodology described by Molina et al. (2023a) and Lee et al. (2023) was replicated. First, we extracted the 12 leading principal components (PCs) from the Z500 fields. Then, we identified the WRs by applying k-means clustering with four centroids over the time series of these 12 PCs (Michelangeli et al. 1995). Following Lee et al. (2023), we created a fifth class called “No WR,” which corresponds to samples that are closer to the origin than to any of the four WR centroids; that is, the PC scores are closer to the (0, 0, 0, ... 0) point on the PC space. Additionally, aiming to achieve higher quality in the clustering, we included in the “No-WR class” the 10% of samples with the highest distance to their corresponding centroids in the PC space.

The average Z500 detrended standardized anomalies associated with the four identified WR classes are shown in Figure 1. Consistent with nomenclature from past studies (e.g., Lee et al. 2019; Robertson et al. 2020; Molina et al. 2023a; Lee et al. 2023), the identified large-scale WRs were named Pacific Trough (PT), Pacific Ridge (PR), Greenland High (GH), and Alaskan Ridge (AR). As shown in previous studies (Molina et al. 2023a; Lee et al. 2023; Jennrich et al. 2024) and in Figure 1e-l, each WR is associated with specific surface temperature and precipitation anomalies. The Pacific Trough regime is associated with increased precipitation over the U.S. Pacific Northwest, dry conditions over the Midwest, and increased temperatures over the central U.S. and Canada. The Pacific Ridge regime is linked to increased precipitation over Alaska and the Midwest, dry conditions over the U.S. Pacific Northwest, colder-than-average temperatures over the U.S. West Coast, and vice versa for the U.S. East Coast. The Greenland High regime is associated with dry conditions over the U.S. Pacific Northwest and eastern Canada, with cold temperature anomalies over the U.S. East Coast and warm temperature anomalies over western Canada. Lastly, the Alaskan Ridge regime presents wet anomalies over British Columbia, dry conditions over Alaska and the U.S. Southeast, cold anomalies over western Canada, and warm anomalies over the western U.S. Therefore, accurate prediction of the WRs has direct implications for these specific regions. The objective of the models described below is to predict the WR present during the majority of the days of the target week. The weeks with no predominant weather regime were labeled as “No WR” weeks.

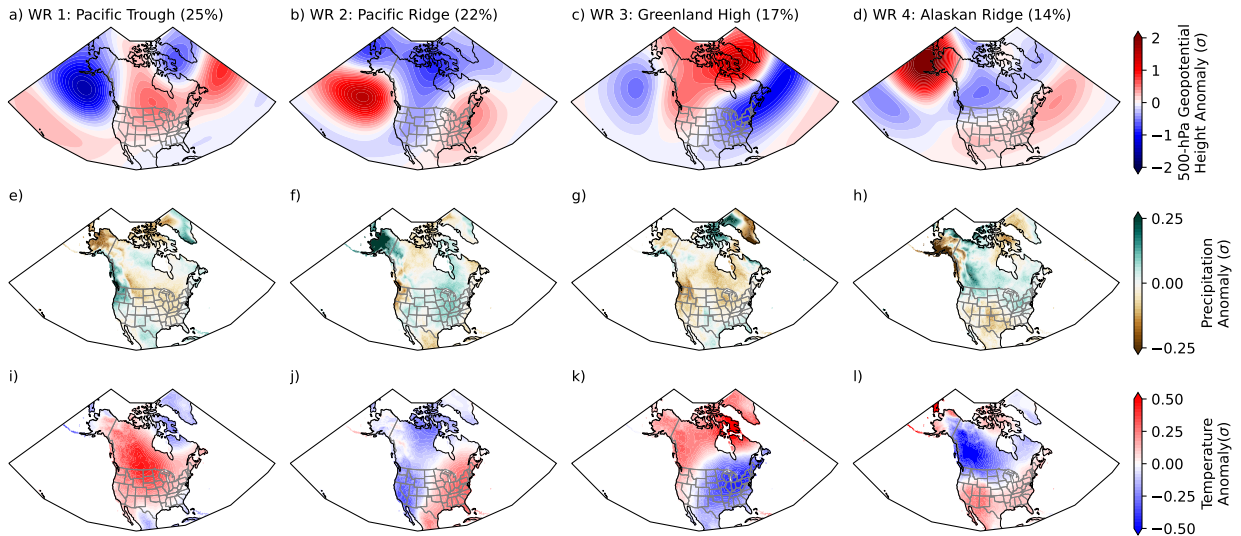


FIG. 1. The upper column (a-d) shows composites of detrended standardized ERA5 500-hPa geopotential height anomalies from 1981 to 2020 for the four weather regimes (WRs) identified. The daily frequency of each weather regime is shown in the subplot titles. The “No WR” class is not shown. The middle and bottom columns show the composited standardized daily precipitation (e-h) and temperature (i-l) anomalies for each one of the four weather regimes sourced from ERA5.

### *b. Sources of Predictability and Input Data for the Models*

Disregarding external forcings, the primary sources of S2S predictability can be classified into two categories: i) the state of recurring or quasi-oscillatory patterns, and ii) anomalies in the initial state of one of the components of the Earth system whose typical timescale of evolution (i.e., persistence time) is similar to the target forecast (National Academies of Sciences et al. 2016). In this study, we start from the (imperfect) assumption that most of the information corresponding to those two categories is contained within a limited set of variables from three Earth system components (ocean, land, and atmosphere). Additionally, we assume that the XGBoost models, in conjunction with a sufficiently large sample size, are skillful enough to at least partially represent the physical laws that determine the spatiotemporal behavior of the processes affecting the future state of the atmosphere within the S2S temporal horizon.

Each model within this study is trained separately using one variable from an Earth system component. For the atmosphere and land, we used data from ERA5 and ERA5-Land (Muñoz-Sabater et al. 2021) reanalyses, respectively. For the ocean, we used data from the Simple Ocean

Data Assimilation reanalysis (SODA; Carton et al. 2018). Taking into account various known sources of predictability originating from tropical and extratropical regions (e.g., Stan et al. 2017; Tseng et al. 2018; Lin et al. 2019), the spatial domain of all input variables spans all longitudes (360° around the prime meridian) and latitudes from 30°S to 90°N. For the input variables, all fields were re-gridded onto a 2° horizontal grid using nearest neighbor interpolation. Finer spatial resolutions were tested without a notable increase in the skill of the models. In the following subsections, we describe the potential sources of predictability we aim to represent and the corresponding input variables.

All input variables were transformed into weekly detrended standardized anomalies following the steps described in Subsection a and averaged each week. However, instead of removing the area average trend and standard deviation, each pixel was treated separately, and no Fourier filtering was performed. After obtaining the weekly detrended standardized anomalies by subtracting the daily climatology and dividing by the standard deviation of each grid cell, we smoothed the fields spatially using a 3x3 moving kernel. Apart from the models trained separately to assess predictability from each individual variable, three other models were trained using atmosphere-only, land-only, and ocean-only variables as inputs. In addition to the fields of standardized anomalies, we added the sine and cosine of the day of the year as another input to the models to improve the representation of seasonal differences in relationships among Earth system components, such as teleconnections or surface-atmosphere interactions.

## 1) ATMOSPHERIC VARIABLES

Four atmospheric variables were used in our study, with three from the troposphere and one from the stratosphere (Table 1). The variables from the troposphere are Z500, outgoing long-wave radiation (OLR), and 200-hPa zonal wind (U200). Z500 describes the mid-latitude atmosphere, including the large-scale atmospheric circulation (Cheng and Wallace 1993), North-Atlantic oscillation (NAO; Wallace and Gutzler 1981; Davini et al. 2012), planetary- and synoptic-scale waves (Blackmon 1976), and blocking episodes (Tibaldi and Molteni 1990; Scherrer et al. 2006). OLR provides information about the strength and spatial coverage of convective clusters, which are related to tropical-extratropical teleconnections (Stan et al. 2017) from the MJO (Hendon and Salby 1994; Wheeler and Kiladis 1999; Wheeler and Hendon 2004), ENSO (Yulaeva and Wallace 1994),

TABLE 1. The four atmospheric variables of the Earth system used for training the ML models sourced from ERA5. The relevance for S2S forecasting is also indicated.

Variable	Relevance to S2S forecasting
500hPa geopotential height (Z500)	Mid-latitude atmosphere (e.g., weather regimes, North-Atlantic oscillation, planetary Rossby waves, and blocking episodes).
Outgoing longwave radiation (OLR)	Clusters of convection and large-scale cloud patterns.
10hPa zonal wind (U10)	Quasi-biennial oscillation and sudden stratospheric warming events.
200hPa zonal wind (U200)	Strength and structure of the walker cell, jet streams, and Indian Ocean dipole.

and Indian Ocean dipole (IOD; Saji et al. 1999). U200 provides information about the shape and location of the polar and subtropical jet streams (Krishnamurti 1961; Palmén 1948; Strong and Davis 2007), in addition to the coupling of the Walker circulation to the phase of ocean-related phenomena, like ENSO or IOD (Philander 1983; Saji et al. 1999).

The stratosphere is a valuable source of S2S predictability given its slowly varying circulation anomalies (National Academies of Sciences et al. 2016). During boreal winter, the stratosphere can interact with the upper troposphere and modify the tropospheric circulation, along with the onset and evolution of WRs (Baldwin and Dunkerton 2001; Baldwin et al. 2003; Gerber et al. 2009; Domeisen et al. 2020). Two primary stratospheric features contributing to mid-latitude predictability have been previously identified, the quasi-biennial oscillation (QBO; Baldwin et al. 2001) and the stratospheric polar vortex (Gerber et al. 2009). The QBO is an easterly-to-westerly reversal of tropical stratospheric winds partly driven by waves originating from the troposphere that can affect the strength of the stratospheric polar vortex (Baldwin et al. 2001). Changes in the stratospheric polar vortex can result in large-scale tropospheric circulation anomalies with a lag of about one month (Thompson and Wallace 2000; Baldwin and Dunkerton 2001). Rapid slowdowns of the stratospheric polar vortex are frequently accompanied by sudden stratospheric warmings that can affect the Northern Annular Mode and contribute to an equatorward shift of the tropospheric midlatitude jet (Gerber et al. 2009; Domeisen et al. 2020). The zonal wind at 10-hPa (U10) is the fourth atmospheric variable we consider because it can help diagnose the QBO phase and the strength of the stratospheric polar vortex (Baldwin et al. 2001; Gerber et al. 2009).

## 2) OCEANIC VARIABLES

The ocean is relevant for S2S prediction because it provides information about natural modes of climate variability (e.g., ENSO; Philander 1983) and it can influence the overlying atmosphere (National Academies of Sciences et al. 2016). Perturbations of the ocean's mean state, particularly within the tropics, can induce organized large-scale convection, modifying the heating distribution through the atmosphere and producing temperature and precipitation anomalies. Organized convection can also change the extratropical circulation due to propagation via Rossby wave-trains (Bjerknes 1969; Horel and Wallace 1981; Trenberth et al. 1998). ENSO is one of the primary modulators of long-lived tropical signals, with positive or negative SST anomalies over the eastern-to-central equatorial Pacific characterizing its warm and cold phases, respectively (Philander 1983).

The atmosphere can also respond to extratropical ocean features (Zhou 2019), such as eddy-mediated processes over the Kuroshio boundary current, which can modulate geopotential heights above the ocean anomaly and generate near-surface atmospheric baroclinicity, modifying the spatial characteristics of overlying storm tracks (Kwon et al. 2010; Kushnir et al. 2002). Although the atmospheric variance associated with extratropical SST anomalies is modest compared to internal atmospheric variability, extratropical SST anomalies can persist for months to years (Chelton and Xie 2010) and the atmosphere can respond within a few months (Ferreira and Frankignoul 2005; Deser et al. 2007).

Ocean data was sourced from SODA reanalysis (Table 2). To capture information about sub-surface conditions that may reach the surface via mixing (Alexander 1992), we used ocean heat content (OHC) integrated from the surface to different depths (50m, 100m, 200m, 300m, and 700m). Sea surface height (SSH) and mixed layer depth (MLD) were also used, which provide information about the upper-ocean circulation, anomalous upwelling or downwelling, and the size, depth, and location of eddies (Chelton et al. 2011; Kurczyn et al. 2012; Uchoa et al. 2023).

Sea ice thickness and concentration anomalies were also used, which are generally slow-evolving and describe sea ice cover (Table 2). Sea ice can significantly reduce heat and moisture fluxes from the ocean to the atmosphere, modifying the surface albedo and increasing how much incoming radiation is reflected (National Academies of Sciences et al. 2016). The presence and persistence of sea ice can also affect the paths of atmospheric lows (Balmaseda et al. 2010; Screen et al. 2011).

TABLE 2. The 10 oceanic variables from the Earth system used for training the ML models sourced from SODA. The relevance for S2S forecasting is also indicated.

Variable	Relevance to S2S forecasting
Sea surface temperature (SST)	ENSO, storage of anomalous surface heat, and eddies.
Ocean heat content (OHC) integrated from the surface to a 50m, 100m, 200m, 300m, or 700m depth.	Magnitude and persistence of anomalous heat.
Mixed-layer depth (MLD)	Anomalous upwelling or downwelling, and upper-ocean circulation structure.
Sea-surface height (SSH)	Anomalous upwelling or downwelling, and upper-ocean circulation structure.
Sea ice thickness (IT)	Anomalous changes of heat and moisture fluxes from the ocean to the atmosphere, and changes in albedo.
Sea ice concentration (IC)	Anomalous changes of heat and moisture fluxes from the ocean to the atmosphere, and changes in albedo.

### 3) LAND VARIABLES

Soil water content (SWC) evolves slowly and can influence surface energy budgets, contributing to anomalies in temperature and precipitation, and is therefore another source of S2S predictability (Koster et al. 2004, 2010; Guo et al. 2011; Koster and Walker 2015; Roundy and Wood 2015; Thomas et al. 2016). ERA5-Land, based on the Tiled ECMWF Scheme for Surface Exchanges over Land (TESSEL; Balsamo et al. 2009) and forced with ERA5 atmospheric fields, provides soil temperature (ST) and SWC for four layers down to a depth of 2.89m. SWC and ST integrated from the surface to the depth of each one of the four reanalysis layers (7cm, 28cm, 1m, and 2.89m) are used as input variables (Table 3). Terrestrial snow is another source of S2S predictability given its radiative effects and liquid water that is released during melting (Peings et al. 2011; Jeong et al. 2013; Orsolini et al. 2013; Thomas et al. 2016). Thus, we used snow depth (SD) as another input variable.

TABLE 3. The nine land variables of the Earth system used for training ML models sourced from ERA5-Land. The relevance for S2S forecasting is also indicated.

Variable	Relevance to S2S forecasting
Soil water content (SWC) integrated from the surface to a 7cm, 28cm, 1m, or 2.89m depth.	Storage of surface water, and proxy for soil and vegetation characteristics.
Soil temperature (STL) averaged from the surface to a 7cm, 28cm, 1m, or 2.89m depth.	Proxy for potential evaporation and proxy for soil and vegetation characteristics.
Snow depth (SD)	Storage of surface water and influence on surface energy budgets.

### *c. Forecasting framework*

Numerous XGBoost models were independently trained to predict weekly WR classes over North America. We trained each model with a different input variable to assess the contribution of the respective predictor at different lead times, ranging from 1 to 8 weeks into the future. Weekly averaged data was used with weeks delineated on Mondays and Thursdays; some overlap in the days of the week was permitted for the weekly averages in exchange for a doubling of sample size. For any weekly forecast that started on Monday ( $t = 0$ ), the input “week 0” contained days from the previous Tuesday and finished on that respective Monday ( $t = -6, \dots, 0$ ), while the first lead time of the forecast (i.e., week 1) included days from the next Tuesday to the following Monday ( $t = 1, \dots, 7$ ), and so on for weeks 2 through 8. Likewise, for weekly forecasts starting on a Thursday ( $t = 0$ ), the weekly input contained days from the previous Friday ending that Thursday.

### *d. XGBoost hyperparameters*

We performed a Bayesian optimization (Snoek et al. 2012) of hyperparameters for each ML model to obtain the combinations that produced the best possible performance within the limits of our study. We emphasize that our objective is not to obtain the best possible S2S forecasts but to assess predictability contributions from different variables and Earth system components. Given the potential presence of certain states of modes of climate variability, a model may perform better when evaluated in one particular period than another. To prevent overfitting to such patterns and unfair evaluation, we used k-fold cross-validation to find the best combination of hyperparameters and cross-testing to obtain performance metrics representative of the whole study period. The “cross-validation and cross-testing” approach has been shown to remove biases when assessing performance while also taking advantage of all samples within the dataset (Korjus et al. 2016). The hyperparameter search is performed independently for each variable and lead time.

We employed cross-validation and cross-testing in the following manner for each one of the variables. The 40-year dataset was divided into four decades (1981-1990, 1991-2000, 2001-2010, and 2011-2020). For cross-testing, we denoted a test subset as one of the four decades. A model predicting a test subset was then trained using the remaining three decades of data. Before training on the remaining three decades, we applied three-fold cross-validation, where each one of the decades served as one “fold.” Three-fold cross-validation allowed us to obtain the combination of

hyperparameters that provided good skill across each held out fold, which involved trying multiple combinations of hyperparameters and training iteratively on two of the three decades. Finally, the combination of hyperparameters yielding the best average performance over the three decades was chosen to predict the respective test subset, and this process was repeated three more times to obtain performance statistics on the four test subsets.

Ten hyperparameters were assessed during optimization (Table 4), including those related to the complexity of the trees (e.g., maximum tree depth) and to prevent overfitting (e.g., L1 and L2 regularization terms). Additionally, we assessed the impact of adding class weights to control for WR class imbalance. The Bayesian algorithm (Snoek et al. 2012) used to perform the hyperparameter search approximates a probabilistic function to the hyperparameter field to make iterative changes in a direction that improves the objective function. We used the F1 score as the objective function, which is similar to “threat score” and “critical success index,” and focuses on the fraction of observed and forecast events that were correctly predicted. The advantage of the F1 score over other objectives lies in that it neglects “true negative” cases that can artificially inflate skill when a class imbalance exists. The F1 score is computed as

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (1)$$

where

$$\text{Precision} = \frac{\sum_c \text{True Positives}_c}{\sum_c (\text{True Positives}_c + \text{False Positives}_c)}, \quad (2)$$

$$\text{Recall} = \frac{\sum_c \text{True Positives}_c}{\sum_c (\text{True Positives}_c + \text{False Negatives}_c)}, \quad (3)$$

and  $c$  represents each of the 5 classes.

60 initial trials with random hyperparameter combinations were run to provide an initial performance baseline and then 15 subsequent iterations were performed using Bayesian optimization to find the best combination. The Bayesian algorithm was executed using the Expected Improvement method (Frazier 2018) with  $\xi = 0$  for prioritizing exploitation after the random search (larger  $\xi$  values produced no improvement in performance). The number of estimators (trees or boosting rounds) in the XGBoost algorithm was 20 for each model. Trials with a higher number of trees were

done without notable improvement in the models' skill. Table 4 shows the optimal hyperparameters identified during the search. The loss function used by the XGBoost algorithm is categorical cross-entropy.

TABLE 4. Set of XGBoost hyperparameters explored during the Bayesian optimization.

Hyperparameter	Search range	Optimal values
Maximum depth of a tree	2-20	12-14
Minimum sum of instance weight needed in a child	1-20	10-12
Percentage of samples used for each tree construction	0.7-0.9	0.8-1
Percentage of features used for each tree construction	0.6-0.9	0-0.25
Percentage of features used for each split	0.75-1	0.83-0.88
Learning rate	0.0001-0.3	0.003-0.015
Minimum loss reduction required to make a further partition on a leaf node of a tree (gamma)	0-5	0-4
L1 regularization term on weights	4-40	6-10
L2 regularization term on weights	1-316	1.5-40
Exponential term for the weights to reduce class imbalance	0-1	0.6-0.8

### 3. Results

#### *a. Sources of predictability based on ML model skill*

Plots a, b, and c in Figure 2 show the F1 score as a function of lead time for each ML model trained using atmospheric, oceanic, or land predictors. To determine whether or not a certain variable (or set of variables) provides meaningful predictability, we compared the skill of its respective model with the skill of three benchmarks: WR persistence (red dash-dot line), random guess based on the annual WR climatology (orange dash-dot line), and random guess based on the seasonal WR climatology (purple dash-dot line). Figure 2 also shows the 5th-95th percentile ranges of the F1 score computed for each benchmark or model using a 1,000-member bootstrap with resampling. The gray shading corresponds to the benchmarks, and the gold (atmosphere), blue (ocean), and green (land) shading corresponds to different models from each component.

Meaningful predictability is obtained by a model if its F1 score is above the confidence interval of the benchmarks, with the null hypothesis being that the observed skill could be obtained by random chance with one of the three baselines. Additionally, in each one of the three plots, the dashed lines (gold for a, blue for b, and green for c) show the F1 score associated with the models trained using

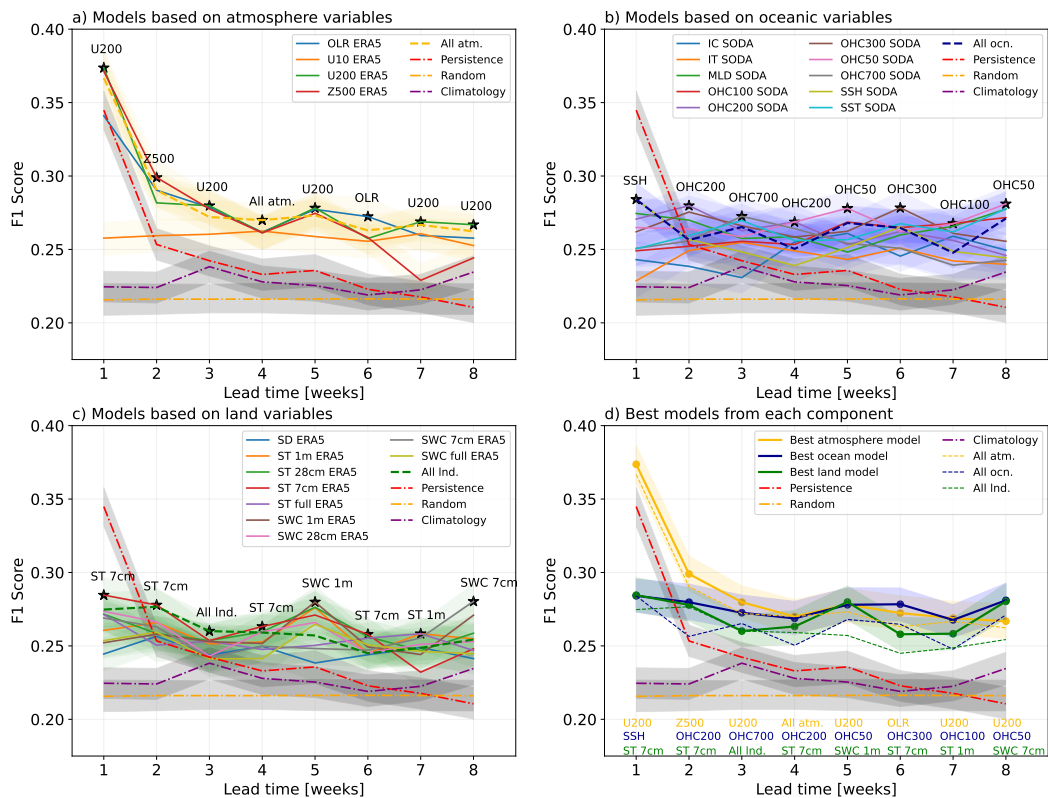


FIG. 2. Predictability from the a) atmosphere, b) land, and c) ocean indicated using F1 score as a function of lead time. The red dash-dot line shows the skill of the persistence forecast, the orange dash-dot line shows the skill of a random guess based on the annual WR climatology, and the purple dash-dot line shows the skill of a random-guess forecast based on the seasonal WR climatology. The yellow, blue, and green dashed lines show the performance of the models trained using all the predictors from the respective Earth system component, and the star markers indicate the predictor that provided, on average, more predictability at the respective lead time. d) F1 score of the best-performing model for each lead time and component, the dashed lines from a, b, and c, are also plotted. The shading in the four plots shows the 5th-95th percentile range of the F1 score from a 1,000-member bootstrap with resampling. The gray shading indicates the uncertainty associated with the benchmarks. Little or no overlap of the model and benchmark uncertainty ranges indicates meaningful predictive skill. The colored text in d) indicates the model with the best F1 score for each component at each lead time.

all the variables from each component. Figure 2a-c also contains annotations indicating the model that presented the best skill for each lead time and each component. As observed in Figure 2, there is high uncertainty regarding which model is the best for each component at every lead time (i.e., there is substantial overlap of model spread). Figure 3 shows the variables whose models displayed

predictability and highlights the models that performed best during each lead time. It's worth noting that persistence presents skill above the other two benchmarks during weeks 1 and 2 of the forecast, remaining as a significant source of atmosphere-based predictability, which highlights the usefulness of the WR framework for S2S forecasting.

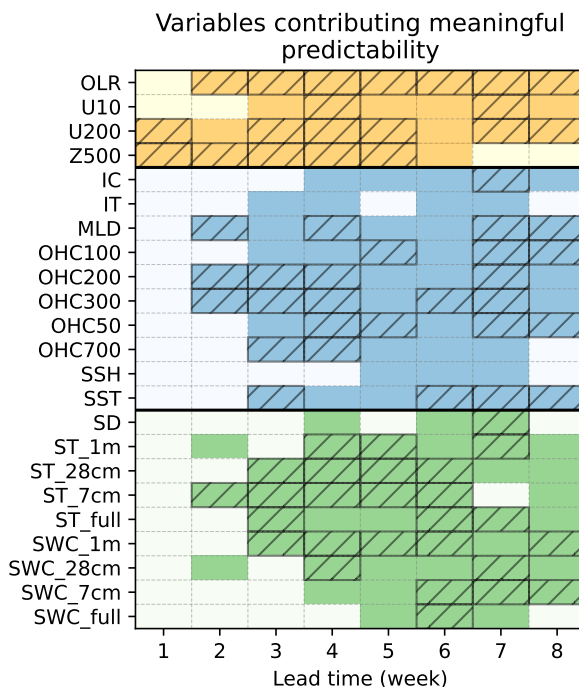


FIG. 3. Variables that contribute meaningful predictability from the atmosphere (yellow), ocean (blue), and land (green) during the different lead times. Darker shading than the background indicates that the F1 score of the model based on the respective variable was higher than the three benchmarks (and outside their confidence interval). The hatching indicates that the F1 score of the model was within the confidence interval of the best model from its respective component, indicating relatively high predictability contributions.

For the atmosphere-only models (Figure 2a), predictability from most of the predictors and the joint model is higher during earlier lead times (weeks 1-4) and then plateaus. The exceptions for this are the U10, which shows nearly constant skill across lead times, and Z500, which decreases continuously and becomes null after week 6. Most of the atmosphere-based models show increased predictability relative to the benchmarks in weeks 2-8 (see Figure 3). We also note that skill was not necessarily improved when using all predictors from the atmosphere, potentially due to competing

predictive signals or limitations associated with partially correlated predictors. This is also true for the other two Earth system components (Figure 2b-c).

The predictability coming from the ocean shows nearly constant skill across lead times, as expected due to its slow evolution, with improvements over the benchmarks after week 2. Models based on OHC perform best during most lead times, and SSH, IC, and IT present predictability during fewer weeks (see Figure 3). Although SST contributes predictability during most lead times, the performance was, on average, lower than the OHC-based models, highlighting the importance of considering subsurface processes to extract the best possible predictive skill from the ocean.

For land-based predictability, the skill of the models is approximately constant across lead times. Skill (although slightly lower on average than the atmosphere and ocean) can be extracted from the land after week 3, coming from several ST- and SWC-based models (see Figure 3). These results underscore the importance of continuing to improve the coverage of soil in situ observations (e.g., Hasan et al. 2020). Snow depth offers limited skill compared to the other predictors. However, it's worth noting that these results consider all seasons, and some Earth system processes exhibit high seasonality. Therefore, a variable or component may provide skill only during certain seasons and not appear important in the year-round analysis. Thus, we stratify results by season in the following subsection (Sect. 3b).

Figure 2d compares the best-performing ML models across Earth system components. Results indicate that the atmosphere provides comparatively greater predictability at 1-2 week lead times compared to the land and ocean. Our results in Figure 2d, although not directly comparable, are similar to those presented in Figure 1 by Richter et al. (2024), where during weeks 3-8, the predictability coming from the three components is of similar magnitude. In our case, only subtle differences are found in weeks 3 and 6, where the best atmosphere-based and ocean-based models, respectively, show higher skill than the best land-based models on average (with overlap among the confidence intervals). Although differences among models are subtle, the three Earth system components provide valuable predictability relative to the benchmarks during weeks 2-8. The following sections dig deeper into the origin of the skill from each component, considering differences depending on the season and the target WR.

*b. Seasonality of the sources of predictability*

Given the seasonality of the large-scale mid-latitude atmosphere and the Earth system in general, certain predictors may provide predictability only during some periods of the year. Some examples include seasonal variations in teleconnections, land heat and moisture fluxes, incident radiation and radiative fluxes, and sea ice thickness and extent. Similar to Figure 2d but stratified by season, Figure 4 shows notable differences in seasonal predictability from the three Earth system components. There is also high variability in the baselines themselves: persistence presents less skill during summer compared to the other seasons, illustrating that baselines can complicate ML model assessments if seasonality is not considered.

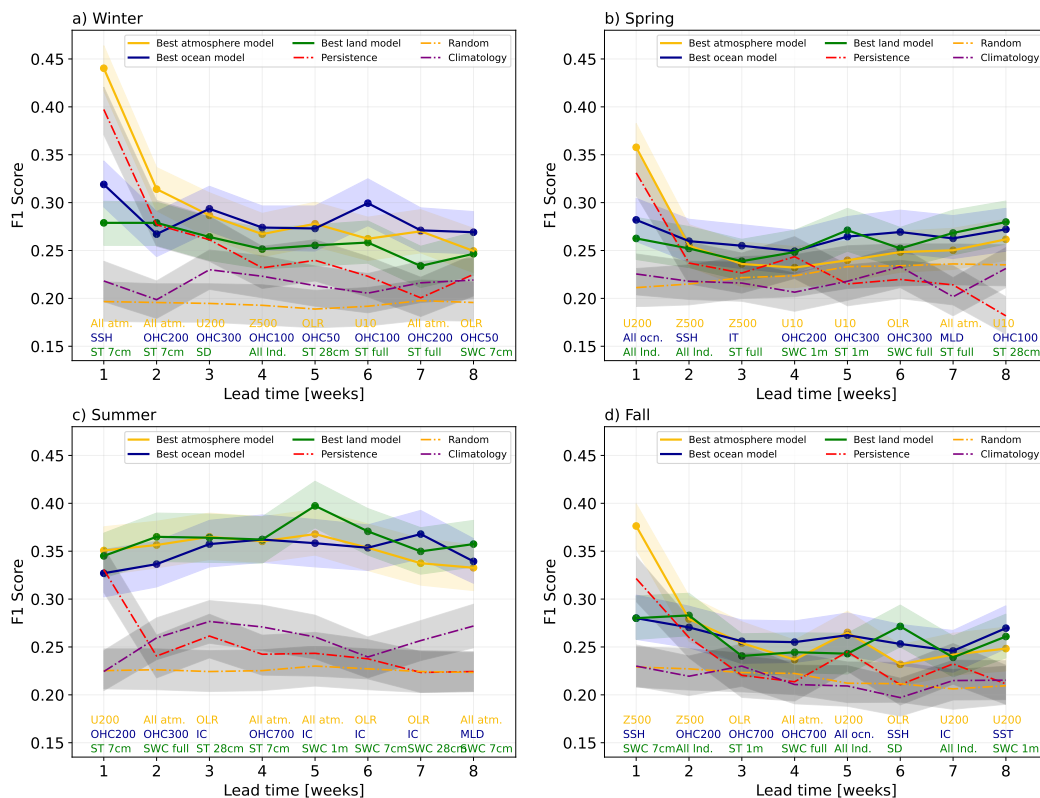


FIG. 4. Same as Figure 2d but for S2S forecasts stratified by the season initialization took place.

When comparing the four seasons in Figure 4, it is evident that increased skill from the three components occurs primarily during winter and summer and secondly during spring and fall. These results were somewhat expected because higher baroclinicity and Rossby wave activity during winter yield a more WR-dominated mid-latitude atmosphere, facilitating the understanding

of the large scale (Kuang et al. 2014). Additionally, teleconnection activity is higher during winter (Madden 1986; Philander 1999), providing additional mechanisms for skillful predictions. During summer, the weekly No-WR frequency is higher (it goes from 18% in the winter to 44% in the summer). The models do a better job than the baselines in distinguishing this pattern, as shown in Figures 5e and 6s. This seasonality is also reflected in the skill of persistence. During winter, persistence shows increased skill above the other two benchmarks during weeks 1-3, while in summer, this only happens for week 1.

During winter (Figure 4a), the atmosphere provides predictability during most lead times, and the ocean-based models show increased skill compared to the benchmarks during weeks 3-8. The atmospheric skill comes from various variables, with the all-atmosphere model performing best on average during weeks 1, 2, and 7. While U200 and Z500 provide meaningful predictability during early lead times, OLR and U10 show increased skill at later lead times (Figure A1a). For the ocean, OHC- and SST-based models offer the highest skill on average after week 3 (Figure A1a). During summer, the predictability coming from the three components is higher than the benchmarks from weeks 2 to 8, and there is a high overlap among the confidence intervals. During the shoulder seasons, the skill of the models relative to the winter and summer decreases notably. However, there are still cases in which predictability comes from different components (with moderate uncertainty). The ocean- and land-based models show increased skill during spring in weeks 5-8, with OHC and SWC/ST providing skill often from their respective components. The skill behavior of the three components is similar during fall. Although the skill obtained during fall and spring is lower, individual WRs have sources of predictability within the Earth system that can be further elucidated when stratifying by regime (Conil et al. 2009; Guo et al. 2012; Thomas et al. 2016; Meehl et al. 2021; Balmaseda et al. 2010; Petrie et al. 2015). These results are discussed in the next Section 3c.

### *c. Variations in predictability sources based on WR*

As shown by Figure 5, sources of predictability differ for the Pacific Trough, Pacific Ridge, and Greenland High regimes (Baldwin and Dunkerton 2001; Büeler et al. 2021; Molina et al. 2023a; Hochman et al. 2021; Kolstad et al. 2022). To further understand the different sources of predictability among regimes, we also explore the seasonal variations in Figure 6. Additionally,

Figure 7 shows the variables that contribute meaningful skill for each season and WR. We omitted the Alaskan Ridge regime from this subsection's discussion since prediction skill is not statistically significant or meaningfully different than the considered baselines (Figure 5d).

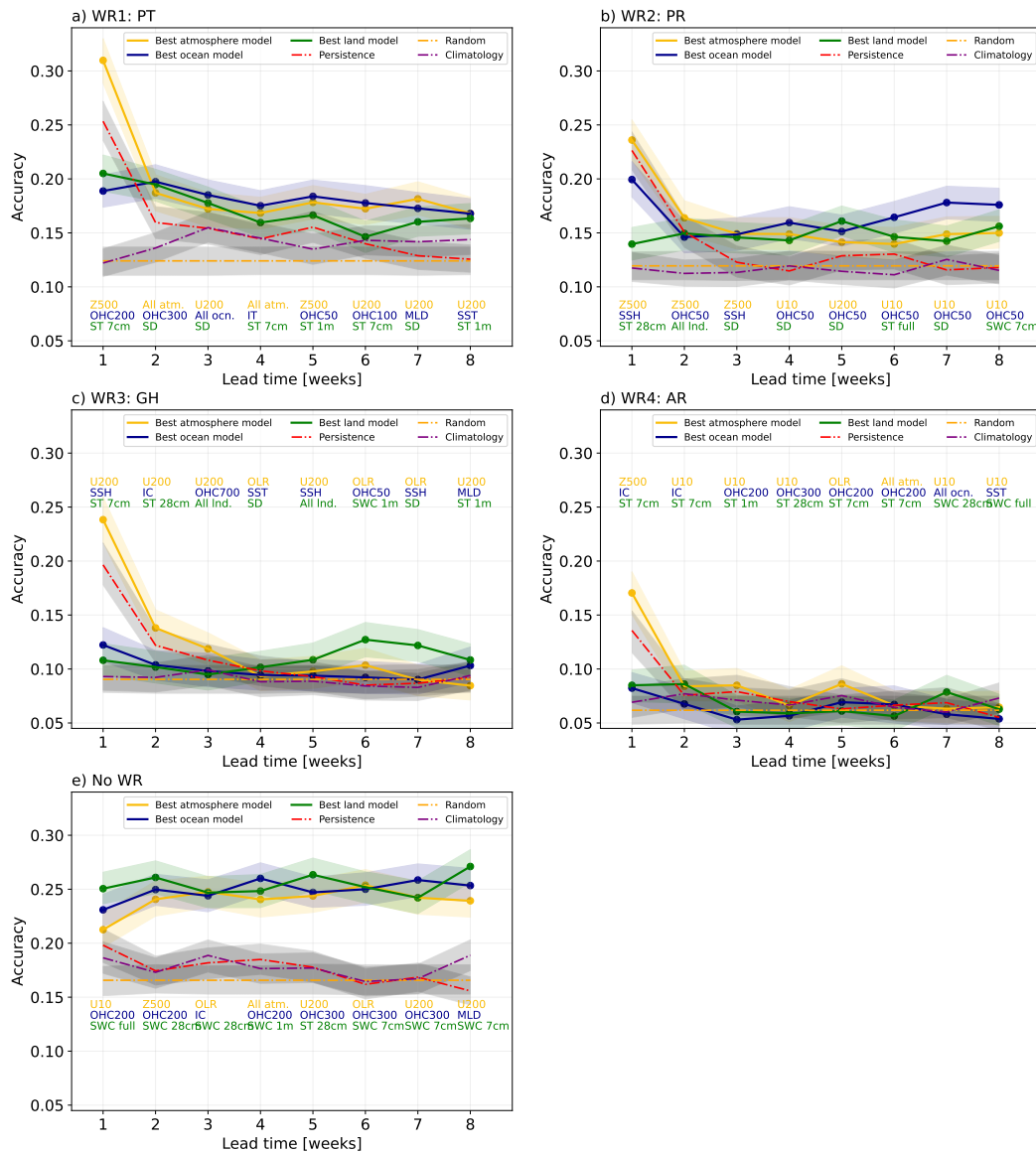


FIG. 5. Same as Figure 3 but for S2S forecasts stratified by target regime. Accuracy is used instead of the F1 score since there is only one regime (i.e., class) being evaluated. Evaluation is performed on a subset of the test set that contains all samples in which either the ground truth or the prediction (or both) are equal to the regime of interest. This accuracy is equivalent to the critical success index, also known as the threat score.

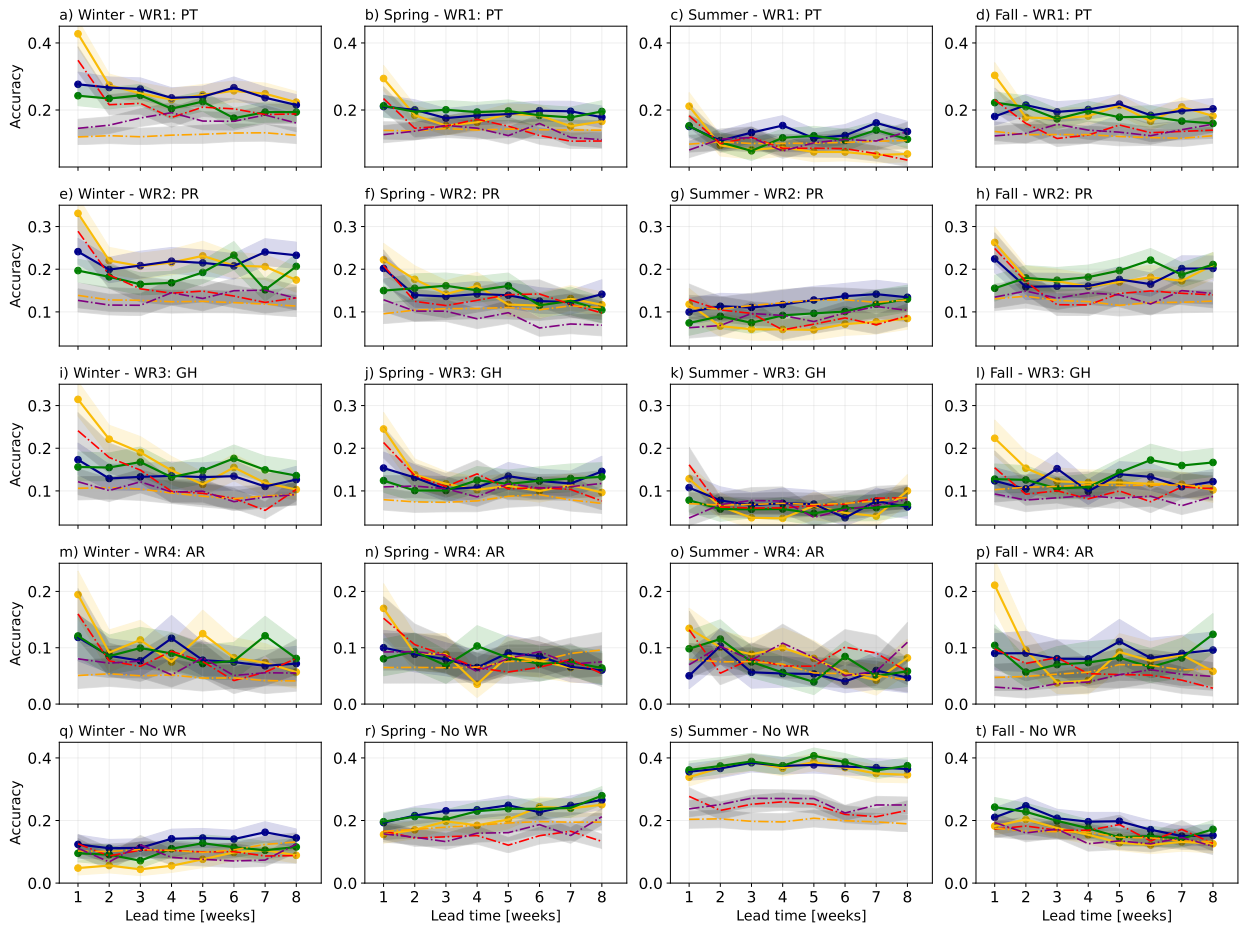


FIG. 6. Same as Figure 4 stratified by both WR and season. Note that y-axis ranges differ among WRs.

## 1) PACIFIC TROUGH

According to Figure 5, the ocean is often (on average) a more substantial contributor to skillful prediction of the Pacific Trough regime from weeks 2 through 8. Additionally, the atmosphere shows skill during most lead times, and the land contributes generally less predictability. According to Figure 6a-d, the increased skill is present predominantly during winter, spring, and fall. During fall (Figure 7d), the most notable oceanic predictability contributors are SST, SSH, and OHC integrated down to different depths. From the atmosphere and land, U200, Z500, and ST show predictability at several lead times. During winter, the oceanic predictability comes only from OHC, while there are atmospheric predictability contributions from Z500, U200, and OLR (Figure 7a). Ocean and land contributions also happen during spring without a clear main contributor (Figure 7b), although we note that ST and SD provide skill during some lead times.

Variables contributing meaningful predictability

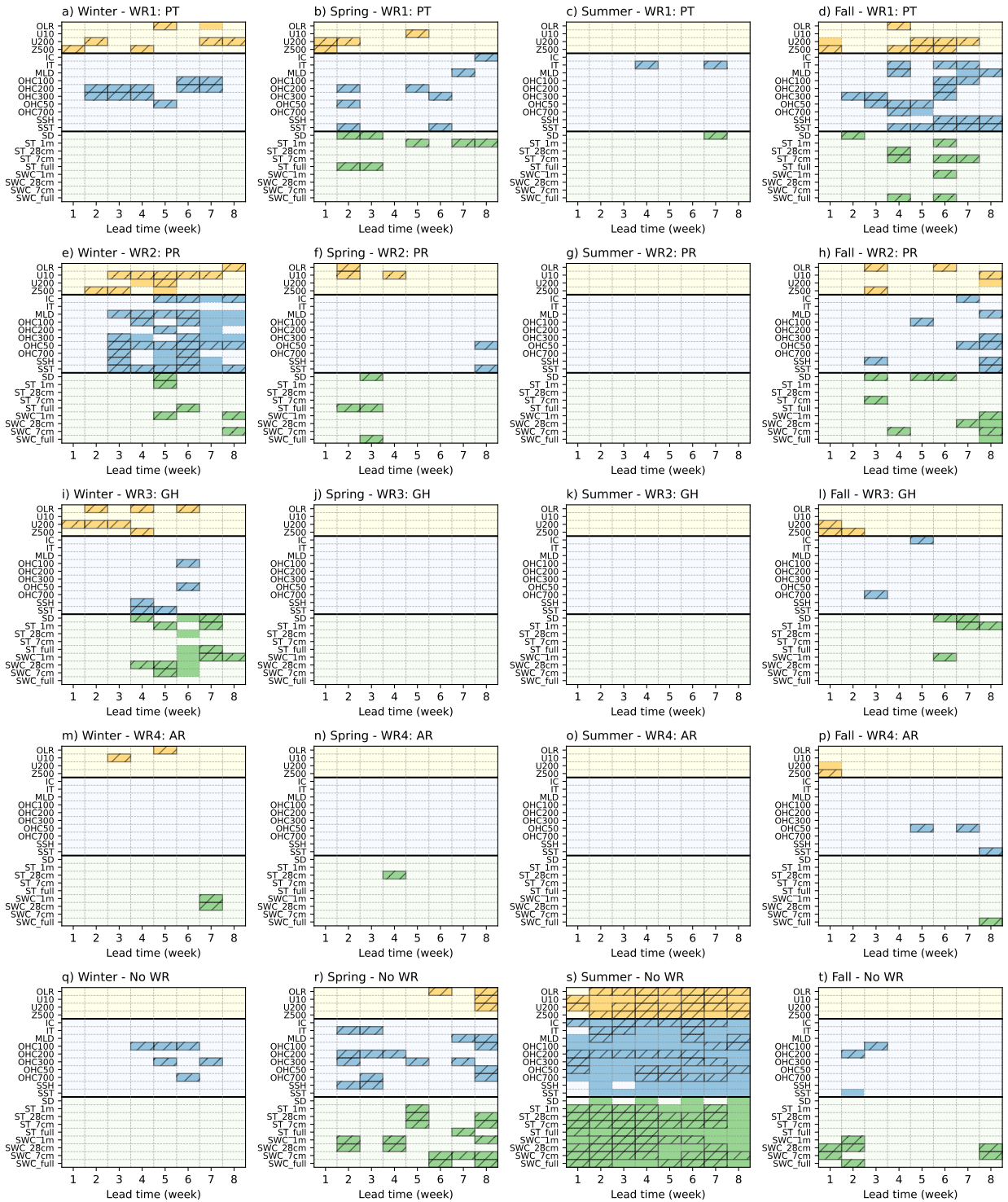


FIG. 7. Same as Figure 3 stratified by both WR and season.

## 2) PACIFIC RIDGE

According to Figure 5b, the Pacific Ridge regime shows ocean-based predictability during weeks 3-8, and significant but generally lower predictability during later weeks from the atmosphere and land. When considering different seasons (Figures 6e-h and 7e-h), there is generally higher atmosphere- and ocean-based predictability during winter. U10 shows relatively high skill during weeks 3-7, which is an indication of stratospheric processes notably affecting the winter occurrence of the Pacific Ridge (Figure 7e). From the ocean, there seem to be numerous relevant variables during winter, including IC (weeks 5-8), MLD (weeks 3-8), OHC (weeks 3-8), and SST (weeks 3-8). Predictability contributions are evident during the fall, with fewer during the spring. During fall (Figure 7e), the SD-based models provided skill at several lead times.

## 3) GREENLAND HIGH

According to Figure 5, the Greenland High regime shows increased atmosphere-based predictability during weeks 1-3 and land-based predictability during weeks 5-8. Figure 6i-l indicates that this predictability behavior occurs primarily in winter and fall. According to Figure 7i, the variables that most contribute to winter predictability are U200 (weeks 1-3) and OLR (weeks 2, 4, and 6). Later during the forecast horizon, the skill comes mostly from land: SD, ST, and SWC. The number of models that show meaningful skill during fall decreases notably (Figure 7l).

## 4) NO-WR CLASS

According to Figure 6q-t, the increased skill in predicting the No-WR class happens primarily in summer, with some variables providing skill in spring, particularly OHC and SWC. However, due to overlapping confidence intervals, it is challenging to identify specific components or predictors as the most relevant during summer (see Figure 7s). The increased skill results from the high No-WR frequency during the summer. The difference in the amount of samples compared to the other classes produces a model tendency to over-predict this class and, despite the false positives (cases in which the No WR class was predicted but did not happen), the accuracy (or critical success index) is higher than the benchmarks, which is not the case for the other seasons.

The results from this section allow for identifying where the skill shown in Figures 2d and 4 comes from. Where the trained models exploit predictability from becomes clearer when considering

each regime independently (Figure 7). Oceanic processes seem to be particularly relevant for the Pacific Trough regime during fall and winter. The occurrence of the Pacific Ridge during winter seems to be influenced by a mix of processes from the stratosphere and the ocean. The structure of the upper troposphere and OLR provide predictability during the first few weeks of the forecast for the Greenland High during winter, while the soil water content and temperature have a lagged contribution at lead times greater than 4 weeks. Finally, snow depth seems to contribute some skill to forecasting the Pacific Ridge during autumn and the Greenland High during fall and winter. In the next section, we further detail where the presented forecasting skill comes from, with an emphasis on the initial states of the Earth system that correspond with accurate forecasts (forecasts of opportunity).

#### *d. Physical processes contributing predictability*

Recent literature has highlighted the importance of taking advantage of particular initial states of the Earth system that offer increased predictability (namely, forecasts of opportunity; Mariotti et al. 2020; Robertson et al. 2020; Mayer and Barnes 2021). In this section, we focus on the initial states in cases when the trained models exhibited significant skill. We do this by compositing standardized anomaly maps of the initial states (week 0) for the variables and associated correct predictions highlighted in Figure 7. Additionally, to identify the regions most relevant to the correct forecasts, we analyze the spatial distribution of contributions to accurate final predictions based on SHapley Additive exPlanations (SHAP) values (Lundberg and Lee 2017; Flora et al. 2024). Figures 8-13 show the average spatially standardized (per instance) change in the probability of the correctly predicted class when conditioning on each grid cell ( $\Delta P$ ; see Appendix B). Thus, if a grid cell (input feature) consistently exhibits high  $\Delta P$  values for accurate predictions of a specific WR, it indicates that the grid cell was particularly important for the model's success on average.

##### 1) ATMOSPHERIC PRECURSORS

As shown in Figure 7, many atmosphere-based models show skill beyond persistence, especially during winter (Figures 4a and A1). U200- and Z500-based models contributed predictability for most WRs during winter, spring, and fall. U200 and Z500 predictability can come from both the persistent anomalies associated with the specific WR and other large-scale features of the

Earth system. Figure 8 shows composited standardized anomalies and  $\Delta P$  values for when several Z500- and U200-based models achieved accurate predictions for different WRs. During early lead times (weeks 2-3), the Pacific Ridge (winter) and Greenland High (winter and fall) regimes show increased predictability from these two variables (Figure 7e,i,l). Correct Pacific Ridge predictions during week 3 were associated with negative Z500 anomalies over the tropics, with positive anomalies over the northeastern Pacific Ocean (Figure 8a). Clusters of positive and high  $\Delta P$  values indicate that both the tropics and the persistent circulation over the northeastern Pacific (Figure 8d) contributed to the accurate predictions. Similar anomalies occur when accurately predicting week 2, but  $\Delta P$  values indicate a higher contribution from persistence at a shorter lead time (not shown).

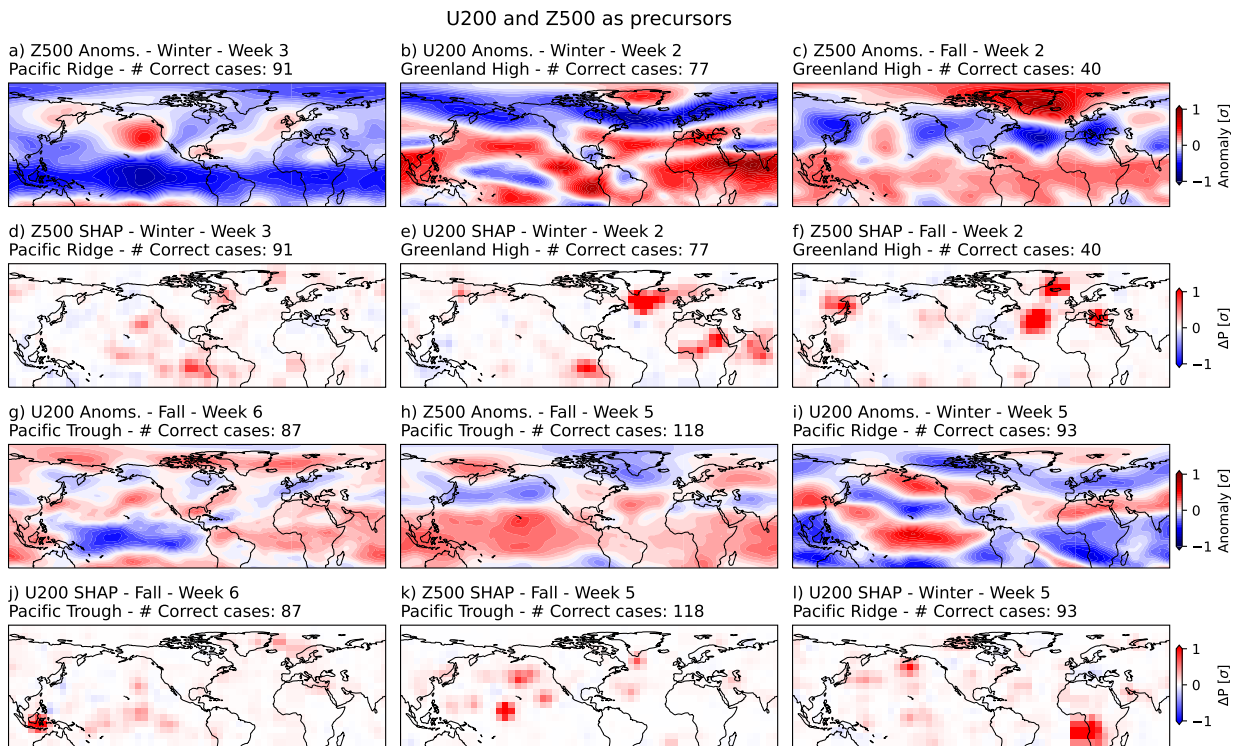


FIG. 8. Initial state standardized anomaly composites of Z500 and U200 corresponding to correct predictions of the specific WR, for the season and lead time indicated in the title of each panel. The average SHAP-based  $\Delta P$  values corresponding to these accurate predictions are also shown.

For the Greenland High regime, U200 showed skill during weeks 1-3 in winter (Figure 7i). According to Figure 8b,e, predictability contributions come from both the Greenland High persistent circulation (negative U200 anomalies south of Greenland) and increased eastward wind

over the eastern tropical Pacific and northern Indian Ocean. For week 3, the  $\Delta P$  spatial pattern is similar but with higher values over the tropics (not shown). For the Greenland High during fall, only the persistent large-scale extratropical structure seems to provide predictability out to week 2 (Figures 7l and 8c,f). For longer lead times in the fall, Z500 and U200 yield skill in predicting the Pacific Trough and Pacific Ridge regimes. Westward anomalies of U200 over the tropical Pacific Ocean and increased Z500 over the tropics are associated with accurate predictions of the Pacific Trough regime during weeks 4-7 (Figures 7d and 8g,h,j,k). The opposite is true for accurate predictions of the Pacific Ridge during winter (Figure 8i), which were associated with positive eastward upper-level flow over the tropical Pacific and upper-level divergence over the maritime continent. These large-scale characteristics related to Pacific Trough and Pacific Ridge occurrence resemble the Walker circulation response for the two opposite states of the ENSO (El Niño and La Niña, respectively).

U10 in Figure 7e consistently shows skill for predicting the Pacific Ridge during winter. The composites in Figure 9a-c and the associated  $\Delta P$  values highlight the westerly anomalies west of North America at around 60°N. Given the accompanying easterly anomalies north of this region and east of North America, the predictability seems to be related to structural changes of the stratospheric polar vortex. The composites also show westerly U10 anomalies along the equator and northern subtropical Pacific Ocean (from around -5°N to 35°N) that weaken with lead time, and easterly anomalies over the Southern Hemisphere tropics (see Figure 9d,e,f). Given the limited domain, it's difficult to infer more about the underlying tropical processes providing predictability. Although the stratospheric winds both at the tropics and extra-tropics show relevance for S2S prediction of the Pacific Ridge, it is difficult to link its occurrence directly with SSW events or specific phases of the QBO. The slowing of the stratospheric polar vortex is not zonally uniform, and the composite tropical U10 anomalies do not follow the classical definition of certain QBO phase (Baldwin et al. 2001). However, the robustness of the relationship (see Figure 7e) justifies future more in-depth research on this predictor.

According to Figure 7a,h,i, OLR provides predictability during some lead times for the Pacific Trough and Greenland High regimes in winter, and for the Pacific Ridge in fall. The composite anomalies and  $\Delta P$  fields show that accurate prediction of the Pacific Trough relates to increased convection over the central tropical Pacific (El-Niño-like signal; Figure 10a,d). In contrast, the

### Stratospheric winds and the Pacific Ridge

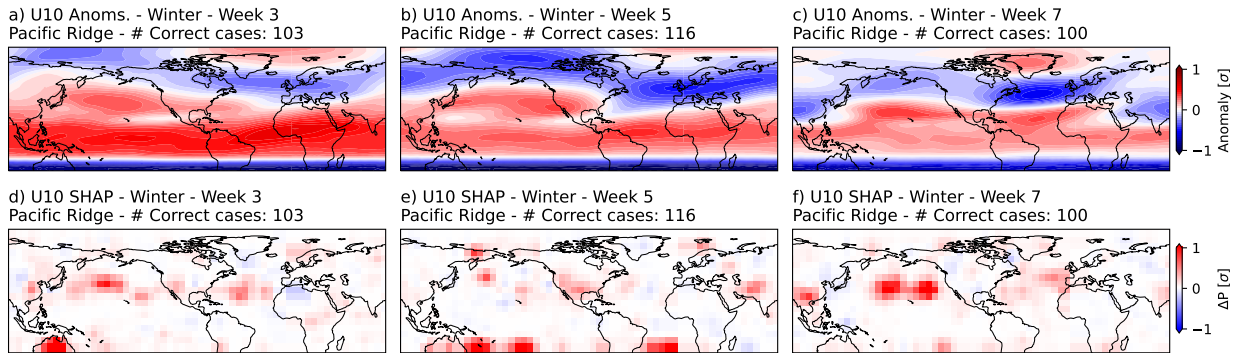


FIG. 9. Initial state standardized anomaly composites of U10 corresponding to correct predictions of the specific WR, for the season and lead time indicated in the title of each panel. The average SHAP-based  $\Delta P$  values corresponding to these accurate predictions are also shown.

OLR pattern is weak for the Pacific Ridge during fall (see Figure 10b,c,e,f). The OLR pattern is mixed for the Greenland High in winter; decreased cloudiness over Greenland (persistence) plays a relevant role for week 2 (Figure 10g,j), while enhanced cloudiness over east Asia provides increased predictability for week 6 (Figure 10i,l). When OLR provides meaningful predictive skill for the Greenland High regime, the tropics show a pattern similar to phases 6 and 7 of the MJO (Wang et al. 2018), with enhanced convection east of the Maritime continent (see Figure 10g-i). These particular MJO phases preceding the onset of Greenland Blocking are consistent with Parker et al. (2018), who showed how the progression of the MJO was related to a negative NAO (and also the onset of a Greenland High regime).

The results associated with U10 and OLR as precursors highlight the complexity of the interactions between the tropical troposphere and the stratospheric vortex that influence the occurrence of weather regimes (Lee et al. 2019; Barnes et al. 2019; Green and Furtado 2019), suggesting that the stratosphere may be more important for the onset of the Pacific Ridge, while tropical OLR may be closer tied to the Greenland High and Pacific Trough regimes.

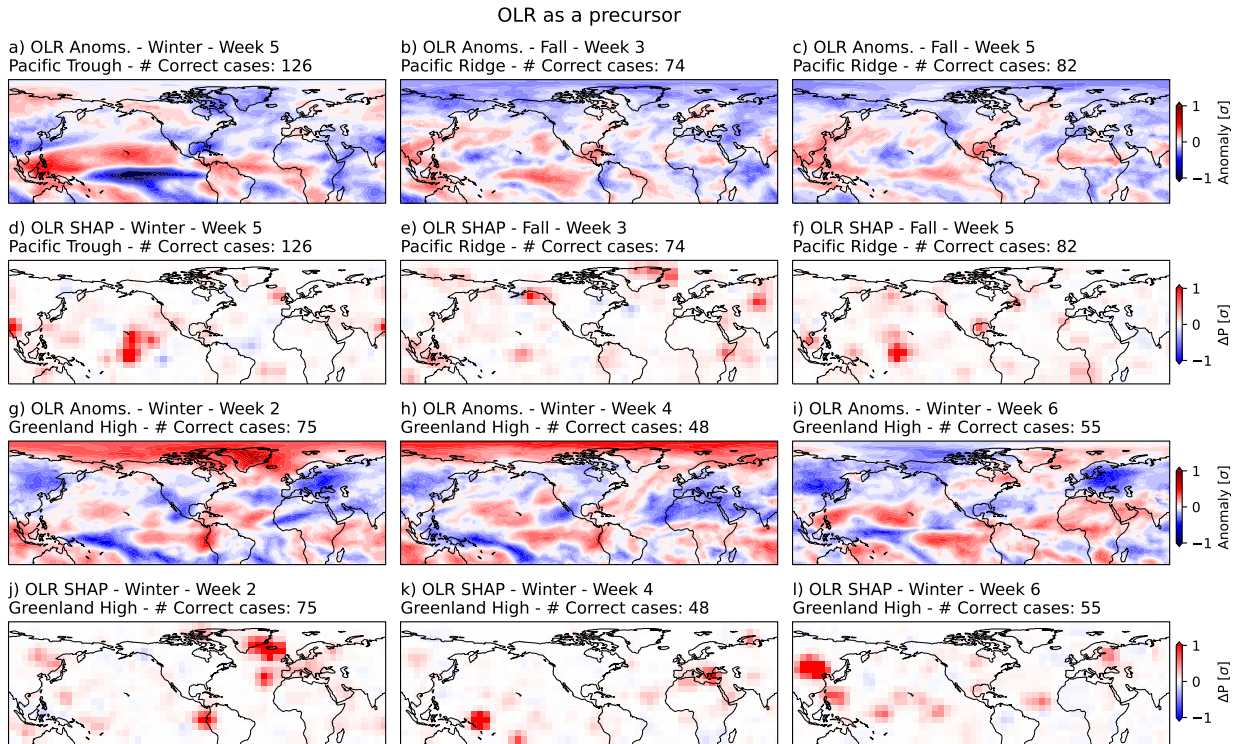


FIG. 10. Initial state standardized anomaly composites of OLR corresponding to correct predictions of the specific WR, for the season and lead time indicated in the title of each panel. The average SHAP-based  $\Delta P$  values corresponding to these accurate predictions are also shown.

## 2) OCEANIC PRECURSORS

The ocean was a significant source of predictability for the Pacific Trough and Pacific Ridge regimes during winter and fall (Figure 7a,d,e,h). Figure 11a,b,g,h shows composites of OHC and SST standardized anomalies for accurate predictions of the Pacific Trough and Pacific Ridge regimes. Although other ocean variables also contributed significant predictability (e.g., SSH, MLD, and OHC at different depths; Figure 7), the associated spatial patterns were similar. As expected (L'Heureux and Thompson 2006; Furtado et al. 2012; Deser et al. 2017), the tropical Pacific Ocean and ENSO play a relevant role (Figure 11a,b,g,h). Consistent with Straus et al. (2007), Vigaud et al. (2018), and Molina et al. (2023a), an El-Niño-like signal during fall and winter helps accurately predict the Pacific Trough regime. Analogously, a La-Niña-like tropical Pacific provides skill for predicting the Pacific Ridge regime. Although ocean-based predictability and ENSO are more consistent sources of predictability for the Pacific Trough regime during fall and

the Pacific Ridge during winter, the spatial distribution of anomalies and  $\Delta P$  is similar for the cold season in general (Figure 11a-b,d-e,g-h,j-k). It is worth noting that ENSO-derived predictability is not only identifiable in the ocean since atmosphere-based models can also accurately predict the Pacific Trough and Pacific Ridge regimes using atmospheric patterns related to ENSO, such as those shown in Figures 8a,g,h,i and 10a.

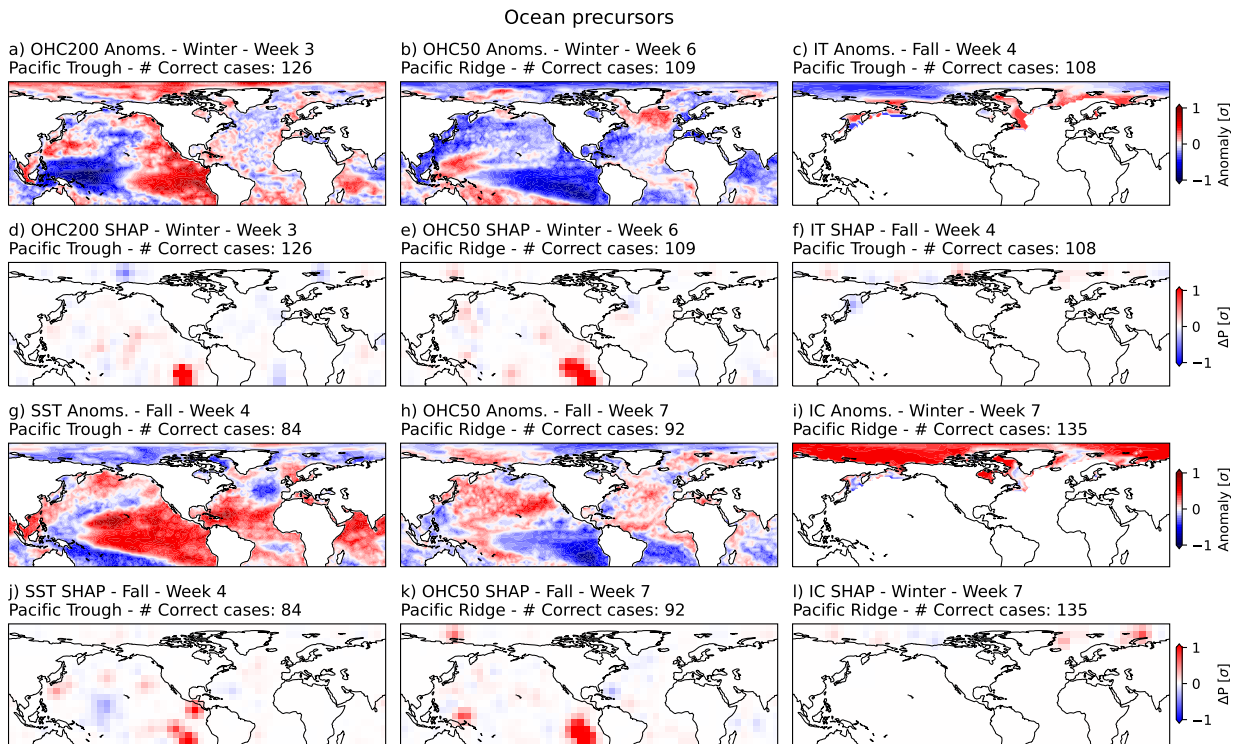


FIG. 11. Initial state standardized anomaly composites of OHC, SST, IT, and IC corresponding to correct predictions of the specific WR, for the season and lead time indicated in the title of each panel. The average SHAP-based  $\Delta P$  values corresponding to these accurate predictions are also shown.

IT provides predictability for the Pacific Trough during fall and IC for the Pacific Ridge during winter (Figure 7d,f). Accurate Pacific Trough predictions in fall are associated with negative anomalies over most of the North Pole (week 4 is shown in Figure 11c,f, but weeks 6 and 7 are similar). Contrarily, increased IC over the whole polar region helps predict the Pacific Ridge in winter (week 7 is shown in Figure 11i,l, but similar results for weeks 5-8). Although high  $\Delta P$  values are not concentrated over one specific region, the composites for the Pacific Trough and

Pacific Ridge regimes are in agreement with Gervais et al. (2024), indicating that less ice over the North Pole favors the occurrence of the Aleutian low.

### 3) LAND PRECURSORS

Figure 7 shows that different land variables provide meaningful predictability during different lead times and seasons for the Pacific Trough, Pacific Ridge, and Greenland High regimes. SWC provides increased skill at several lead times when predicting the Pacific Trough during fall (Figure 7d), the Pacific Ridge during winter and spring (Figure 7e,f), and the Greenland High during winter (Figure 7i). For the Greenland High in winter, models based on SWC averaged down to 28cm and 1m provide predictive skill during weeks 4-6 and 6-8, respectively. Both layers and the various lead times have a dipole of SWC anomalies over southeastern Asia, with high  $\Delta P$  values over that region, although the anomaly composites are noisy (Figure 12a,b,d,e). The anomalies and SHAP fields are more consistent across lead times for SWC-1m than for SWC-28cm (not shown). For the Pacific Ridge during winter and fall, negative anomalies of SWC over the western U.S. and positive anomalies over the Amazon rainforest are notable (Figure 12g,h,j,k). However,  $\Delta P$  values are inconclusive (Figure 12j,k). The SWC-full anomalies associated with increased predictability for the Pacific Trough in fall seem to be associated with increased SWC over North America (see Figure 12c,f). Although earlier results in Section 3c indicate that there is predictive skill from SWC, further exploration of these potential links is needed because our approach offers limited interpretation.

ST contributes to skillful prediction of the Pacific Trough (spring and fall) and Greenland High (winter and fall; see Figure 7b,d,i,l). Pacific Trough's increased predictability is associated with warmer soil temperatures over tropical South America and Africa and cooler temperatures over most of the extratropics (see Figure 12i,l,m,p). This land temperature pattern, given its global extent, could be a footprint of the El Niño phase of ENSO. ST-1m-based models are skillful when predicting the Greenland High during weeks 5 and 7 during winter and weeks 7 and 8 during fall (Figure 7i,l). The composites in these cases show negative anomalies over western North America, southeastern Asia, and western Russia, with positive anomalies over the Amazon (see Figure 12n,o,q,k). The pattern shown for ST in Figure 12n slightly resembles the one shown in Figure 10h,i, indicating that the predictability from OLR and ST may have the same process as its

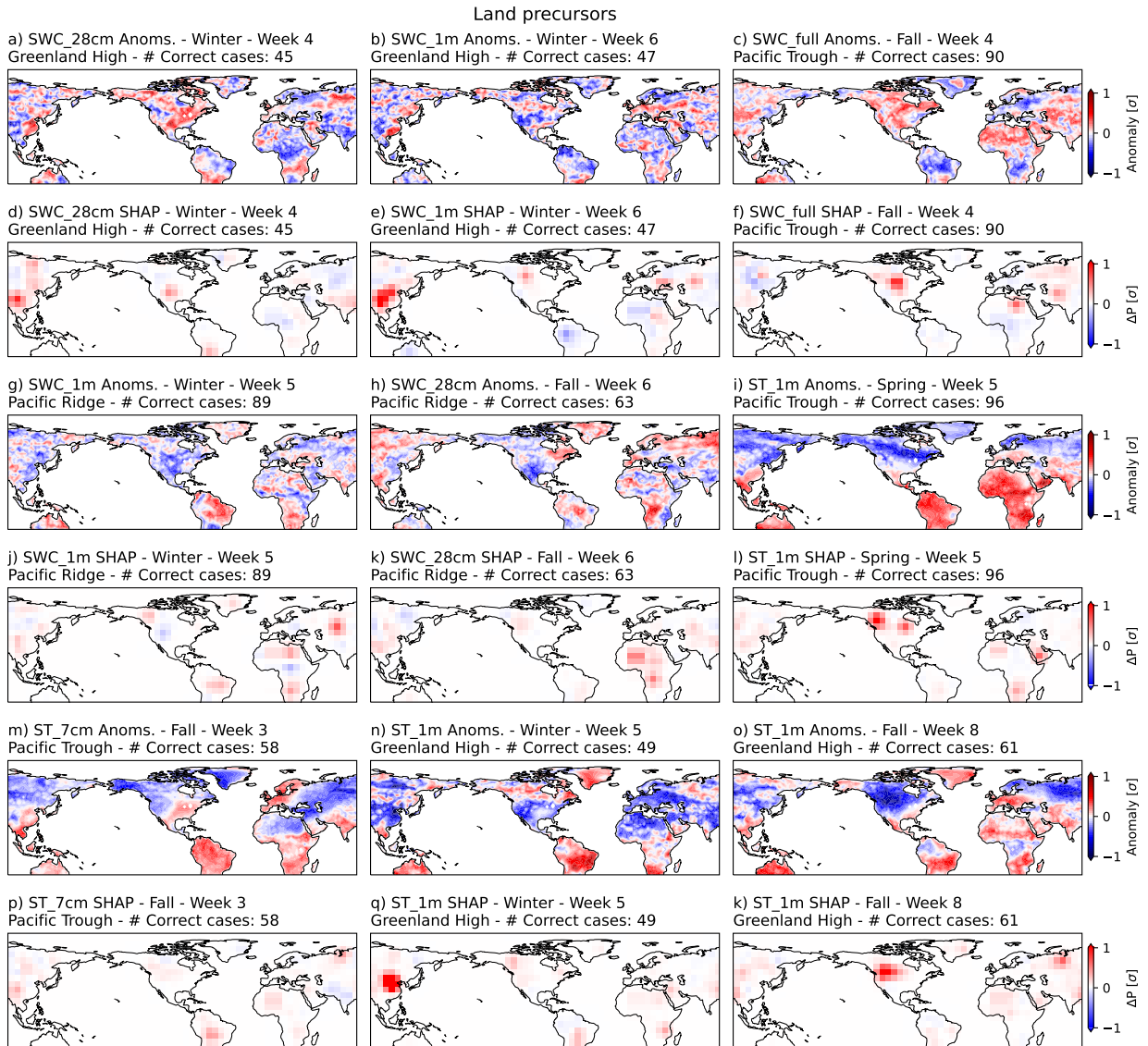


FIG. 12. Initial state standardized anomaly composites of SWC and ST corresponding to correct predictions of the specific WR, for the season and lead time indicated in the title of each panel. The average SHAP-based  $\Delta P$  values corresponding to these accurate predictions are also shown.

origin, such as the MJO-extratropic-tropospheric pathways described by Lee et al. (2019), Green and Furtado (2019), and Barnes et al. (2019).

Another predictability contributor for North American WRs is snow depth (SD). According to Figure 7b,h,i,l, SD-based models provided skill for the Pacific Trough in spring (weeks 2 and 3), the Pacific Ridge in fall (weeks 3, 5, and 6), and the Greenland High in winter (weeks 4, 6, and

7) and fall (weeks 6 and 7). The increased predictability of the Pacific Trough during spring is associated with increased SD over Alaska and Northern Canada and decreased SD over the U.S. Pacific Northwest (Figure 13a,d). For the Pacific Ridge in fall, the increased predictive skill is derived from negative SD anomalies over most of the Northern extra-tropics, mainly over Asia (Figures 13b,c,e,f). For the Greenland High, increased predictability is associated with increased SD over Asia, and decreased SD over northern Russia (Figure 13g,h,i,j,k,l). These results support studies that have shown a significant lagged influence of snowpack over the mid-latitude large-scale atmosphere (Peings et al. 2011; Orsolini et al. 2013; Thomas et al. 2016). However, given the sparsity of  $\Delta P$  values, further research on the mechanisms associated with this potential source of predictability is needed.

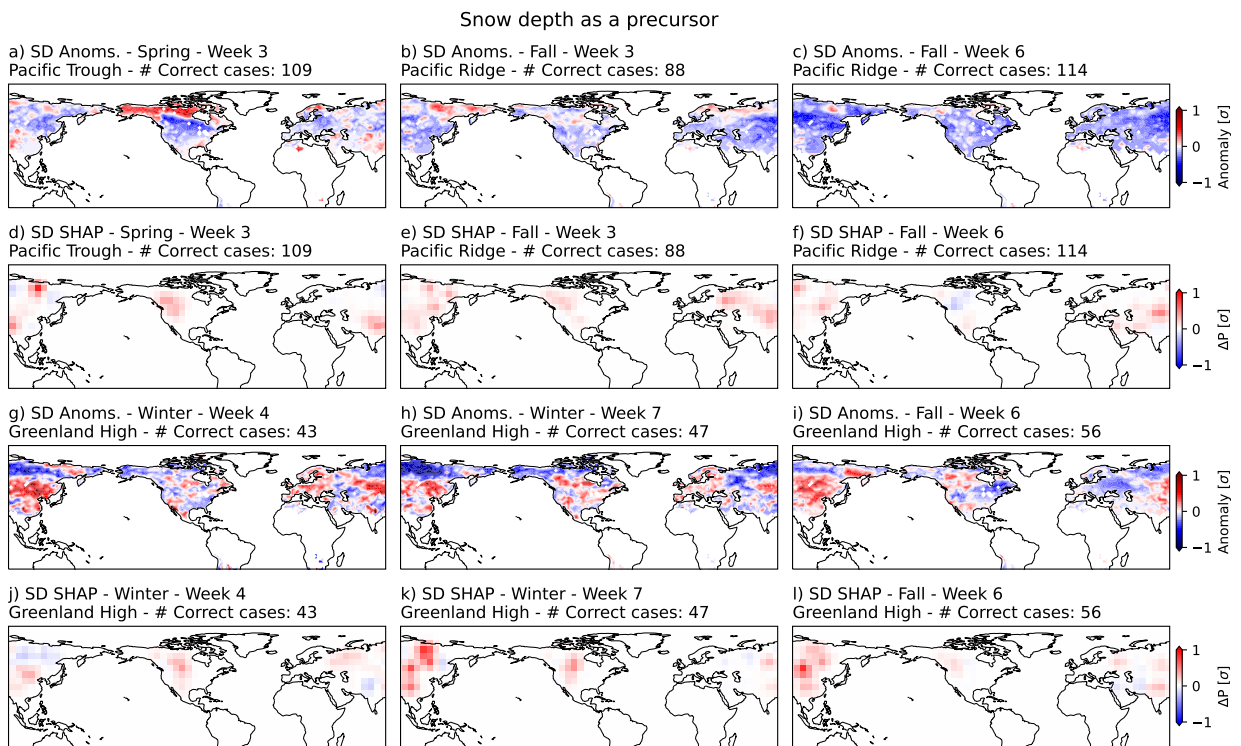


FIG. 13. Initial state standardized anomaly composites of SD corresponding to correct predictions of the specific WR, for the season and lead time indicated in the title of each panel. The average SHAP-based  $\Delta P$  values corresponding to these accurate predictions are also shown.

## 4. Conclusions

In this study, we used ML to learn more about the sources of predictability of North American WRs at S2S lead times. We trained several XGBoost models with an input variable from one of three Earth system components to assess their relative importance at various lead times and seasons. Similar to the results from Richter et al. (2024), we found that atmospheric variables were better predictors during weeks 1-2 of the forecast, with the ocean providing only slightly higher predictability than the other two components on average after week 3. On average, OHC variables had higher skill than SSTs on S2S lead times, highlighting the importance of sub-surface processes. In general, the ML models that included all variables from the land or ocean did not necessarily present the best skill compared to the one-variable models, possibly due to competing predictive signals or limitations in using correlated variables. Thus, while convenient, it is not advisable to train an ML model that includes all predictors without special treatment or weighting. Contrary to our expectations, fewer input features could provide more predictability when using ML, particularly for slower-evolving land and ocean processes.

Our results also show how highly seasonally dependent the sources of predictability are. We demonstrate that the S2S WR predictability increases during winter compared to spring and fall. For winter, the ocean and atmosphere provide more predictability than the land, while during spring, the ocean- and land-based models show higher skill than the atmosphere-based models. When initialized in summer, the models have little to no skill in predicting the specific WR happening. The forecast skill and the sources of predictability were also dependent on the regime being predicted. The models showed higher skill when predicting the Pacific Trough and Pacific Ridge regimes compared to the Greenland High and Alaskan Ridge. When stratifying by WR and season in Section 3c, results indicate that the ocean provides predictability for the Pacific Trough and Pacific Ridge regimes during fall and winter. Stratospheric processes are especially relevant for the Pacific Ridge during winter. OLR-related signals help forecast the Greenland High during some weeks in winter. The structure of the upper troposphere corresponding to both persistence and other large-scale features was relevant for the Pacific Trough, Pacific Ridge, and Greenland High regimes during early lead times. Additionally, some land-based models show significant skill during winter, spring, and fall.

In Section 3d, we provide a deeper analysis of the initial conditions associated with accurate forecasts and local feature contributions using the SHAP method. The results indicate that persistent tropospheric winds and convection, at times associated with weather regimes, were useful during early lead times (weeks 2-3), while the Walker circulation and tropical and subtropical convection, potentially associated with ENSO and MJO, were relevant features during later lead times. Stratospheric wind anomalies over the tropics and extra-tropics provide predictability for the Pacific Ridge. Although these patterns may be related to a phase of the QBO and/or sudden stratospheric warming events, future research is needed to understand their influence. From the ocean, most of the predictability is associated with distinct phases of the ENSO. El Niño provides skill for the Pacific Trough and La Niña for the Pacific Ridge. Additionally, ice thickness and concentration provided skill for predicting the Pacific Trough during fall and the Pacific Ridge during winter, respectively. From land, the spatial patterns of SWC associated with correct predictions are sparse and inconclusive; thus, further research is needed to better understand the influence of soil moisture as a source of predictability for WRs. ST provides some predictability, potentially acting as a footprint of ENSO. SD also shows predictive skill for the Pacific Trough and Pacific Ridge regimes, but more research on the physical mechanisms is needed. ENSO's footprint was evident in the three Earth system components, and when comparing MJO and ENSO contributions, our results agree with Mayer et al. (2024) that ENSO is a more meaningful contributor to S2S predictability.

The sources of predictability identified in this study are relevant beyond large-scale circulation patterns. The sources of predictability also present potential for improving the forecasting skill of surface weather, given how closely related regimes are to temperature and precipitation anomalies (see Figure 1), storm tracks, and extreme events (Lee et al. 2023; Jennrich et al. 2024). Future work should focus on more deeply understanding these sources of predictability, including testing causal hypotheses and quantifying how they may change in a warming climate. This study could also be expanded to incorporate other Earth system components and processes, delve further into the predictors herein, or include data from Earth system numerical models. Our study is subject to limitations associated with clustering methods, wherein the regimes may be sensitive to the length of the observational record; a limitation that future work should also explore. Combined with domain knowledge, ML continues to present opportunities to elucidate sources of predictability.

*Acknowledgments.* JSPC and MJM were supported by a University of Maryland Grand Challenges Seed Grant. MJM was also supported by the National Science Foundation (NSF) under Grant Number 2425735. Computing and data storage resources were provided by the Computational and Information Systems Laboratory at the NSF National Center for Atmospheric Research (NCAR), which is a major facility sponsored by the NSF under Cooperative Agreement No. 1852977. We acknowledge the use of OpenAI's ChatGPT for assistance in developing and refining the code used for analysis and visualization. We also thank the three anonymous reviewers for their valuable insights. Finally, we thank Simon H. Lee for providing clarifications and comments regarding WR computation and Vanessa Robledo for her ideas on results visualization.

*Data availability statement.* ERA5 and SODA reanalyses can be obtained from the NSF NCAR Research Data Archive (<https://doi.org/10.5065/D6X34W69> and <https://doi.org/10.5065/HBTB-R521>). Software developed for this study is available as open source at the GitHub repository: [https://github.com/jhayron-perez/WR\\_Predictability](https://github.com/jhayron-perez/WR_Predictability).

## APPENDIX A

### Variables that contributed meaningful predictability depending on the season and weather regime, separately

Figures A1 and A2 show the variables contributing meaningful and high predictability during different seasons and when aiming to predict different weather regimes. These are shown for completeness, but a more detailed version is shown in Figure 7.

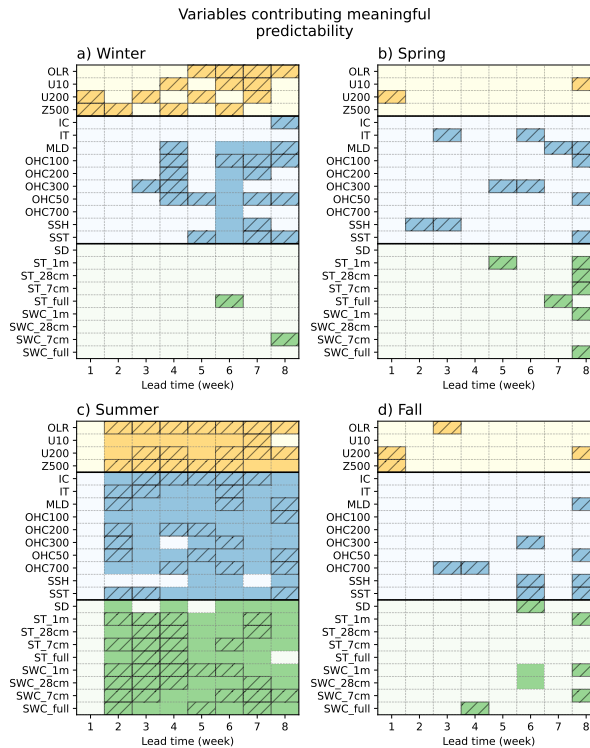


FIG. A1. Same as Figure 3 but stratified by the season in which the forecasts were initialized.

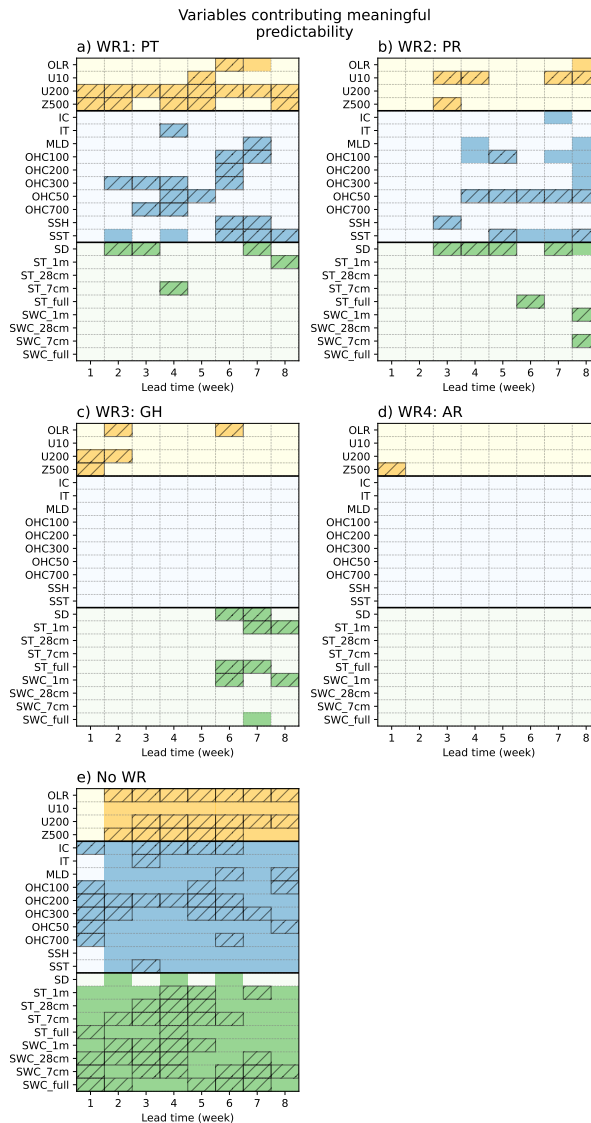


FIG. A2. Same as Figure 3 but stratified by target weather regime.

## APPENDIX B

### Processing of the SHapley Additive exPlanations (SHAP)

In Section 3d and Figures 8-13, we used SHAP values to identify regions in the input fields that were relevant to making accurate predictions. These values quantify the contribution of each input feature to the final predictions (Lundberg and Lee 2017; Flora et al. 2024), and the XGBoost model allows for their exact estimation. In XGBoost, the probability assigned to a class  $i$  in the final prediction is

$$P(i) = \frac{e^{\sum_n \text{SHAP}_i + \text{base SHAP for class } i}}{\sum_j e^{\sum_n \text{SHAP}_j + \text{base SHAP for class } j}}, \quad (\text{B1})$$

where  $j$  refers to each one of the 5 predicted classes and  $n$  refers to each input feature. Consequently, for each individual prediction (instance), and for each input feature, positive SHAP values indicate that the feature contributed to an increase in the probability of that class being predicted. After obtaining the SHAP values in log-odds scale (as in Equation B1), we converted the SHAP values into probability anomalies from the base probability as

$$\Delta P_{n,i} = \frac{e^{\sum \text{SHAP}_n + \text{base value}_i}}{\sum_j e^{\sum \text{SHAP}_j + \text{base value}_j}} - P_i(\text{base SHAP}). \quad (\text{B2})$$

where the base probability is  $20\% \pm 5\%$  (mean and standard deviation for all classes and models), depending on the class and model. This way, additivity is preserved, meaning that the final probability prediction for class  $i$  equals the base probability for the class plus the sum of the  $\Delta P_{n,i}$  from all the input features:

$$P(i) = P_i(\text{base SHAP}) + \sum_n \Delta P_{n,i}, \quad (\text{B3})$$

with

$$P_i(\text{base SHAP}) = \frac{e^{\text{base SHAP}_i}}{\sum_j e^{\text{base SHAP}_j}}. \quad (\text{B4})$$

Finally, since our interest is to investigate the local importance of each feature (regardless of the final predicted probability or the base frequency of the predicted class), and in order to have comparable units among different models and classes, we divide the  $\Delta P_{n,i}$  by the standard deviation of all  $\Delta P_{n,i}$  for each instance (cosine-latitude-weighted spatial standard deviation). The average in time of those spatially standardized values (for times in which the model correctly predicted class  $i$ ) is shown in Figures 8-13 as an indication of the local importance of each feature for the final prediction. Taking advantage of the additivity of the  $\Delta P_{n,i}$ , we aggregated values over  $6^\circ \times 6^\circ$  grid cells, and in order to remove discontinuities due to the random sampling from XGBoost, we smoothed the fields using a Gaussian filter ( $\sigma = 0.75$ ).

## References

- Alexander, M. A., 1992: Midlatitude atmosphere–ocean interaction during El Niño. Part I: the North Pacific Ocean. *J. Climate*, **5** (9), 944–958, [https://doi.org/10.1175/1520-0442\(1992\)005<0944:MAIDEN>2.0.CO;2](https://doi.org/10.1175/1520-0442(1992)005<0944:MAIDEN>2.0.CO;2).
- Baldwin, M., and Coauthors, 2001: The quasi-biennial oscillation. *Rev. Geophys.*, **39** (2), 179–229, <https://doi.org/10.1029/1999RG000073>.
- Baldwin, M. P., and T. J. Dunkerton, 2001: Stratospheric harbingers of anomalous weather regimes. *Science*, **294** (5542), 581–584, <https://doi.org/10.1126/science.1063315>.
- Baldwin, M. P., D. B. Stephenson, D. W. 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, <https://doi.org/10.1126/science.1087143>.
- Balmaseda, M. A., L. Ferranti, F. Molteni, and T. N. Palmer, 2010: Impact of 2007 and 2008 Arctic ice anomalies on the atmospheric circulation: Implications for long-range predictions. *Quart. J. Roy. Meteor. Soc.*, **136** (652), 1655–1664, <https://doi.org/10.1002/qj.661>.
- Balsamo, G., A. Beljaars, K. Scipal, P. Viterbo, B. van den Hurk, M. Hirschi, and A. K. Betts, 2009: A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the Integrated Forecast System. *Journal of hydrometeorology*, **10** (3), 623–643, <https://doi.org/10.1175/2008JHM1068.1>.
- Barnes, E. A., S. M. Samarasinghe, I. Ebert-Uphoff, and J. C. Furtado, 2019: Tropospheric and stratospheric causal pathways between the mjo and nao. *Journal of Geophysical Research: Atmospheres*, **124** (16), 9356–9371, <https://doi.org/10.1029/2019JD031024>, URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JD031024>, <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019JD031024>.
- Becker, E. J., B. P. Kirtman, M. L’Heureux, Á. G. Muñoz, and K. Pegion, 2022: A decade of the North American Multimodel Ensemble (NMME): Research, application, and future directions. *Bull. Amer. Meteor. Soc.*, **103** (3), E973–E995, <https://doi.org/10.1175/BAMS-D-20-0327.1>.

- Bi, K., L. Xie, H. Zhang, X. Chen, X. Gu, and Q. Tian, 2023: Accurate medium-range global weather forecasting with 3d neural networks. *Nature*, **619** (7970), 533–538, <https://doi.org/10.1038/s41586-023-06185-3>.
- Bjerknes, J., 1969: Atmospheric teleconnections from the equatorial Pacific. *Mon. Wea. Rev.*, **97** (3), 163–172, [https://doi.org/10.1175/1520-0493\(1969\)097<0163:ATFTEP>2.3.CO;2](https://doi.org/10.1175/1520-0493(1969)097<0163:ATFTEP>2.3.CO;2).
- Blackmon, M. L., 1976: A climatological spectral study of the 500 mb geopotential height of the Northern Hemisphere. *J. Atmos. Sci.*, **33** (8), 1607–1623, [https://doi.org/10.1175/1520-0469\(1976\)033<1607:ACSSOT>2.0.CO;2](https://doi.org/10.1175/1520-0469(1976)033<1607:ACSSOT>2.0.CO;2).
- Büeler, D., L. Ferranti, L. Magnusson, J. F. Quinting, and C. M. Grams, 2021: Year-round sub-seasonal forecast skill for atlantic–european weather regimes. *Quarterly Journal of the Royal Meteorological Society*, **147** (741), 4283–4309, <https://doi.org/10.1002/qj.4178>, URL <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.4178>, <https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.4178>.
- Carton, J. A., G. A. Chepurin, and L. Chen, 2018: SODA3: A new ocean climate reanalysis. *J. Climate*, **31** (17), 6967–6983, <https://doi.org/10.1175/JCLI-D-18-0149.1>.
- Chelton, D. B., M. G. Schlax, and R. M. Samelson, 2011: Global observations of nonlinear mesoscale eddies. *Progress in oceanography*, **91** (2), 167–216, <https://doi.org/10.1016/j.pocean.2011.01.002>.
- Chelton, D. B., and S.-P. Xie, 2010: Coupled ocean-atmosphere interaction at oceanic mesoscales. *Oceanography*, **23** (4), 52–69, URL <https://www.jstor.org/stable/24860862>.
- Chen, L., and Coauthors, 2024: A machine learning model that outperforms conventional global subseasonal forecast models. *Nature Communications*, **15** (1), 6425, <https://doi.org/10.1038/s41467-024-50714-1>.
- Chen, T., and C. Guestrin, 2016: Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794, <https://doi.org/10.1145/2939672.2939785>.

- Cheng, X., and J. M. Wallace, 1993: Cluster analysis of the Northern Hemisphere wintertime 500-hPa height field: Spatial patterns. *J. Atmos. Sci.*, **50** (16), 2674–2696, [https://doi.org/10.1175/1520-0469\(1993\)050<2674:CAOTNH>2.0.CO;2](https://doi.org/10.1175/1520-0469(1993)050<2674:CAOTNH>2.0.CO;2).
- Conil, S., H. Douville, and S. Tyteca, 2009: Contribution of realistic soil moisture initial conditions to boreal summer climate predictability. *Climate Dyn.*, **32**, 75–93, <https://doi.org/10.1007/s00382-008-0375-9>.
- Davini, P., C. Cagnazzo, R. Neale, and J. Tribbia, 2012: Coupling between Greenland blocking and the North Atlantic Oscillation pattern. *Geophys. Res. Lett.*, **39** (14), <https://doi.org/10.1029/2012GL052315>.
- DelSole, T., L. Trenary, M. K. Tippett, and K. Pegion, 2017: Predictability of Week-3–4 Average Temperature and Precipitation over the Contiguous United States. *J. Climate*, **30** (10), 3499 – 3512, <https://doi.org/10.1175/JCLI-D-16-0567.1>.
- Deser, C., I. R. Simpson, K. A. McKinnon, and A. S. Phillips, 2017: The northern hemisphere extratropical atmospheric circulation response to enso: How well do we know it and how do we evaluate models accordingly? *Journal of Climate*, **30** (13), 5059 – 5082, <https://doi.org/10.1175/JCLI-D-16-0844.1>, URL <https://journals.ametsoc.org/view/journals/clim/30/13/jcli-d-16-0844.1.xml>.
- Deser, C., R. A. Tomas, and S. Peng, 2007: The transient atmospheric circulation response to North Atlantic SST and sea ice anomalies. *J. Climate*, **20** (18), 4751–4767, <https://doi.org/10.1175/JCLI4278.1>.
- Domeisen, D. I., and Coauthors, 2020: The role of the stratosphere in subseasonal to seasonal prediction: 2. Predictability arising from stratosphere-troposphere coupling. *Journal of Geophysical Research: Atmospheres*, **125** (2), e2019JD030923, <https://doi.org/10.1029/2019JD030923>.
- Dong, J., W. Zeng, L. Wu, J. Huang, T. Gaiser, and A. K. Srivastava, 2023: Enhancing short-term forecasting of daily precipitation using numerical weather prediction bias correcting with XGBoost in different regions of China. *Engineering Applications of Artificial Intelligence*, **117**, 105 579, <https://doi.org/10.1016/j.engappai.2022.105579>.

- Fatima, S., A. Hussain, S. B. Amir, S. H. Ahmed, and S. M. H. Aslam, 2023: XGBoost and Random Forest Algorithms: An in Depth Analysis. *Pakistan Journal of Scientific Research*, **3** (1), 26–31, <https://doi.org/doi.org/10.57041/pjosr.v3i1.946>.
- Ferreira, D., and C. Frankignoul, 2005: The transient atmospheric response to midlatitude SST anomalies. *J. Climate*, **18** (7), 1049–1067, <https://doi.org/10.1175/JCLI-3313.1>.
- Flora, M. L., C. K. Potvin, A. McGovern, and S. Handler, 2024: A Machine Learning Explainability Tutorial for Atmospheric Sciences. *Artificial Intelligence for the Earth Systems*, **3** (1), e230018, <https://doi.org/10.1175/AIES-D-23-0018.1>, URL <https://journals.ametsoc.org/view/journals/aies/3/1/AIES-D-23-0018.1.xml>, place: Boston MA, USA Publisher: American Meteorological Society.
- Frazier, P. I., 2018: A tutorial on bayesian optimization. URL <https://arxiv.org/abs/1807.02811>, <https://doi.org/10.48550/arXiv.1807.02811>, 1807.02811.
- Friedman, J. H., 2001: Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 1189–1232, URL <https://www.jstor.org/stable/2699986>.
- Furtado, J. C., E. Di Lorenzo, B. T. Anderson, and N. Schneider, 2012: Linkages between the north pacific oscillation and central tropical pacific ssts at low frequencies. *Climate Dynamics*, **39**, 2833–2846, <https://doi.org/10.1007/s00382-011-1245-4>.
- Gerber, E., C. Orbe, and L. M. Polvani, 2009: Stratospheric influence on the tropospheric circulation revealed by idealized ensemble forecasts. *Geophys. Res. Lett.*, **36** (24), <https://doi.org/10.1029/2009GL040913>.
- Gervais, M., L. Sun, and C. Deser, 2024: Impacts of projected arctic sea ice loss on daily weather patterns over north america. *Journal of Climate*, **37** (3), 1065 – 1085, <https://doi.org/10.1175/JCLI-D-23-0389.1>, URL <https://journals.ametsoc.org/view/journals/clim/37/3/JCLI-D-23-0389.1.xml>.
- Green, M. R., and J. C. Furtado, 2019: Evaluating the joint influence of the madden-julian oscillation and the stratospheric polar vortex on weather patterns in the northern hemisphere. *Journal of Geophysical Research: Atmospheres*, **124** (22), 11 693–11 709, <https://doi.org/10.1029/>

- 2019JD030771, URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JD030771>, <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019JD030771>.
- Guo, Z., P. A. Dirmeyer, and T. DelSole, 2011: Land surface impacts on subseasonal and seasonal predictability. *Geophys. Res. Lett.*, **38** (24), <https://doi.org/10.1029/2011GL049945>.
- Guo, Z., P. A. Dirmeyer, T. DelSole, and R. D. Koster, 2012: Rebound in atmospheric predictability and the role of the land surface. *J. Climate*, **25** (13), 4744–4749, <https://doi.org/10.1175/JCLI-D-11-00651.1>.
- Hartmann, D. L., 2015: Pacific sea surface temperature and the winter of 2014. *Geophys. Res. Lett.*, **42** (6), 1894–1902, <https://doi.org/10.1002/2015GL063083>.
- Hasan, S. S., L. Zhen, M. G. Miah, T. Ahamed, and A. Samie, 2020: Impact of land use change on ecosystem services: A review. *Environmental Development*, **34**, 100 527, <https://doi.org/10.1016/j.envdev.2020.100527>.
- Hendon, H. H., and M. L. Salby, 1994: The life cycle of the Madden–Julian oscillation. *J. Atmos. Sci.*, **51** (15), 2225–2237, [https://doi.org/10.1175/1520-0469\(1994\)051<2225:TLCOTM>2.0.CO;2](https://doi.org/10.1175/1520-0469(1994)051<2225:TLCOTM>2.0.CO;2).
- Herman, G. R., and R. S. Schumacher, 2018: Money doesn't grow on trees, but forecasts do: Forecasting extreme precipitation with random forests. *Mon. Wea. Rev.*, **146** (5), 1571–1600, <https://doi.org/10.1175/MWR-D-17-0250.1>.
- Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146** (730), 1999–2049, <https://doi.org/10.1002/qj.3803>.
- Hochman, A., G. Messori, J. F. Quinting, J. G. Pinto, and C. M. Grams, 2021: Do atlantic-european weather regimes physically exist? *Geophysical Research Letters*, **48** (20), e2021GL095 574, <https://doi.org/10.1029/2021GL095574>, URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021GL095574>, e2021GL095574 2021GL095574, <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2021GL095574>.
- Horel, J. D., and J. M. Wallace, 1981: Planetary-scale atmospheric phenomena associated with the Southern Oscillation. *Mon. Wea. Rev.*, **109** (4), 813–829, [https://doi.org/10.1175/1520-0493\(1981\)109<0813:PSAPAW>2.0.CO;2](https://doi.org/10.1175/1520-0493(1981)109<0813:PSAPAW>2.0.CO;2).

- Jennrich, G., D. Straus, M. Chelliah, and C. Baggett, 2024: Ensemble predictability of week 3/4 precipitation and temperature over the united states via cluster analysis of the large-scale circulation. *Weather and Forecasting*, **39** (11), 1531 – 1544, <https://doi.org/10.1175/WAF-D-23-0065.1>, URL <https://journals.ametsoc.org/view/journals/wefo/39/11/WAF-D-23-0065.1.xml>.
- Jeong, J.-H., H. W. Linderholm, S.-H. Woo, C. Folland, B.-M. Kim, S.-J. Kim, and D. Chen, 2013: Impacts of snow initialization on subseasonal forecasts of surface air temperature for the cold season. *J. Climate*, **26** (6), 1956–1972, <https://doi.org/10.1175/JCLI-D-12-00159.1>.
- 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. *Nature Geoscience*, **8** (6), 433–440, <https://doi.org/10.1038/ngeo2424>.
- Kolstad, E. W., S. H. Lee, A. H. Butler, D. I. V. Domeisen, and C. O. Wulff, 2022: Diverse surface signatures of stratospheric polar vortex anomalies. *Journal of Geophysical Research: Atmospheres*, **127** (20), e2022JD037422, <https://doi.org/10.1029/2022JD037422>, URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022JD037422>, e2022JD037422 2022JD037422, <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2022JD037422>.
- Korjus, K., M. N. Hebart, and R. Vicente, 2016: An efficient data partitioning to improve classification performance while keeping parameters interpretable. *PloS one*, **11** (8), e0161788, <https://doi.org/10.1371/journal.pone.0161788>.
- Koster, R., and G. Walker, 2015: Interactive vegetation phenology, soil moisture, and monthly temperature forecasts. *J. Hydrometeor.*, **16** (4), 1456–1465, <https://doi.org/10.1175/JHM-D-14-0205.1>.
- Koster, R. D., and Coauthors, 2004: Regions of strong coupling between soil moisture and precipitation. *Science*, **305** (5687), 1138–1140, <https://doi.org/10.1126/science.1100217>.
- Koster, R. D., and Coauthors, 2010: Contribution of land surface initialization to subseasonal forecast skill: First results from a multi-model experiment. *Geophys. Res. Lett.*, **37** (2), <https://doi.org/10.1029/2009GL041677>.

- Koster, R. D., and Coauthors, 2011: The Second Phase of the Global Land–Atmosphere Coupling Experiment: Soil Moisture Contributions to Subseasonal Forecast Skill. *J. Hydrometeorol.*, **12** (5), 805 – 822, <https://doi.org/10.1175/2011JHM1365.1>.
- Krishnamurti, T. N., 1961: The subtropical jet stream of winter. *J. Atmos. Sci.*, **18** (2), 172–191, [https://doi.org/10.1175/1520-0469\(1961\)018<0172:TSJSOW>2.0.CO;2](https://doi.org/10.1175/1520-0469(1961)018<0172:TSJSOW>2.0.CO;2).
- Kuang, X., Y. Zhang, Y. Huang, and D. Huang, 2014: Spatial differences in seasonal variation of the upper-tropospheric jet stream in the northern hemisphere and its thermal dynamic mechanism. *Theoretical and applied climatology*, **117**, 103–112, <https://doi.org/10.1007/s00704-013-0994-x>.
- Kurczyn, J., E. Beier, M. F. Lavín, and A. Chaigneau, 2012: Mesoscale eddies in the northeastern Pacific tropical-subtropical transition zone: Statistical characterization from satellite altimetry. *Journal of Geophysical Research: Oceans*, **117** (C10), <https://doi.org/10.1029/2012JC007970>.
- Kushnir, Y., W. Robinson, I. Bladé, N. Hall, S. Peng, and R. Sutton, 2002: Atmospheric GCM response to extratropical SST anomalies: Synthesis and evaluation. *J. Climate*, **15** (16), 2233–2256, [https://doi.org/10.1175/1520-0442\(2002\)015<2233:AGRTES>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<2233:AGRTES>2.0.CO;2).
- Kwon, Y.-O., M. A. Alexander, N. A. Bond, C. Frankignoul, H. Nakamura, B. Qiu, and L. A. Thompson, 2010: Role of the Gulf Stream and Kuroshio–Oyashio systems in large-scale atmosphere–ocean interaction: A review. *J. Climate*, **23** (12), 3249–3281, <https://doi.org/10.1175/2010JCLI3343.1>.
- Lang, S., and Coauthors, 2024: Aifs – ecmwf’s data-driven forecasting system. URL <https://arxiv.org/abs/2406.01465>, <https://doi.org/10.48550/arXiv.2406.01465>, 2406.01465.
- Lee, S., J. Furtado, and A. Charlton-Perez, 2019: Wintertime North American weather regimes and the Arctic stratospheric polar vortex. *Geophys. Res. Lett.*, **46** (24), 14 892–14 900, <https://doi.org/10.1029/2019GL085592>.
- Lee, S. H., and G. Messori, 2024: The dynamical footprint of year-round north american weather regimes. *Geophysical Research Letters*, **51** (2), e2023GL107 161, <https://doi.org/10.1029/2023GL107161>, URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/>

10.1029/2023GL107161, e2023GL107161 2023GL107161, <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2023GL107161>.

Lee, S. H., M. K. Tippett, and L. M. Polvani, 2023: A new year-round weather regime classification for North America. *J. Climate*, **36** (20), 7091–7108, <https://doi.org/10.1175/JCLI-D-23-0214.1>.

Lin, H., J. Frederiksen, D. Straus, and C. Stan, 2019: Tropical-extratropical interactions and teleconnections. *Sub-Seasonal to Seasonal Prediction*, Elsevier, 143–164, <https://doi.org/10.1016/B978-0-12-811714-9.00007-3>.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20** (2), 130–141, [https://doi.org/10.1175/1520-0469\(1963\)020<0130:DNF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2).

Lundberg, S., and S.-I. Lee, 2017: A unified approach to interpreting model predictions. URL <https://arxiv.org/abs/1705.07874>, <https://doi.org/10.48550/arXiv.1705.07874>, 1705.07874.

L'Heureux, M. L., and D. W. J. Thompson, 2006: Observed relationships between the el niño–southern oscillation and the extratropical zonal-mean circulation. *Journal of Climate*, **19** (2), 276 – 287, <https://doi.org/10.1175/JCLI3617.1>, URL <https://journals.ametsoc.org/view/journals/clim/19/2/jcli3617.1.xml>.

Madden, R. A., 1986: Seasonal variations of the 40-50 day oscillation in the tropics. *Journal of Atmospheric Sciences*, **43** (24), 3138–3158.

Mariotti, A., and Coauthors, 2020: Windows of opportunity for skillful forecasts subseasonal to seasonal and beyond. *Bull. Amer. Meteor. Soc.*, **101** (5), E608–E625, <https://doi.org/10.1175/BAMS-D-18-0326.1>.

Mayer, K. J., and E. A. Barnes, 2021: Subseasonal forecasts of opportunity identified by an explainable neural network. *Geophys. Res. Lett.*, **48** (10), e2020GL092092, <https://doi.org/10.1029/2020GL092092>.

Mayer, K. J., W. E. Chapman, and W. A. Manriquez, 2024: Exploring the relative importance of the mjo and enso to north pacific subseasonal predictability. *Geophysical Research Letters*, **51** (10), e2024GL108479, <https://doi.org/10.1029/2024GL108479>, URL <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024GL108479>.

wiley.com/doi/abs/10.1029/2024GL108479, e2024GL108479 2024GL108479, <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2024GL108479>.

- Meehl, G. A., and Coauthors, 2021: Initialized Earth system prediction from subseasonal to decadal timescales. *Nature Reviews Earth & Environment*, **2** (5), 340–357, <https://doi.org/10.1038/s43017-021-00155-x>.
- Merryfield, W. J., and Coauthors, 2020: Current and emerging developments in subseasonal to decadal prediction. *Bull. Amer. Meteor. Soc.*, **101** (6), E869–E896, <https://doi.org/10.1175/BAMS-D-19-0037.1>.
- Michelangeli, P.-A., R. Vautard, and B. Legras, 1995: Weather regimes: Recurrence and quasi stationarity. *J. Atmos. Sci.*, **52** (8), 1237–1256, [https://doi.org/10.1175/1520-0469\(1995\)052<1237:WRRASQ>2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052<1237:WRRASQ>2.0.CO;2).
- Molina, M. J., J. H. Richter, A. A. Glanville, K. Dagon, J. Berner, A. Hu, and G. A. Meehl, 2023a: Subseasonal representation and predictability of North American weather regimes using cluster analysis. *Artificial Intelligence for the Earth Systems*, 1–54, <https://doi.org/10.1175/AIES-D-22-0051.1>.
- Molina, M. J., and Coauthors, 2023b: A review of recent and emerging machine learning applications for climate variability and weather phenomena. *Artificial Intelligence for the Earth Systems*, 1–46, <https://doi.org/10.1175/AIES-D-22-0086.1>.
- Mostajabi, A., D. L. Finney, M. Rubinstein, and F. Rachidi, 2019: Nowcasting lightning occurrence from commonly available meteorological parameters using machine learning techniques. *Npj Climate and Atmospheric Science*, **2** (1), 41, <https://doi.org/10.1038/s41612-019-0098-0>.
- Muñoz-Sabater, J., and Coauthors, 2021: ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, **13** (9), 4349–4383, <https://doi.org/10.5194/essd-13-4349-2021>.
- National Academies of Sciences, E., B. o. A. S. Medicine, D. o. E. Climate, Ocean Studies Board, and L. Studies, 2016: *Next generation Earth system prediction: Strategies for subseasonal to seasonal forecasts*. National Academies Press, Washington, D.C., <https://doi.org/10.17226/21873>.

- Orsolini, Y., R. Senan, G. Balsamo, F. Doblas-Reyes, F. Vitart, A. Weisheimer, A. Carrasco, and R. Benestad, 2013: Impact of snow initialization on sub-seasonal forecasts. *Climate Dyn.*, **41**, 1969–1982, <https://doi.org/10.1007/s00382-013-1782-0>.
- Palmén, E., 1948: On the distribution of temperature and wind in the upper westerlies. *J. Atmos. Sci.*, **5** (1), 20–27, [https://doi.org/10.1175/1520-0469\(1948\)005<0020:OTDOTA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1948)005<0020:OTDOTA>2.0.CO;2).
- Parker, T., T. Woollings, and A. Weisheimer, 2018: Ensemble sensitivity analysis of greenland blocking in medium-range forecasts. *Quarterly Journal of the Royal Meteorological Society*, **144** (716), 2358–2379, <https://doi.org/10.1002/qj.3391>, URL <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3391>, <https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3391>.
- Pegion, K., and Coauthors, 2019: The Subseasonal Experiment (SubX): A multimodel subseasonal prediction experiment. *Bull. Amer. Meteor. Soc.*, **100** (10), 2043–2060, <https://doi.org/10.1175/BAMS-D-18-0270.1>.
- Peings, Y., H. Douville, R. Alkama, and B. Decharme, 2011: Snow contribution to springtime atmospheric predictability over the second half of the twentieth century. *Climate Dyn.*, **37**, 985–1004, <https://doi.org/10.1007/s00382-010-0884-1>.
- Petrie, R. E., L. C. Shaffrey, and R. T. Sutton, 2015: Atmospheric response in summer linked to recent arctic sea ice loss. *Quarterly Journal of the Royal Meteorological Society*, **141** (691), 2070–2076, <https://doi.org/10.1002/qj.2502>, URL <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2502>, <https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.2502>.
- Philander, S. G., 1999: A review of tropical ocean–atmosphere interactions. *Tellus B*, **51** (1), 71–90.
- Philander, S. G. H., 1983: El Niño–Southern Oscillation phenomena. *Nature*, **302** (5906), 295–301, <https://doi.org/10.1038/302295a0>.
- Qian, Q. F., X. J. Jia, and H. Lin, 2020: Machine learning models for the seasonal forecast of winter surface air temperature in North America. *Earth and Space Science*, **7** (8), e2020EA001140, <https://doi.org/10.1029/2020EA001140>.

- Reinhold, B. B., and R. T. Pierrehumbert, 1982: Dynamics of weather regimes: Quasi-stationary waves and blocking. *Mon. Wea. Rev.*, **110** (9), 1105–1145, [https://doi.org/10.1175/1520-0493\(1982\)110<1105:DOWRQS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1982)110<1105:DOWRQS>2.0.CO;2).
- Richter, J. H., and Coauthors, 2024: Quantifying sources of subseasonal prediction skill in CESM2. *npj Climate and Atmospheric Science*, **7** (1), 59, <https://doi.org/10.1038/s41612-024-00595-4>.
- Robertson, A., and F. Vitart, 2018: *Sub-seasonal to seasonal prediction: the gap between weather and climate forecasting*. Elsevier.
- Robertson, A. W., N. Vigaud, J. Yuan, and M. K. Tippett, 2020: Toward identifying subseasonal forecasts of opportunity using North American weather regimes. *Mon. Wea. Rev.*, **148** (5), 1861–1875, <https://doi.org/10.1175/MWR-D-19-0285.1>.
- Roundy, J. K., and E. F. Wood, 2015: The attribution of land–atmosphere interactions on the seasonal predictability of drought. *J. Hydrometeorol.*, **16** (2), 793–810, <https://doi.org/10.1175/JHM-D-14-0121.1>.
- Saji, N., B. N. Goswami, P. Vinayachandran, and T. Yamagata, 1999: A dipole mode in the tropical Indian Ocean. *Nature*, **401** (6751), 360–363, <https://doi.org/10.1038/43854>.
- Scherrer, S. C., M. Croci-Maspoli, C. Schwierz, and C. Appenzeller, 2006: Two-dimensional indices of atmospheric blocking and their statistical relationship with winter climate patterns in the Euro-Atlantic region. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, **26** (2), 233–249, <https://doi.org/10.1002/joc.1250>.
- Schmitt, M., 2022: Deep learning vs. gradient boosting: Benchmarking state-of-the-art machine learning algorithms for credit scoring. *arXiv preprint arXiv:2205.10535*, <https://doi.org/10.48550/arXiv.2205.10535>.
- Screen, J. A., I. Simmonds, and K. Keay, 2011: Dramatic interannual changes of perennial Arctic sea ice linked to abnormal summer storm activity. *Journal of Geophysical Research: Atmospheres*, **116** (D15), <https://doi.org/10.1029/2011JD015847>.
- Sengupta, A., B. Singh, M. J. DeFlorio, C. Raymond, A. W. Robertson, X. Zeng, D. E. Waliser, and J. Jones, 2022: Advances in subseasonal to seasonal prediction relevant to water management

- in the western United States. *Bull. Amer. Meteor. Soc.*, **103** (10), E2168–E2175, <https://doi.org/10.1175/BAMS-D-22-0146.1>.
- Snoek, J., H. Larochelle, and R. P. Adams, 2012: Practical Bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, **25**.
- Stan, C., D. M. Straus, J. S. Frederiksen, H. Lin, E. D. Maloney, and C. Schumacher, 2017: Review of tropical-extratropical teleconnections on intraseasonal time scales. *Rev. Geophys.*, **55** (4), 902–937, <https://doi.org/10.1002/2016RG000538>.
- Straus, D. M., S. Corti, and F. Molteni, 2007: Circulation regimes: Chaotic variability versus sst-forced predictability. *Journal of Climate*, **20** (10), 2251 – 2272, <https://doi.org/10.1175/JCLI4070.1>, URL <https://journals.ametsoc.org/view/journals/clim/20/10/jcli4070.1.xml>.
- Strong, C., and R. E. Davis, 2007: Winter jet stream trends over the Northern Hemisphere. *Quart. J. Roy. Meteor. Soc.*, **133** (629), 2109–2115, <https://doi.org/10.1002/qj.171>.
- Teng, H., G. Branstator, A. B. Tawfik, and P. Callaghan, 2019: Circumglobal response to prescribed soil moisture over North America. *J. Climate*, **32** (14), 4525 – 4545, <https://doi.org/10.1175/JCLI-D-18-0823.1>.
- Thomas, J. A., A. A. Berg, and W. J. Merryfield, 2016: Influence of snow and soil moisture initialization on sub-seasonal predictability and forecast skill in boreal spring. *Climate Dyn.*, **47**, 49–65, <https://doi.org/10.1007/s00382-015-2821-9>.
- Thompson, D. W., and J. M. Wallace, 2000: Annular modes in the extratropical circulation. Part I: Month-to-month variability. *J. Climate*, **13** (5), 1000–1016, [https://doi.org/10.1175/1520-0442\(2000\)013<1000:AMITEC>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<1000:AMITEC>2.0.CO;2).
- Tibaldi, S., and F. Molteni, 1990: On the operational predictability of blocking. *Tellus A*, **42** (3), 343–365, <https://doi.org/10.1034/j.1600-0870.1990.t01-2-00003.x>.
- Trenberth, K. E., G. W. Branstator, D. Karoly, A. Kumar, N.-C. Lau, and C. Ropelewski, 1998: Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. *Journal of Geophysical Research: Oceans*, **103** (C7), 14 291–14 324, <https://doi.org/10.1029/97JC01444>.

- Tseng, K.-C., E. Barnes, and E. Maloney, 2018: Prediction of the midlatitude response to strong Madden-Julian Oscillation events on S2S time scales. *Geophys. Res. Lett.*, **45** (1), 463–470, <https://doi.org/10.1002/2017GL075734>.
- Uchoa, I., I. T. Simoes-Sousa, and I. C. da Silveira, 2023: The Brazil Current mesoscale eddies: Altimetry-based characterization and tracking. *Deep Sea Research Part I: Oceanographic Research Papers*, **192**, 103 947, <https://doi.org/10.1016/j.dsr.2022.103947>.
- Vigaud, N., A. W. Robertson, and M. K. Tippett, 2018: Predictability of recurrent weather regimes over North America during winter from submonthly reforecasts. *Mon. Wea. Rev.*, **146** (8), 2559–2577, <https://doi.org/10.1175/MWR-D-18-0058.1>.
- Vitart, F., A. W. Robertson, and D. L. Anderson, 2012: Subseasonal to seasonal prediction project: Bridging the gap between weather and climate. *Bulletin of the World Meteorological Organization*, **61** (2), 23.
- Vitart, F., and Coauthors, 2022: Outcomes of the WMO Prize Challenge to improve subseasonal to seasonal predictions using artificial intelligence. *Bull. Amer. Meteor. Soc.*, **103** (12), E2878–E2886, <https://doi.org/10.1175/BAMS-D-22-0046.1>.
- Wallace, J. M., and D. S. Gutzler, 1981: Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.*, **109** (4), 784–812, [https://doi.org/10.1175/1520-0493\(1981\)109<0784:TITGHF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2).
- Wang, F., W. Tian, F. Xie, J. Zhang, and Y. Han, 2018: Effect of madden–julian oscillation occurrence frequency on the interannual variability of northern hemisphere stratospheric wave activity in winter. *Journal of Climate*, **31** (13), 5031 – 5049, <https://doi.org/10.1175/JCLI-D-17-0476.1>, URL <https://journals.ametsoc.org/view/journals/clim/31/13/jcli-d-17-0476.1.xml>.
- Weyn, J. A., D. R. Durran, and R. Caruana, 2019: Can machines learn to predict weather? using deep learning to predict gridded 500-hpa geopotential height from historical weather data. *Journal of Advances in Modeling Earth Systems*, **11** (8), 2680–2693, <https://doi.org/https://doi.org/10.1029/2019MS001705>, URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001705>, <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019MS001705>.

- Weyn, J. A., D. R. Durran, R. Caruana, and N. Cresswell-Clay, 2021: Sub-seasonal forecasting with a large ensemble of deep-learning weather prediction models. *Journal of Advances in Modeling Earth Systems*, **13** (7), e2021MS002502, <https://doi.org/10.1029/2021MS002502>.
- Wheeler, M., and G. N. Kiladis, 1999: Convectively coupled equatorial waves: Analysis of clouds and temperature in the wavenumber–frequency domain. *J. Atmos. Sci.*, **56** (3), 374–399, [https://doi.org/10.1175/1520-0469\(1999\)056<0374:CCEWAO>2.0.CO;2](https://doi.org/10.1175/1520-0469(1999)056<0374:CCEWAO>2.0.CO;2).
- Wheeler, M. C., and H. H. Hendon, 2004: An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Mon. Wea. Rev.*, **132** (8), 1917–1932, [https://doi.org/10.1175/1520-0493\(2004\)132<1917:AARMMI>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2).
- White, C. J., and Coauthors, 2017: Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteorological Applications*, **24** (3), 315–325, <https://doi.org/10.1002/met.1654>.
- Yulaeva, E., and J. M. Wallace, 1994: The signature of ENSO in global temperature and precipitation fields derived from the microwave sounding unit. *J. Climate*, **7** (11), 1719–1736, [https://doi.org/10.1175/1520-0442\(1994\)007<1719:TSEOIG>2.0.CO;2](https://doi.org/10.1175/1520-0442(1994)007<1719:TSEOIG>2.0.CO;2).
- Zhou, G., 2019: Atmospheric response to sea surface temperature anomalies in the mid-latitude oceans: A brief review. *Atmosphere-Ocean*, **57** (5), 319–328, <https://doi.org/10.1080/07055900.2019.1702499>.